Brazilian Journal of Probability and Statistics
2015, Vol. 29, No. 3, 608–639
DOI: 10.1214/14-BJPS236

Bayesian analysis and diagnostic of overdispersion models for binomial data

Carolina C. M. Paraíba, Carlos A. R. Diniz and Rubiane M. Pires

UFSCar

Abstract. In the present paper, we focus our attention on the multiplicative binomial model, the double binomial model and the beta-binomial model considering the Bayesian perspective, modeling both the probability of success and the dispersion parameters. A Bayesian methodology is considered for estimation and diagnostic under these three overdispersed binomial regression models. A teratology data set is analyzed using the considered methodology. We present a simulation study, based on data sets generated mimicking the characteristics of the teratology data to assess the quality of Bayesian estimates and to assess the performance of the considered Bayesian diagnostic tools under each regression model. An extended study based on simulated data is also performed to compare the logit and probit link functions in a setting of overdispersed binomial data. We also consider simulated data sets to illustrate how to detect overdispersion using posterior predictive checks.

1 Introduction

One possible approach to binary data modeling in the presence of extra-binomial variation is through distributions which generalize the binomial model. Among these distributions, we focus on the multiplicative binomial (Altham, 1978), the double binomial (Efron, 1986) and the beta-binomial distribution (Skellam, 1948). The multiplicative binomial and double binomial distributions are two-parameter models belonging to the exponential family and allowing for both over and underdispersion. On the other hand, the beta-binomial distribution is not a member of the exponential family and originally only allows for overdispersion. However, Prentice (1986) extended the beta-binomial distribution and established the necessary conditions for the model to accommodate underdispersion as well. Both the multiplicative binomial and the double binomial models have intractable normalizing constants which, according to Lindsey and Altham (1998), may be the reason why they are not widely used.

A Bayesian parametric approach for the double exponential family was proposed by Dey et al. (1997), while Nott (2006) considered a nonparametric Bayesian estimation procedure for the mean and variance functions of response

Key words and phrases. Bayesian analysis, diagnostic, overdispersion, multiplicative binomial, double binomial, beta-binomial.

Received May 2012; accepted January 2014.

variables following a distribution in the double exponential family. Cottet et al. (2008) considered nonparametric Bayesian estimation and variable selection for the double binomial regression model. Lee and Sabavala (1987) developed a Bayesian estimation procedure for the beta-binomial parameters and Kahn and Raftery (1996) generalized the logistic regression model for the beta-binomial distribution and performed a hierarchical Bayesian analysis for the proposed model.

As argued in Gelman et al. (1996), Bayesian data analysis is conditioned on the considered probability model and can be misleading if the assumed model is incorrect (or not plausible enough). Thus, in addition to considering appropriate estimation procedures for overdispersed binomial data, it is also important to conduct diagnostics to check the underlying model assumptions, outliers and influent observations. Gelman et al. (1996) and Gelman et al. (2000) proposed posterior predictive checks, which are aimed to assess the goodness of fit of the proposed models by detecting differences between the fitted model and the observed data. The discrepancy variables can be compared using the posterior predictive p-value proposed by Meng (1994). Albert and Chib (1993) defined a Bayesian residual for binary data which has a real-valued posterior distribution and may be used for outliers detection. Albert and Chib (1995) also proposed two types of Bayesian residual for binary regression analysis. The authors stress that since these residuals are functions of the parameters, the precision of the knowledge about the parameters is reflected in the precision of the residuals. Spiegelhalter et al. (2002) derived a metric for the effective number of parameters in a model. They noted that the contribution of each observation turns out to be its leverage, defined as the relative influence of each observation on its fitted value. Therefore, the authors suggested the posterior mean deviance as a Bayesian of fit or adequacy. Moreover, Spiegelhalter et al. (2002) noted that the influence of each observation in parameter estimates can be assessed by means of a Bayesian deviance residual against leverages plot.

Therefore, this paper sets out not only to present a simple, yet effective, Bayesian alternative to otherwise complex classical estimation procedures for the multiplicative binomial, double binomial and beta-binomial regression models, but also to consider Bayesian diagnostic metrics for model assessment under these three overdispersed binomial regression models. Bayesian parameter estimation is validated using simulation results, which are also used to assess the effectiveness of different Bayesian techniques for model diagnostic. Moreover, we present the analysis of a low-iron teratology study data to illustrate the considered Bayesian estimation and diagnostic procedures.

The paper is organized as follows. In Section 2, we review Altham's multiplicative binomial model, Efron's double binomial model and Skellam's beta-binomial model. In Section 3, we present the proposed Bayesian methodology for estimation and diagnostic for binomial overdispersion with the multiplicative binomial, the double binomial and the beta-binomial as the sampling distribution. In Section 4, results based on simulated data sets are presented to assess the quality of

Bayesian estimates and to investigate the effectiveness of the Bayesian model diagnostic techniques addressed. We also present a discussion regarding the use of posterior predictive checks to detect overdispersion and a comparison between the results obtained with the logit and probit link functions when dealing with overdispersed binomial data. In Section 5, a low-iron teratology data set by Shepard et al. (1980) is analyzed using the proposed Bayesian methodology. Finally, in Section 6, we give a few brief concluding remarks.

2 Generalizations of the binomial model

Consider Y a response variable denoting the sum of n binary random variables (r.v.'s), $U_1, \ldots, U_n$, and $\mathbf{x} = (x_1, \ldots, x_q)'$ a vector of q covariates. Let p be the probability of success and γ the dispersion parameter of the binary variables. Thus, it is possible to specify regression models, $p(\mathbf{x}_1; \boldsymbol{\beta})$ and $\gamma(\mathbf{x}_2; \boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{q_1})$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{q_2})$ are unknown vectors of parameters to be estimated, with $\mathbf{x}_1$ and $\mathbf{x}_2$ being subsets of the covariate vector, $\mathbf{x}$. Let $\boldsymbol{\eta}_1 = \mathbf{x}_1'\boldsymbol{\beta}$ and $\boldsymbol{\eta}_2 = \mathbf{x}_2'\boldsymbol{\alpha}$ be the linear predictors of p and γ. We note that the linear predictors of p and γ may or may not contain an intercept parameter which would be denoted by β_0 and α_0, with $x_0 = 1$. In this paper, we consider the complementary log-log, logit and probit link functions for p and the log link function for γ. The link functions for p will be generically denoted by $g^{-1}(\cdot)$ and the log link considered for γ will be denoted by $h^{-1}(\cdot)$.

The multiplicative binomial distribution (Altham, 1978), denoted by MB(n, p, γ), is given by

$$P(Y = y|p, \gamma) = \binom{n}{y} p^y (1-p)^{n-y} \gamma^{y(n-y)} \times \left[\sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} \gamma^{j(n-j)} \right]^{-1}, \tag{2.1}$$

where $n \in \mathbb{N}$, $p \in (0, 1)$, $\gamma > 0$, and $y = 0, \ldots, n$.

The expectation and variance of Y, $Y \sim \text{MB}(n, p, \gamma)$, are

$$E(Y) = np \frac{\kappa_{n-1}(p, \gamma)}{\kappa_n(p, \gamma)} \tag{2.2}$$

and

$$\text{Var}(Y) = np \left[\frac{\kappa_{n-1}(p, \gamma)}{\kappa_n(p, \gamma)} - p \frac{n\kappa_{n-1}^2(p, \gamma) - (n-1)\kappa_{n-2}(p, \gamma)\kappa_n(p, \gamma)}{\kappa_n^2(p, \gamma)} \right], \tag{2.3}$$

where $\kappa_{n-a}(p,\gamma) = \sum_{j=0}^{n-a} \binom{n-a}{j} p^j (1-p)^{n-a-j} \gamma^{(n-a-j)(j+a)}$.

Consider $Y \sim \text{Bin}(n,p)/n$, $n \in \mathbb{N}$, $p \in (0,1)$, a rescaled binomial r.v. Then, the double binomial distribution (Efron, 1986) may be written as

$$\begin{aligned}\tilde{P}(Y=y|p,\gamma) = c(p,\gamma,n)\gamma^{1/2} p^{ny\gamma}(1-p)^{n\gamma(1-y)} \\ \times y^{ny(1-\gamma)}(1-y)^{n(1-y)(1-\gamma)},\end{aligned} \tag{2.4}$$

where $\gamma > 0$, and $c(p,\gamma,n) = 1 + (1/12n)((1-\gamma)/\gamma)(1-1/p(1-p))$ is the approximation for the normalizing constant. Efron (1986) showed that as the n increases $c(p,\gamma,n) \approx 1$, so that (2.4) can by approximated by

$$P(Y=y|p,\gamma) = \gamma^{1/2} p^{ny\gamma}(1-p)^{n\gamma(1-y)} y^{ny(1-\gamma)}(1-y)^{n(1-y)(1-\gamma)}, \tag{2.5}$$

which is the unnormalized form of the double binomial distribution. As discussed in Lee and Nelder (2000), (2.5) works quite well for moderate overdispersed data but can be poor for extremely overdispersed data when p is small. We shall denote a double binomial r.v. by $\text{DB}(n,p,\gamma)$.

It can be shown (see Efron, 1986) that the mean and variance of a double binomial r.v. can be approximated by

$$E(Y) = p, \tag{2.6}$$

and

$$\text{Var}(Y) = \frac{p(1-p)}{n\gamma}. \tag{2.7}$$

The beta-binomial distribution, first proposed by Skellam (1948), was derived considering $Y \sim \text{Bin}(n,p)$, $n \in \mathbb{N}$, $p \in (0,1)$ and assuming that the $p \sim \text{Beta}(a,b)$, $a,b > 0$. Then,

$$P(Y=y|a,b) = \binom{n}{y} \frac{B(a+x, b+n-x)}{B(a,b)},$$

is the beta-binomial distribution, where $B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1}\,dy$ is the beta function.

Letting, $p = a(a+b)^{-1}$ and $\gamma = (a+b)^{-1}$, $\gamma > 0$, we have that $a = p\gamma^{-1}$ and $b = (1-p)\gamma^{-1}$. Thus, the beta-binomial can be reparameterized as

$$P(Y=y|p,\gamma) = \binom{n}{y} \frac{B(p\gamma^{-1}+y, (1-p)\gamma^{-1}+n-y)}{B(p\gamma^{-1},(1-p)\gamma^{-1})}. \tag{2.8}$$

The expectation and variance of a beta-binomial r.v., denoted by $\text{BB}(n,p,\gamma)$, are given by

$$E(Y) = np, \tag{2.9}$$

and

$$\mathrm{Var}(Y) = np(1-p)\left[1 + (n-1)\left(\frac{\gamma}{1+\gamma}\right)\right], \tag{2.10}$$

respectively.

Since p and γ in (2.8) are reparametrization of parameters a and b, they are not the same as in the multiplicative binomial model (2.1) and double binomial model (2.4).

Suppose $Y_1, \ldots, Y_m$ are independent r.v.'s, $i = 1, \ldots, m$. Let $\mathbf{y} = (y_1, \ldots, y_m)'$ be a vector of observed values of $\mathbf{Y} = (Y_1, \ldots, Y_m)'$, with $\mathbf{n} = (n_1, \ldots, n_m)'$, $\mathbf{p} = (p_1, \ldots, p_m)'$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)'$, where $p_i = g^{-1}(\eta_{1i}) = g^{-1}(\mathbf{x}_{1i}'\boldsymbol{\beta})$ and $\gamma_i = h^{-1}(\eta_{2i}) = h^{-1}(\mathbf{x}_{2i}'\boldsymbol{\alpha})$, $i = 1, \ldots, m$. Given the data set $D = (m, \mathbf{n}, \mathbf{y}, \mathbf{x})$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ the vector of unknown parameters to be estimated, then:

$$\begin{aligned} L(\boldsymbol{\theta}|D) = \prod_{i=1}^{m} \Bigg\{ & \binom{n_i}{y_i} \big(g^{-1}(\eta_{1i})\big)^{y_i} \big(1 - g^{-1}(\eta_{1i})\big)^{n_i - y_i} \big(h^{-1}(\eta_{2i})\big)^{y_i(n_i - y_i)} \\ & \times \Bigg[\sum_{j=0}^{n_i} \binom{n_i}{j} \big(g^{-1}(\eta_{1i})\big)^{j} \big(1 - g^{-1}(\eta_{1i})\big)^{n_i - j} \\ & \times \big(h^{-1}(\eta_{2i})\big)^{j(n_i - j)} \Bigg]^{-1} \Bigg\}, \end{aligned}$$

is the likelihood function obtained assuming $Y_i \sim \mathrm{MB}(n_i, p_i, \gamma_i)$, $i = 1, \ldots, m$.

$$\begin{aligned} L(\boldsymbol{\theta}|D) = \prod_{i=1}^{m} \big\{ & \big(h^{-1}(\eta_{2i})\big)^{1/2} \big(g^{-1}(\eta_{1i})\big)^{n_i y_i (h^{-1}(\eta_{2i}))} \big(1 - g^{-1}(\eta_{1i})\big)^{n_i (h^{-1}(\eta_{2i}))(1 - y_i)} \\ & \times y_i^{n_i y_i (1 - h^{-1}(\eta_{2i}))} (1 - y_i)^{n_i (1 - y_i)(1 - h^{-1}(\eta_{2i}))} \big\}, \end{aligned}$$

is the likelihood function where each Y_i follows a $\mathrm{DB}(n_i, p_i, \gamma_i)$, $i = 1, \ldots, m$.

$$\begin{aligned} L(\boldsymbol{\theta}|D) = \prod_{i=1}^{m} \binom{n_i}{y_i} & B\big(g^{-1}(\eta_{1i})\big(h^{-1}(\eta_{2i})\big)^{-1} + y_i, \\ & \big(1 - g^{-1}(\eta_{1i})\big)\big(h^{-1}(\eta_{2i})\big)^{-1} + n_i - y_i\big) \\ & / B\big(g^{-1}(\eta_{1i})\big(h^{-1}(\eta_{2i})\big)^{-1}, \big(1 - g^{-1}(\eta_{1i})\big)\big(h^{-1}(\eta_{2i})\big)^{-1}\big), \end{aligned}$$

is the likelihood function for $Y_i \sim \mathrm{BB}(n_i, p_i, \gamma_i)$, $i = 1, \ldots, m$.

3 Bayesian analysis

Consider a data set $D = (m, \mathbf{n}, \mathbf{y}, \mathbf{x})$ and let $P(\mathbf{Y}|\boldsymbol{\theta})$ be the sampling distribution of $\mathbf{y}$, with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ its vector of indexing parameters. The Bayesian analysis will

be carried out assuming independence between the parameters and a non informative normal prior, $\pi(\boldsymbol{\theta})$, for $\boldsymbol{\theta}$, $\boldsymbol{\theta} \sim N_{q_1+q_2}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_{q_1}, \tau_1, \ldots, \tau_{q_2})$ and $\sigma_1, \ldots, \sigma_{q_1}, \tau_1, \ldots, \tau_{q_2}$ are fixed known hyperparameters.

The posterior distributions are given by

$$
\begin{aligned}
&\pi(\boldsymbol{\theta}|D) \\
&\quad \propto \exp\left\{-\frac{1}{2}\left[\sum_{k=1}^{q_1}\frac{\beta_K^2}{\sigma_k} + \sum_{t=1}^{q_2}\frac{\alpha_t^2}{\tau_t}\right]\right\} \\
&\qquad \times \prod_{i=1}^{m}\left\{\binom{n_i}{y_i}\left(g^{-1}(\boldsymbol{\eta}_{1i})\right)^{y_i}\left(1-g^{-1}(\boldsymbol{\eta}_{1i})\right)^{n_i-y_i}\left(h^{-1}(\boldsymbol{\eta}_{2i})\right)^{y_i(n_i-y_i)}\right. \\
&\qquad\qquad \left.\times\left[\sum_{j=0}^{n_i}\binom{n_i}{j}\left(g^{-1}(\boldsymbol{\eta}_{1i})\right)^{j}\left(1-g^{-1}(\boldsymbol{\eta}_{1i})\right)^{n_i-j}\left(h^{-1}(\boldsymbol{\eta}_{2i})\right)^{j(n_i-j)}\right]^{-1}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&\pi(\boldsymbol{\theta}|D) \\
&\quad \propto \exp\left\{-\frac{1}{2}\left[\sum_{k=1}^{q_1}\frac{\beta_K^2}{\sigma_k} + \sum_{t=1}^{q_2}\frac{\alpha_t^2}{\tau_t}\right]\right\} \\
&\qquad \times \prod_{i=1}^{m}\left\{\left(h^{-1}(\boldsymbol{\eta}_{2i})\right)^{1/2}\left(g^{-1}(\boldsymbol{\eta}_{1i})\right)^{n_i y_i (h^{-1}(\boldsymbol{\eta}_{2i}))}\right. \\
&\qquad\qquad \times\left(1-g^{-1}(\boldsymbol{\eta}_{1i})\right)^{n_i(h^{-1}(\boldsymbol{\eta}_{2i}))(1-y_i)} \\
&\qquad\qquad \left.\times\, y_i^{n_i y_i(1-h^{-1}(\boldsymbol{\eta}_{2i}))}(1-y_i)^{n_i(1-y_i)(1-h^{-1}(\boldsymbol{\eta}_{2i}))}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
\pi(\boldsymbol{\theta}|D) \propto \exp&\left\{-\frac{1}{2}\left[\sum_{k=1}^{q_1}\frac{\beta_K^2}{\sigma_k} + \sum_{t=1}^{q_2}\frac{\alpha_t^2}{\tau_t}\right]\right\} \\
&\times \prod_{i=1}^{m}\binom{n_i}{y_i} B\left(g^{-1}(\boldsymbol{\eta}_{1i})(h^{-1}(\boldsymbol{\eta}_{2i}))^{-1} + y_i,\right. \\
&\qquad \left.\left(1-g^{-1}(\boldsymbol{\eta}_{1i})\right)\left(h^{-1}(\boldsymbol{\eta}_{2i})\right)^{-1} + n_i - y_i\right) \\
&\qquad / B\left(g^{-1}(\boldsymbol{\eta}_{1i})\left(h^{-1}(\boldsymbol{\eta}_{2i})\right)^{-1}, \left(1-g^{-1}(\boldsymbol{\eta}_{1i})\right)\left(h^{-1}(\boldsymbol{\eta}_{2i})\right)^{-1}\right),
\end{aligned}
$$

for the multiplicative binomial, double binomial and beta-binomial regression models, respectively.

We note that the posterior distributions and the full conditionals (not shown) of each model are not analytically tractable. Bayesian inference will be performed using Markov chain Monte Carlo methods (MCMC) such as Metropolis within Gibbs algorithm (Robert and Casella, 2004).

3.1 Conditional predictive ordinate

The conditional predictive ordinate (CPO) (Gelfand et al., 1992), defined as the predictive density of the ith case given the data without the ith case ($D_{(-i)}$), for $i = 1, \ldots, m$, is given by

$$CPO_i = \left[\int_{\Theta} \frac{1}{P(y_i|\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|D)\, d\boldsymbol{\theta} \right]^{-1}. \tag{3.1}$$

The CPO is interpreted as the posterior probability of observing the ith observation when the model is fitted to $D_{(-i)}$. The larger the value of the CPO_i, the better the model fits y_i. On the other hand, small values of CPO_i indicate that the ith observation may be an outlier.

If a MCMC sample of size r of the posterior distribution, $\pi(\boldsymbol{\theta}|D)$, is available, then the predicted value, $\tilde{y}$, of an observation y may be computed by a Monte Carlo approximation of (3.1). An algorithm derived by Pires and Diniz (2012) to determine the predicted value $\tilde{y}_i$ of y_i is described in Appendix A.1.

3.2 Posterior predictive checks

Posterior predictive checks (Gelman et al., 2000) for goodness of fit assessment are performed by generating replicated data sets, $\mathbf{y}^{\text{rep}}$, from the posterior predictive distribution and comparing them to the observed data set by means of discrepancy variables, $T(\cdot)$, which can be any function of data and model parameters. In this paper we consider the following discrepancy variables: the mean and variance of the response variable, and the model deviance. Once a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$ is obtained, the discrepancy variable may be computed by the procedure described in Appendix A.2.

Discrepancy variables can be shown graphically, as a histogram, or summarized by a posterior predictive p-value (Gelman et al., 2003), which is estimated as

$$\hat{p}\text{-value} = \frac{\#(T(\mathbf{y}^{\text{rep}}) \geq T(\mathbf{y}))}{r}. \tag{3.2}$$

If the discrepancy variable does not depend on model parameters, the graphical representation consists of a histogram of the values of the discrepancy variable computed for each replication of the response variable. It is common to add a vertical line to this histogram representing the value obtained for this same discrepancy variable computed with the observed data. If the model agrees with the data, then the vertical line will be next to the histogram's peak.

If the discrepancy variable depends on the model parameters, then we shall compute the difference between the value of discrepancy obtained for the observed data and its value obtained for the replicated data. These differences are graphically shown in a histogram which should contain the zero so that the model is in agreement with the data.

The posterior predictive p-value is interpreted as the probability of observing data as extreme as that which is actually observed, conditional on the model (Lynch and Western, 2004). If the posterior predictive p-value is close to 0 or 1, it may suggest that the observed data has an extreme discrepancy variable and the model may be inappropriate.

3.3 Bayesian residuals and influence measures

We consider three types of Bayesian residuals to check for model adequacy, outliers, and influent observations: standardized residuals based on the CPO (Pires and Diniz, 2012); standardized residuals based on the posterior distribution of model parameters (Albert and Chib, 1993, 1995); and standardized Bayesian deviance residuals (Spiegelhalter et al., 2002), which may be used to detect outliers and as an influence measure of the observations in parameter fitting.

As influence diagnostics metrics, we consider: leverages (Spiegelhalter et al., 2002), which represent the relative influence of each observation on its fitted value; and the Kullback–Leibler divergence calibration based on the CPO (Peng and Dey, 1995; Cho et al., 2009).

If a model is well-fitted to a data set, then the points in a residuals vs. predicted values plot are expected to be randomly distributed around zero. Moreover, an observation may be considered as an outlier if its residual is distant from zero (or if it is large). Since we are considering standardized residuals, they represent how much predicted values deviate from real values. Thus, we can set limits for the residuals based on the amount of deviation that we are willing to tolerate between the real value and the predicted value. In this paper, we shall consider an observation as an outlier if its standardized residual is larger than 3 or smaller than -3.

The standardized residual based on the CPO, for $i = 1, \ldots, m$, is written as

$$r_i^{\text{pred}} = \frac{y_i - \tilde{y}_i}{\sqrt{\text{Var}(Y_i | D, \hat{\boldsymbol{\theta}})}},$$

where y_i is the observed value for the ith response, $\tilde{y}_i$ its predicted value obtained by the CPO procedure described in Appendix A.1. We consider the posterior mean as the Bayesian estimate, $\hat{\boldsymbol{\theta}}$, of the parameter vector $\boldsymbol{\theta}$. Expressions for $\text{Var}(Y_i | D, \hat{\boldsymbol{\theta}})$ are given by (2.3), (2.7) and (2.10) for each of the considered models.

The standardized residual based on the posterior distribution of model parameters, for $i = 1, \ldots, m$, is given by

$$r_i^{\text{post}} = \frac{y_i - E(Y_i | D, \boldsymbol{\theta})}{\sqrt{\text{Var}(Y_i | D, \boldsymbol{\theta})}}. \tag{3.3}$$

An algorithm to compute a sample of (3.3), based on a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$, is described in Appendix A.3. Expressions for $E(Y_i | D, \boldsymbol{\theta})$ are given by (2.2), (2.6) and (2.9) and expressions for $\text{Var}(Y_i | D, \boldsymbol{\theta})$ are given by (2.3), (2.7) and (2.10).

The standardized Bayesian deviance residual, for $i = 1, \ldots, m$, is

$$r_i^{\text{dev}} = \frac{\text{sgn}(y_i - E(Y_i|\hat{\boldsymbol{\theta}}))\sqrt{D_i(\hat{\boldsymbol{\theta}}) + \widehat{p_{D_i}}}}{\sqrt{\text{Var}(Y_i|D, \hat{\boldsymbol{\theta}})}},$$

where

$$D_i(\hat{\boldsymbol{\theta}}) = -2\log P(Y_i|\hat{\boldsymbol{\theta}}) \tag{3.4}$$

is the contribution of the ith observation for the overall deviance,

$$\widehat{p_{D_i}} = -2\left\{\frac{1}{r}\sum_{j=1}^{r}\left[\log\frac{\pi(\boldsymbol{\theta}_j|D_i)}{\pi(\boldsymbol{\theta}_j)}\right] - \log\frac{\pi(\hat{\boldsymbol{\theta}}|D_i)}{\pi(\hat{\boldsymbol{\theta}})}\right\} \tag{3.5}$$

is the Monte Carlo approximation of the relative influence of each observation on its fitted value (the leverage of each observation), based on a sample of size r of the posterior distribution. $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ is its Bayesian estimate, that is, the posterior mean of $\boldsymbol{\theta}$. $E(Y_i|D, \hat{\boldsymbol{\theta}})$ is given by (2.2), (2.6) or (2.9), and $\text{Var}(Y_i|D, \hat{\boldsymbol{\theta}})$ is given by (2.3), (2.7) or (2.10).

According to Spiegelhalter et al. (2002), adding curves of the form $x^2 + y = c$ to the plot of deviance residuals against leverages allow us to identify the contribution c of the ith observation to the overall deviance information criterion value. Therefore, leverages can be used to assess the influence of observations in parameter estimates and the closer the value of (3.5) is to one, the more influential the observation is on its fitted value. Based on the simulation results discussed in Section 4.2, we shall consider an observation as an influential case if its leverage is greater than 0.8.

The Kullback–Leibler (KL) divergence calibration for influence diagnostic based on the CPO uses a Bayesian perspective to case deletion diagnostic and evaluates the influence of a given observation in parameter estimates.

Following Peng and Dey (1995) and Cho et al. (2009), the calibration, for $i = 1, \ldots, m$, is written as

$$p_i = 0.5\{1 + \sqrt{1 - \exp[-2\hat{K}(\pi(\boldsymbol{\theta}|D), \pi(\boldsymbol{\theta}|D_{(-i)}))]}\}, \tag{3.6}$$

where $\hat{K}(\pi(\boldsymbol{\theta}|D), \pi(\boldsymbol{\theta}|D_{(-i)})) = \log\{\frac{1}{r}\sum_{j=1}^{r}\frac{1}{P(y_i|\boldsymbol{\theta}_j)}\} + \frac{1}{r}\sum_{j=1}^{r}\log P(y_i|\boldsymbol{\theta}_j)$ is the Monte Carlo approximation of the KL divergence between the posterior distribution with the complete data and the posterior distribution with the ith observation deleted based on a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$.

According to Cho et al. (2009), if $p_i \gg 0.5$ then the ith case is an influential outlier, since its deletion changes the posterior distribution as much as describing an observed event as having probability p_i when the correct probability is 0.5. Based on the simulation results discussed in Section 4.2, we shall consider an observation as an influential case if its calibration is greater than 0.9.

3.4 Bayesian model selection

The metric known as sum of log-CPO and written as

$$\text{log-CPO} = \sum_{i=1}^{m} \log(CPO_i), \tag{3.7}$$

is an estimator of the logarithm of the marginal likelihood and can be used as model selection criterion. The model to be selected is the one providing the larger value of (3.7) (Carlin and Louis, 2009).

Model selection can also be accomplished using the deviance information criterion (DIC) (Spiegelhalter et al., 2002), which is given by

$$DIC = \bar{D} + p_D, \tag{3.8}$$

where $\bar{D} = \sum_{i=1}^{m} D_i(\hat{\boldsymbol{\theta}}) + p_{D_i}$ and $p_D = \sum_{i=1}^{m} p_{D_i}$, with $D_i(\hat{\boldsymbol{\theta}})$ given in (3.4) and p_{D_i} replaced by its Monte Carlo approximation (3.5). The model to be selected is the one which provides the smaller value of (3.8).

We also consider the Bayes factor (Kass and Raftery, 1995) for model selection. Given two models M_0 and M_1 with sampling distributions $P_0(D|\boldsymbol{\theta})$ and $P_1(D|\boldsymbol{\theta})$ and prior distribution $\pi_0(\boldsymbol{\theta})$ and $\pi_1(\boldsymbol{\theta})$, respectively, the Bayes factor in support of model M_0 compared with model M_1 can be approximated by

$$BF = \frac{\pi(D|M_0)}{\pi(D|M_1)}, \tag{3.9}$$

where, based on a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$, $\pi(D|M_k)$ is approximated by $\frac{1}{r}\sum_{j=1}^{r} L_{M_K}(\boldsymbol{\theta}_j|D)$ (Aitkin, 1991), with $L_{M_K}(\boldsymbol{\theta}_j|D)$ denoting the likelihood function considering model M_k, $k = 0, 1$.

4 Results based on simulated data sets

4.1 Model fit

Although results based on simulation studies cannot be generalized, they can be regarded as an internal control of the considered methodology. Thus, in this section, we present results based on simulation studies to assess the quality of Bayesian estimates for each proposed model. Our main aim is to assess the quality of estimates and study their frequentist properties. This simulation study was conducted based on data sets generated using the Bayesian parameter estimates obtained from the application example, namely, the low-iron rat teratology data to be analyzed in Section 5. The teratology data features $m = 58$ observations from pregnant rats where n_i is the litter size, y_i is the number of dead fetuses in each litter and x_i is a covariate accounting for the hemoglobin level of the rat, $i = 1, \ldots, m$.

Let $\eta_{1i} = \beta_0 + \beta_1 x_i$ and $\eta_{2i} = \alpha_1 n_i$ be the linear predictors of p_i and γ_i, respectively, with n_i the number of dead fetuses and x_i the hemoglobin level of the rats

Table 1 *Parameter estimates obtained for the low-iron rat teratology data under each regression model*

	Parameter estimate		
Model	β_0	β_1	α_1
MBC	0.84	−0.19	−0.01
MBL	1.63	−0.26	−0.01
MBP	0.96	−0.15	−0.01
DBC	1.64	−0.31	−0.09
DBL	2.58	−0.37	−0.10
DBP	1.42	−0.20	−0.10
DBC approx.	2.23	−0.42	−0.12
DBL approx.	3.84	−0.59	−0.12
DBP approx.	2.09	−0.31	−0.13
BBC	2.00	−0.37	−0.07
BBL	3.25	−0.48	−0.07
BBP	1.74	−0.25	−0.06

both taken from the teratology data set, for $i = 1, \ldots, m$ and $m = 58$. Assume a $N_3(\mathbf{0}, 10^4 I_3)$ prior distribution for the unknown parameter vector $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma_1)$. Then, each observation y_i is simulated from a $\text{MB}(n_i, p_i, \gamma_i)$, $\text{DB}(n_i, p_i, \gamma_i)$, and $\text{BB}(n_i, p_i, \gamma_i)$, where $p_i = g^{-1}(\boldsymbol{\eta}_{1i}) = g^{-1}(\beta_0 + \beta_1 x_i)$ and $\gamma_i = h^{-1}(\eta_{2i}) = h^{-1}(\alpha_1 n_i)$. Since the generated data is set to mimic the application data, the responses are simulated with the real value of parameters β_0, β_1 and γ_1 set at their Bayesian estimate, given by the posterior mean, obtained fitting each considered regression models to the low-iron rat teratology data.

Table 1 shows the Bayesian parameter estimates of $\boldsymbol{\theta} = (\beta_0, \beta_1, \gamma_1)$ for all considered models fitted to the teratology data set. Models are coded as follows: MBC, MBL, and MBP stand for the multiplicative binomial regression model with complementary log-log, logit and probit link for p, respectively; DBC, DBL, and DBP stand for the normalized double binomial regression model with complementary log-log, logit and probit link for p, respectively; DBC approx., DBL approx., and DBP approx. stand for the unnormalized double binomial regression model with complementary log-log, logit and probit link for p, respectively; and BBC, BBL, and BBP stand for the beta-binomial regression model with complementary log-log, logit and probit link for p, respectively.

A total of 100 samples were computed for each model. MCMC samples of the posterior distributions were obtained using a Gibbs–Metropolis type algorithm with candidates generated by random walk from a multivariate normal distribution with covariance matrix given by minus the Hessian matrix evaluated at the maximum likelihood estimator of $\boldsymbol{\theta}$. Convergence was checked using Geweke's criterion (Geweke, 1992). The Gibbs–Metropolis algorithm for each model was written

in R software (R Core Team, 2013). Chain sizes were 100,000 with burn-in periods of 20,000 and a thinning interval of size 200 for all models. Rejection rates were: 0.55 for the multiplicative binomial with complementary log-log, logit and probit; 0.54 for both the normalized and unnormalized double binomial regression models with complementary log-log, logit and probit; and 0.55 for the beta-binomial regression models with complementary log-log, logit and probit link for p.

For each model, the results summarized in Table 2 are the following: the row called Mean refers to the mean of the posterior means obtained for each simulated sample; the Bias-m row presents the mean of the bias of the posterior means; the MSE-m row provides the mean of the mean square error (MSE) computed for all posterior means; the row named Median refers to the median of the posterior medians obtained for each simulated sample; the Bias-md row presents the median of the bias of the posterior medians; the MSE-md row provides the median of the MSE computed for all posterior medians; the IQ-CP and HPD-CP rows provide the estimated coverage probability of 95% interquantile and HPD credibility intervals, respectively.

The results in Table 2 show that for the multiplicative binomial regression model, the Bayesian estimation procedure provided quite precise estimates for the unknown parameters β_0, β_1, and γ_1. For all three parameters, bias mean and MSE mean are very small. We notice that the posterior mean and posterior median are quite similar and that both estimate the real parameters values quite precisely. Moreover, from Table 2, we notice that the estimated coverage probability approaches the nominal expected of 95%.

As mentioned in Section 2, from the frequentist perspective, working with the unnormalized form of (2.5) may lead to a more convenient implementation of the inference procedures. On the other hand, from the Bayesian perspective there is no computational or methodological gain in considering the unnormalized double binomial distribution, since both forms of the probability function can be treated conversely. Moreover, a simulation conducted using (2.5) revealed that the bias mean and MSE mean of β_0, β_1 and α_1 are much larger than those observed for these same parameters when using the normalized form of the distribution (2.4). Moreover, considering (2.5), model parameter estimates are not as accurate as the ones obtained in the normalized form of the double binomial distribution. It can also be observed that the estimated coverage probabilities are far from the expected nominal one. Thus, we proceed with the analysis using (2.4) and consider the results shown in Table 2 for the normalized double binomial regression model, which shows that the posterior mean and median are very close, providing similar results, and parameter estimates are accurate with both the mean bias and MSE mean quite small. Furthermore, the normalized form of the double binomial model provides coverage probabilities close to the expected 95%.

For the beta-binomial regression model, the Bayesian estimation procedure (summarized in Table 2) provided quite precise estimates for the unknown parameters β_0, β_1, and γ_1. For all three parameters, bias mean and MSE mean are

Table 2 *Summarized simulation results of the regression models considered*

Link for p	Complementary log-log			Logit			Probit		
Link for γ	Log			Log			Log		
Parameter	β_0	β_1	α_1	β_0	β_1	α_1	β_0	β_1	α_1
				Multiplicative binomial					
Real value	0.84	−0.19	−0.01	1.63	−0.26	−0.01	0.96	−0.15	−0.01
Mean	0.95	−0.21	−0.01	1.80	−0.28	−0.01	1.04	−0.16	−0.01
Bias-m	0.11	−0.02	0.00	0.18	−0.03	0.00	0.08	−0.01	0.00
MSE-m	0.12	0.00	0.00	0.25	0.01	0.00	0.07	0.00	0.00
Median	0.89	−0.21	−0.01	1.73	−0.27	−0.01	0.99	−0.16	−0.01
Bias-md	0.05	−0.01	0.00	0.10	−0.02	0.00	0.03	−0.01	0.00
MSE-md	0.08	0.00	0.00	0.13	0.00	0.00	0.04	0.00	0.00
IQ-CP	0.88	0.91	0.92	0.91	0.92	0.92	0.94	0.94	0.92
HPD-CP	0.89	0.92	0.92	0.89	0.90	0.93	0.94	0.94	0.91
				Normalized double binomial					
Real value	1.64	−0.31	−0.09	2.58	−0.37	−0.10	1.42	−0.20	−0.10
Mean	1.52	−0.29	−0.09	2.42	−0.35	−0.09	1.31	−0.18	−0.10
Bias-m	−0.12	0.02	0.00	−0.16	0.02	0.00	−0.11	0.02	0.00
MSE-m	0.15	0.00	0.00	0.31	0.01	0.00	0.10	0.00	0.00
Median	1.50	−0.29	−0.09	2.46	−0.35	−0.09	1.34	−0.19	−0.10
Bias-md	−0.14	0.02	0.00	−0.11	0.02	0.00	−0.09	0.01	0.00
MSE-md	0.11	0.00	0.00	0.23	0.00	0.00	0.08	0.00	0.00
IQ-CP	0.91	0.93	0.95	0.95	0.94	0.97	0.94	0.95	0.97
HPD-CP	0.91	0.93	0.95	0.95	0.93	0.96	0.94	0.94	0.96
				Unnormalized double binomial					
Real value	2.23	−0.42	−0.12	3.84	−0.59	−0.12	2.09	−0.31	−0.13
Mean	1.86	−0.36	−0.10	3.31	−0.51	−0.10	1.77	−0.27	−0.11
Bias-m	−0.37	0.06	0.02	−0.53	0.09	0.02	−0.31	0.05	0.02
MSE-m	0.32	0.01	0.00	0.82	0.02	0.00	0.25	0.01	0.00
Median	1.84	−0.35	−0.10	3.25	−0.50	−0.10	1.78	−0.26	−0.11
Bias-md	−0.39	0.07	0.02	−0.59	0.09	0.02	−0.31	0.05	0.02
MSE-md	0.24	0.01	0.00	0.64	0.02	0.00	0.18	0.00	0.00
IQ-CP	0.72	0.77	0.86	0.78	0.76	0.87	0.78	0.74	0.89
HPD-CP	0.73	0.75	0.85	0.79	0.72	0.84	0.76	0.69	0.86
				Beta-binomial					
Real value	2.00	−0.37	−0.07	3.25	−0.48	−0.07	1.74	−0.25	−0.06
Mean	2.08	−0.39	−0.08	3.38	−0.50	−0.07	1.74	−0.25	−0.06
Bias-m	0.07	−0.02	−0.01	0.13	−0.02	0.00	0.00	0.00	0.00
MSE-m	0.23	0.01	0.00	0.65	0.01	0.00	0.15	0.00	0.00
Median	2.07	−0.38	−0.08	3.34	−0.49	−0.07	1.73	−0.25	−0.06
Bias-md	0.06	−0.01	−0.01	0.09	−0.02	0.00	−0.02	0.00	0.00
MSE-md	0.16	0.00	0.00	0.42	0.01	0.00	0.12	0.00	0.00
IQ-CP	0.93	0.93	0.94	0.92	0.93	0.92	0.98	0.94	0.92
HPD-CP	0.92	0.93	0.95	0.91	0.92	0.95	0.99	0.94	0.91

very small. It can be observed that the posterior mean and posterior median are quite similar and that both estimate the real parameters values quite precisely. The estimated coverage probability approaches the nominal expected of 95%.

4.2 Model diagnostic

For model diagnostic, we take one sample of each model and compute the Bayesian residuals and metrics described before. The main idea is to assess the effectiveness of the different model diagnostics considered and to see if they behave as would be expected when the true model is fitted to the observed data set. Since the observed results for each link function were similar for each model, we have chosen to show only one case of the multiplicative binomial, double binomial and beta-binomial regression models considered in this paper. Thus, we present the results obtained for the multiplicative binomial regression model with complementary log-log link function, the results of the double binomial regression model with probit link function and the results of the beta-binomial regression model with logit link function. It can be noted that all diagnostic metrics were computed to several simulated data sets for each regression model and the same pattern of behavior presented next was observed.

Posterior predictive checks were computed for the discrepancies described in Section 3.2. The estimated posterior predictive p-values were: 0.53, 0.44, and 0.54 for the multiplicative binomial regression model with complementary log-log link function; 0.50, 0.47, and 0.60 for the double binomial regression model with probit link function; 0.60, 0.52, and 0.37 for the beta-binomial regression model with logit link function. It can be observed that no extreme posterior predictive p-values were observed, which suggest that the pattern presented by the observed data is not different from the pattern of replications of the data, therefore indicating that the model fits well to the data. This is consistent with Figures 1(a), (b), 2(a), (b) and 3(a), (b) where the dashed vertical line, indicating the value of the discrepancy variable for the observed data, and the histogram, showing the values of the discrepancy variable for the replicated data, are in agreement. The histograms in Figures 1(c), 2(c), and 3(c) contain the value zero, once again indicating agreement between the data and the replicated data. Therefore, it is possible to say that the model is well fitted to the data based on the mean, variance and deviance discrepancies.

For the multiplicative binomial regression model, it can be seen that the standardized Bayesian deviance residuals (Figure 1(e)) behave quite well and, as expected, these residuals are around zero. On the other hand, the standardized residuals based on the CPO and the standardized residuals based on the posterior distribution of parameters are not around zero and an almost linear pattern can be observed (Figures 1(d) and (f)). As a consequence of the poor performance of Bayesian deviance residuals, the points in the Bayesian residuals against leverages plots (Figure 2(g)) do not cluster around zero as would be expected for simulated

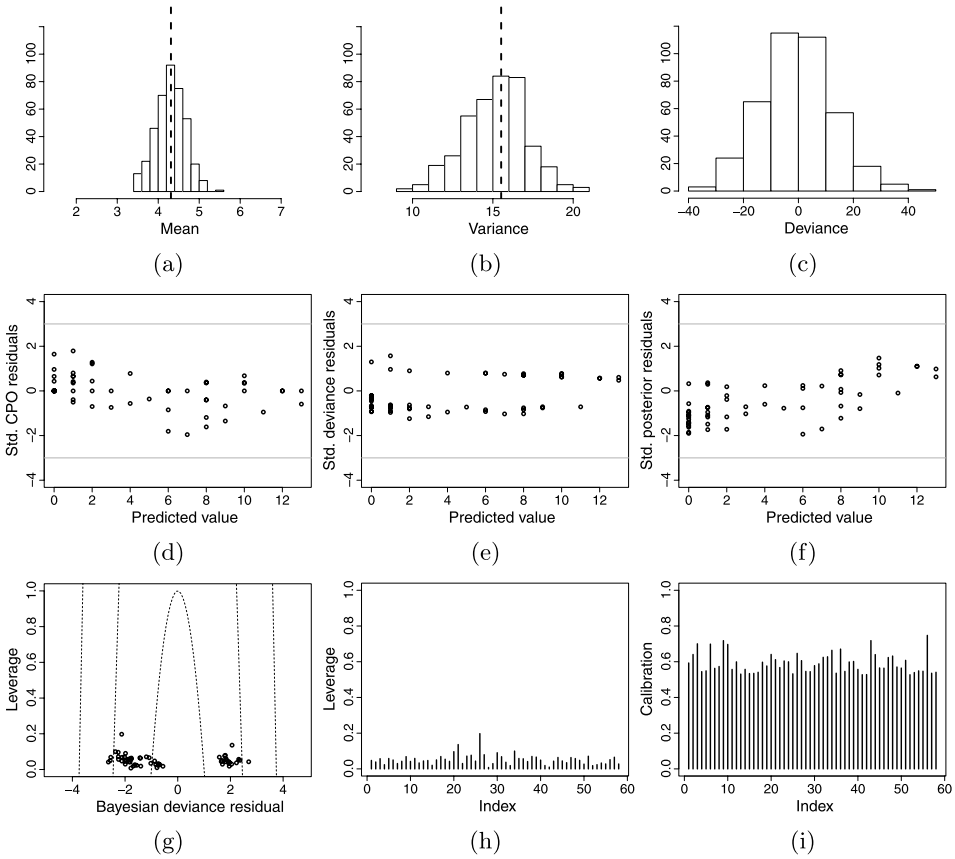

Figure 1 *Simulated data*: *diagnostics for the multiplicative binomial regression model with complementary log-log link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *posterior predictive check for the mean* (a), *for the variance* (b), *and for the model deviance* (c); *standardized residuals based on the CPO* (d); *standardized Bayesian deviance residuals* (e); *standardized residuals based on the posterior distribution of parameters* (f); *Bayesian deviance residuals against leverages* (g); *leverage* (h); *calibration* (i).

data with no perturbation. However, inspection of the leverages and calibration metrics (Figures 1(h), (i)), clearly show that no influential outliers are present in the data sets.

In the double binomial regression model, standardized residuals based on the CPO (Figure 2(d)) do not behave as expected for simulated data sets and it is clear that they are not around zero (Figure 2(d)). On the other hand, the standardized Bayesian deviance residuals and standardized residuals based on the posterior distribution of parameters (Figures 2(e) and (f)) behave as expected and are clustered around zero (Figure 2(f)). It can also be observed that inspection of the leverages and calibration metrics (Figures 2(h) and (i)) do not reveal influential outliers.

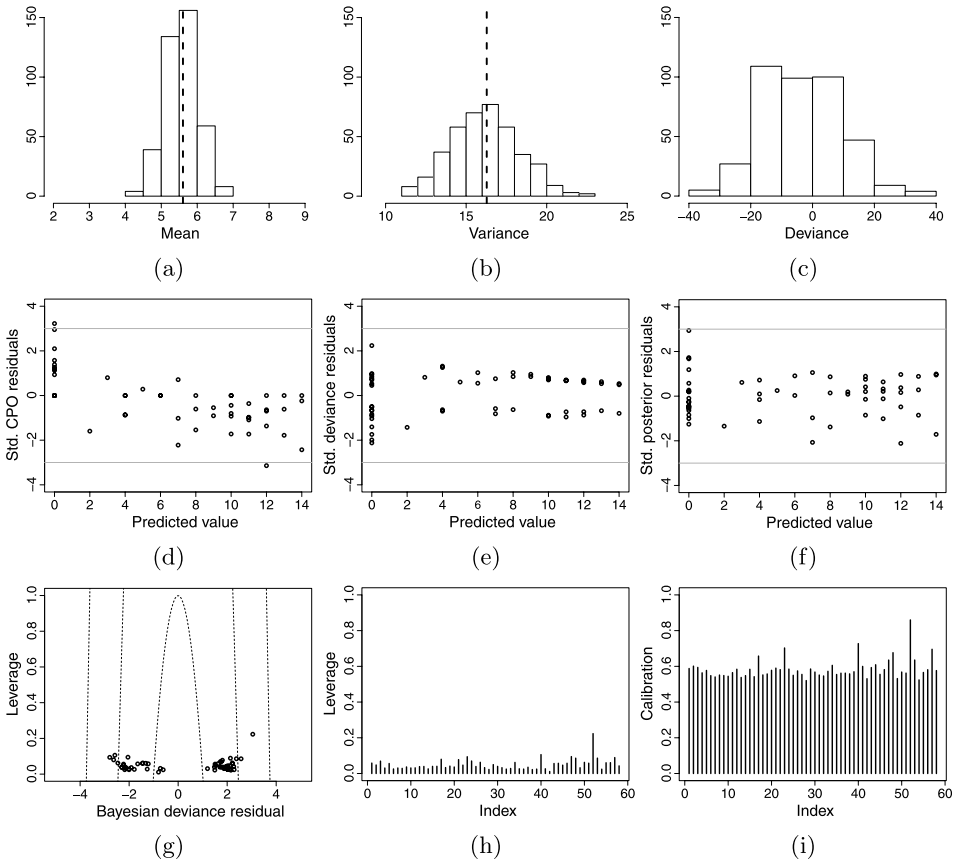

Figure 2 *Simulated data*: *diagnostics for the double binomial regression model with probit link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *posterior predictive check for the mean* (a), *for the variance* (b), *and for the model deviance* (c); *standardized residuals based on the CPO* (d); *standardized Bayesian deviance residuals* (e); *standardized residuals based on the posterior distribution of parameters* (f); *Bayesian deviance residuals against leverages* (g); *leverage* (h); *calibration* (i).

From Figure 3(d), it can be seen that the standardized residuals based on the CPO are not randomly distributed around zero. The standardized Bayesian deviance residuals (Figure 3(e)) and the residuals based on the posterior distribution of parameters (Figure 3(f)) are randomly distributed around zero as would be expected for simulated data. Nevertheless, residuals based on the posterior distribution of parameters show a greater dispersion than the Bayesian deviance residuals. Since the residuals based on the posterior distribution of parameters for the beta-binomial models tend to be quite large when compared to the Bayesian deviance residuals for these same models, we believe that caution should be taken when interpreting these residuals and a large value should not be considered as an outlier unless its Bayesian deviance residual is too indicated as an outlier.

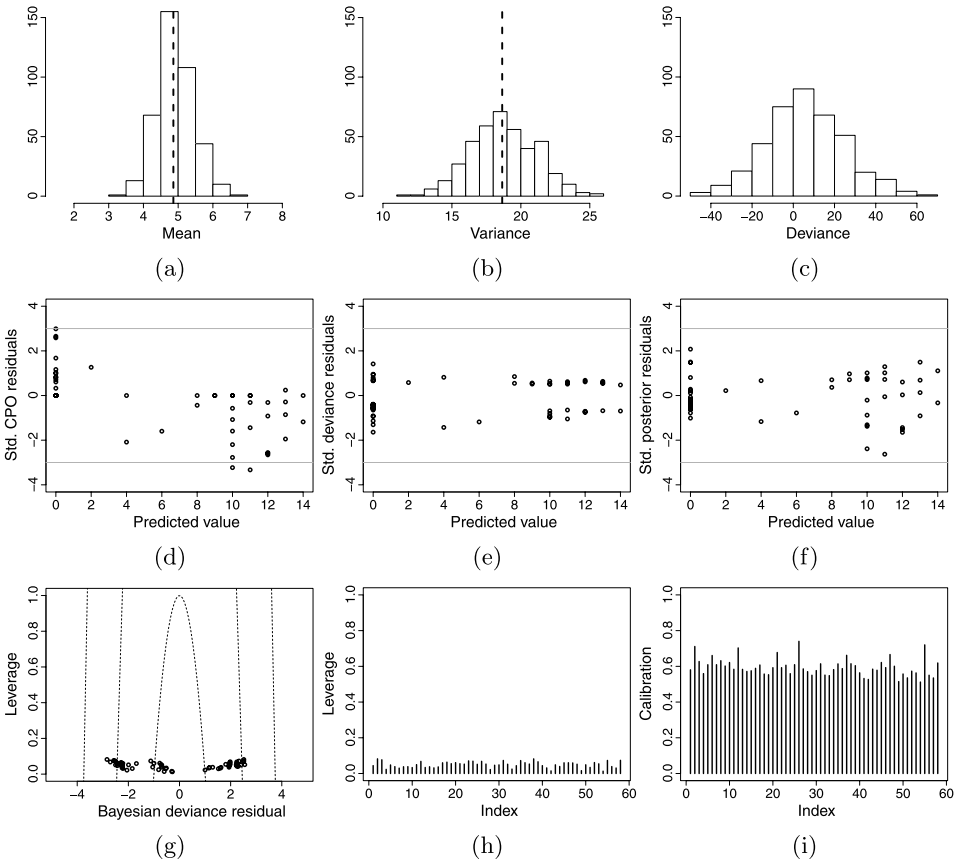

Figure 3 *Simulated data*: *diagnostics for the beta-binomial regression model with logit link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *posterior predictive check for the mean* (a), *for the variance* (b), *and for the model deviance* (c); *standardized residuals based on the CPO* (d); *standardized Bayesian deviance residuals* (e); *standardized residuals based on the posterior distribution of parameters* (f); *Bayesian deviance residuals against leverages* (g); *leverage* (h); *calibration* (i).

4.3 Detecting overdispersion through posterior predictive checks

To illustrate how posterior predictive checks can be used to detect overdispersion in a data set while performing a goodness of fit test, a beta-binomial data set with complementary log-log link for p and log link for γ was generated in the settings described in Section 4. Then, the usual binomial regression model with complementary log-log link was fitted to this simulated data using the Bayesian methodology described in Section 3. Next, we computed the posterior predictive checks discussed in Section 3.2. This same procedure was repeated to a few data sets and for the other regression models discussed in this paper as well and similar results were obtained.

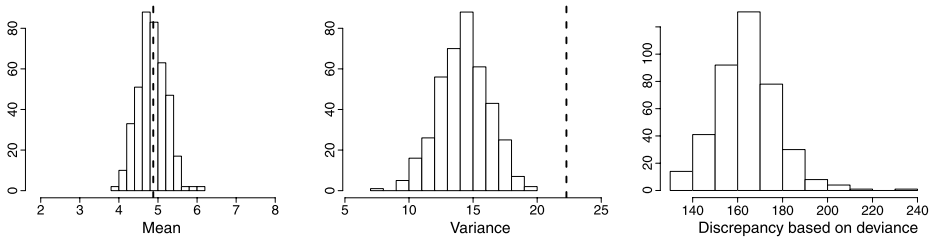

Figure 4 *Simulated data*: *posterior predictive checks for the usual binomial regression model with complementary log-log link for* p: *posterior predictive check for the mean* (a), *for the variance* (b), *and for the model deviance* (c).

Figure 4(a) reveals that the mean of the replicated data sets obtained by the fitted model are in agreement with the mean of the simulated data, which would be expected since both models have the same mean. This is also supported by the estimated posterior predictive p-value of 0.45 obtained for the discrepancy based on the mean. However, Figure 4(b) shows that the variance of the replicated data sets do not coincide with the variance of the simulated data and the estimated posterior predictive p-value was zero, which suggests that the pattern presented by the replicated data is different from the one observed for the simulated data thus, indicating lack of fit. Moreover, it can be seem that variance in the replicated data is considerably smaller than the variance of the simulated data, thus indicating overdispersion. Similar conclusions can be taken from the discrepancy based on the model deviance, whose histogram in Figure 4(c) do not contain the zero and are quite large. The estimated posterior predictive p-value was zero, which indicates lack of fit.

4.4 Comparison between the logit and probit link under overdispersed binomial models

It was pointed out by one referee that for usual binomial data, Chambers and Cox (1967) showed that both the logit and probit link functions provide quite similar results. Therefore, we have conducted a simulation study to compare the results provided by these two links when applied to overdispersed binomial data.

The simulation study consisted on generating data sets considering the logit and probit link combined with each of the three models for overdispersed binomial data, for example, the multiplicative binomial, the double binomial and the beta-binomial model. After the data was generated, both models were fitted to it and model selection criteria were used to select among them. The comparison was performed for models within the same class. Since the model selection criteria considered are known to perform well, the idea of this procedure was to see if the true model would be preferred for most of the data sets and, if so, we would be

Table 3 *Comparison between the logit and probit link*: *proportion of samples for which the true model was selected*

	Bayes factor H_0 : true model					
Simulated	<3	3 to 20	20 to 150	>150	log-CPO	DIC
MBL	0.59	0.02	0.00	0.39	0.42	0.48
MBP	0.47	0.01	0.00	0.52	0.52	0.57
DBL	0.37	0.01	0.00	0.62	0.60	0.60
DBP	0.48	0.00	0.00	0.52	0.51	0.53
BBL	0.48	0.00	0.00	0.52	0.57	0.53
BBP	0.33	0.00	0.01	0.66	0.60	0.62

able to conclude that the link provided different results. On the other hand, if the model selection criteria would fail to select the true model most of the time, then we would be able to conclude that their no evidence that the two link functions provided different results. We considered the model selection criteria described in Section 3.4.

Results based on 100 samples simulated in the same settings described in Section 4.1 are shown in Table 3, where MBL and MBP stand for the multiplicative binomial regression models with logit and probit link, DBL and DBP represent the double binomial regression models with logit and probit link, and BBL and BBP are the beta-binomial regression models with logit and probit link.

From Table 3 it can be observed that both the sum of log-CPO and DIC provided similar results. Furthermore, the highest proportion of samples for which the true model was selected is 0.62, and therefore this two model selection criteria indicates that the two links are similar. Moreover, the proportion of samples for which the Bayes factor (Table 3) showed very little evidence against the null hypothesis is quite high with the smallest being above 30%, thus indicating no difference between the results provided by the two links. Therefore, we conclude that the results provided by the logit and probit link may be regarded as non different since all the model selection criteria used did not show the true model as the preferred one.

5 Low-iron rat teratology data

Teratology studies are concerned with the understanding of abnormalities of physiological development. In a teratology study, it is usual to expose pregnant animals to environmental agents and check the fetuses for abnormalities. In the low-iron rat teratology data, the aim is to study the effects of iron deficient diets in rats' litters. Iron deficiency is known to be one of the most usual nutritional problems and is more frequent among pregnant females (Cook et al., 1994). The original low-iron rat teratology data set is presented in Shepard et al. (1980). In Moore and Tsiatis

Table 4 *Low-iron rat teratology data*

Group	x	**n**	**y**	Group	x	**n**	**y**	Group	x	**n**	**y**
1	4.1	10	1	1	4.8	14	5	2	8.9	4	0
1	3.2	11	4	1	6.7	10	10	2	11.1	1	0
1	4.7	12	9	1	5.2	12	10	2	9	12	0
1	3.5	4	4	1	4.3	13	8	3	11.2	8	0
1	3.2	10	10	1	3.9	10	10	3	11.5	11	1
1	5.9	11	9	1	6.3	14	3	3	12.6	14	0
1	4.7	9	9	1	4.4	13	13	3	9.5	14	1
1	4.7	11	11	1	5.2	4	3	3	9.8	11	0
1	3.5	10	10	1	3.9	8	8	4	16.6	3	0
1	4.8	10	7	1	7.7	13	5	4	14.5	13	0
1	4.3	12	12	1	5	12	12	4	15.4	9	2
1	4.1	10	9	2	8.6	10	1	4	14.5	17	2
1	3.2	8	8	2	11.1	3	1	4	14.6	15	0
1	6.3	11	9	2	7.2	13	1	4	16.5	2	0
1	4.3	6	4	2	8.8	12	0	4	14.8	14	1
1	3.1	9	7	2	9.3	14	4	4	13.6	8	0
1	3.6	14	14	2	9.3	9	2	4	14.5	6	0
1	4.1	12	7	2	8.5	13	2	4	12.4	17	0
1	4.8	11	9	2	9.4	16	1				
1	4.7	13	8	2	6.9	11	0				

(1991), the data set is used to illustrate a method of moments estimation procedure with a variance correction factor when there is presence of extra binomial variation. The data is also presented in Agresti (2002) (page 152, Table 4.5) to illustrate the presence of overdispersion in binomial data.

In the low-iron rat teratology data set (Table 4), $m = 58$ female rats were divided into four groups and given iron-deficient diets. Group 1 received a placebo injection, group 2 received iron-supplement injections on days 7 and 10, group 3 received injections on days 0 and 7, and group 4 received iron-injections weekly. The rats were made pregnant and sacrificed 3 weeks later, and the total number of fetuses, n_i, the number of dead fetuses, y_i, in each litter $i = 1, \ldots, m$, and a covariate, x_i, accounting for the hemoglobin level of the rats were computed. The number of dead fetuses, y_i, in each litter of size n_i may be treated as a binomial data, nevertheless as argued in Agresti (2002), in teratology experiments genetic variability and unobserved covariates may cause the probability of death to vary from litter to litter within a particular treatment group, leading to extra-binomial variability.

We consider two models for the teratology data set: the hemoglobin model which is given by $\eta_{1i} = \beta_0 + \beta_1 x_{1i}$ as the linear predictor of p_i and $\eta_{2i} = \alpha_1 n_i$ as the linear predictor of γ_i, $i = 1, \ldots, m$; and the treatment model which is given by $\eta_{1i} = \beta_0 + \beta_2 t_{2i} + \beta_3 t_{3i} + \beta_4 t_{4i}$ as the linear predictor of p_i and $\eta_{2i} = \alpha_1 n_i$ as

the linear predictor of γ_i, $i = 1, \ldots, m$, with t_2, t_3, and t_4 being the dummy variables of groups 2, 3, and 4 respectively. Thus, both the hemoglobin level covariate and the means for each treatment group are used to model the probability of success of the responses. We notice that these choices of linear predictors for $\mathbf{p}$ and $\boldsymbol{\gamma}$ are not final, as other regression structures could possibly be considered. We also point out that this teratology data set has been widely studied and different statistical model formulation have been proposed for it (e.g. Moore and Tsiatis, 1991; Agresti, 2002). Moreover, according to Moore and Tsiatis (1991), the choice of a specific model depends on the biological questions of interest.

For the hemoglobin model, we assume a non informative normal prior $N_3(\mathbf{0}, 10^4 I_3)$ for the unknown parameter vector $\boldsymbol{\theta} = (\beta_0, \beta_1, \alpha_1)$. Similarly, for the treatment model, a non informative normal prior $N_5(\mathbf{0}, 10^4 I_5)$ is assumed for the unknown parameter vector $\boldsymbol{\theta} = (\beta_0, \beta_2, \beta_3, \beta_4, \alpha_1)$. In both cases, MCMC samples of the posterior distributions were obtained using a Gibbs–Metropolis type algorithm with candidates generated by random walk from a multivariate normal distribution with covariance matrix given by minus the Hessian matrix evaluated at the maximum likelihood estimator of $\boldsymbol{\theta}$. For each model, chain sizes were set to 100,000 with burn-in periods of 20,000 and a thinning interval of size 200. Rejection rates were 0.55 for the multiplicative binomial, double binomial and beta-binomial regression models fitted with the hemoglobin linear predictor for $\mathbf{p}$ and 0.70 for the multiplicative binomial, double binomial and beta-binomial regression models fitted with the treatment linear predictor for $\mathbf{p}$. Rejection rates of usual binomial regression models were 0.45 and 0.62 with the hemoglobin linear predictor and treatment linear predictor, respectively. Convergence was checked using Geweke's criterion (Geweke, 1992).

Table 5 shows values of DIC and log-CPO obtained for the multiplicative binomial, double binomial, and beta-binomial regression models with complementary log-log, logit, and probit links for $\mathbf{p}$, and log link for $\boldsymbol{\gamma}$ and linear predictors as defined for the hemoglobin model. We also present DIC and log-CPO values for usual binomial regression models. It can be seen that, based on these two criteria, the model to be selected is the beta-binomial hemoglobin regression model with complementary log-log link function for $\mathbf{p}$. The Bayes factor model selection criteria presented in Table 6 also shows the beta-binomial hemoglobin model with complementary log-log link for $\mathbf{p}$ as the most favorable model.

In Table 5, values of p_D show the effective number of parameters to be in agreement with the effective number of parameters expected for each hemoglobin model fitted to the teratology data.

We also computed estimated posterior predictive p-values (3.2) for discrepancies based on the mean and variance of the response variable and model deviance for each regression model. The obtained values are shown in Table 5. For multiplicative binomial regression models MBC, MBL and MBP the computed $\hat{p}$-value for the discrepancy based on model deviance are all equal to zero. Double binomial regression models DBC, DBL and DBP show considerably small values of

Table 5 *Low-iron rat teratology data—hemoglobin model*: *DIC*, *log-CPO*, *effective number of parameters* (p_D) *and estimated posterior predictive p-value obtained of candidate models*

				$\hat{p}$-value		
Model	DIC	log-CPO	p_D	Mean	Variance	Deviance
MBC	232.06	−117.92	3.07	0.48	0.35	0
MBL	232.84	−118.28	3.12	0.5	0.4	0
MBP	233.45	−118.45	3.15	0.53	0.42	0
DBC	224.18	−112.60	2.86	0.81	0.02	0.73
DBL	228.50	−114.29	3.03	0.81	0.01	0.77
DBP	231.37	−115.59	2.96	0.86	0.01	0.85
BBC	201.63	−101.04	3.20	0.7	0.55	0.45
BBL	204.20	−102.41	3.08	0.74	0.52	0.46
BBP	207.20	−104.02	3.30	0.75	0.56	0.54
BINC	280.21	−143.75	2.15	0.48	0	0
BINL	281.45	−143.94	2.12	0.55	0	0
BINP	292.23	−149.97	2.02	0.66	0	0

Table 6 *Low-iron rat teratology data—hemoglobin model*: *Bayes factor values of candidate models*

H1\H0	MBC	MBL	MBP	DBC	DBL	DBP	BBC	BBL	BBP	BINC	BINL	BINP
MBC	–	1.38	1.95	0.02	0.18	0.77	0.00	0.00	0.00	>150	>150	>150
MBL	0.72	–	1.41	0.02	0.13	0.55	0.00	0.00	0.00	>150	>150	>150
MBP	0.51	0.71	–	0.01	0.09	0.39	0.00	0.00	0.00	>150	>150	>150
DBC	43.56	60.27	85.01	–	7.71	33.39	0.00	0.00	0.00	>150	>150	>150
DBL	5.65	7.82	11.03	0.13	–	4.33	0.00	0.00	0.00	>150	>150	>150
DBP	1.30	1.81	2.55	0.03	0.23	–	0.00	0.00	0.00	>150	>150	>150
BBC	>150	>150	>150	>150	>150	>150	–	4.11	14.97	>150	>150	>150
BBL	>150	>150	>150	>150	>150	>150	0.24	–	3.65	>150	>150	>150
BBP	>150	>150	>150	>150	>150	>150	0.07	0.27	–	>150	>150	>150
BINC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–	1.96	>150
BINL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.51	–	>150
BINP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–

$\hat{p}$-value for the discrepancy based on the variance of the responses. The usual binomial regression models, BINC, BINL and BINP present extreme small values of $\hat{p}$-value for both the variance discrepancy and the model deviance discrepancy. On the other hand, for the beta-binomial models BBC, BBL and BBP no extreme values of $\hat{p}$-value for the mean, variance and deviance discrepancy are observed. Extreme $\hat{p}$-values indicate that the pattern of the observed data is different of the pattern of the replicated data, i.e., the pattern of the observed data is not likely to be seen under the assumed model. Therefore, based on estimated posterior predictive p-values, it could be argued that the beta-binomial regression models seem to be

Table 7 *Low-iron rat teratology data—hemoglobin model*: *posterior model summaries for the beta-binomial regression model with complementary log-log link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$

Model	Parameter	Mean	Median	Std. dev.	95% C.I.		95% HPD	
BBC	β_0	2.00	2.02	0.12	1.31	2.62	1.34	2.65
	β_1	−0.37	−0.37	0.00	−0.47	−0.25	−0.47	−0.25
	α_1	−0.07	−0.07	0.00	−0.13	−0.01	−0.13	−0.01

a more plausible choice for the data set, as no extreme $\hat{p}$-values were observed for these models, which indicates that the pattern of the observed data is more likely to occur under these models.

Based on the considered model selection criteria, on the p_D values and on the estimated posterior predictive p-values, we present the posterior model fit summaries for the beta-binomial hemoglobin model with complementary log-log link for $\mathbf{p}$ and log link for $\boldsymbol{\gamma}$ in Table 7.

Bayesian diagnostic analysis for the selected hemoglobin model is shown in Figure 5. The posterior predictive check histogram in Figures 5(a)–(c) show the discrepancies based on the mean, variance and deviance as indicating that the beta-binomial hemoglobin regression model with complementary log-log link for p_i, log link for γ_i and linear predictors $\eta_{1i} = \beta_0 + \beta_1 x_i$ and $\eta_{2i} = \alpha_1 n_i$, $i = 1, \ldots, m$, provides a good fit for the low-iron rat teratology data. The vertical dashed line in Figures 5(a) and (b) indicate that the mean and variance of the observed data coincide with the peak of the mean and variance histogram obtained for the data replicated from the fitted model. The histogram for the model deviance in Figure 5(c) also reveals that the replicated data is in agreement with the observed data since the value zero is contained in it. Estimated posterior predictive (Table 5) p-values for the discrepancies were 0.70, 0.55 and 0.45 for mean, variance and model deviance, respectively. Thus, posterior predictive checks suggest that the beta-binomial hemoglobin regression model with complementary log-log link for $\mathbf{p}$ and log link for $\boldsymbol{\gamma}$ provides a good fit for the low-iron rat teratology data.

The standardized Bayesian deviance residuals shown in Figure 5(d) reveal observation 51 as an outlier. On the other hand, standardized residuals based on the posterior distribution of parameters (Figure 5(e)) suggests observation 2 as an outlier. Nevertheless, neither the leverages (Figure 5(g)) nor the calibration (Figure 5(h)) uncover these observations as being influential. Moreover, both residuals are randomly clustered on zero indicating that the model is well fitted to the data.

Figure 6 presents the standardized deviance residuals and the standardized residuals based on the posterior distribution against the hemoglobin level covariate. From Figures 6(a) and (b), we notice that there is no pattern and the residuals are clustered around zero. Thus, there is no evidence that a high order term for the hemoglobin level covariate should be introduced in the model.

Figure 7(a) presents the estimated probability of death $\hat{p}_i$ in each litter against the observed hemoglobin level of the mothers. It can be seen that, as the rat's

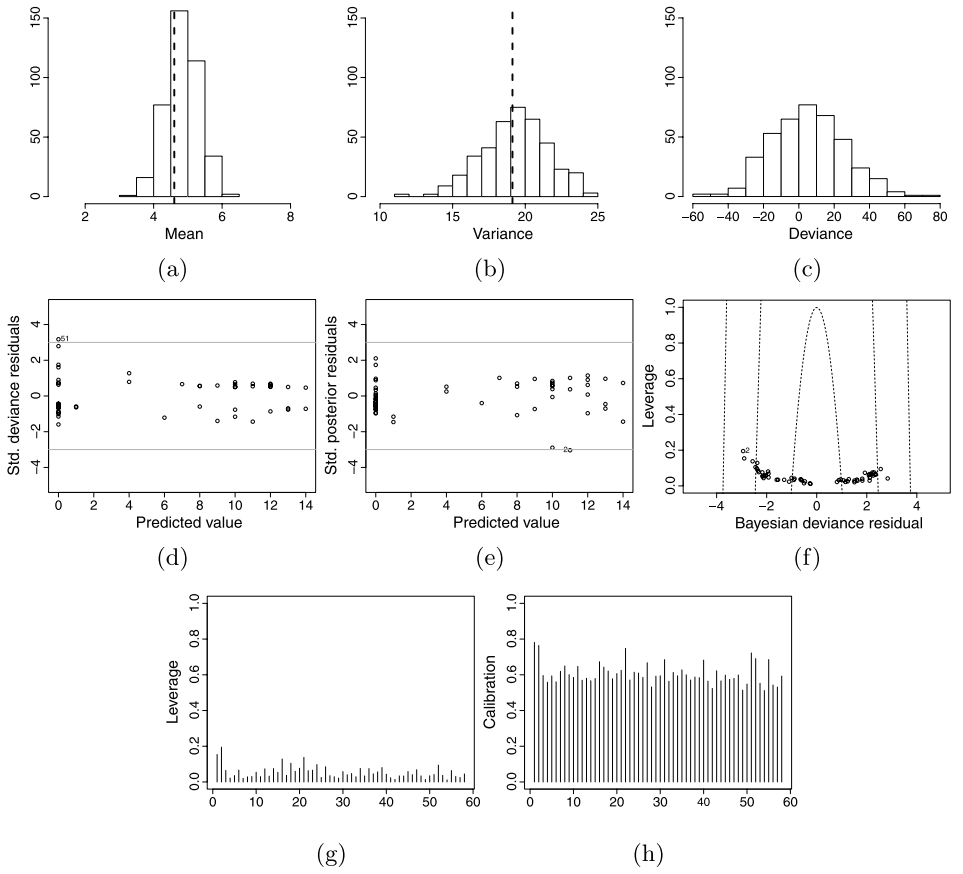

Figure 5 *Low-iron rat teratology data—hemoglobin model*: *diagnostics for the beta-binomial regression model with complementary log-log link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *posterior predictive check for the mean* (a), *for the variance* (b), *and for the model deviance* (c); *standardized Bayesian deviance residuals* (d); *standardized residuals based on the posterior distribution of parameters* (e); *Bayesian deviance residuals against leverages* (f); *leverage* (g); *calibration* (h).

hemoglobin level increases, the estimated probability of death fetuses decreases. In addition, the hemoglobin level seems to be related to the iron deficient diet. Figure 7(b) shows the rats in the placebo group (group 1) as the ones with lower hemoglobin levels, while rats in the control group (group 4), whose iron intake was normal, are the ones with higher levels of hemoglobin.

Table 8 shows the values of DIC and log-CPO obtained for the multiplicative binomial, double binomial, beta-binomial, and usual binomial regression models with complementary log-log, logit, and probit links for $\mathbf{p}$, and log link for $\boldsymbol{\gamma}$ and linear predictors as defined for the treatment model. Based on these two criteria, the selected model is the beta-binomial treatment regression model with logit link function for p. The Bayes factor model selection criteria presented in Table 9

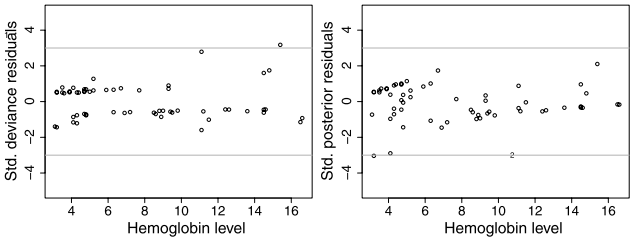

Figure 6 *Low-iron rat teratology data—hemoglobin model*: *diagnostics for the beta-binomial regression model with complementary log-log link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *standardized Bayesian deviance residuals against hemoglobin level* (a); *standardized residuals based on the posterior distribution of parameters against hemoglobin level* (b).

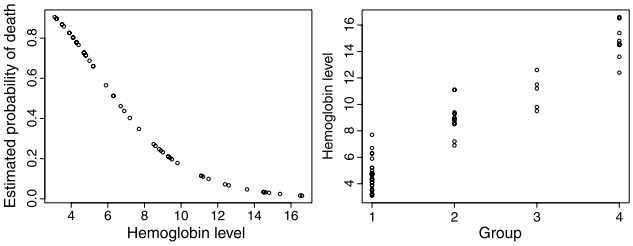

Figure 7 *Low-iron rat teratology data—hemoglobin model*: *estimated probability of death against hemoglobin level* (a); *hemoglobin level against treatment group* (b).

Table 8 *Low-iron rat teratology data—treatment model*: *DIC*, *log-CPO*, *effective number of parameters* (p_D) *and estimated posterior predictive p-value obtained of candidate models*

				$\hat{p}$-value		
Model	DIC	log-CPO	p_D	Mean	Variance	Deviance
MBC	230.95	−118.82	4.84	0.51	0.49	0
MBL	231.29	−118.90	5.06	0.51	0.5	0
MBP	230.58	−118.02	4.83	0.55	0.54	0
DBC	211.53	−105.44	4.69	0.79	0.2	0.53
DBL	211.66	−105.37	4.75	0.84	0.17	0.52
DBP	211.12	−104.97	4.47	0.83	0.17	0.59
BBC	197.35	−99.04	4.89	0.78	0.7	0.44
BBL	197.28	−98.82	4.90	0.8	0.74	0.44
BBP	197.81	−99.01	5.21	0.8	0.77	0.44
BINC	253.09	−129.67	3.99	0.47	0.08	0
BINL	253.75	−130.49	4.33	0.5	0.1	0
BINP	253.17	−129.17	4.09	0.52	0.12	0

Table 9 *Low-iron rat teratology data—treatment model*: *Bayes factor values of candidate models*

H1\H0	MBC	MBL	MBP	DBC	DBL	DBP	BBC	BBL	BBP	BINC	BINL	BINP
MBC	–	0.94	0.86	0.00	0.00	0.00	0.00	0.00	0.00	>150	>150	>150
MBL	1.07	–	0.92	0.00	0.00	0.00	0.00	0.00	0.00	>150	>150	>150
MBP	1.16	1.09	–	0.00	0.00	0.00	0.00	0.00	0.00	>150	>150	>150
DBC	>150	>150	>150	–	1.04	1.00	0.00	0.00	0.00	>150	>150	>150
DBL	>150	>150	>150	0.96	–	0.96	0.00	0.00	0.00	>150	>150	>150
DBP	>150	>150	>150	1.00	1.04	–	0.00	0.00	0.00	>150	>150	>150
BBC	>150	>150	>150	>150	>150	>150	–	1.04	1.07	>150	>150	>150
BBL	>150	>150	>150	>150	>150	>150	0.96	–	1.04	>150	>150	>150
BBP	>150	>150	>150	>150	>150	>150	0.93	0.97	–	>150	>150	>150
BINC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–	1.15	1.04
BINL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.87	–	0.91
BINP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	1.10	–

also shows the beta-binomial model with logit link for $\mathbf{p}$ as the most favorable model. Moreover, values of p_D presented in Table 8 show the effective number of parameters to be in agreement with the effective number of parameters expected for each treatment model fitted to the teratology data.

Estimate posterior predictive p-values (Table 8) obtained for treatment models are quite similar to those obtained for hemoglobin models. Computed $\hat{p}$-value for the discrepancy based on model deviance are all equal to zero for multiplicative binomial models, whereas double binomial regression models show small values of $\hat{p}$-value for the discrepancy based on the variance of the responses, and the usual binomial regression models present extreme small values of $\hat{p}$-value for both the variance discrepancy and the model deviance discrepancy. On the other hand, for the beta-binomial models no extreme values of $\hat{p}$-value for the mean, variance and deviance discrepancy are observed.

Based on the considered model selection criteria, on the p_D values and on the estimated posterior predictive p-values, we present the posterior model fit summaries for the beta-binomial treatment model with logit link for $\mathbf{p}$ and log link for $\boldsymbol{\gamma}$ in Table 10.

Bayesian diagnostic analysis for the selected treatment model is shown in Figure 8. The posterior predictive check histogram in Figures 8(a)–(c) show the discrepancies based on the mean, variance and deviance as indicating that the beta-binomial treatment regression model with logit link for p_i, log link for γ_i and linear predictors $\eta_{1i} = \beta_0 + \beta_2 t_{2i} + \beta_3 t_{3i} + \beta_4 t_{4i}$ and $\eta_{2i} = \alpha_1 n_i$, $i = 1, \ldots, m$, provides a good fit for the low-iron rat teratology data. The vertical dashed line in Figures 8(a) and (b) indicate that the mean and variance of the observed data coincide with the peak of the mean and variance histogram obtained for the data replicated from the fitted model. The histogram for the model deviance in Figure 8(c) also reveals that the replicated data is in agreement with the observed data since the value zero is

Table 10 *Low-iron rat teratology data—treatment model*: *posterior model summaries for the beta-binomial regression model with logit link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$

Model	Parameter	Mean	Median	Std. dev.	95% C.I.		95% HPD	
BBL	β_0	1.27	1.27	0.06	0.79	1.73	0.82	1.74
	β_2	−3.06	−3.06	0.25	−4.06	−2.12	−4.02	−2.11
	β_3	−4.01	−3.94	0.81	−5.84	−2.56	−5.76	−2.49
	β_4	−3.95	−3.93	0.49	−5.44	−2.69	−5.45	−2.69
	α_1	−0.09	−0.09	0.00	−0.15	−0.04	−0.15	−0.04

contained in it. The estimated posterior predictive p-values (Table 8) for the discrepancies were 0.8, 0.74 and 0.44; therefore, posterior predictive checks suggest that the beta-binomial treatment regression model with logit link for $\mathbf{p}$ and log link for $\boldsymbol{\gamma}$ provides a good fit for the low-iron rat teratology data.

Both the standardized Bayesian deviance residuals and the standardized residuals based on the posterior distribution of parameters, shown in Figures 8(d) and (e), respectively, reveal no observation as an outlier. Both residuals are randomly clustered on zero indicating that the model provides a good fit for the data. Moreover, no observation is shown as influent by the leverages (Figure 8(g)) or by the calibration metric (Figure 8(h)).

Figure 9 presents the standardized deviance residuals and the standardized residuals based on the posterior distribution against treatment group. From both Figures 9(a) and (b) no patterns are observed, therefore indicating no evidence of departure from the linear assumption.

Figure 10 presents the estimated probability of death $\hat{p}_i$ in each litter against treatment group. It can be seen that the estimated probability of death is greater in groups with low iron supplement.

6 Conclusions

In Section 4.1 we presented results based on simulated data sets for each regression model considered. It could be seen that the Bayesian methodology worked quite well, providing accurate estimates for the parameters. We also would like to stress that, contrary to frequentist methods, Bayesian inference procedures do not need to be modified or adapted in the presence of a normalizing constant which, for instance, enables us to consider the normalized probability distribution of the double binomial model. As a consequence of considering such a normalized form, parameter estimates of the double binomial regression models are considerably improved.

In Section 4.2, application of different diagnostic techniques provided a guide or indicative of which ones may be considered in practical situations where the

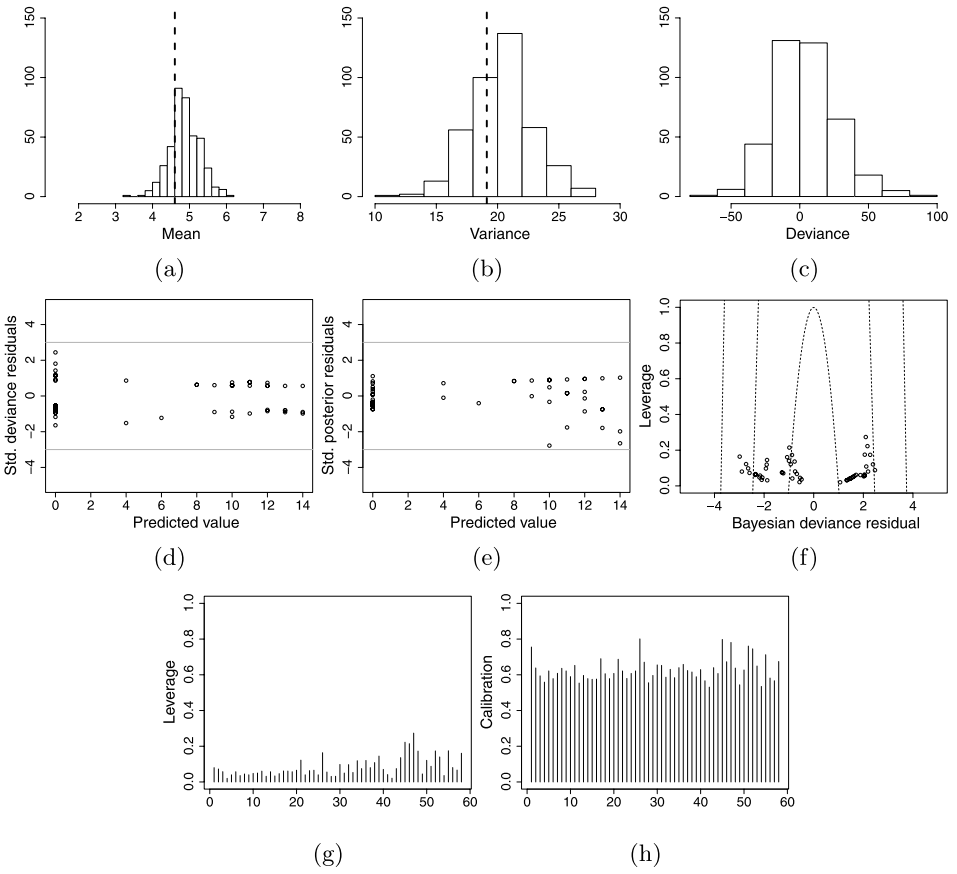

Figure 8 *Low-iron rat teratology data—treatment model*: *diagnostics for the beta-binomial regression model with logit link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *posterior predictive check for the mean* (a), *for the variance* (b), *and for the model deviance* (c); *standardized Bayesian deviance residuals* (d); *standardized residuals based on the posterior distribution of parameters* (e); *Bayesian deviance residuals against leverages* (f); *leverage* (g); *calibration* (h).

multiplicative, the double binomial or the beta-binomial regression models are fitted to the data. Since it is usual to consider more than one metric or graphical display for model assessment, we have tried to approach as many methods as feasible though it is possible that some other interesting ideas for model check and diagnostic may have been overlooked. From the results presented in that section, there are indications that residuals based on the CPO and residuals based on the posterior distribution of parameters do not seem to be accurate when applied to the multiplicative binomial regression models, the residuals based on the CPO and Bayesian deviance residuals do not seem accurate when applied to the double binomial regression models, and the residuals based on the CPO do not seem ap-

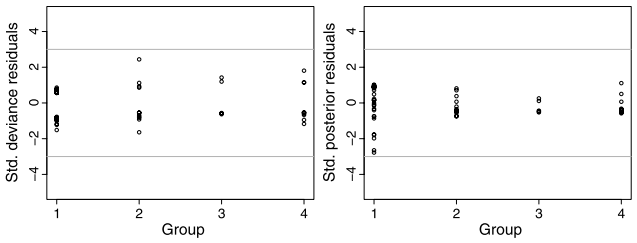

Figure 9 *Low-iron rat teratology data—treatment model*: *diagnostics for the beta-binomial regression model with logit link for* $\mathbf{p}$ *and log link for* $\boldsymbol{\gamma}$: *standardized Bayesian deviance residuals against hemoglobin level* (a); *standardized residuals based on the posterior distribution of parameters against hemoglobin level* (b).

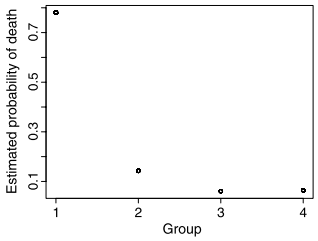

Figure 10 *Low-iron rat teratology data—treatment model*: *estimated probability of death against treatment group*.

propriate for the beta-binomial regression model. Even so, we note that the reason why these problems occurred should be investigated further.

In Section 5, both the hemoglobin level covariate and the means for each diet group were used to model the probability of dead fetuses. The analysis presented indicated that the probability of death of fetuses can be related to the mother's hemoglobin level and to the diet group, thus indicating that a iron-deficient diet has a negative effect on the litters. Once more, we highlight that the choice of linear predictors for $\mathbf{p}$ and $\boldsymbol{\gamma}$ are not final and other regression structures could possibly be considered. Moreover, both models tested were shown to be good fits for the low-iron teratology data and to choose a specific model we would have to rely on the biological questions of interest.

In summary, we have illustrated Bayesian estimation and Bayesian diagnostic for the multiplicative binomial, double binomial and beta-binomial regression models. The proposed methodology aims to explore techniques of Bayesian inference and diagnostics developed by several authors and to draw attention to the fact that a more detailed analysis of residuals is crucial for a plausible model specification. Moreover, considering these models for overdispersed binomial data analysis provide an alternative approach that would be useful in practical problems.

Appendix: Algorithms

A.1 Algorithm to compute predicted values based on the CPO

Given a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$:

1. For each $\boldsymbol{\theta}_j$, $j = 1, \ldots, r$, compute $p_{ij} = g^{-1}(\mathbf{x}'_{1i}\boldsymbol{\beta}_j)$ and $\gamma_{ij} = h^{-1}(\mathbf{x}'_{2i}\boldsymbol{\beta}_j)$.
2. For each $\tilde{y}_i$ in $\{0, \ldots, n_i\}$, $i = 1, \ldots, m$, compute the Monte Carlo estimate of CPO_i by $\widehat{CPO_i} = [\frac{1}{r}\sum_{j=1}^{r} \frac{1}{P(Y_i|p_{ij},\gamma_{ij})}]^{-1}$.
3. Set as the predicted value of y_i the value $\tilde{y}_i$ in $\{0, \ldots, n_i\}$ which maximizes $\widehat{CPO_i}$, $i = 1, \ldots, m$.

A.2 Algorithm to compute posterior predictive checks

Given a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$:

1. For each $\boldsymbol{\theta}_j$, $j = 1, \ldots, r$, sample $\mathbf{y}_j^{\text{rep}}$ from $P(Y|\boldsymbol{\theta}_j)$.
2. Compute the discrepancy variable, $T(\mathbf{y})$, for the observed data.
3. For each $\mathbf{y}_j^{\text{rep}}$, $j = 1, \ldots, r$, compute the discrepancy variable $T(\mathbf{y}_j^{\text{rep}})$.

A.3 Algorithm to compute a sample of the residuals based on the posterior distribution of parameters

Given a MCMC sample of size r of $\pi(\boldsymbol{\theta}|D)$:

1. For each $\boldsymbol{\theta}_j$, $j = 1, \ldots, r$, compute $p_{ij} = g^{-1}(\mathbf{x}'_{1i}\boldsymbol{\beta}_j)$ and $\gamma_{ij} = h^{-1}(\mathbf{x}'_{2i}\boldsymbol{\alpha}_j)$, $i = 1, \ldots, m$.
2. For $i = 1, \ldots, m$, compute (3.3) to obtain a sample of size r of the residuals.

Acknowledgments

The research of Carolina C. M. Paraíba and Rubiane M. Pires are supported by the Brazilian organization CAPES.

The authors would like to thank four anonymous referees for their helpful comments and suggestions which greatly improved this paper.

Algorithms are available upon request to Carolina C. M. Paraíba.

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed. *Wiley Series in Probability and Statistics*. New York: Wiley-Interscience. MR1914507

Aitkin, M. (1991). Posterior Bayes factor (with discussion). *Journal of the Royal Statistical Society, Series B* **53**, 111–142.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679. MR1224394

Albert, J. H. and Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747–759. MR1380812

Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society, Series C* **27**, 162–167. MR0506839

Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*, 3rd ed. Boca Raton: Chapman & Hall/CRC. MR2442364

Chambers, E. A. and Cox, D. R. (1967). Discrimination between alternative binary response models. *Biometrika* **54**, 573–578. MR0223058

Cho, H., Ibrahim, J. G., Sinha, D. and Zhu, H. (2009). Bayesian case in influence diagnostics for survival models. *Biometrics* **65**, 116–124. MR2665851

Cook, J. D., Skikne, B. S. and Baynes, R. D. (1994). Iron deficiency: The global perspective. *Advances in Experimental Medicine and Biology* **356**, 219–228.

Cottet, R., Kohn, R. J. and Nott, D. J. (2008). Variable selection and model averaging in semiparametric overdispersed generalized linear models. *Journal of the American Statistical Association* **103**, 661–671. MR2524000

Dey, K., Gelfand, A. E. and Peng, F. (1997). Overdispersed generalized linear models. *Journal of Statistical Planning and Inference* **64**, 93–107.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**, 709–721. MR0860505

Gelfand, A. E., Dey, D. and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *Bayesian Statistics 4* (Bernardo, J. M., Berger, J., Dawid, A. P. and Smith, J. F. M., eds.) 147–167. Oxford: Oxford Univ. Press. MR1380275

Gelman, A., Meng, X. L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**, 733–807. MR1422404

Gelman, A., Goegebeur, Y., Tuerlinckx, F. and Van Mechelen, I. (2000). Diagnostic checks for discrete data regression models using posterior predictive simulations. *Journal of the Royal Statistical Society, Series C* **49**, 247–268. MR1821324

Gelman, A., Carlin, J., Stern, H. and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd ed. *Texts in Statistical Science*. London: Chapman & Hall. MR1385925

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (Bernardo, J. M., Berger, J., Dawid, A. P. and Smith, J. F. M., eds.). Oxford: Oxford Univ. Press. MR1380276

Kahn, M. J. and Raftery, A. E. (1996). Discharge rates of medicare stroke patients to skilled nursing facilities: Bayesian logistic regression with unobserved heterogeneity. *Journal of the American Statistical Association* **91**, 29–41.

Kass, R. E. and Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association* **90**, 773–795.

Lee, J. C. and Sabavala, D. J. (1987). Bayesian estimation and prediction for the beta-binomial model. *Journal of Business & Economic Statistics* **5**, 357–367.

Lee, Y. and Nelder, J. A. (2000). The relationship between double-exponential families and extended quasilikelihood families, with application to modelling Geissler's human sex ratio data. *Journal of the Royal Statistical Society, Series C* **49**, 413–419. MR1824549

Lindsey, J. K. and Altham, P. M. E. (1998). Analysis of the human sex ratio by using overdispersion models. *Journal of the Royal Statistical Society, Series C* **1**, 149–157.

Lynch, S. M. and Western, B. (2004). Bayesian posterior predictive checks for complex models. *Sociological Methods & Research* **32**, 301–335. MR2037797

Meng, X. L. (1994). Posterior predictive p-values. *The Annals of Statistics* **22**, 1142–1160. MR1311969

Moore, D. F. and Tsiatis, A. (1991). Robust estimation of the variance in moment methods for extra-binomial and extra-Poisson. *Biometrics* **47**, 383–401. MR1132537

Nott, D. (2006). Semiparametric estimation of the mean and variance functions for non-Gaussian data. *Computational Statistics* **21**, 603–620. MR2312636

Peng, F. and Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics* **23**, 199–213.

Pires, R. M. and Diniz, C. A. R. (2012). Correlated binomial regression models. *Computational Statistics and Data Analysis* **56**, 2513–2525. MR2910066

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81**, 321–327.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org/.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. *Springer Texts in Statistics*. New York: Springer. MR2080278

Shepard, T. H., Mackler, B. and Finch, C. A. (1980). Reproductive studies in the iron-deficient rat. *Teratology* **22**, 329–334.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society, Series B* **10**, 257–261. MR0028539

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639. MR1979380

Departamento de Estatística
Universidade Federal de São Carlos
C.P. 676, 13565-905
São Carlos-SP
Brasil
E-mail: carolina_paraiba@yahoo.com.br
dcad@ufscar.br
rubianemariapires@gmail.com