Bayesian Analysis (2013) **8**, Number 2, pp. 381–410

Parameter Interpretation in Skewed Logistic Regression with Random Intercept

Cristiano C. Santos*, Rosangela H. Loschi† and Reinaldo B. Arellano-Valle‡

Abstract. This paper aims at providing the prior and posterior interpretations for the parameters in the logistic regression model with random or cluster-level intercept when univariate and multivariate classes of skew normal distributions are assumed to model the random effects behavior. We obtain the prior distributions for the odds ratio and their medians under skew normality for the random effects. Original results related to linear combinations of skew-normal distributions are obtained as a by-product and, in the univariate case, a new class of log-skew-normal distribution is introduced. Robust results are obtained whenever a class of multivariate skew-normal distribution is assumed. We also evaluate the effect of the misspecification of the random effects distributions in the odds ratio estimation. We consider both simulated and the Teratogenic activity experiment datasets. The latter was previously analysed in the literature. We concluded that the misspecification of the random effects distribution yields poor odds ratios estimates and that the median odds ratio is not necessarily the best measure of heterogeneity among the clusters as suggested in the literature.

Keywords: Cluster, mixed models, random odds ratio, skew normal distribution

1 Introduction

Logistic regression has been the standard method to analyze binary response data which occur in several fields of research. However, it fails to model binary data from clustered, multi-level and longitudinal studies, for instance, because independence among the observations is assumed in its construction. Inspired by the theory of linear normal models, random effects were included in the linear predictor of the logistic regression model in order to allow for correlated responses (see Larsen et al. 2000; Diggle et al. 2002; Paulino et al. 2005, for instance). The obtained model accounts for the covariance among the measures in a relatively parsimonious way. Another important feature of random effect models is their flexibility in representing the effect of important non-observed or latent variables (Gibbons et al. 1994) as well as in accommodating outliers (Souza and Migon 2010), overdispersion (Schall 1991) and any degree of imbalance in longitudinal data (Fitzmaurice et al. 2004).

Most research related to logistic regression with random effects (LRRE) or mixed logistic regression have focused on the estimation of the parameters. Classical analysis of correlated binary data is usually not straightforward. The difficulty in estimating the parameters arises because the likelihood function involves multiple integrals with

*Universidade Federal de Minas Gerais, Pampulha, Brasil, cristcarvalhosan@yahoo.com.br
†Universidade Federal de Minas Gerais, Pampulha, Brasil, loschi@est.ufmg.br
‡Pontificia Universidad Católica de Chile, Santiago, Chile, reivalle@mat.puc.cl

 DOI:10.1214/13-BA813

solutions that are frequently non-analytical even under the assumption of independent and normally distributed random effects (Zeger and Karim 1991; Larsen et al. 2000; Diggle et al. 2002; Larsen and Merlo 2005). Numerical approximations, such as the ones considered by Breslow and Clayton (1993) and McCulloch and Searle (2001), are therefore considered to solve the integrals. The non-normal case requires even more work. According to Nelson et al. (2006) and Liu and Yu (2008), a transformation in the joint (data and random effects) likelihood is needed in order to make possible the use of the regular numerical approximations. An efficient expectation-maximization (EM) algorithm for maximum likelihood estimation in nonlinear mixed model can be found in Kuhn and Lavielle (2005).

The Bayesian paradigm provides a natural approach to inference in mixed models (see Paulino et al. 2005, for instance). The random effects are treated as parameters to be estimated. If the posterior distributions can not be obtained analytically, Markov chain Monte Carlo (MCMC) methods can be used to approximate them. This easier way of handling the inference problem under mixed models allows the use of more realistic distributions for the random effects such as the Student-t or finite mixtures of normal distributions as in Souza and Migon (2010) or the skew-normal and non-parametric prior distributions as in Liu and Dey (2008). An empirical Bayes approach for LRRE models is presented by Ten Have and Localio (1999).

Despite its great flexibility in modeling binary correlated data, LRRE does not inherit the interpretational features of the standard logistic model (Larsen and Merlo 2005). Larsen et al. (2000) first consider parameter interpretation in the LRRE. Assuming normally distributed random effects, Larsen et al. (2000) prove that the odds ratio (OR) in LRRE will depend on the random effects. Thus, under the classical approach considered in Larsen et al. (2000), the OR is a random quantity in a great number of possible comparisons. Larsen et al. (2000) discuss different measures of heterogeneity and suggest using the median (MOR) and the quantile intervals (IOR) of the odds ratio distribution as the best alternatives. In addition, both measures (MOR and IOR) have nice interpretations in terms of probability. In some cases, the MOR also allows a simple interpretation in terms of the well-known odds ratio, that greatly facilitates communication between the data analyst and the subject-matter researcher. The IOR is not a classical confidence interval thus can not be used to make decisions about the OR significance.

This paper focuses attention on the parameter interpretation in logistic regression with random or cluster-level intercept. We assume independent and correlated random effects considering, respectively, the univariate (Azzalini 1985) and the multivariate (Azzalini and Dalla-Valle 1996) classes of skew-normal (SN) distributions to model the random effects behavior. We extend some results in Larsen et al. (2000) by obtaining the prior distributions of the odds ratio and their medians under skew normality for the random effects. Some results related to the linear combination of SN random variables are also introduced and, in the univariate case, a new class of log-skew-normal distribution is defined. From the Bayesian point of view, the results in Larsen et al. (2000) are useful in the construction of more appropriate prior distributions for the random effects since they provide a nice prior interpretation of the OR, conditionally on the fixed effects. Posterior interpretations for the OR as suggested in Martins-Filho et al. (2010) are provided. Besides the posterior medians and modes, highest posterior density (HPD) intervals for the OR are also obtained and with this a tool for testing hypotheses about the OR is built. Alonso et al. (2008) and Litière et al. (2008) show that, under maximum likelihood theory, the misspecification of the random effects distributions produces inconsistent estimators for the fixed effects (but the magnitude of the bias is small) and severely affect the power and type I error rate. Since the OR depends on both the fixed and random effects, it is also our goal to evaluate the impact of the misspecification of the random effects distribution in the OR estimation. We consider simulated datasets.

Biological, environmental and financial data are often skewed and heavy tailed. Such behaviors are usually not well fitted by the normal distribution. The attractiveness of using the SN family of distributions in these cases is its flexibility in fitting densities of different shapes such as a perfectly symmetric distribution (like the normal distribution when the skewness parameter is zero) or a distribution that reveals a strong degree of asymmetry (like the half-normal distribution). In addition, the SN families of distributions preserve some nice properties of the normal family. Because of this, as a motivation, we reanalyze the dataset reported in Liu and Dey (2008) from a teratogenic activity experiment which was performed to study the interaction between teratogenic activity of the two niacin analogs *6-aminonicoti-namide*(6AN) and *3-acetylpyridine*(3AP). Liu and Dey (2008) assumed a mixed logistic regression model with different prior specifications for the random effects. They concluded that the skew-normal mixed logistic model is better than the normal and the non-parametric ones. However, Liu and Dey (2008) elicited a point mass probability prior for the skewness parameter. By doing this, it is assumed that the random effects are necessarily all positive, or all negative values.

This paper is organized as follows. Section 2 presents the LRRE in a general setting and summarizes the results obtained by Larsen et al. (2000) for normal random effects. Extensions of Larsen et al. (2000)'s results as well as the Bayesian approach for the OR interpretation are obtained in Section 3 under skew-normally distributed random effects. Section 4 discusses the posterior inference assuming independent and correlated SN random effects. In Section 5, we evaluate the influence of the misspecification of the random effects distributions in the OR as well as fixed and random effects estimates. In Section 6 we consider an application to the teratogenic activity experiment analyzed first by Liu and Dey (2008). We close the paper in Section 7 with some conclusions and discussions about the limitations of the proposed models.

2 Odds ratio interpretation in logistic regression model with random intercept

Following Larsen et al. (2000), suppose that the population is divided in k clusters and that a sample of size n_i is selected into the ith cluster. Let y_{ij} be the response variable (y_{ij} is 1 if a success and 0 otherwise) for individual j in the cluster i, $i = 1, \ldots, k$ and

$j = 1, \ldots, n_i$. Let $\mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2}, ..., x_{ijp})^t$ be the $(p+1) \times 1$ vector of covariates for the individual j in the ith cluster. Assume that $\mathbf{X}$ is the $N \times (p+1)$ matrix with the information related to the covariates for all $N = \sum_{i=1}^{k} n_i$ observed individuals.

Define $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_k)^t \in \mathbb{R}^k$ as the vector of random effects, where γ_i denotes the random effect for the ith cluster. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_p)^t \in \mathbb{R}^{p+1}$ be the vector of fixed effects. Denote by $\eta_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \gamma_i$ the linear predictor. Consequently, in the mixed logistic regression, $\pi_{ij} = P(y_{ij} = 1 | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X}) = \exp\{\eta_{ij}\}[1 + \exp\{\eta_{ij}\}]^{-1}$.

Assume that for a sample of N individuals it follows that, given $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\mathbf{X}$, the responses $y_{ij} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X} \overset{ind}{\sim} Ber(\pi_{ij})$. As a consequence, the likelihood function is

$$f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{X}) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \left[\frac{\exp\{\eta_{ij}\}}{1 + \exp\{\eta_{ij}\}} \right]^{y_{ij}} \left[\frac{1}{1 + \exp\{\eta_{ij}\}} \right]^{1-y_{ij}}, \tag{1}$$

where $\mathbf{y} = (y_{11}, ..., y_{1n_1}, ..., y_{k1}, ..., y_{kn_k})^t$.

In the usual logistic regression model, the fixed effects have an interpretation in terms of the so-called odds ratio between the highest and the lowest risk individuals. That interpretation makes the communication with researchers from other areas easy and thus can help in the construction of more appropriate prior distributions for the fixed effects. The interpretation of the fixed effects in the LRRE model was first discussed by Larsen et al. (2000). Larsen et al. (2000) show that the odds ratio depends on both the fixed and random effects and, consequently, several useful interpretations arise from such a quantity. Reviewing Larsen et al. (2000)'s results, let j_1 and j_2 be two individuals in the different clusters i_1 and i_2, respectively. The odds ratio becomes

$$OR_{i_1 j_1, i_2 j_2} = \exp\left\{(\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t)\boldsymbol{\beta} + \gamma_{i_1} - \gamma_{i_2}\right\}. \tag{2}$$

If the comparison is between individuals in the same cluster - say, $i_1 = i_2 = i$ - but having different covariates, $OR_{ij_1, ij_2} = \exp\left\{(\mathbf{x}_{ij_1}^t - \mathbf{x}_{ij_2}^t)\boldsymbol{\beta}\right\}$ depends only on the fixed effects and is exactly the same as for the usual logistic regression model. To quantify the random effects, the comparison is done assuming that two individuals, j_1 and j_2, have the same covariate vectors and are in different clusters; that is, the individual j_k belongs to cluster i_k, $k = 1, 2$. In this case, the odds ratio depends on the random effects only and is $OR_{i_1 j_1, i_2 j_2} = \exp\{\gamma_{i_1} - \gamma_{i_2}\}$. The odds ratio in (2) permits also the comparison between the individuals with the highest risk in two different clusters, among others.

Under the classical approach for inference, the OR is a random quantity only when the comparison depends on the random effects. Since it is random, Larsen et al. (2000) propose to interpret the OR in terms of the median of its distribution. The so-called median odds ratio is named here MOR. According to Larsen et al. (2000), the MOR quantifies appropriately the heterogeneity among the different clusters. For the general case, whenever we are comparing individuals with different covariates in different clusters the median odds ratio is defined as

$$MOR_{i_1 j_1, i_2 j_2} = \text{med}\{\exp\left\{(\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t)\boldsymbol{\beta} + (\gamma_{i_1} - \gamma_{i_2}) \mid \boldsymbol{\beta}, \mathbf{X}\right\}\}.$$

It is important to recall that the distribution of $\gamma_{i_1} - \gamma_{i_2}$ is symmetric around zero under the assumption of independent and identically distributed (i.i.d.) random effects. Therefore, independently of the clusters we are comparing, we have the same value for the MOR when the individuals have the same covariates. To properly quantifying the effect of clusters in these cases, Larsen et al. (2000) consider the median of $\exp\{|\gamma_{i_1} - \gamma_{i_2}|\}$, that is, the median odds ratio becomes

$$MOR_{i_1j_1,i_2j_2} = \text{med}\{\exp\{|\gamma_{i_1} - \gamma_{i_2}| \mid \boldsymbol{\beta}, \mathbf{X}\}\},$$

where $|A|$ denotes the absolute value of A. To simplify the notation, through this section let $OR_{12} = OR_{i_1j_1,i_2j_2}$ and $MOR_{12} = MOR_{i_1j_1,i_2j_2}$. Also, consider $\kappa_{12} = (\mathbf{x}^t_{i_1j_1} - \mathbf{x}^t_{i_2j_2})\boldsymbol{\beta}$ and $W_{12} = \gamma_{i_1} - \gamma_{i_2}$.

Larsen et al. (2000) show that if $\boldsymbol{\gamma} \sim N_k(\mathbf{0}, \boldsymbol{\Sigma})$, then $W_{12} \sim N(0, \sigma^2_{12})$, where σ^2_{12} denotes the variance of W_{12}. Consequently, the odds ratio in (2) can be rewritten as $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$. Thus the random variable OR_{12} has a lognormal distribution with location and scale parameters κ_{12} and σ^2_{12}, respectively, and with probability density function (p.d.f.) given by

$$f_{OR_{12}|\boldsymbol{\beta},\boldsymbol{\Sigma}}(r) = \frac{1}{\sqrt{2\pi} r \sigma_{12}} \exp\left\{-\frac{1}{2\sigma^2_{12}}(\ln r - \kappa_{12})^2\right\},\ r \in \mathbb{R}_+, \tag{3}$$

and the MOR_{12} is $\text{med}\{\exp\{\kappa_{12} + W_{12}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{X}\}\} = \exp\{\kappa_{12}\}$. This OR interpretation was considered from the classical point of view and also provides a nice prior interpretation for the OR, given the matrix $\mathbf{X}$, the fixed effects $\boldsymbol{\beta}$ and the hyperparameters for the prior of $\boldsymbol{\gamma}$. From the Bayesian point of view, however, the OR is a random quantity which depends on $\boldsymbol{\beta}$ and/or $\boldsymbol{\gamma}$. Thus, in the next section, we propose to use the posterior summaries of OR in order to interpret the fixed effects. We also extend results in Larsen et al. (2000) by considering independent and correlated SN random effects.

3 A Bayesian OR interpretation under skew normality

To simplify the inferential process in the logistic regression with random intercept, the random effects are usually assumed to be independent with a common normal distribution. Such an assumption, however, is questionable in some biological data as shown in Liu and Dey (2008), for instance. In this section, we consider independent (Azzalini 1985) and correlated (multivariate) (Azzalini and Dalla-Valle 1996) skew-normal random effects. Both SN families include the normal one as a special case but, differently from what is observed for the normal distribution, the independent case does not follow straightforwardly from the multivariate case by assuming a diagonal scale matrix. However, independence follows from the multivariate case if the scale matrix is diagonal and there is only one non-null component in the vector of skewness parameters.

Throughout this paper denote by $\phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the p.d.f. associated with the multivariate $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, and by $\Phi_n(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the corresponding cumulative distribution function (c.d.f.). If $\boldsymbol{\mu} = \mathbf{0}$ (respectively $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_n$) these functions will be denoted by $\phi_n(\mathbf{y} \mid \boldsymbol{\Sigma})$ and $\Phi_n(\mathbf{y} \mid \boldsymbol{\Sigma})$ (respectively $\phi_n(\mathbf{y})$ and $\Phi_n(\mathbf{y})$). For simplicity, $\phi(\mathbf{y})$ and $\Phi(\mathbf{y})$ will be used in the univariate case.

3.1 Prior results for the OR - Independent case

Let us assume that, given $\sigma_i^2 \in \mathbb{R}_+$ and ξ_i, $\lambda_i \in \mathbb{R}$, $\gamma_i \overset{ind.}{\sim} SN(\xi_i, \sigma_i^2, \lambda_i)$ for $i = 1, \ldots, k$, that is, $f(\gamma_i \mid \xi_i, \sigma_i^2, \lambda_i) = \frac{2}{\sigma_i}\phi\left(\frac{\gamma_i - \xi_i}{\sigma_i}\right)\Phi\left(\lambda_i \frac{\gamma_i - \xi_i}{\sigma_i}\right)$, $\gamma_i \in \mathbb{R}$.

As pointed out in (2), the OR_{12} depends on the random effects through the random variable $W_{12} = \gamma_{i_1} - \gamma_{i_2}$. Thus, in order to obtain the distribution of OR_{12}, we need first to find the distribution of linear combinations of independent SN random variables. As before, let $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$, where $\kappa_{12} = (\mathbf{x}_{i_1 j_1}^t - \mathbf{x}_{i_2 j_2}^t)\boldsymbol{\beta}$. Also, let $\delta_i = \lambda_i[1 + \lambda_i^2]^{-1/2}$. It follows from Proposition 3 in Appendix A that W_{12} has the unified skew-normal (SUN) p.d.f. given by

$$f_{W_{12}|\boldsymbol{\xi},\boldsymbol{\Lambda},\boldsymbol{\Omega}_{12}^*}(w) = 4\phi(w \mid \xi_{i_1} - \xi_{i_2}, \sigma_{i_1}^2 + \sigma_{i_2}^2) \times\Phi_2\left(\frac{w - \xi_{i_1} + \xi_{i_2}}{\sigma_{i_1}^2 + \sigma_{i_2}^2}\boldsymbol{\epsilon}|\mathbf{I}_2 - \frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^t}{\sigma_{i_1}^2 + \sigma_{i_2}^2}\right), \quad (4)$$

where $\boldsymbol{\epsilon} = (\sigma_{i_1}\delta_{i_1}, -\sigma_{i_2}\delta_{i_2})^t$. The distribution in (4) is denoted by $W_{12} \sim SUN_{1,2}(\xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \boldsymbol{\Omega}_{12}^*)$ where $\boldsymbol{\Omega}_{12}^* = \begin{pmatrix} \mathbf{I}_2 & \boldsymbol{\Gamma}_{12}^t \\ \boldsymbol{\Gamma}_{12} & 1 \end{pmatrix}$ and $\boldsymbol{\Gamma}_{12} = [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{-1/2}\boldsymbol{\epsilon}^t$. The SUN family of distributions was introduced by Arellano-Valle and Azzalini (2006) and its definition is in Appendix A.

The next proposition provides the prior distribution for the OR_{12} and its median, given the hyperparameters $\boldsymbol{\theta}_{12} = (\xi_{i_1}, \sigma_{i_1}^2, \lambda_{i_1}, \xi_{i_2}, \sigma_{i_2}^2, \lambda_{i_2})^t$ and the fixed effects $\boldsymbol{\beta}$.

Proposition 1. *If $\gamma_{i_l} \overset{ind}{\sim} SN(\xi_{i_l}, \sigma_{i_l}^2, \lambda_{i_l})$, $l = 1, 2$, then, given the fixed effects $\boldsymbol{\beta}$ and the hyperparameters $\boldsymbol{\theta}_{12} = (\xi_{i_1}, \sigma_{i_1}^2, \lambda_{i_1}, \xi_{i_2}, \sigma_{i_2}^2, \lambda_{i_2})^t$, it follows that*

(i) the prior distribution for the OR_{12} is the log-skew-normal distribution with p.d.f.

$$f_{OR_{12}|\boldsymbol{\beta},\boldsymbol{\theta}_{12}}(r) = \frac{4}{r}\phi(\ln r|\kappa_{12} + \xi_{i_1} - \xi_{i_2}, \sigma_{i_1}^2 + \sigma_{i_2}^2) \times\Phi_2\left(\frac{\ln r}{\sigma_{i_1}^2 + \sigma_{i_2}^2}\boldsymbol{\epsilon}|\frac{\kappa_{12} + \xi_{i_1} - \xi_{i_2}}{\sigma_{i_1}^2 + \sigma_{i_2}^2}\boldsymbol{\epsilon}, \mathbf{I}_2 - \frac{\boldsymbol{\Omega}_{12}}{\sigma_{i_1}^2 + \sigma_{i_2}^2}\right), \quad (5)$$

for $r \in \mathbb{R}_+$, where $\boldsymbol{\Omega}_{12} = \begin{pmatrix} \sigma_{i_1}^2\delta_{i_1}^2 & -\sigma_{i_1}\sigma_{i_2}\delta_{i_1}\delta_{i_2} \\ -\sigma_{i_1}\sigma_{i_2}\delta_{i_1}\delta_{i_2} & \sigma_{i_2}^2\delta_{i_2}^2 \end{pmatrix}$;

(ii) the median odds ratio is given by

$$MOR_{12} = \exp\{\kappa_{12}\}\exp\left\{\Phi_{SUN_{1,2}}^{-1}\left(0.5|\xi_{i_1} - \xi_{i_2}, \mathbf{0}, [\sigma_{i_1}^2 + \sigma_{i_2}^2]^{1/2}, \boldsymbol{\Omega}_{12}^*\right)\right\}, \quad (6)$$

where $\Phi_{SUN_{1,2}}$ denotes the c.d.f of the SUN p.d.f. given in (4).

The proof of Proposition 1 is straightforward from Proposition 3 and some well-known results of probability calculus. Thus it will be omitted.

To reduce the number of parameters to be estimated, a common practice in mixed model is to assume i.i.d. random effects. If we assume that, given $\boldsymbol{\beta}$, σ^2 and λ, $\gamma_{i_l} \overset{iid}{\sim} SN(\xi, \sigma^2, \lambda)$, then *a priori* the random quantity OR_{12} has log-skew-normal distribution with p.d.f.

$$\begin{aligned} f_{OR_{12}|\boldsymbol{\beta},\boldsymbol{\theta}_{12}}(r) &= \frac{4}{r}\phi(\ln r|\kappa_{12}, 2\sigma^2) \\ &\times\Phi_2\left(\frac{\delta \ln r}{2\sigma}\begin{pmatrix} 1 \\ -1 \end{pmatrix} | \frac{\delta\kappa_{12}}{2\sigma}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{I}_2 - \frac{\delta^2}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right), \quad (7) \end{aligned}$$

for $r \in \mathbb{R}_+$, where $\mathbf{1}_k$ is the $k \times 1$ vector of ones. By assuming i.i.d. random effects we have that $W_{12} \sim SUN_{1,2}\left(0, \mathbf{0}, (2\sigma^2)^{1/2}, \begin{pmatrix} \mathbf{I}_2 & \frac{\delta(1,-1)^t}{\sqrt{2}} \\ \frac{\delta(1,-1)}{\sqrt{2}} & 1 \end{pmatrix}\right)$ which is symmetric around zero (see Figure 1). Consequently, the median of the distribution in (7) is

$$MOR_{12} = \exp\{\kappa_{12}\}. \quad (8)$$

Results in (7) and (8) still follow if we consider $\gamma_l \overset{iid}{\sim} SN(0, \sigma^2, \lambda)$. Figure 1 presents the p.d.f of W_{12} and OR in the i.i.d case for different values of λ and assuming $\sigma^2 = 1$ and $\kappa_{12} = 0$. Positive and negative values of λ provide the same density. We also notice that by eliciting high values for $|\lambda|$, the OR tends to be higher than assumed in Larsen et al. (2000)'s model.

Figure 1 shows that the normal distribution ($\lambda = 0$) for the random effects reveals that, most probably, the odds ratio to compare two individuals with the same characteristics but that are in different clusters is up to 1. However, normality is not a reasonable assumption for the random effects, if the prior opinion of the expert discloses that $OR > 0.5$ with high probability. To obtain a better fit, in this case, appropriate prior distributions of OR must have heavier tails than that induced by the assumption of normality for the random effects. From Figure 1, we notice that we reach this proposal if *a priori* we assume, for instance, that $\gamma_i \sim SN(0, 1, -8.5)$.

It is noteworthy from (8) that by assuming i.i.d. SN random effects the prior interpretation of the fixed effects through the MOR will not depend on the clusters the individuals are in. That is not observed if the random effects are not identically distributed. A similar result was obtained by Larsen et al. (2000) in a more general setting where the random effects can have different normal distributions.

The distributions of OR_{12} given in (5) and (7) belong to a class of log-skew-normal distributions different from and with heavier tails than the log-skew-normal class introduced by Marchenko and Genton (2010). In the class of distributions defined in Marchenko and Genton (2010), the skewing function is the c.d.f of a univariate normal distribution while in (7) the skewing function is the c.d.f of a bivariate normal distribution. The distribution in (7) is obtained by skewing the distribution in (3) and we return to Larsen et al. (2000)'s results if we let $\lambda = 0$ in (7).

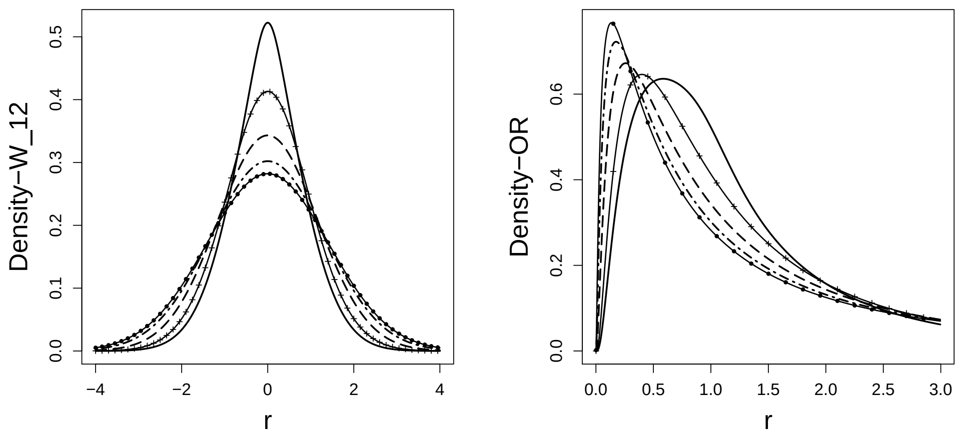

Figure 1: Prior distribution of W_{12} (left) and OR (right) for the independent case, $\sigma^2 = 1$, $\kappa_{12} = 0$ and $\lambda = -8.5$ (solid line), -2 (+), -1 (dashed line), -0.5 (dotdashed line), 0 (•, Larsen et al. (2000)'s model).

3.2 Prior results for the OR - Correlated case

Assume that, given $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\alpha}$, $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ with density

$$f(\boldsymbol{\gamma} \mid \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) = 2\phi_k(\boldsymbol{\gamma} \mid \boldsymbol{\xi}, \boldsymbol{\Omega})\, \Phi\left(\boldsymbol{\alpha}^t \boldsymbol{\omega}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\xi})\right), \ \boldsymbol{\gamma} \in \mathbb{R}^k, \tag{9}$$

where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)^t \in \mathbb{R}^k$ is the location vector, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^t \in \mathbb{R}^k$ is the vector of skewness parameters, $\boldsymbol{\Omega}$ is a $k \times k$ positive definite covariance matrix and $\boldsymbol{\omega} = diag\{\sigma_1, \ldots, \sigma_k\}$ is a diagonal matrix formed by the square root of elements in the diagonal of $\boldsymbol{\Omega}$ and is such that $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$ where $\bar{\boldsymbol{\Omega}}$ is a correlation matrix.

In order to obtain the distribution of the random quantity OR_{12}, the distribution of linear combinations of the random effects is needed. As in the previous section, it is of particular interest to find the distribution of $W_{12} = \gamma_{i_1} - \gamma_{i_2}$. If $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ with p.d.f given in (9), it follows from Proposition 4 in Appendix A that

$$W_{12} \sim SN(\xi_{i_1} - \xi_{i_2}, \sigma^2_{i_1} + \sigma^2_{i_2} - 2\sigma_{i_1 i_2}, \lambda^*_{12}), \tag{10}$$

where $\sigma_l^2 = \sigma_{ll}$, σ_{lm} is the (l, m)th entry of $\boldsymbol{\Omega}$, α_l is the lth coordinate of $\boldsymbol{\alpha}$ and $\lambda^*_{12} = \frac{\sum_{l=1}^k \frac{\alpha_l}{\sigma_l}[\sigma_{i_1 l} - \sigma_{i_2 l}]}{\left[\left(1+\sum_{l,m=1}^k \frac{\alpha_m \alpha_l \sigma_{lm}}{\sigma_l \sigma_m}\right)\left(\sigma^2_{i_1} + \sigma^2_{i_2} - 2\sigma_{i_1 i_2}\right) - \left(\sum_{l=1}^k \frac{\alpha_l}{\sigma_l}[\sigma_{i_1 l} - \sigma_{i_2 l}]\right)^2\right]^{1/2}}$.

Simpler structures for the matrix $\boldsymbol{\Omega}$ can be assumed and hence more parsimonious models are built. Also, depending on the structure assumed for $\boldsymbol{\Omega}$, we preserve the interpretation of the odds ratio obtained under normality. This invariance occurs when in (10) the distribution of W_{12} is normal, i.e., for $\lambda_{12}^* = 0$. For instance, consider in (9) that $\mathbf{\Omega} = diag(\sigma_1^2, \ldots, \sigma_k^2)$, $\sigma_i^2 \in \mathbb{R}_+$. For any vector $\mathbf{c} \in \mathbb{R}^k$ that is orthogonal to the vector $\mathbf{\Omega}\boldsymbol{\omega}^{-1}\boldsymbol{\alpha}$, say, such that $\sum_{i=1}^k \alpha_i \sigma_i c_i = 0$, we have that

$$\boldsymbol{c}^t\boldsymbol{\gamma} \sim N\left(\sum_{i=1}^k \xi_i c_i, \sum_{i=1}^k c_i^2\sigma_i^2\right), \tag{11}$$

which is equivalent to the result obtained under the assumption of independent and normally distributed random effects.

A natural choice is to consider $\sigma_i^2 = \sigma^2$ for all i. If $\sum_{i=1}^k \alpha_i c_i \neq 0$ and $\mathbf{\Omega} = \sigma^2\mathbf{I}_k$, the distribution for $\boldsymbol{c}^t\boldsymbol{\gamma}$ is

$$\boldsymbol{c}^t\boldsymbol{\gamma} \sim SN\left(\sum_{i=1}^k \xi_i c_i, \sigma^2\sum_{i=1}^k c_i^2, \lambda^*\right), \tag{12}$$

where $\lambda^* = [\sum_{i=1}^k \alpha_i c_i][(1+\sum_{i=1}^k \alpha_i^2)(\sum_{i=1}^k c_i^2) - (\sum_{i=1}^k \alpha_i c_i)^2]^{-1/2}$. Under this particular structure for $\mathbf{\Omega}$, we have two interesting results related to the distribution of W_{12}. If $\boldsymbol{\alpha}$ discloses different degree of skewness through the clusters, it follows that $W_{12} \sim SN(\xi_{i_1} - \xi_{i_2}, 2\sigma^2, [\alpha_{i_1} - \alpha_{i_2}][2(1+\sum_{i=1}^k \alpha_i^2) - (\alpha_{i_1} - \alpha_{i_2})^2]^{-1/2})$. However, we return to normality if we assume $\boldsymbol{\alpha} = \alpha\mathbf{1}_k$, for $\alpha \in \mathbb{R}$, that is, if $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$. In this case, we have that $W_{12} \sim N(\xi_{i_1} - \xi_{i_2}, 2\sigma^2)$.

Therefore the results established in Larsen et al. (2000) for the OR_{12} and MOR_{12} also follow if $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$ and under the conditions that lead to the result in (11). In all these cases, the OR_{12} has the same log-normal distribution we observed for normal random effects given in (3). In Proposition 2, we provide the distribution of OR_{12} and its median for the general case.

Proposition 2. *If the random effects $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \mathbf{\Omega}, \boldsymbol{\alpha})$ with p.d.f given in (9) then, given $\boldsymbol{\beta}$, $\mathbf{\Omega}$ and $\boldsymbol{\alpha}$, it follows that:*

(i) the random variable $OR_{12} = \exp\{\kappa_{12} + W_{12}\}$ has p.d.f.

$$\begin{aligned} f_{OR_{12}|\boldsymbol{\beta},\boldsymbol{\alpha},\mathbf{\Omega}}(r) &= \frac{2}{r}\phi\left(\ln r \mid (\kappa_{12} + \xi_{i_1} - \xi_{i_2}), \sigma^{*2}\right) \\ &\quad \times\Phi\left(\frac{\lambda_{12}^*\left(\ln r - (\kappa_{12} + \xi_{i_1} - \xi_{i_2})\right)}{\sigma^*}\right), \end{aligned} \tag{13}$$

$r \in \mathbb{R}_+$, where $\sigma^{*2} = \sigma_{i_1}^2 + \sigma_{i_2}^2 - 2\sigma_{i_1 i_2}$ and λ_{12}^* is as defined in (10), and

(ii) the median of the distribution in (13) is

$$\begin{aligned} MOR_{12} &= med\{\exp\{\kappa_{12} + W_{12}\} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{\Omega}\} \\ &= \exp\{\kappa_{12}\}\exp\{\Phi_{SN}^{-1}(0.5)\sigma^* + \xi_{i_1} - \xi_{i_2}\}, \end{aligned} \tag{14}$$

where $\Phi_{SN}(.)$ denotes the c.d.f of the standard SN distribution with skewness parameter λ_{12}^* given in (10).

(iii) if $\xi_{i_1} - \xi_{i_2} = 0$ *it follows that*

(1) $med\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}\} = \exp\{med\{|W_{12}| \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}\}\} = \exp\{\sigma^* \Phi^{-1}(0.75)\}$

(2) $med\{\exp\{-|W_{12}|\} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}\} = [med\{\exp\{|W_{12}|\} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Omega}\}]^{-1}$.

The proof of Proposition 2 is omitted. It follows from Proposition 1.6.1 in Dalla-Valle (2004) and some well-known results of probability calculus.

In particular, assume that $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \sigma^2 \mathbf{I}_k, \boldsymbol{\alpha})$, where as before we have $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)^t$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^t$. Consequently, given $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\alpha}$, the random variable OR_{12} has p.d.f

$$
\begin{aligned}
f_{OR_{12}|\boldsymbol{\beta},\boldsymbol{\alpha},\sigma^2}(r) \quad = \quad & \frac{\sqrt{2}}{\sigma r} \phi\left(\frac{1}{\sqrt{2}\sigma}(\ln r - (\kappa_{12} + \xi_{i_1} - \xi_{i_2}))\right) \\
& \Phi\left(\frac{\lambda^*_{12}}{\sqrt{2}\sigma}\left(\ln r - (\kappa_{12} + \xi_{i_1} - \xi_{i_2})\right)\right), \ r \in \mathbb{R}_+, \qquad (15)
\end{aligned}
$$

with median given by

$$
MOR_{12} = \exp\{\kappa_{12}\} \exp\{\Phi^{-1}_{SN}(0.5)\sqrt{2}\sigma + \xi_{i_1} - \xi_{i_2}\}, \qquad (16)
$$

where $\lambda^*_{12} = [\alpha_{i_1} - \alpha_{i_2}][2(1 + \sum_{i=1}^{k} \alpha_i^2) - (\alpha_{i_1} - \alpha_{i_2})^2]^{-1/2}$.

Figure 2 shows the p.d.f. of both W_{12} and OR_{12} when we assume two individuals with the same values for the covariates ($\kappa_{12} = 0$), $\sigma^2 = 1$ and $\alpha_{i_2} = 0$ and -5. Different values of α_{i_1} are considered. If the prior opinion of an expert reveals that OR_{12} is, most probably, higher than one, from Figure 2 we perceive that the normal distribution ($\alpha_{i_1} = \alpha_{i_2} = 0$) will not be an appropriate choice to describe the uncertainty about the random effects. In this case, the multivariate SN distribution with $\alpha_{i_1} = 5$ and $\alpha_{i_2} = -5$ yields a better description of her/his opinion.

The distributions in (13) and (15) belong to the log-skew-elliptical class of distributions introduced by Marchenko and Genton (2010). As for the independent case, the MOR_{12} depends on the fixed effects through $\exp\{\kappa_{12}\}$ which is the odds ratio if the random effects are not included in the model. Moreover, in both cases, $\exp\{\kappa_{12}\}$ is multiplied by a term which depends on the c.d.f. of some SN distribution. If for the multivariate case in (16), the skewness parameter for the cluster i_2 is $\alpha_{i_2} = 0$ and for the cluster i_1 we have a strong degree of skewness ($\alpha_{i_1} \to \infty$), then the MOR_{12} is higher than that observed for normal random effects. On the other hand, if $\alpha_{i_1} = 0$ and $\alpha_{i_2} \to \infty$, then the MOR_{12} is smaller than that observed for the normal case. Figure 3 shows the behavior of $MOR_{12} \times \alpha_{i_1}$ for $\kappa_{12} = 0$ and $\sigma = 1$ assuming different values for the skewness parameter in the cluster i_2. For simplicity, in Figures 2 and 3 we assume $\alpha_{i_j} = 0$ for all j except for $j = 1, 2$.

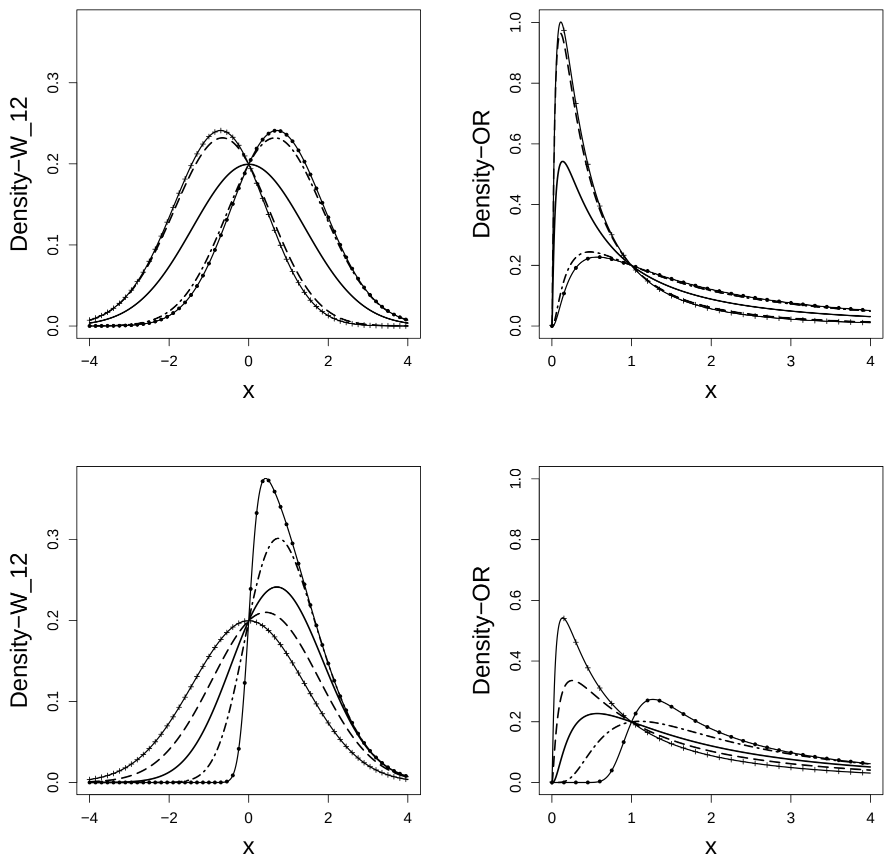

Figure 2: Prior distribution of W_{12} (left) and OR (right) for the dependent case, $\alpha_{i_2} = 0$ (top) and -5 (bottom), $\sigma^2 = 1$, $\kappa_{12} = 0$ and $\alpha_{i_1} = 0$ (solid line), -5 (+), -2 (dashed line), 2 (dotdashed line), 5 (•).

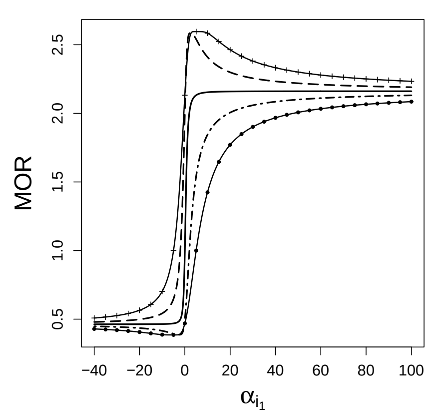

Figure 3: $MOR_{12} \times \alpha_{i_1}$, for $\alpha_{i_2} = -5$ (+), $\alpha_{i_2} = -2$ (dashed line), $\alpha_{i_2} = 0$ (solid line), $\alpha_{i_2} = 2$ (dotdashed line), $\alpha_{i_2} = 5$ (•).

3.3 Posterior Odds Ratio

The OR is a random quantity which depends on both the parameters $\boldsymbol{\beta}$ and the random effects $\boldsymbol{\gamma}$. Due to the complexity of the posterior distribution in the LRRE, it is a hard task to analytically obtain the posterior distribution of OR. On the other hand, computational approaches such as MCMC methods provide good approximations for the posterior distributions. Thus, we can approximate the posterior of OR_{12} as follows:

Step 1: Generate a sample $(\boldsymbol{\beta}^l, \boldsymbol{\gamma}^l)$ from the posterior distribution;

Step 2: Compute ${OR_{12}}^l = \exp\{(\mathbf{x}_{i_1 j_1} - \mathbf{x}_{i_2 j_2})^t \boldsymbol{\beta}^l + \gamma^l_{i_1} - \gamma^l_{i_2}\}$.

Steps 1 and 2 must be repeated until convergence has been attained for all parameters and we obtain samples of the posterior distributions of reasonable sizes. The expression in Step 2 is changed according to the specific interest of the researchers. One advantage of such an approach is that the posterior summaries of OR, such as the median, mode, and mean as well as the HPD intervals, are easily obtained (Martins-Filho et al. 2010). The posterior modal odds ratio is an appealing estimator of the OR since it is in the highest density region. Thus, informally speaking, we can conclude that when comparing two subjects of two randomly selected clusters, the odds ratio will be, most probably, the modal odds ratio. Median and HPD intervals also have natural interpretations in terms of probability and they are similar to the ones proposed by Larsen et al. (2000) for the MOR and IOR, respectively. However, using the posterior of OR we can perform hypothesis tests about OR using the HPD intervals. In this case, our main interest relies in testing the null hypothesis $H_0 : OR_{12} = 1$ which will be accepted if the value one is in the HPD interval. Performing such a test there is a gain in the analysis since we can take into consideration only comparisons that are shown to be significant.

4 Bayesian inference in logistic regression models with random intercept

From the Bayesian point of view, inference for mixed models is simpler since the random effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)^t$, $\gamma_i \in \mathbb{R}^k$, are considered as unknown quantities to be estimated.

Assume the likelihood in (1). To complete the model specification, we should elicit prior distributions for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. We consider two different hierarchical models. We assume first i.i.d random effects with univariate SN distribution, say, $\gamma_1, \ldots, \gamma_k \overset{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$, given σ^2 and λ, where $\delta = \lambda[1 + \lambda^2]^{-1/2}$. We also consider skewed correlated random effects such that, given σ^2 and α, $\boldsymbol{\gamma} \sim SN(-\Delta\sigma\sqrt{2/\pi}\mathbf{1}_k, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$ where $\Delta = \alpha[1 + k\alpha^2]^{-1/2}$. In both models, we center the SN distributions on zero to avoid nonidentifiability. As prior distributions for the fixed effects we assume $\boldsymbol{\beta} \sim N_{p+1}(\mathbf{m}, b^2\mathbf{I}_{p+1})$, and for the hyperparameters we consider $\sigma^2 \sim IG(a, d)$, $\lambda \sim SN(h, \tau^2, \theta)$ (Azzalini 1985) and $\alpha \sim N(h, \tau^2)$, where $\mathbf{m} \in \mathbb{R}^{p+1}$, h, $\theta \in \mathbb{R}$, and τ^2, b^2, a and d are real positive numbers. $IG(a, d)$ denotes the inverted gamma distribution with $\mathrm{E}(\sigma^2) = d(a-1)^{-1}$ and $\mathrm{V}(\sigma^2) = d[(a-1)^2(a-2)]^{-1}$.

Let $\mathbf{D} = (\mathbf{y}, \mathbf{X})$. As a consequence of the previous assumptions, for the i.i.d. case the joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \lambda)$ is

$$
\begin{aligned}
f(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \lambda|\mathbf{D}) \;\propto\; & \left[\prod_{i=1}^{k}\prod_{j=1}^{n_i} (\pi_{ij})^{y_{ij}} \left(1-\pi_{ij}\right)^{1-y_{ij}}\right] \left(\frac{1}{\sigma^2}\right)^{(d+2)/2} \left(\frac{1}{\tau^2}\right)^{1/2} \\
& \times \exp\left\{-\frac{a}{2\sigma^2} - \frac{(\boldsymbol{\beta}-\mathbf{m})^t(\boldsymbol{\beta}-\mathbf{m})}{2b^2} - \frac{(\lambda-h)^2}{2\tau^2}\right\} \\
& \times \exp\left\{-\frac{(\boldsymbol{\gamma}+\delta\sigma\sqrt{2/\pi}\mathbf{1}_k)^t(\boldsymbol{\gamma}+\delta\sigma\sqrt{2/\pi}\mathbf{1}_k)}{2\sigma^2}\right\} \\
& \times \Phi_k\left(\frac{\lambda\left(\boldsymbol{\gamma}+\mathbf{1}_k\delta\sigma\sqrt{\frac{2}{\pi}}\right)}{\sigma}\right) \Phi\left(\frac{\theta(\lambda-h)}{\tau}\right). \qquad (17)
\end{aligned}
$$

The posterior distribution in (17) has no closed form. An MCMC scheme will be considered to sample from it. To obtain the posterior full conditional distributions (f.c.d.) of the random effects, we consider Henze's stochastic representation for the univariate SN distribution. This well-known result establishes that, if $T_i \sim SN(\lambda)$ then $T_i \stackrel{d}{=} \delta|U_i| + (1-\delta^2)^{1/2}V_i$, where U_i and V_i are i.i.d. random variables with the standard normal distribution. For simplicity, let $\psi_i = |U_i|$. Since $\gamma_i \sim SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$, given σ^2 and λ, it follows that $\gamma_i \stackrel{d}{=} \delta\sigma\left(\psi_i - \sqrt{2/\pi}\right) + \sigma\left(1-\delta^2\right)^{\frac{1}{2}} V_i$. Similarly, if $\lambda \sim SN(h, \tau^2, \theta)$ we obtain that, given h, τ^2 and θ, $\lambda \stackrel{d}{=} h + \frac{\theta\tau}{\sqrt{1+\theta^2}} M_3 + \frac{\tau}{\sqrt{1+\theta^2}} M_2$, where $M_1 \stackrel{d}{=} M_2 \sim N(0,1)$ and $M_3 = |M_1|$. Assuming such transformations, the posterior f.c.d. are

$$
\begin{aligned}
f(\boldsymbol{\beta}|\mathbf{D},\boldsymbol{\gamma},\sigma^2,\boldsymbol{\psi},\lambda) &\propto \left[\prod_{i=1}^{k}\prod_{j=1}^{n_i}\left[\frac{\exp\{\eta_{ij}\}}{1+\exp\{\eta_{ij}\}}\right]^{y_{ij}}\left[\frac{1}{1+\exp\{\eta_{ij}\}}\right]^{1-y_{ij}}\right] \\
&\quad\times\exp\left\{-\frac{(\boldsymbol{\beta}-\mathbf{m})^t(\boldsymbol{\beta}-\mathbf{m})}{2b^2}\right\}, \\
f(\boldsymbol{\gamma}|\mathbf{D},\boldsymbol{\beta},\sigma^2,\boldsymbol{\psi},\lambda) &\propto \left[\prod_{i=1}^{k}\prod_{j=1}^{n_i}\left[\frac{\exp\{\eta_{ij}\}}{1+\exp\{\eta_{ij}\}}\right]^{y_{ij}}\left[\frac{1}{1+\exp\{\eta_{ij}\}}\right]^{1-y_{ij}}\right] \\
&\quad\times\exp\left\{-\frac{\sum_{i=1}^{n}\left(\gamma_i-\sigma\delta(\psi_i-\sqrt{2/\pi})\right)^2}{2\sigma^2(1-\delta^2)}\right\}, \\
f(\boldsymbol{\psi}|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\lambda) &\propto \prod_{i=1}^{k}\exp\left\{-\frac{\left(\psi_i-(\gamma_i+\sigma\delta\sqrt{2/\pi})\frac{\delta}{\sigma}\right)^2}{2(1-\delta^2)}\right\}\mathbf{1}_{(0,\infty)}(\psi_i), \\
f(\sigma^2|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\psi},\lambda) &\propto \left(\frac{1}{\sigma^2}\right)^{\frac{k}{2}+a+1}\exp\left\{-\frac{\sum_{i=1}^{n}\left(\gamma_i-\sigma\delta(\psi_i-\sqrt{2/\pi})\right)^2}{2\sigma^2(1-\delta^2)}-\frac{d}{\sigma^2}\right\}, \\
f(\lambda|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\boldsymbol{\psi},M_3) &\propto (1+\lambda^2)^{\frac{k}{2}}\exp\left\{-\frac{(1+\lambda^2)}{2\sigma^2}\sum_{i=1}^{n}\left(\gamma_i-\frac{\lambda\sigma(\psi_i-\sqrt{2/\pi})}{\sqrt{1+\lambda^2}}\right)^2\right\}, \\
&\quad\times\exp\left\{-\frac{(1+\theta^2)\left(\lambda-\left(\frac{\theta\tau M_3}{\sqrt{1+\theta^2}}+h\right)\right)^2}{2\tau^2}\right\}, \\
f(M_3|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\boldsymbol{\psi},\lambda) &\propto \exp\left\{-\frac{(1+\theta^2)}{2}\left(M_3-\frac{\theta}{\sqrt{1+\theta^2}}\frac{(\lambda-h)}{\tau}\right)^2\right\}\mathbf{1}_{[0,\infty)}\{M_3\},
\end{aligned}
$$

where $\mathbf{1}_A\{x\}$ is the indicator function assuming 1 if $x \in A$. Notice that the posterior full conditional distribution for each component ψ_i of $\boldsymbol{\psi}$ is the normal distribution truncated below zero, with mean $(\gamma_i+\sigma\delta\sqrt{2/\pi})\delta(\sigma)^{-1}$ and variance $1-\delta^2$.

For the dependent case, we also have that the posterior distribution has no closed form. In this case, it is given by

$$f(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\alpha|\mathbf{D}) \propto \left[\prod_{i=1}^{k}\prod_{j=1}^{n_i}(\pi_{ij})^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}}\right]\left(\frac{1}{\sigma^2}\right)^{\frac{k+2a+2}{2}}\exp\left\{-\frac{d}{\sigma^2}\right\}$$
$$\times\exp\left\{-\frac{\sum_{i=0}^{p}(\beta_i-m_i)^2}{2b^2}-\frac{\sum_{i=1}^{k}\left(\gamma_i+\Delta\sigma\sqrt{\frac{2}{\pi}}\right)^2}{2\sigma^2}-\frac{(\alpha-h)^2}{2\tau^2}\right\}$$
$$\times\Phi\left(\frac{\alpha}{\sigma}\sum_{i=1}^{k}\left(\gamma_i+\Delta\sigma\sqrt{\frac{2}{\pi}}\right)\right).$$

To obtain the posterior f.c.d. we also consider the stochastic representation for a multivariate SN distribution (Azzalini and Dalla-Valle 1996) which establishes that if $\boldsymbol{\gamma}\sim SN_k(\mathbf{0}_k,\mathbf{I}_k,\alpha\mathbf{1}_k)$ then $\boldsymbol{\gamma}\overset{d}{=}\Delta|U|+\mathbf{V}$, where $U\sim N(0,1)$ and $\mathbf{V}\sim N_k(\mathbf{0}_k,\mathbf{I}_k-\left(1+k\alpha^2\right)^{-1}\alpha^2\mathbf{1}_k\mathbf{1}_k^t)$. Let us assume $T=|U|$. Consequently, if $\boldsymbol{\gamma}\sim SN_k(-\Delta\sigma\sqrt{2/\pi}$, $\sigma^2\mathbf{I}_k$, $\mathbf{1}_k\alpha)$, then $\boldsymbol{\gamma}\overset{d}{=}\Delta\sigma\left(T-\sqrt{2/\pi}\right)+\sigma\mathbf{V}$. Let $\mathbf{F}=\sigma\left(T-\sqrt{2/\pi}\right)$. The posterior f.c.d. are thus given by

$$f(\boldsymbol{\gamma}|\mathbf{D},\boldsymbol{\beta},\sigma^2,T,\alpha) \propto \left[\prod_{i=1}^{k}\prod_{j=1}^{n_i}\left[\frac{\exp\{\eta_{ij}\}}{1+\exp\{\eta_{ij}\}}\right]^{y_{ij}}\left[\frac{1}{1+\exp\{\eta_{ij}\}}\right]^{1-y_{ij}}\right]$$
$$\times\exp\left\{-\frac{1}{2\sigma^2}[\boldsymbol{\gamma}-\Delta\mathbf{F}]'\left[\mathbf{I}_k+\frac{\Delta^2}{1-k\Delta^2}\mathbf{1}_k\mathbf{1}_k'\right][\boldsymbol{\gamma}-\Delta\mathbf{F}]\right\},$$

$$f(T|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,\alpha) \propto \exp\left\{\frac{-\left[T^2-2T\left(\frac{\Delta}{\sigma}\sum_{i=1}^{k}\gamma_i+k\Delta^2\sqrt{2/\pi}\right)\right]}{2(1-k\Delta^2)}\right\}\mathbf{1}_{(0,\infty)}(T),$$

$$f(\alpha|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},\sigma^2,T) \propto (1+k\alpha^2)^{\frac{1}{2}}\exp\left\{-\frac{(\alpha-h)^2}{2\tau^2}-\frac{1}{2\sigma^2}A'\left[\mathbf{I}_k+\alpha^2\mathbf{1}_k\mathbf{1}_k'\right]A\right\},$$

$$f(\sigma^2|\mathbf{D},\boldsymbol{\beta},\boldsymbol{\gamma},T,\alpha) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{k}{2}+a+1}\exp\left\{-\frac{d}{\sigma^2}\right\}$$
$$\times\exp\left\{-\frac{1}{2\sigma^2}[\boldsymbol{\gamma}-\Delta\mathbf{F}]'\left[\mathbf{I}_k+\frac{\Delta^2}{1-k\Delta^2}\mathbf{1}_k\mathbf{1}_k'\right][\boldsymbol{\gamma}-\Delta\mathbf{F}]\right\},$$

where $A=\boldsymbol{\gamma}-\frac{\alpha\mathbf{F}}{\sqrt{1+k\alpha^2}}\mathbf{1}_k$.

The posterior f.c.d. for $\boldsymbol{\beta}$ is as that obtained in the independent case, previously discussed. In both cases considered here, the posterior f.c.d. for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are log-concave and the adaptive rejection sampling (ARS) algorithm can be used to generate from their posteriors. The rejection method and the adaptive rejection Metropolis sampling

(ARMS) algorithm are also used to sample from the other posterior distributions. For details of these algorithms, see Gilks et al. (1995) and Gilks and Wild (1992) for instance.

Remark: If $\boldsymbol{\gamma} \sim SN_k(\mathbf{0}, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$, then any $r \times 1$ sub-vector of $\boldsymbol{\gamma}$ has an r-variate $SN_r(\mathbf{0}, \sigma^2\mathbf{I}_r, \alpha^*\mathbf{1}_r)$ distribution, where $\alpha^* = \alpha[1 + \alpha^2(k-r)]^{-1/2}$. Consequently, the marginal distribution of any γ_i tends to a normal distribution if there is a high number of clusters. Moreover, the degree of skewness in the marginal distribution belongs to the interval $(-[k-1]^{-1/2}; [k-1]^{-1/2})$. For any two components γ_l and γ_j of $\boldsymbol{\gamma}$, it also follows that $Corr(\gamma_l, \gamma_j) = -2(\alpha^*)^2[\pi + 2(\alpha^*)^2(\pi-2)]^{-1}$, where $\alpha^* = \alpha[1+\alpha^2(k-r)]^{-1/2}$. See details in Azzalini and Dalla-Valle (1996).

5 Simulation studies

In this section, we have as the main goal to evaluate the effect of the misspecification of the random effects distributions in the posterior odds ratio. We generate four datasets from mixed logistic regression models. We consider 25 clusters as well as two covariates and their interaction. The datasets (Data 1, Data 2, Data 3 and Data 4) are in Appendix B. The difference among the datasets lies in the distribution used to generate the random effects. For Data 1 and Data 2, the random effects were generated from i.i.d. normal distributions, centered around zero and with variance 1 and 4, respectively. For Data 3 and Data 4, we assumed i.i.d. skew-normal distributions which put most of their mass in positive values, that is the random effects are positive values with very high probability. We assume $\boldsymbol{\gamma} \overset{ind}{\sim} SN(-1.32, 2.75, 35)$ and $\boldsymbol{\gamma} \overset{ind}{\sim} SN(-2.64, 10.99, 35)$ for datasets 3 and 4, respectively. For comparison proposal, such skew-normal distributions have both mean equal to zero and variance equal to 1, for Data 3, and 4, for Data 4.

We also fit four different models by assuming different prior distributions for the random effects. For simplicity, we assume only i.i.d. distributions for the random effects. In Model 1 we consider that $\gamma_i \overset{iid}{\sim} N(0, \sigma^2)$. Models 2, 3 and 4 assume that $\gamma_i \overset{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$. However, we assume different prior specifications for the skewness parameter: $\lambda \sim N(0, 1000)$, in Model 2, $\lambda \sim SN(0, 2578.21, 5)$, in Model 3, and in Model 4 we consider $\lambda \sim SN(0, 2751.46, 100)$. In all cases the prior variance for λ is 1,000. Moreover, the prior distributions for λ in Models 3 and 4 put most of their mass in positive values. The prior mode and mean for λ in Model 3 (Model 4) are 18.81 (1.96) and 39.73 (41.85), respectively. In our analysis, we also assume few informative prior distributions for the fixed effects ($\beta_i \sim N(0, 10)$, $i = 0, \ldots, 3$) and for σ^2 ($\sigma^2 \sim IG(2.001, 1)$). Consequently, *a priori*, we have that $E(\sigma^2) = 1$ and $V(\sigma^2) = 1,000$.

For the MCMC, we ran a chain of size 30,000 (50,000) for the normal (skew-normal) cases, discarded the first 20,000 samples as the burn-in period and used a lag of 10 (30) steps to avoid serial correlation obtaining a sample of size 1,000. The algorithm was implemented using the Ox software.

Although the deviance information criterion (DIC) and the conditional predictive ordinate (CPO) in Table 1 point out that Model 1 must be preferred for Data 1 and Data

2 and Models 2 and 3 are the best for Data 4 and Data 3, respectively, such measures do not provide substantial evidence to support any of these models. Moreover, we observe from Table 2 that the posterior point estimates for the fixed effects and for σ^2 do not substantially differ. On the other hand, the posterior HPD interval proves to be a good auxiliary tool for model selection. For all parameters, the posterior 95% HPD intervals disclose that the posterior uncertainty about the parameters is smaller than *a priori*. Moreover, if the random effects distribution is not well specified, the magnitude of the HPD intervals for the fixed effects and σ^2 tends to be high. The skewness parameter is in general poorly estimated mainly if we have a strong degree of asymmetry in the random effects. The posterior 95% HPD intervals reveal that the posterior uncertainty about λ is still very high. It is also noticeable that, for almost all cases, λ is not significantly different from zero, exceptions occur for Data 3, under Models 3 and 4, and for Data 4 under Model 4. The mean square error (MSE) presented in Table 1 shows that the

Table 1: Model comparison

	MSE for γ				
Model	Mean	Median	Mode	CPO	DIC
	Data 1				
1	3.05	2.95	2.98	-1631.84	3263.18
2	2.78	2.93	3.73	-1632.14	3263.91
3	3.61	3.85	4.86	-1632.17	3263.89
4	3.83	3.83	4.09	-1631.82	3263.21
	Data 2				
1	12.69	12.30	12.456	-1402.09	2803.93
2	20.13	19.66	19.12	-1402.46	2804.41
3	14.22	14.67	15.72	-1402.57	2804.86
4	13.35	13.26	14.43	-1402.77	2804.98
	Data 3				
1	5.03	5.12	5.30	-1568.09	3136.27
2	3.53	3.41	3.37	-1567.46	3135.01
3	3.17	3.01	3.12	-1566.81	3133.63
4	3.35	3.33	3.56	-1567.81	3135.63
	Data 4				
1	16.30	16.76	18.79	-1527.95	3056.05
2	12.01	11.40	11.30	-1526.32	3052.88
3	9.92	9.77	9.38	-1526.37	3053.15
4	10.26	10.68	11.74	-1526.61	3053.60

misspecification of the random effects distributions provide poor posterior estimates for the random effects (see also Figure 4) and, because the OR depends on the $\boldsymbol{\gamma}$s, for the odds ratio as well (see Table 3). Figure 4 shows the plug-in estimates for the p.d.f. of the random effects assuming the posterior means (column 1), medians (column 2) and modes (column 3) for all models and datasets. The graphics in line k correspond to the estimates for Data k, $k = 1, \ldots, 4$. It is noticeable from column 3 that, using the posterior modes, all models provide good approximations for the true density for Data 1 and Data 2. That is an expected result since the skew-normal family of distributions includes the normal one as a particular case. However, that is not necessarily observed for the plug-in estimates using the posterior means and medians. Using the posterior means, for instance, Models 3 and 4 for Data 1 and Model 2, for Data 2, point out great degrees of skewness for the random effects distributions. For Data 3 and Data 4, Model 1 works poorly (which is expected, since the normal distribution does not capture asymmetric behavior). The plug-in estimates under Models 2, 3 and 4 using the posterior mean and median tend to be closer to the true density than the ones obtained using the posterior mode.

Table 3 presents some posterior summaries of the odds ratios that compare individuals in different clusters with the same covariates (OR_1, OR_2 and OR_3) as well as with different covariates (OR_4, OR_5 and OR_6). The posterior estimates for the OR can truly differ from one model to another. The posterior distribution for the OR can be strongly asymmetric since the posterior summaries for its distribution can be quite far from each other (e.g., for the OR_6 in Data 2). Contrary to what is suggested by Larsen et al. (2000), the median of the OR distribution does not necessarily provide the best estimate for the OR. If we have strong variation between clusters the posterior means tend to provide better estimates for the OR, if there is strong asymmetry between clusters (Data 4). However if we have symmetry between clusters (Data 2), the posterior means (medians) provide better estimates for individual with the same (different) covariates. For datasets with weak variation between clusters (Data 1 and 3), the posterior median usually is not the best estimator for the OR. In these cases, the posterior means and modes tend to provide better estimates. If we have no (strong) asymmetry between clusters the posterior modes (means) of the OR tend to be better if we are comparing individuals with the same covariates.

Table 2: Posterior estimates for $\boldsymbol{\beta}$, σ^2 and λ, all models and datasets

	True	Mean	Median	Mode	HPD	True	Mean	Median	Mode	HPD
		Data 1					**Data 2**			
		Model 1					Model 1			
β_0	-0.10	-0.15	-0.15	-0.17	[-0.52,0.35]	-0.10	-0.12	-0.12	-0.06	[-0.87,0.68]
β_1	1.16	1.32	1.32	1.36	[0.86,1.73]	1.16	1.17	1.17	1.10	[0.26,2.00]
β_2	0.48	0.32	0.32	0.34	[-0.14,0.76]	0.48	0.96	0.96	0.94	[-0.05,2.09]
β_3	-0.78	-0.76	-0.77	-0.80	[-1.26,-0.32]	-0.78	-1.29	-1.27	-1.25	[-2.27,-0.49]
$V(\gamma)$	1	1.11	1.04	0.95	[0.54,1.76]	4	4.31	4.01	3.60	[1.79,7.38]
		Model 2					Model 2			
β_0	-0.10	-0.17	-0.18	-0.19	[-0.58,0.29]	-0.10	-0.24	-0.23	-0.22	[-1.11,0.58]
β_1	1.16	1.32	1.33	1.34	[0.84,1.83]	1.16	1.02	1.00	0.90	[0.15,1.82]
β_2	0.48	0.36	0.36	0.39	[-0.10,0.82]	0.48	1.06	1.04	0.99	[0.14,2.11]
β_3	-0.78	-0.74	-0.74	-0.77	[-1.31,-0.28]	-0.78	-1.34	-1.34	-1.17	[-2.38,-0.40]
λ	0	0.48	0.33	0.19	[-52.41,39.83]	0	-9.95	-2.26	-0.70	[-43.10,3.08]
$V(\gamma)$	1	1.22	1.15	1.06	[0.55,2.07]	4	4.87	4.50	3.62	[2.35,8.79]
		Model 3					Model 3			
β_0	-0.1	-0.22	-0.23	-0.22	[-0.72,0.24]	-0.10	-0.26	-0.27	-0.21	[-1.07,0.46]
β_1	1.16	1.30	1.30	1.25	[0.84,1.71]	1.16	1.11	1.10	1.04	[0.33,2.15]
β_2	0.48	0.28	0.26	0.19	[-0.08,0.73]	0.48	1.00	0.99	0.94	[0.18,1.86]
β_3	-0.78	-0.70	-0.69	-0.67	[-1.23,-0.21]	-0.78	-1.28	-1.27	-1.17	[-2.15,-0.32]
λ	0	16.97	4.03	1.12	[-7.25,74.12]	0	-0.53	-0.54	-0.06	[-12.80,3.35]
$V(\gamma)$	1	1.23	1.14	1.01	[0.50,2.07]	4	4.59	4.28	3.75	[1.93,8.05]
		Model 4					Model 4			
β_0	-0.10	-0.20	-0.20	-0.19	[-0.68,0.22]	-0.10	-0.28	-0.28	-0.43	[-1.11,0.48]
β_1	1.16	1.32	1.33	1.33	[0.84,1.78]	1.16	1.16	1.14	1.17	[0.36,2.01]
β_2	0.48	0.28	0.27	0.24	[-0.12,0.79]	0.48	0.95	0.94	0.86	[0.10,1.78]
β_3	-0.78	-0.69	-0.68	-0.68	[-1.13,-0.18]	-0.78	-1.39	-1.35	-1.28	[-2.42,-0.55]
λ	0	3.37	1.70	0.84	[-0.94,14.21]	0	0.74	0.57	0.45	[-0.76,2.63]
$V(\gamma)$	1	1.17	1.09	0.98	[0.52,1.99]	4	4.37	4.08	3.62	[2.08,7.34]
		Data 3					**Data 4**			
		Model 1					Model 1			
β_0	-0.1	-0.04	-0.04	-0.04	[-0.42,0.37]	-0.10	-0.22	-0.24	-0.26	[-0.88,0.42]
β_1	1.16	1.50	1.50	1.53	[1.10,1.94]	1.16	1.03	1.04	1.00	[0.34,1.79]
β_2	0.48	0.82	0.82	0.86	[0.40,1.29]	0.48	0.92	0.92	0.96	[0.25,1.68]
β_3	-0.78	-0.72	-0.71	-0.69	[-1.20,-0.22]	-0.78	-0.35	-0.38	-0.53	[-1.29,0.59]
$V(\gamma)$	1	0.91	0.85	0.79	[0.39,1.55]	4	2.81	2.61	2.24	[1.34,4.95]
		Model 2					Model 2			
β_0	-0.1	-0.06	-0.07	-0.13	[-0.39,0.39]	-0.10	-0.25	-0.26	-0.25	[-0.94,0.42]
β_1	1.16	1.45	1.44	1.38	[1.10,1.83]	1.16	1.23	1.24	1.12	[0.64,1.95]
β_2	0.48	0.68	0.66	0.64	[0.32,1.07]	0.48	0.84	0.83	0.76	[0.29,1.48]
β_3	-0.78	-0.69	-0.69	-0.69	[-1.05,-0.27]	-0.78	-0.33	-0.37	-0.40	[-0.98,0.50]
λ	35	22.45	19.9	8.59	[-2.03,64.28]	35	24.63	20.81	5.78	[-1.15,64.28]
$V(\gamma)$	1	0.83	0.78	0.70	[0.41,1.38]	4	2.90	2.71	2.37	[1.30,4.96]
		Model 3					Model 3			
β_0	-0.1	-0.05	-0.06	-0.10	[-0.35,0.36]	-0.10	-0.26	-0.27	-0.37	[-0.74,0.34]
β_1	1.16	1.40	1.40	1.39	[1.10,1.72]	1.16	1.18	1.20	1.34	[0.56,1.84]
β_2	0.48	0.64	0.63	0.60	[0.36,0.98]	0.48	0.82	0.79	0.73	[0.29,1.46]
β_3	-0.78	-0.65	-0.66	-0.70	[-1.03,-0.28]	-0.78	-0.36	-0.35	-0.34	[-0.93,0.32]
λ	35	48.63	42.63	21.22	[1.48,113.22]	35	43.08	38.14	23.00	[-0.51,102.06]
$V(\gamma)$	1	0.84	0.79	0.74	[0.40,1.41]	4	2.84	2.73	2.80	[1.28,4.71]
		Model 4					Model 4			
β_0	-0.1	-0.05	-0.07	-0.08	[-0.42,0.29]	-0.10	-0.20	-0.23	-0.23	[-0.82,0.57]
β_1	1.16	1.42	1.42	1.42	[1.10,1.77]	1.16	1.27	1.27	1.25	[0.52,1.94]
β_2	0.48	0.68	0.66	0.66	[0.34,1.02]	0.48	0.82	0.80	0.77	[0.22,1.43]
β_3	-0.78	-0.69	-0.69	-0.70	[-1.07,-0.28]	-0.78	-0.35	-0.36	-0.32	[-0.98,0.32]
λ	35	12.62	12.96	20.33	[1.64,21.43]	35	15.25	16.86	21.39	[2.80,21.82]
$V(\gamma)$	1	0.83	0.78	0.66	[0.41,1.50]	4	3.02	2.78	2.49	[1.28,5.10]

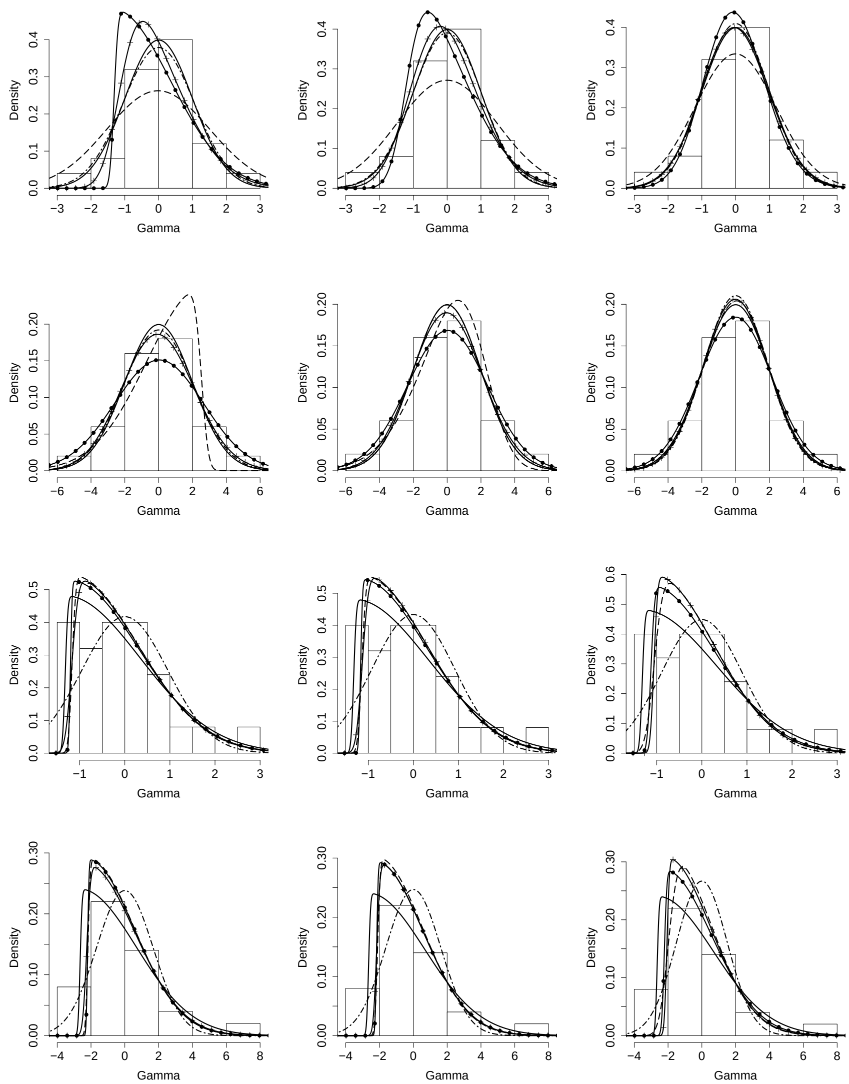

Figure 4: Histogram of the generated random effects, the true (solid line) and the plug-in estimates for the random effect densities using the posterior mean (1st column), median (2nd column) and mode (3rd column), under Model 1 (dotdashed line), Model 2 (dashed line), Model 3 (•) and Model 4 (+).

Table 3: Posterior estimates for the odds ratio

	True	Mean	Median	Mode	HPD	True	Mean	Median	Mode	HPD
		Data 1					**Data 2**			
		Model 1					Model 1			
OR_1	68.05	43.77	39.96	36.78	[15.06,79.95]	1.75	1.75	1.65	1.51	[0.91,3.00]
OR_2	2.91	3.69	3.53	3.35	[1.73,5.72]	14.41	15.94	15.20	13.10	[7.34,25.70]
OR_3	1.66	2.18	2.06	1.90	[1.13,3.60]	6.09	5.34	5.11	4.83	[2.91,8.46]
OR_4	5.69	5.97	5.22	4.26	[1.92,11.93]	1.60	9.61	2.51	0.66	[0,34.18]
OR_5	108.77	150.72	120.06	92.63	[34.77,338.70]	244.40	109.84	81.31	61.50	[19.73,282.15]
OR_6	1.58	1.63	1.57	1.49	[0.97,2.40]	538.73	1233.35	480.02	203.35	[54.56,4986.74]
		Model 2					Model 2			
OR_1	68.05	39.61	36.87	31.00	[15.67,69.72]	1.75	1.79	1.67	1.47	[0.78,2.88]
OR_2	2.91	3.79	3.58	3.34	[1.82,6.32]	14.41	16.16	15.36	13.39	[7.44,26.32]
OR_3	1.66	2.24	2.15	1.88	[1.00,3.53]	6.09	5.51	5.37	5.43	[2.92,8.41]
OR_4	5.69	5.70	5.08	4.43	[1.50,11.06]	1.60	136.09	2.04	0.35	[0.00,101.42]
OR_5	108.77	160.26	126.26	90.71	[28.89,386.59]	244.40	82.20	61.73	44.89	[14.12,200.95]
OR_6	1.58	1.60	1.55	1.45	[0.90,2.34]	538.73	1937.18	529.76	223.08	[27.57,5203.58]
		Model 3					Model 3			
OR_1	68.05	39.49	36.64	31.40	[14.96,69.86]	1.75	1.79	1.69	1.60	[0.87,3.04]
OR_2	2.91	3.74	3.55	3.37	[1.71,5.96]	14.41	15.82	15.01	13.48	[7.79,26.24]
OR_3	1.66	2.20	2.09	1.96	[1.08,3.68]	6.09	5.50	5.27	4.81	[2.69,8.45]
OR_4	5.69	5.80	5.17	4.40	[1.50,11.88]	1.60	21.43	2.49	0.56	[0.00,38.13]
OR_5	108.77	164.98	131.58	92.36	[21.66,395.93]	244.40	100.73	74.16	42.66	[15.25,266.41]
OR_6	1.58	1.61	1.55	1.45	[0.92,2.34]	538.73	1457.14	514.93	220.62	[40.66,4162.15]
		Model 4					Model 4			
OR_1	68.05	42.13	39.13	32.71	[17.05,75.78]	1.75	1.75	1.65	1.50	[0.79,2.88]
OR_2	2.91	3.73	3.57	3.39	[1.69,5.97]	14.41	15.89	15.09	13.24	[7.67,26.82]
OR_3	1.66	2.18	2.08	1.87	[1.07,3.65]	6.09	5.36	5.17	5.22	[2.96,8.50]
OR_4	5.69	5.92	5.14	4.27	[1.71,12.16]	1.60	24.91	3.03	0.79	[0.02,42.67]
OR_5	108.77	152.70	122.66	103.22	[24.92,358.69]	244.40	110.21	82.40	62.18	[15.17,265.10]
OR_6	1.58	1.61	1.56	1.40	[0.92,2.34]	538.73	1449.51	511.03	202.53	[39.81,4308.76]
		Data 3					**Data 4**			
		Model 1					Model 1			
OR_1	4.03	4.29	4.05	3.68	[1.86,7.05]	1.98	1.13	1.09	1.00	[0.63,1.71]
OR_2	2.34	2.26	2.22	2.11	[1.28,3.24]	59.34	81.02	71.50	52.95	[24.99,166.57]
OR_3	1.85	1.31	1.23	1.09	[0.59,2.14]	1160.22	133.73	68.14	43.64	[9.58,439.28]
OR_4	53.48	28.57	26.29	21.42	[12.12,52.84]	2.10	1.88	1.77	1.72	[0.85,3.01]
OR_5	9.30	17.17	16.12	14.80	[7.18,30.10]	2.21	1.55	1.28	0.96	[0.16,3.51]
OR_6	15.77	21.12	20.28	19.75	[11.62,32.39]	1.82	1.96	1.85	1.66	[0.89,3.38]
		Model 2					Model 2			
OR_1	4.03	4.33	4.13	4.00	[1.98,7.24]	1.98	1.14	1.11	1.11	[0.64,1.78]
OR_2	2.34	2.20	2.16	2.00	[1.38,3.26]	59.34	64.75	58.72	52.84	[26.77,123.12]
OR_3	1.85	1.32	1.24	1.12	[0.60,2.22]	1160.22	331.00	115.55	55.39	[10.30,1194.05]
OR_4	53.48	30.93	28.10	25.10	[11.95,57.29]	2.10	1.87	1.79	1.70	[0.93,3.01]
OR_5	9.30	17.45	16.00	14.14	[6.32,29.29]	2.21	1.73	1.54	1.28	[0.52,3.37]
OR_6	15.77	19.63	18.76	17.20	[10.59,30.35]	1.82	1.97	1.91	1.89	[1.00,3.19]
		Model 3					Model 3			
OR_1	4.03	4.44	4.24	4.00	[2.21,7.19]	1.98	1.14	1.09	0.98	[0.61,1.72]
OR_2	2.34	2.17	2.13	2.11	[1.27,3.12]	59.34	60.24	54.93	45.56	[24.14,110.90]
OR_3	1.85	1.32	1.26	1.15	[0.65,2.18]	1160.22	1179.08	133.44	53.33	[13.74,3093.65]
OR_4	53.48	31.27	29.32	27.41	[13.73,57.79]	2.10	1.88	1.81	1.66	[0.86,2.98]
OR_5	9.30	17.48	16.49	14.18	[7.71,30.54]	2.21	1.79	1.65	1.50	[0.47,3.32]
OR_6	15.77	19.26	18.58	16.64	[10.31,29.25]	1.82	2.01	1.92	1.62	[0.96,3.30]
		Model 4					Model 4			
OR_1	4.03	4.31	4.14	3.58	[2.01,7.10]	1.98	1.11	1.09	1.08	[0.61,1.68]
OR_2	2.34	2.17	2.13	2.16	[1.34,3.20]	59.34	64.34	58.40	48.15	[22.01,117.00]
OR_3	1.85	1.32	1.24	1.07	[0.54,2.15]	1160.22	979.37	132.36	50.01	[13.28,3209.89]
OR_4	53.48	31.02	28.21	25.50	[12.98,59.37]	2.10	1.86	1.77	1.58	[0.89,3.03]
OR_5	9.30	17.36	16.23	13.36	[7.46,30.70]	2.21	1.74	1.55	1.32	[0.46,3.60]
OR_6	15.77	19.52	18.57	16.81	[10.66,30.74]	1.82	2.03	1.93	1.80	[0.96,3.46]

6 Case Study: Teratogenic activity experiment

In this section we consider the dataset reported in Liu and Dey (2008). The goal was to study the interaction between teratogenic activity of the two niacin analogs *6-aminonicotinamide*(6AN) and *3 acetylpyridine*(3AP). Eggs were incubated and, after 96 hours of incubation, both substances were injected at the same time and dissolved together. The experiment was done in four independent months.

We consider a Bernoulli model where the response variable assumes value 1 if the egg generates an abnormal fowl at the end of the experiment. The covariates are 6AN, 3AP and their interaction. The covariates are standarized. We consider two definitions for the clusters, the months (Data 1) as in Liu and Dey (2008) and the slots (Data 2). The total numbers of clusters k in Data 1 and Data 2 are 4 and 16, respectively.

We fit the five models that differ from each other because different distributions for the random effects are assumed. In Model 1 we consider $\gamma_i \overset{iid}{\sim} N(0, \sigma^2)$. Models 2 and 3 assume i.i.d. skew-normal random effects, that is, $\gamma_i \overset{iid}{\sim} SN(-\delta\sigma\sqrt{2/\pi}, \sigma^2, \lambda)$, with different prior information about the skewness parameter - for Model 2, $\lambda \sim N(-8.5, 100)$ and for Model 3, $\lambda \sim N(0, 100)$. Correlated skew-normal random effects are considered in Models 4 and 5 for which we assume $\boldsymbol{\gamma} \sim SN(-\Delta\sigma\sqrt{2/\pi}\mathbf{1}_k, \sigma^2\mathbf{I}_k, \alpha\mathbf{1}_k)$, where $\Delta = \alpha[1 + k\alpha^2]^{-1/2}$, and different prior distributions for α, say, $\alpha \sim N(-8.5, 100)$ for Model 4 and $\alpha \sim N(0, 100)$ for Model 5. For comparison proposal, we considered the posterior information provided by Liu and Dey (2008) in order to build the prior distributions of the skewness parameters in Models 2 and 4. In these cases we are assuming *a priori* that λ and α are negative with high probability and, consequently, we have a strong evidence in favor of negative asymmetry for the random effects. In all cases, however, we consider flat prior distributions for the skewness parameter. To complete the model specification, we assume flat prior distributions for β_i and σ^2, that is, we assume $\beta_i \sim N(0, 10)$, for all $i = 0, \ldots, 3$, and $\sigma^2 \sim IG(2.001, 1)$. Notice that under these prior specifications for Models 1, 3 and 5, given σ^2 and κ_{12}, we are assuming that the prior distributions of OR have similar behavior and put substantial probability mass on values up to one (see case $\lambda = 0$ in Figure 1).

For the MCMC, we ran chains of sizes 30,000, 50,000 and 70,000, respectively, for the normal, independent skew-normal and correlated skew-normal cases. We discarded the first 20,000 (normal and independent skew-normal cases) and 40,000 (correlated skew-normal) samples as the burn-in periods and used lags of 10 (Normal case) and 30 (independent and correlated skew-normal cases) steps to avoid serial correlation obtaining samples of size 1,000 in all models.

Table 4 presents the posterior point estimates and the HPD intervals with 95% probability for the fixed effects, the variance for the random effects and the skewness parameter. Comparing the models, for both datasets, the posterior point estimates for the fixed effects obtained under the SN models are similar to the ones obtained under the normal model, but the estimates for the variance of the random effect distribution $V(\gamma_i)$ differ. The posterior estimates indicate that both niacin analogs, 6AN and 3AP, have positive effects in the teratogenic activity and their interaction has negative one. Also, the intercept is not relevant to explain the teratogenic activity. The variances of the random effect distributions $V(\gamma_i)$ are smaller than 1.0, with high posterior probability, in all cases. To fairly compare the models, assume only the models in which the posterior information provided by Liu and Dey (2008) is not considered. In these cases, the 95.0% HPD intervals for the fixed effects and $V(\gamma_i)$ are tighter under Models 3 and 5 for both datasets. Thus, we have evidence against normality for the random effects. The distributions adopted to describe the prior uncertainty about the skewness parameter do not strongly influence the posterior inference for the variances of the random effects (for instance, the posterior mode for $V(\gamma_i)$ is 0.14 under Model 2, and is 0.13 under Model 3). However, the estimates for $V(\gamma_i)$ tend to increase with the number of clusters. To give an example, under Model 2, the posterior modes for $V(\gamma_i)$ in Data 1 and Data 2 are 0.14 and 0.25, respectively. Similar behaviors are observed for the fixed effects and the skewness parameter. A great reduction in the uncertainty about the skewness parameters λ and α was observed *a posteriori* for both datasets and all models.

Table 4: Posterior estimates for the parameters, Teratogenic activity experiment

	Mean	Median	Mode	HPD	Mean	Median	Mode	HPD
	Data 1				**Data 2**			
	Model 1				Model 1			
	DIC=2453.353, CPO= -1226.711				DIC=2361.254, CPO= -1180.818			
β_0	-0.22	-0.22	-0.29	[-0.78,0.37]	-0.30	-0.30	-0.32	[-0.64,0.06]
β_1	1.07	1.07	1.07	[0.91,1.23]	1.26	1.26	1.30	[0.86,1.67]
β_2	0.56	0.56	0.56	[0.42,0.70]	0.67	0.67	0.67	[0.30,1.05]
β_3	-0.71	-0.70	-0.69	[-0.93,-0.53]	-0.78	-0.78	-0.76	[-1.15,-0.38]
$V(\gamma_i)$	0.44	0.37	0.27	[0.10,0.97]	0.44	0.40	0.38	[0.17,0.79]
	Model 2				Model 2			
	DIC=2453.154, CPO= -1226.613				DIC=2360.162, CPO= -1180.198			
β_0	-0.28	-0.28	-0.29	[-0.83,0.16]	-0.29	-0.28	-0.24	[-0.60,-0.01]
β_1	1.08	1.07	1.07	[0.92,1.24]	1.32	1.32	1.34	[1.02,1.65]
β_2	0.57	0.57	0.55	[0.44,0.71]	0.68	0.68	0.69	[0.41,0.99]
β_3	-0.72	-0.72	-0.72	[-0.93,-0.56]	-0.77	-0.77	-0.84	[-1.10,-0.41]
λ	-7.88	-7.57	-6.61	[-27.89,11.58]	-11.83	-11.02	-11.84	[-27.55,3.19]
$V(\gamma_i)$	0.24	0.19	0.14	[0.04,0.54]	0.32	0.28	0.25	[0.12,0.60]
	Model 3				Model 3			
	DIC=2453.725, CPO= -1226.914				DIC=2360.621, CPO= -1180.419			
β_0	-0.22	-0.22	-0.29	[-0.68,0.18]	-0.29	-0.29	-0.27	[-0.59,0.06]
β_1	1.07	1.07	1.04	[0.90,1.23]	1.27	1.28	1.28	[0.82,1.63]
β_2	0.56	0.56	0.57	[0.42,0.69]	0.65	0.66	0.66	[0.26,0.96]
β_3	-0.72	-0.71	-0.71	[-0.93,-0.53]	-0.77	-0.76	-0.79	[-1.12,-0.41]
λ	1.14	2.03	5.68	[-19.11,20.06]	-6.26	-5.54	-3.78	[-22.07,9.34]
$V(\gamma_i)$	0.22	0.18	0.13	[0.05,0.50]	0.35	0.31	0.24	[0.11,0.70]
	Model 4				Model 4			
	DIC=2453.436, CPO= -1226.778				DIC=2361.036, CPO= -1180.699			
β_0	-0.26	-0.27	-0.30	[-0.58,0.18]	-0.30	-0.31	-0.34	[-0.59,-0.04]
β_1	1.07	1.07	1.09	[0.92,1.23]	1.26	1.26	1.24	[0.87,1.66]
β_2	0.57	0.56	0.55	[0.44,0.71]	0.64	0.65	0.68	[0.25,1.07]
β_3	-0.72	-0.71	-0.68	[-0.90,-0.51]	-0.76	-0.76	-0.74	[-1.14,-0.36]
α	-8.28	-8.61	-12.07	[-28.67,9.69]	-8.78	-8.75	-8.98	[-29.46,10.41]
$V(\gamma_i)$	0.38	0.31	0.24	[0.09,0.85]	0.43	0.39	0.33	[0.17,0.82]
	Model 5				Model 5			
	DIC=2453.095, CPO= -1226.604				DIC=2360.549, CPO= -1180.454			
β_0	-0.23	-0.24	-0.27	[-0.72,0.21]	-0.30	-0.30	-0.29	[-0.55,-0.06]
β_1	1.07	1.07	1.09	[0.92,1.24]	1.23	1.24	1.25	[0.84,1.59]
β_2	0.56	0.56	0.55	[0.42,0.69]	0.66	0.66	0.64	[0.19,1.04]
β_3	-0.72	-0.71	-0.73	[-0.89,-0.51]	-0.77	-0.76	-0.71	[-1.15,-0.35]
α	0.05	-0.07	2.14	[-17.9,20.76]	0.26	0.22	-0.85	[-18.43,20.28]
$V(\gamma_i)$	0.40	0.32	0.23	[0.08,0.91]	0.42	0.38	0.33	[0.16,0.76]

Considering the 95% HPD intervals and point estimates of λ and α, we have the highest evidence in favor of a negative degree of skewness in the random effects distributions for Data 1, under Models 2 and 4, and for Data 2, under Models 2, 3 and 4. That is in agreement with the conclusions in Liu and Dey (2008). For Data 2 and all models we also noticed that the asymmetry is higher *a posteriori*. However, the HPD intervals also disclose that λ and α are not significatively different from zero pointing to a normal distribution for the random effects. Despite this, taking into consideration the DIC and the CPO, skew-normal models are selected as the best models (Model 2 for Data 2, and Model 5 for Data 1). The CPO and the DIC also indicate an improvement in the model fitting by assuming the slots as the clusters.

Table 5 provides posterior summaries of the odds ratio distributions for some particular comparisons: OR_1 compares the eggs of slots 4 and 5, November 1961; OR_2 compares the eggs of slots 2 and 3, April 1962; OR_3 compares the eggs of slot 3, January 1962, and of slot 5, November 1961; OR_4 compares the eggs of slot 3, April 1962, and of slot 5, November 1961; OR_5 compares the eggs of slot 1, January 1962, and of slot 1, December 1961; and, finally, OR_6 compares the eggs of slot 2, December 1961, and of slot 5, April 1962. The posterior means, modes and medians for all odds ratios

Table 5: Posterior summaries for some OR, Teratogenic activity experiment

	Mean	Median	Mode	HPD	Mean	Median	Mode	HPD
	Data 1				**Data 2**			
	Model 1				Model 1			
OR_1	1.48	1.48	1.48	[1.34,1.63]	1.32	1.27	1.15	[0.70,2.03]
OR_2	4.28	4.26	4.22	[3.51,5.34]	1.55	1.49	1.42	[0.79,2.33]
OR_3	2.15	2.12	2.15	[1.46,2.87]	1.85	1.78	1.53	[1.01,2.72]
OR_4	1.29	1.28	1.25	[0.95,1.57]	0.97	0.93	0.86	[0.50,1.55]
OR_5	109.18	98.4	79.89	[38.59,206.49]	170.55	143.00	103.94	[45.99,367.00]
OR_6	1.50	1.48	1.48	[1.09,1.97]	0.77	0.75	0.74	[0.43,1.17]
	Model 2				Model 2			
OR_1	1.48	1.48	1.48	[1.33,1.61]	1.34	1.28	1.13	[0.69,2.07]
OR_2	4.32	4.30	4.30	[3.39,5.27]	1.51	1.46	1.37	[0.78,2.22]
OR_3	2.04	2.01	1.97	[1.37,2.69]	1.78	1.73	1.59	[1.04,2.62]
OR_4	1.28	1.28	1.25	[0.97,1.59]	0.97	0.94	0.89	[0.51,1.53]
OR_5	109.31	99.08	79.19	[34.48,197.56]	171.77	148.71	108.70	[36.37,354.32]
OR_6	1.49	1.47	1.41	[1.09,1.98]	0.79	0.77	0.69	[0.44,1.17]
	Model 3				Model 3			
OR_1	1.48	1.48	1.48	[1.33,1.63]	1.34	1.27	1.17	[0.72,2.10]
OR_2	4.29	4.24	4.11	[3.38,5.26]	1.52	1.49	1.54	[0.80,2.37]
OR_3	2.07	2.04	2.00	[1.39,2.72]	1.80	1.73	1.60	[1.08,2.69]
OR_4	1.28	1.26	1.24	[1.01,1.66]	0.96	0.94	0.84	[0.50,1.51]
OR_5	108.93	97.99	86.50	[36.48,198.44]	168.80	146.97	122.37	[42.14,348.33]
OR_6	1.49	1.48	1.39	[1.05,1.96]	0.77	0.75	0.75	[0.43,1.19]
	Model 4				Model 4			
OR_1	1.48	1.48	1.48	[1.34,1.61]	1.34	1.31	1.28	[0.73,2.14]
OR_2	4.27	4.26	4.32	[3.42,5.14]	1.53	1.50	1.52	[0.80,2.35]
OR_3	2.14	2.10	1.96	[1.51,2.84]	1.86	1.76	1.69	[1.12,2.94]
OR_4	1.29	1.28	1.26	[0.98,1.63]	0.97	0.93	0.90	[0.45,1.54]
OR_5	110.91	98.41	85.80	[36.59,211.24]	164.35	141.74	113.97	[44.90,344.45]
OR_6	1.49	1.48	1.45	[1.05,1.95]	0.77	0.74	0.72	[0.41,1.16]
	Model 5				Model 5			
OR_1	1.48	1.48	1.46	[1.34,1.62]	1.31	1.26	1.08	[0.72,2.02]
OR_2	4.28	4.25	4.00	[3.33,5.17]	1.57	1.51	1.31	[0.82,2.38]
OR_3	2.14	2.11	1.99	[1.41,2.82]	1.85	1.80	1.78	[1.04,2.71]
OR_4	1.29	1.28	1.25	[0.95,1.58]	0.96	0.93	0.84	[0.48,1.54]
OR_5	108.75	98.66	85.25	[42.09,202.69]	168.96	142.91	108.15	[45.51,371.59]
OR_6	1.51	1.48	1.44	[1.13,1.96]	0.76	0.73	0.70	[0.43,1.16]

(except for OR_5), all models and for the two datasets tend to be similar indicating only a smooth degree of asymmetry in their posterior distributions. Such similarity is stronger in Data 2. For Data 1, under almost all models (except Model 3), we can conclude that the eggs in slot 3, April 1962, are comparable to the ones in slot 5, November 1961 (see OR_4). In this case, the value 1 belongs to the 95% HPD interval. Thus, the eggs in slot 3, incubated in April 1962, and in slot 5, incubated in November 1961, have the same chance of generating abnormal fowls. Similar conclusion can be drawn from OR_1, OR_2, OR_4 and OR_6, for Data 2, under all models.

Let us consider Data 1 and only the best model (Model 5). Taking into consideration the modal OR_3, for instance, we conclude that, for eggs that did not receive any dosage of 6AN and a dosage of 500 of 3AP, the chance of generating abnormal fowls for eggs incubated in January 1962 is, most probably, 1.99 as likely as for eggs incubated in November 1961. If we consider the median OR, our conclusion is that such a chance is higher than 2.11 with posterior probability of 50.0%. For OR_5, the posterior estimates greatly differ. Considering eggs that did not receive any dosage of 3AP, the expected and the most probable chances of those incubated in January, 1962 which received a dosage of 10 of 6AN generating an abnormal fowls are, respectively, 108.8 and 85.3 as likely as those incubated in December 1961 which received a dosage of 4 of 6AN. The other comparisons follow similarly.

7 Conclusions

In this paper, we extended previous work by providing an odds ratio interpretation in logistic regression with random or cluster-level intercept under the Bayesian paradigm. We assume skew-normal distributions for the random effects. We considered independent and dependent skew-normal random effects and, given the fixed effects, we obtained the prior distributions for the odds ratio and their medians. As a by-product, we also obtained results related to linear combinations of skew-normal random variables which, as far as we know, have not been considered in the literature yet. One advantage of the posterior odds ratio interpretation is that we also provided tools to decide about its significance. We analyzed simulated datasets and the dataset reported in Liu and Dey (2008) from a teratogenic activity experiment.

In summary, although the point estimates for the fixed effects are not influenced by the random effects distributions, such distributions do influence the estimates for the odds ratio as well as the magnitude of the HPD intervals. We concluded that the misspecification of the random effects distributions can lead to poor estimates for the odds ratios and leads to HPD intervals with higher magnitude. We also observed the median odds ratio is not necessarily the best measure of heterogeneity as suggested by Larsen et al. (2000). In several cases, the posterior mean and mode prove to be better estimators for the odds ratio. In addition, the posterior mode for the OR also has an appealing interpretation in terms of probability. For the teratogenic activity experiment we considered more flexible skew-normal models than that in Liu and Dey (2008). The posterior estimates for the skewness parameter were negative under the majority of the models which agrees with the conclusions in Liu and Dey (2008).

Acknowledgments

We would like to express our gratitude to the Associate Editor and two anonymous referees whose comments and suggestions have contributed to the improvement of the paper. The authors also thank Prof. Dipak Dey for his comments on the first draft of the paper and Prof. Fabio Demarqui for his help with the ARS and ARMS algorithms. C. C. Santos and R. H. Loschi acknowledge, respectively, CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*) and CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil (CNPq-grants 473163/2010-1, 306085/2009-7, 304505/2006-4) for a partial support of their researches. The research of R. B. Arellano-Valle was partially sponsored by FONDECYT (Chile), grant 1120121.

References

Alonso, A., Litière, S. and Molemberghs, G. (2008). A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. *Computational Statistics and Data Analysis*, **52**, 4474–4486. 383

Arellano-Valle, R.B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, **33**, 561–574. 386, 408

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171-178. 382, 385, 392

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715–726. 382, 385, 395, 396

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9–25. 382

Dalla-Valle, A., (2004). The skew-normal distribution. In *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, M. G. Genton (ed.), Chapman & Hall / CRC, Boca Raton, FL, pp. 3–24. 390, 409

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, New York: Oxford University Press. 381, 382

Fitzmaurice, G. M., Laird, N. and Ware, J. H. (2004). *Applied Longitudinal Analysis*, New York: John Wiley. 381

Gibbons, R. D., Hedeker, D., Charles, S. C. and Frisch, P. (1994). A random-effects probit model for predicting medical malpractice claims. *Journal of the American Statistical Association* **89**, 760–767. 381

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Applied Statatistics* **44**: 455–472. 396

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statatistics* **41**: 337–348. 396

Kuhn, E. and Lavielle, M.(2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, **49**, 1020–1038. 382

Larsen, K. and Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology*, 161, 81–88. 382

Larsen, K., Petersen, J. H., Budtz-Jørgensen, E. and Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics*, 56, 909–914. 381, 382, 383, 384, 385, 387, 388, 389, 392, 398, 405

Litière, S., Alonso, A. and Molemberghs, G. (2008). The impact of a misspecified random-effect distribution on the estimation and the performance of inferential procedures in the generalized linear mixed models. *Statistics in Medicine*, **27**, 3125–3144. 383

Liu, J. and Dey, D. K.(2008). Skew random effects in multilevel binomial models: an alternative to nonparametric approach. *Statistical Modelling*, 8(3), 221–241. 382, 383, 385, 402, 403, 404, 405, 406

Liu, L. and Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* **27**, 3105–3124. 382

Marchenko, Y.V. and Genton, M. G. (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data, *Environmetrics*, 21, 318–340. 387, 390

Martins-Filho, S., Loschi, R.H. and Colosimo, E.A. (2010). Bayesian interpretation for the parameters of the mixed logistic model: A seed germination application. *Unplublished manuscript*. 383, 392

McCulloch, C. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*, Wiley: New York. 382

Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M. and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* **15**, 39–57. 382

Paulino, C. D., Silva, G. and Achcar, J. A. (2005). Bayesian analysis of correlated misclassified binary data. *Computational Statistics and Data Analysis*, **49**, 1120–1131. 381, 382

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–727. 381

Souza, A. D. P. and Migon, H. S. (2010). Bayesian outlier analysis in binary regression. *Journal of Applied Statistics* **37**, 1355–1368. 381, 382

Ten Have, T. R. and Localio, A. R. (1999). Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models. *Biometrics*, 55, 1022–1029. 382

Zeger, S. L. and Karim, M. (1991). Generalized linear mixed models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79–86. 382

Appendix A: Linear combination of SN random variables

Assume that the distribution of a random quantity $\mathbf{Y} \in \mathbb{R}^d$ belongs to the unified SN (SUN) family of distributions introduced by Arellano-Valle and Azzalini (2006), that is, $\mathbf{Y} \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\gamma}, \bar{\boldsymbol{\omega}}, \boldsymbol{\Omega}^*)$, where $\bar{\boldsymbol{\omega}} = \boldsymbol{\omega}\mathbf{1}_d$, $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$ and $\boldsymbol{\Omega}^* = \begin{pmatrix} \boldsymbol{\Delta} & \boldsymbol{\Gamma}^t \\ \Gamma & \bar{\boldsymbol{\Omega}} \end{pmatrix}$, if its p.d.f. and moment generating function (m.g.f.) are, respectively, given by

$$\begin{aligned} f(\mathbf{y}) &= \phi_d(\mathbf{y} - \boldsymbol{\xi}; \boldsymbol{\Omega}) \frac{\Phi_m(\boldsymbol{\gamma} + \boldsymbol{\Gamma}^t \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \boldsymbol{\Delta} - \boldsymbol{\Gamma}^t \bar{\boldsymbol{\Omega}} \boldsymbol{\Gamma})}{\Phi_m(\boldsymbol{\gamma} \mid \boldsymbol{\Delta})}, \quad \mathbf{y} \in \mathbb{R}^d, \quad (18) \\ M_{\mathbf{Y}}(\mathbf{r}) &= \exp\left\{\boldsymbol{\xi}^t \mathbf{r} + \frac{1}{2} \mathbf{r}^t \Omega t\right\} \Phi_m(\boldsymbol{\gamma} + \boldsymbol{\Gamma}^t \boldsymbol{\omega} \mathbf{r} \mid \boldsymbol{\Delta}) [\Phi_m(\boldsymbol{\gamma} \mid \boldsymbol{\Delta})]^{-1}, \quad \mathbf{r} \in \mathbb{R}^d. \end{aligned}$$

Consider the linear combination $\boldsymbol{c}^t\boldsymbol{\gamma}$ where $\boldsymbol{c}^t \in \mathbb{R}^k$ is a constant vector. If the coordinates of vector $\boldsymbol{\gamma}$ are such that $\gamma_i \overset{ind}{\sim} SN(\xi_i, \sigma_i^2, \lambda_i)$, for all $i = 1, \ldots k$, then we prove next that the distribution of the linear combination $\boldsymbol{c}^t\boldsymbol{\gamma}$ belongs to the SUN family given in (18).

Proposition 3. *If* $\gamma_i \overset{ind}{\sim} SN(\xi_i, \sigma_i^2, \lambda_i)$, *for all* $i = 1, \ldots, k$, *and* $\mathbf{c}^t = (c_1, \ldots, c_k)$ *is a* $1 \times k$ *constant vector, then* $\mathbf{c}^t\boldsymbol{\gamma} \sim SUN_{1,k}(\boldsymbol{c}^t\boldsymbol{\xi}, \mathbf{0}, (\boldsymbol{c}^t\boldsymbol{\Lambda}^t\boldsymbol{\Lambda}\boldsymbol{c})^{1/2}, \boldsymbol{\Omega}^*)$, *where* $\boldsymbol{\xi}^t = (\xi_1, \ldots, \xi_k)$, $\boldsymbol{\Lambda} = diag\{\sigma_1, \ldots, \sigma_k\}$, $\boldsymbol{\Delta} = diag\{\delta_1, \ldots, \delta_k\}$, $\delta_i = \lambda_i[1 + \lambda_i^2]^{-1/2}$, $\Gamma^t = (\boldsymbol{c}^t\boldsymbol{\Lambda}^t\boldsymbol{\Lambda}\boldsymbol{c})^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Delta}\boldsymbol{c}$, $\boldsymbol{\Omega}^* = \begin{pmatrix} \mathbf{I}_k & \boldsymbol{\Gamma}^t \\ \boldsymbol{\Gamma} & 1 \end{pmatrix}$ *and its p.d.f. is*

$$\begin{aligned} f_{\mathbf{c}^t\boldsymbol{\gamma}|\boldsymbol{\xi},\boldsymbol{\Lambda},\boldsymbol{\Omega}^*}(x) &= 2^k \phi(x \mid \boldsymbol{c}^t\boldsymbol{\xi}, \boldsymbol{c}^t\boldsymbol{\Lambda}^t\boldsymbol{\Lambda}\boldsymbol{c}) \\ &\quad \times \Phi_k\left(\frac{\boldsymbol{\Lambda}\boldsymbol{\Delta}\boldsymbol{c}}{\boldsymbol{c}^t\boldsymbol{\Lambda}^t\boldsymbol{\Lambda}\boldsymbol{c}}(x - \boldsymbol{c}^t\boldsymbol{\xi}) \mid \mathbf{I}_k - \frac{\boldsymbol{\Lambda}\boldsymbol{\Delta}\boldsymbol{c}\boldsymbol{c}^t\boldsymbol{\Delta}^t\boldsymbol{\Lambda}^t}{\boldsymbol{c}^t\boldsymbol{\Lambda}^t\boldsymbol{\Lambda}\boldsymbol{c}}\right). \qquad (19) \end{aligned}$$

Proof: To prove (19) we consider that if $\gamma \sim SN(\xi, \sigma^2, \lambda)$ then its m.g.f. is given by $M_\gamma(r) = 2\exp\left\{r\xi + \frac{r^2\sigma^2}{2}\right\}\Phi(\delta\sigma r)$ where $\delta = \lambda[1 + \lambda^2]^{-1/2}$. Therefore, using well-known results on the m.g.f. for linear combinations of independent random variables, we have that

$$\begin{aligned}
M_{\boldsymbol{c}^t\boldsymbol{\gamma}}(g) &= 2^k \exp\left\{\sum_{i=1}^{k} g c_i \xi_i + \frac{1}{2}\sum_{i=1}^{k} g^2 c_i^2 \sigma_i^2\right\}\prod_{i=1}^{k}\Phi\left(\frac{\lambda_i}{(1+\lambda_i^2)^{\frac{1}{2}}}\sigma_i c_i g\right) \\
&= 2^k \exp\left\{g\boldsymbol{c}^t\boldsymbol{\xi} + \frac{r^2}{2}(\boldsymbol{\Lambda}\boldsymbol{c})^t(\boldsymbol{\Lambda}\boldsymbol{c})\right\}\Phi_k\left(\boldsymbol{\Lambda}\boldsymbol{\Delta}\boldsymbol{c}g\right) \\
&= \exp\left\{g\boldsymbol{c}^t\boldsymbol{\xi} + \frac{r^2}{2}\boldsymbol{c}^t\boldsymbol{\Lambda}^t\boldsymbol{\Lambda}\boldsymbol{c}\right\}\Phi_k(g\boldsymbol{\Lambda}\boldsymbol{\Delta}\boldsymbol{c})[\Phi_k(\mathbf{0})]^{-1}, \qquad (20)
\end{aligned}$$

which is the m.g.f. of the SUN distribution pointed out in (19).

Assume now that the coordinates of vector $\boldsymbol{\gamma}$ are correlated. The following proposition provides the distribution of a linear combination of the components of $\boldsymbol{\gamma}$ assuming that $\boldsymbol{\gamma}$ has the multivariate skew-normal distribution.

Proposition 4. *If* $\boldsymbol{\gamma} \sim SN_k(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ *with p.d.f. given in (9), then it follows that* $\boldsymbol{c}^t\boldsymbol{\gamma} \sim SN(\boldsymbol{\xi}^t\boldsymbol{c}, \boldsymbol{c}^t\boldsymbol{\Omega}\boldsymbol{c}, \lambda)$, *where* $\lambda = [\boldsymbol{\alpha}^t\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}\boldsymbol{c}][(1+\boldsymbol{\alpha}^t\bar{\boldsymbol{\Omega}}\boldsymbol{\alpha})(\boldsymbol{c}^t\boldsymbol{\Omega}\boldsymbol{c}) - (\boldsymbol{\alpha}^t\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}\boldsymbol{c})^2]^{-1/2}$.

The proof of Proposition 4 is omitted. It follows straightforwardly from the moment generating function. A similar result can be found in Dalla-Valle (2004) (Proposition 1.6.1) where it is proved that if $\mathbf{C}$ is a $k\times k$ non-singular matrix then $\mathbf{C}\boldsymbol{\gamma}$ has a k-variate skew normal distribution.

Appendix B: Simulated datasets

Table 6: Simulated Datasets

	X_1	X_2	n_i	Number of success Data 1	Data 2	Data 3	Data 4
1	-0.429	-1.519	110	6	0	15	43
2	-0.429	0.261	108	30	35	24	90
3	-1.194	0.261	107	8	89	12	5
4	-0.429	0.854	107	14	0	76	56
5	-1.194	0.854	103	17	32	37	68
6	0.031	-1.519	98	41	3	18	24
7	0.031	0.261	102	87	5	64	64
8	0.031	0.854	106	37	82	41	20
9	1.868	-1.519	203	195	201	194	148
10	1.868	0.854	175	137	151	152	160
11	-1.194	0.854	180	87	157	36	115
12	0.337	-1.519	156	122	43	67	131
13	0.337	0.261	152	27	131	132	79
14	-1.194	0.261	148	21	134	21	27
15	0.337	0.854	151	90	53	143	42
16	-1.194	0.854	148	63	88	66	54
17	1.868	0.261	110	92	100	94	91
18	0.031	-1.519	108	27	47	53	6
19	0.337	0.854	107	68	79	43	36
20	-0.429	0.261	103	68	23	60	6
21	-1.194	-1.519	98	3	1	0	1
22	-0.429	0.854	102	90	42	70	102
23	0.031	0.261	106	81	104	96	42
24	0.337	-1.519	203	142	69	51	21
25	1.868	0.854	175	144	83	156	171