A robust Bayesian mixed effects approach for zero inflated and highly skewed longitudinal count data emanating from the zero inflated discrete Weibull distribution

Divan Aristo Burger[1], **Robert Schall**[2], **Johannes Theodorus Ferreira**[1], **Ding-Geng Chen**[1]

[1]*Department of Statistics, University of Pretoria, Pretoria, South Africa*
[2]*Department of Mathematical Statistics & Actuarial Science, University of the Free State, Bloemfontein, South Africa*

ABSTRACT

This paper proposes a Bayesian mixed effects zero inflated discrete Weibull (ZIDW) regression model for zero inflated and highly skewed longitudinal count data, as an alternative to mixed effects regression models that are based on the negative binomial, zero inflated negative binomial and conventional discrete Weibull (DW) distributions. The mixed effects ZIDW regression model is an extension of a recently introduced model based on the DW distribution, and uses the log-link function to specify the relationship between the linear predictors and the median counts. The ZIDW approach offers a more robust characteristic of central tendency, compared to the mean count, when there is skewness in the data. A matrix generalized half-t (MGH-t) prior distribution is specified for the random effects covariance matrix as an alternative to the widely used Wishart prior distribution. The methodology is applied to a longitudinal dataset from an epilepsy clinical trial. In a data contamination simulation study we show that the mixed effects ZIDW regression model is more robust than the competing mixed effects regression models when the data contain excess zeros or outliers. The performance of the ZIDW regression model is also assessed in a simulation study under the specification of respectively the MGH-t and Wishart prior distributions for the random effects covariance matrix. It turns out that the highest posterior density intervals under the MGH-t prior for the fixed effects maintain nominal coverage when the true variability between random slopes over time is small, whereas those under the Wishart prior are generally conservative.

KEYWORDS: Bayesian; zero inflation; longitudinal; discrete Weibull; matrix generalized half-t

---

Address correspondence to Divan Aristo Burger, Department of Statistics, University of Pretoria, Pretoria 0028, South Africa; Email: divanaburger@gmail.com

1 INTRODUCTION

Longitudinal count data in medical applications (for example substance abuse and microbiome data[1,2]) are often modeled through generalized mixed effects regression models. Here the underlying distributions belong to the exponential family such as the Poisson and negative binomial (NB) distributions,[3,4] and their zero inflated extensions, namely the zero inflated Poisson and zero inflated negative binomial (ZINB) distributions.[5,6] Alternatives include mixed effects regression models that are based on the generalized Poisson,[7,8] Conway-Maxwell Poisson[9] and Poisson-inverse Gaussian distribution.[10] These distributions, however, lack robustness when modeling regression functions for longitudinal count data that exhibit high skewness. As an alternative, Luyts et al.[11] recently introduced a flexible approach that uses the discrete Weibull (DW) distribution in a mixed effects regression model in order to accommodate underdispersed, overdispersed and highly skewed data using the `SAS`® procedure `NLMIXED`.[12]

The mixed effects DW regression model is flexible in the sense that it allows one to model the conditional median as an alternative to the conditional mean, which in turn offers a robust characteristic of central tendency when the distribution of the data is skew.[13] The mixed effects DW regression model of Luyts et al.[11] is an extension of the fixed effects DW regression models of Klakattawi et al.[13] and Haselimashhadi et al.[14] who respectively made use of maximum likelihood and Bayesian inference techniques to estimate the model parameters.

Luyts et al.[11] argued that the DW distribution, according to the zero inflation index by Puig and Valero[15], is able to accommodate count data that are zero inflated and zero deflated, and therefore suggested that the DW distribution can be fitted to count data "without the need for introducing zero-inflated or hurdle components". However, as shown in the supplementary material of this paper, the zero inflation index by Puig and Valero[15] would also suggest that the *conventional* NB distribution is able to accommodate zero inflated counts while it is well-known that this distribution cannot handle all forms of excess zeros (e.g. structural zeros). These observations suggest that the conventional DW distribution is not necessarily fully robust to excess zeros. In the present paper we therefore consider a generalization of the conventional DW distribution which includes an additional parameter in order to fully accommodate structural and random zeros in the data. Thereby we investigate an extension of the mixed effects DW regression model by Luyts et al.[11] for hierarchically structured longitudinal count data. In particular, our regression model specifies the *zero inflated* discrete Weibull (ZIDW) distribution for longitudinal count data from a Bayesian perspective. The likelihood of the mixed effects ZIDW regression model is reparameterized to assess the effect of the covariates on the median counts.[16]

The ZIDW distribution has previously been applied in a regression context by Fortin and DeBlois[17], Kalktawi[16] and Peluso[18]. However, to the best of the authors' knowledge, no literature is available on ZIDW regression models with random effects, and the ZIDW model has not been considered in either the longitudinal or Bayesian contexts. The contribution of the present paper is therefore threefold: 1) The implementation of the ZIDW distribution in a mixed effects regression modeling framework (which emanates from the DW distribution); 2) the robust modeling of median counts; and 3) the Bayesian implementation. The methodology is illustrated using an epilepsy dataset, along with a simulation study.

In mixed effects regression modeling the choice of an appropriate prior distribution for random effects covariance matrices remains challenging. The variability between random slopes over time can be quite small (close to zero). It is known that inferences for close-to-zero variance components can be sensitive to the specification of "too" vague prior distributions.[19–22] Such prior misspecification can produce variance component estimates that are biased upwards, and as a result, cause the coverage of the confidence intervals of the corresponding fixed effects to be too high.[23] We therefore specify the matrix generalized half-t (MGH-t) prior distribution[24] for the random effects covariance matrix as an alternative to the widely used Wishart prior distribution.[25]

The paper is organized as follows: In Section 2 we present the motivation from a longitudinal epilepsy clinical trial. Section 3 introduces four candidate Bayesian mixed effects regression models for longitudinal count data, namely the NB, ZINB, DW and ZIDW regression models. In Section 4 we compare the competing models applied to the longitudinal seizure count dataset. Section 5 presents simulation studies to investigate the robustness of the four mixed effects regression models to excess zero counts and outliers (data contamination), as well as assessing the performance of the ZIDW regression model. Section 6 presents a discussion of the results and findings of the paper.

2 MOTIVATING DATA

The natural variation in seizure counts of patients with epilepsy may incorrectly classify patients as positive or negative responders.[26] Furthermore, the mean seizure count may greatly exceed the median count when the distribution of the data is highly skewed and overdispersed, so that the median count might be a more appropriate characteristic for the effectiveness of epilepsy drugs than the mean. It is also well known that models for count data may yield incorrect statistical inferences when zero inflation in the data is not appropriately taken care of.[27] For this reason we investigate the ZIDW distribution for this type of data, given its robustness to overdispersion, high skewness and zero inflation, as well as its ability to conveniently model the conditional median as an alternative to the conditional mean. We apply the methods proposed in this paper in a reanalysis of data collected from the two-period crossover trial of Leppik et al.[28]. In this trial, patients with epilepsy were randomly assigned to receive either placebo or progabide. Here, only data collected during the first period of the trial (that is, before crossover) are considered for the data analysis. The dataset consists of the number of seizures experienced over bi-weekly intervals for each patient at four post-randomization clinic visits, and baseline characteristics such as age and the number of seizures experienced during the eight weeks before the start of treatment (that is, baseline seizure count). Thall and Vail[29] and Booth et al.[3] analyzed the same dataset by respectively exploring the fit of mixed effects Poisson and NB regression models to the seizure counts over time. For our analysis we chose the same linear predictor as that of Thall and Vail[29] and Booth et al.[3]. We therefore specify the linear predictor for the modeling of seizure count y_{ij} for patient i at Visit j as follows:

$$\beta_0 + \beta_1 \mathrm{Trt}_i + \beta_2 \mathrm{Visit}_{ij} + \beta_3 \mathrm{Age}_i + \beta_4 \mathrm{Base}_i + \beta_5 \mathrm{Base}_i \mathrm{Trt}_i + u_{0i} + u_{1i} \mathrm{Visit}_{ij} \tag{1}$$

where $i = 1,\ldots,59$, $j = 1,\ldots,4$, $\mathrm{Trt}_i = 0$ if patient i received progabide (31 patients), $\mathrm{Trt}_i = 1$ if patient i received placebo (28 patients), and $\mathrm{Visit}_{ij} = (j-2.5)/5$. Furthermore, Age_i is the logarithm of the age (in years) of patient i, and Base_i is the logarithm of $\frac{1}{4}$ the baseline seizure count of patient i.

Figure 1 shows the observed seizure counts over time by treatment. A visual inspection of the seizure count versus time profiles suggests that the distribution of the seizure counts may be skewed due to the presence of a few large outliers. In a preliminary analysis we estimated the skewness measure of the seizure counts by fitting the NB distribution to the data via the `R` package `MASS`.[30] The sample and estimated skewness measure (see supplementary material) are presented in Table 1 by treatment and visit. We observe that the sample skewness of the data is larger than the estimated skewness for all four visits under progabide treatment, and for two visits under placebo treatment. This finding can motivate the fit of the DW regression model as an alternative to the fit of standard count models (such as the NB and ZINB regression models).

In another preliminary analysis we fitted the NB and ZINB regression models to the seizure counts using the `SAS`® procedure `COUNTREG`.[12] The regression models considered fixed effects only. We used the log-link function to model the relationship between the linear predictor (Equation (1)) and the mean counts. Figure 2 presents the average predicted count probability for 0 to 10 seizure counts under each regression model. Under the NB regression model, the predicted proportion of zero seizure counts is 8.25%, which slightly underestimates the observed proportion (9.75%). The ZINB regression model more closely estimates the proportion of zeros (9.82%). This preliminary investigation, therefore, does not suggest significant evidence of zero inflation. Nevertheless, in this paper we use the epilepsy dataset 1) in order to illustrate the suggested robust modeling approach by fitting the mixed effects ZIDW regression model to the data and compare the results to those obtained from mixed effects NB, ZINB and DW regression models, and 2) as the basis for a data contamination simulation study (see Section 5.1), mimicking the data of the epilepsy trial, to investigate the robustness of the ZIDW regression model to excess zero counts and outliers.

3 BAYESIAN MIXED EFFECTS REGRESSION MODELS

For the mixed effects regression models under consideration in this paper (namely regression models NB, ZINB, DW and ZIDW) we use the log-link function to model mean and median counts over time.

Suppose that y_{ij} is the count outcome for patient $i = 1, \ldots, N$ at timepoint $j = 1, \ldots, T_i$. Furthermore, $\boldsymbol{\beta}$ & $\boldsymbol{u_i}$ are respectively vectors of fixed and patient-specific random effects, and $\boldsymbol{x}_{ij}$ & $\boldsymbol{z}_{ij}$ are covariate vectors respectively containing baseline characteristics and measurement times. Assume the $\boldsymbol{u_i}$ follow multivariate normal distributions with mean $\mathbf{0}$ and d-dimensional unstructured covariance matrix $\boldsymbol{\Sigma}$, such that $\boldsymbol{u_i} \sim N_d\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$.

3.1 Negative binomial distributions

The probability mass function (PMF) of the ZINB regression model for a given count y_{ij} over time is written as:

$$f\left(y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right) = \pi I\left(y_{ij} = 0\right) + (1-\pi)\left[\binom{y_{ij}+\rho-1}{y_{ij}}\left(\frac{\rho}{e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{'}\boldsymbol{u_i}}+\rho}\right)^{\rho}\left(\frac{e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{'}\boldsymbol{u_i}}}{e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{'}\boldsymbol{u_i}}+\rho}\right)^{y_{ij}}\right]$$

where $I\left(a\right)$ denotes an indicator function taking the value 1 if a is true, and 0 otherwise. Here, ρ and π are respectively the dispersion parameter and zero inflation probability of the ZINB distribution. The mean of the y_{ij} under the ZINB regression model is given by:

$$E\left(y_{ij}\right) = (1-\pi)\, e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{'}\boldsymbol{u_i}}$$

As a special case, the PMF of the ZINB regression model reduces to that of the NB regression model when $\pi = 0$ with corresponding mean $E\left(y_{ij}\right) = e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+z_{ij}^{'}\boldsymbol{u_i}}$.

3.2 Discrete Weibull distributions

The appendix to this paper provides the key properties of the ZIDW distribution, and its reparameterization in terms of its median.

The PMF of the ZIDW regression model for a given count y_{ij} over time is written as:

$$f\left(y_{ij}|\boldsymbol{\beta},\boldsymbol{u}_i,\rho,\pi\right) = \pi I\left(y_{ij}=0\right)+(1-\pi)\times \left[\exp\left(-\log(2)\left[\frac{y_{ij}}{e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+z_{ij}^{'}\boldsymbol{u_i}}}\right]^{\rho}\right)-\exp\left(-\log(2)\left[\frac{y_{ij}+1}{e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+z_{ij}^{'}\boldsymbol{u_i}}}\right]^{\rho}\right)\right]$$

Here, ρ and π are respectively the shape parameter and zero inflation probability of the ZIDW distribution.[16] As per Equation (8) (see appendix), the median of the y_{ij} under the ZIDW regression model is given by:

$$M\left(y_{ij}\right)+1 = \left(\frac{\log\left[\frac{0.5}{1-\pi}\right]}{\log(0.5)}\right)^{1/\rho} e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+z_{ij}^{'}\boldsymbol{u_i}}$$

As a special case, the PMF of the ZIDW regression model reduces to that of the DW regression model when $\pi = 0$ with corresponding median $M\left(y_{ij}\right)+1 = e^{\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta}+z_{ij}^{'}\boldsymbol{u_i}}$.[16]

3.3 Bayesian specification

The prior distributions are specified in such a way to assure vagueness with regard to prior belief on the model parameters.

A normal prior distribution, namely $N(0,10000)$, is specified for each component of $\boldsymbol{\beta}$. The dispersion and shape parameters are assigned a gamma prior distribution, namely $G(0.5,0.5)$, whereas the zero inflation probability is assigned a uniform prior distribution, namely $U(0,1)$.

As Schuurman et al.[20] state, the Wishart distribution “tends to be informative when variances are close to zero”. For this reason we specify the MGH-t prior distribution for the variance-covariance matrix (i.e. $\boldsymbol{\Sigma}$) as a more appropriate alternative to the widely used Wishart prior distribution. The MGH-t prior distribution of $\boldsymbol{\Sigma}$ is expressed as a mixture representation of $G\left(0.5,1/A^2\right)$ for the diagonal entries of diagonal matrix $\boldsymbol{\Omega} = \text{diag}(\omega_1, \ldots, \omega_k, \ldots, \omega_d)$, and a Wishart distribution with inverse scale matrix $2v\boldsymbol{\Omega}$ and degrees of freedom $v+d-1$, namely $W(2v\boldsymbol{\Omega}, v+d-1)$, with corresponding quantities $A = 10000$ and $v = 2$.[24] This mixture representation results in the specification of the half-t prior distribution, namely half-$t(v, A)$, for the standard deviation terms in $\boldsymbol{\Sigma}$, and the uniform prior distribution, namely $U(-1, 1)$, for the correlation terms in $\boldsymbol{\Sigma}$. The probability density function of $\boldsymbol{\Sigma}$ is written as $P(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(v+2d)/2} \prod_{k=1}^{d} \left[v\left(\boldsymbol{\Sigma}^{-1}\right)_{kk} + 1/A^2\right]^{-(v+d)/2}$ where $\boldsymbol{\Sigma} > \mathbf{0}$. Here, $\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}$ denotes the k^{th} diagonal entry of $\boldsymbol{\Sigma}^{-1}$.

Let $\boldsymbol{y}_i$ denote $T_i \times 1$ vectors containing $\left(y_{i1}, \ldots, y_{ij}, \ldots, y_{iT_i}\right)'$. The resulting joint posterior distribution of the model parameters can be written as:

$$
\begin{aligned}
&P(\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi, \boldsymbol{\Sigma}, \omega_k, i = 1, \ldots, N, k = 1, \ldots, d | \boldsymbol{y}) \\
&\propto \left[\prod_{i=1}^{N} \prod_{j=1}^{T_i} f\left(y_{ij} | \boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right)\right] P(\boldsymbol{\beta}) \left[\prod_{i=1}^{N} P(\boldsymbol{u}_i | \boldsymbol{\Sigma})\right] P(\rho) P(\pi) P\left(\boldsymbol{\Sigma}^{-1} | \boldsymbol{\Omega}\right) \left[\prod_{k=1}^{d} P(\omega_k)\right]
\end{aligned}
$$

where $\boldsymbol{y}$ denotes the $\sum_{i=1}^{N} T_i \times 1$ vector containing $\boldsymbol{y}_i$ for all $i = 1, \ldots, N$. The corresponding probability density functions are written as:

$$
\begin{aligned}
P(\boldsymbol{\beta}) &\propto \exp\left(-0.00005\boldsymbol{\beta}'\boldsymbol{\beta}\right) \\
P(\boldsymbol{u}_i | \boldsymbol{\Sigma}) &\propto |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{u}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_i\right) \\
P(\rho) &\propto \rho^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\rho\right) \\
P(\pi) &\propto 1 \\
P\left(\boldsymbol{\Sigma}^{-1} | \boldsymbol{\Omega}\right) &\propto \exp\left[-2\text{tr}\left(\boldsymbol{\Omega}\boldsymbol{\Sigma}^{-1}\right)\right]; \boldsymbol{\Omega} = \text{diag}(\omega_1, \ldots, \omega_k, \ldots, \omega_d) \\
P(\omega_k) &\propto \omega_k^{-\frac{1}{2}} \exp\left(-\frac{1}{10^8}\omega_k\right)
\end{aligned}
$$

The MCMC Gibbs sampling algorithm can be used to draw samples from the joint posterior distribution of the model parameters.[31] The conditional posterior distributions of the model parameters are derived from the joint posterior distribution by ignoring terms that do not include the relevant model parameter.

3.4 Model discrimination

The deviance information criterion (DIC) statistics[32] conditional on the random effects of the regression models, and compound Laplace-Metropolis marginal likelihoods (CLMMLs)[33,34] were calculated to discriminate between the candidate regression models.

3.4.1 Deviance information criterion statistic

The DIC is defined under Model M as follows:

$$\mathrm{DIC}(M) = 2\overline{\mathrm{D}\left(\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right)} - \mathrm{D}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{u}}_i, \hat{\rho}, \hat{\pi}\right)$$

where $\mathrm{D}\left(\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right) = -2\log\left(\prod_{i=1}^{N}\prod_{j=1}^{T_i} f\left(y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right)\right)$ is the deviance measure. Here, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{u}}_i$, $\hat{\rho}$ and $\hat{\pi}$ are respectively the mean of the posterior distribution of $\boldsymbol{\beta}$, $\boldsymbol{u}_i$, ρ and π, and $\overline{\mathrm{D}\left(\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right)}$ is the mean of the posterior distribution of $\mathrm{D}\left(\boldsymbol{\beta}, \boldsymbol{u}_i, \rho, \pi\right)$. Models with small DICs are favored.

3.4.2 Compound Laplace-Metropolis marginal likelihood

The Laplace-Metropolis approximation of the marginal likelihoods of $\boldsymbol{y}$ (that is, the CLMML) under Model M can be written as[35]:

$$\begin{aligned}\log(f[\boldsymbol{y}|M]) = &\frac{1}{2}\log(2\pi)p + \frac{1}{2}\log\left|R_{(\boldsymbol{\beta},\rho,\pi)}\right| + s_{(\boldsymbol{\beta},\rho,\pi)} + \\ &\sum_{i=1}^{N}\log\left[\int P\left(\boldsymbol{u}_i|\hat{\boldsymbol{\Sigma}}\right)\prod_{j=1}^{T_i} f\left(y_{ij}|\hat{\boldsymbol{\beta}}, \boldsymbol{u}_i, \hat{\rho}, \hat{\pi}\right)\mathrm{d}\boldsymbol{u}_i\right] + \\ &\log\left[P\left(\hat{\boldsymbol{\beta}}\right)\right] + \log\left[P\left(\hat{\rho}\right)\right] + \log\left[P\left(\hat{\pi}\right)\right] + \log\left[P\left(\hat{\boldsymbol{\Sigma}}\right)\right]\end{aligned}$$

where p is the number of parameters among $\boldsymbol{\beta}$, ρ, π and $\boldsymbol{\Sigma}$. Here, $\hat{\boldsymbol{\beta}}$, $\hat{\rho}$, $\hat{\pi}$ and $\hat{\boldsymbol{\Sigma}}$ are respectively the mean of the posterior distribution of $\boldsymbol{\beta}$, ρ, π and $\boldsymbol{\Sigma}$. $\left|R_{(\boldsymbol{\beta},\rho,\pi)}\right|$ and $s_{(\boldsymbol{\beta},\rho,\pi)}$ respectively denote the determinant of the correlation matrix and the sum of the logarithm of the standard deviations of the posterior distributions of $\boldsymbol{\beta}$, ρ and π. Models with large CLMMLs are favored.

4 DATA ANALYSIS

4.1 Model implementation and computational issues

The above mixed effects regression models were implemented according to the model specifications discussed in Section 3. That is, the NB, ZINB, DW and ZIDW regression models were fitted to the seizure counts over time.

The terms contained in the linear predictor $\boldsymbol{x}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{'}\boldsymbol{u_i}$ are therefore defined as follows: $\boldsymbol{x}_{ij} = \left(1, \text{Trt}_i, \text{Visit}_{ij}, \text{Age}_i, \text{Base}_i, \text{Base}_i\text{Trt}_i\right)'$ & $\boldsymbol{z}_{ij} = \left(1, \text{Visit}_{ij}\right)'$ are the applicable covariate vectors, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_5)'$ & $\boldsymbol{u_i} = (u_{0i}, u_{1i})'$ are respectively the applicable vectors of fixed and random effects. Furthermore, $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$ is the unstructured covariance matrix of $\boldsymbol{u_i}$.

We also investigated the influence of the alternative prior specifications of the random effects covariance matrix on the inference about the regression model parameters. As an alternative to the MGH-t prior distribution, $\boldsymbol{\Sigma}^{-1}$ was assigned the standard Wishart prior distribution with inverse scale identity matrix $\boldsymbol{I}$ and degrees of freedom d, namely $W(\boldsymbol{I}, d)$.

The regression models were fitted using `JAGS`[36] via the package `runjags`[37] of the `R` project.[38] Posterior samples were monitored and convergence was confirmed using iteration and autocorrelation plots, and Brooks-Gelman-Rubin statistics of parallel chains.[39]

The `R` code for the implementation of the ZIDW regression model is included in the supplementary material of this paper.

The `R` project was called remotely from `SAS`®,[12] and accordingly, posterior samples were exported back to `SAS`® for further computation. For each regression model, 7550000 samples were simulated from the joint posterior distribution for 7 parallel chains. Among those 7550000 samples (per chain), the initial 50000 samples were discarded (burn-in). The thinning factor was set to 50000 to reduce autocorrelation among the samples. The run time of the ZINB and ZIDW regression models (in `JAGS`) was similar.

The multidimensional integration library `cubature` of the R project was used to approximate the Laplace integrals (CLMMLs).[40]

4.2 Regression fits

Figure 3 presents iteration plots of posterior samples of regression model ZIDW from seven parallel chains.

The posterior estimates (PEs) and 95% highest posterior density (HPD) intervals of the model parameters are presented in Table 2. These are presented visually for the treatment effect (β_1) in Figure 4. The analyses suggest that the seizure counts are slightly zero inflated (3% and 6% respectively under regression models ZINB and ZIDW). The results for regression models NB and ZINB suggest that the treatment effect (based on the mean seizure count) is statistically significant, seizure counts of patients on placebo being higher than those of patients on progabide, with PEs for the treatment effect of 0.93 and 0.88, respectively. In contrast, under regression models DW and ZIDW PEs for the treatment effect are somewhat lower, namely 0.72 and 0.66, respectively, and of the respective treatment effects (based on the median seizure count) are no longer statistically significant. Furthermore, the statistically significant covariate effects (in all models) suggest that patients with higher counts at baseline tend to have higher seizure counts post-baseline. In contrast, the model terms age, "time" and treatment & baseline interaction are not statistically significant. The sensitivity analysis further suggests that the ZIDW regression model is sensitive to the choice of prior distributions for $\mathbf{\Sigma}$. More specifically, under the specification of the Wishart prior distribution, the variance component estimate of the random slopes over time (i.e. σ_2^2) is noticeably larger than that of the MGH-t prior distribution, whereas the estimates of the fixed effects (under the two prior specifications) are similar.

In an additional sensitivity analysis we investigated the influence of the prior specification of ρ and π on the estimation of the model parameters. We specified the half-Cauchy (or half-$t\,(1,1)$) and beta prior distributions for ρ and π, namely $\rho \sim$ half-$t\,(1,1)$ and $\pi \sim \text{Beta}\,(0.5,0.5)$. The PEs and 95% HPD intervals are very similar under these prior specifications of ρ and π [data not shown]. We therefore conclude that the statistical inferences of the model parameters are not sensitive to the two types of prior specifications of ρ and π.

4.3 Model comparison

Model comparison statistics for the fitted mixed effects regression models are provided in Table 2. The DIC statistic selects the ZIDW regression model over the NB, ZINB and DW regression models, while it indicates no distinct preference between the NB, ZINB and DW regression models;

the CLMMLs indicate no distinct preference between any of the regression models (though, under the specification of the MGH-t prior distribution, the CLMMLs slightly favor the ZIDW regression model over the competing regression models).

5 SIMULATION STUDIES

5.1 Data contamination

We performed a simulation study to investigate the robustness of the proposed regression models to excess zero counts and outliers.[41] Datasets were simulated from regression models NB, ZINB, DW and ZIDW where model parameters were chosen to mimic the seizure count dataset in Section 4. For simplicity, the covariates age and baseline seizure count were disregarded. The linear predictor for the modeling of seizure count y_{ij} for patient i at Visit j was specified as follows:

$$\beta_0 + \beta_1 \text{Trt}_i + \beta_2 \text{Visit}_{ij} + u_{0i} + u_{1i} \text{Visit}_{ij}$$

The model parameter values for datasets simulated from regression models NB and ZINB were chosen as $\beta_0 = 1.51$, $\beta_1 = 0.30$, $\beta_2 = -0.22$, $\rho = 8.77$, $\sigma_1^2 = 0.90$, $\sigma_2^2 = 0.05$ and $\sigma_{12} = -0.001$, whereas those simulated from regression models DW and ZIDW were chosen as $\beta_0 = 1.66$, $\beta_1 = 0.28$, $\beta_2 = -0.20$, $\rho = 2.85$, $\sigma_1^2 = 0.72$, $\sigma_2^2 = 0.05$ and $\sigma_{12} = -0.01$. Furthermore, $\pi = 0.05$ was chosen for datasets simulated from regression models ZINB and ZIDW. Each dataset consisted of 30 patient profiles per treatment.

The dataset was randomly contaminated by replacing y_{ij} with 0 at a rate of 5%, or with 1.25 the maximum y_{ij} at a rate of 5% (hence an overall contamination rate of 10%). The candidate regression models were fitted to both the uncontaminated and contaminated versions of the simulated datasets (hence a total of 8 datasets were fitted by each regression model), and the DIC statistic was calculated for each fitted model. Figure 5 shows the simulated seizure counts over time of the simulated datasets (uncontaminated and contaminated).

The DIC statistics for the mixed effects regression models fitted to the simulated datasets (uncontaminated and contaminated)[41] are provided in Table 3. The results presented in Table 3 suggest the following:

- If *uncontaminated* datasets are generated from:

- Regression model NB: The fits of regression models ZINB, DW and ZIDW are similar to the fit of regression model NB.
- Regression model ZINB: The fit of regression model ZINB is similar to the fit of regression model ZIDW, and both fits are better than those of regression models NB and DW.
- Regression model DW: The fit of regression model ZIDW is similar to the fit of regression model DW, and both fits are better than those of regression models NB and ZINB.
- Regression model ZIDW: The fit of regression model ZIDW is better than the fits of regression models NB, ZINB and DW.

- If *contaminated* datasets are generated (either from regression models NB, ZINB, DW or ZIDW), then the fit of regression model ZIDW is always better than the fits of regression models NB, ZINB and DW.

In summary, based on the fits of the simulated datasets (uncontaminated and contaminated), regression model ZIDW is always either competitive with or performs better than regression models NB, ZINB and DW.

5.2 Model performance

We assessed the performance of the ZIDW regression model in a simulation study, under the specification of both the MGH-t and Wishart prior distributions for the random effects covariance matrix. Datasets were simulated from the ZIDW regression model where the model parameter values for β_0, β_1, β_2, ρ, σ_1^2 and σ_{12} were chosen as in Section 5.1. The parameter scenarios were investigated for each combination of the following zero inflation probabilities and variance components: $\pi = 0.05$, $\pi = 0.10$, $\sigma_2^2 = 0.05$ and $\sigma_2^2 = 0.75$ (hence, a total of 4 parameter scenarios). Parameter values $\pi = 0.05$ and $\pi = 0.10$ respectively represent cases of low and moderate amounts of zero inflation in the data, whereas $\sigma_2^2 = 0.05$ and $\sigma_2^2 = 0.75$ respectively represent cases of low and high variability in the random slopes over time.

The accuracy and precision characteristics bias, standard error (SE), and root mean square error (RMSE) of the posterior estimates of the regression models were calculated for each parameter scenario, as was the empirical coverage probability of the associated 95% HPD intervals. The candidate regression models were fitted to 1000 simulated datasets, each dataset consisting of 10 patient profiles per treatment. The `autorun.jags` function of the `runjags` package[37] was used to guarantee the successful convergence of the posterior samples for each fitted dataset.

From Table 4 we observe that the bias of the fixed effects estimates is small under both the specification of the MGH-t and Wishart prior distributions for the random effects covariance matrix, whereas the bias of the variance component estimates under the Wishart prior distribution is considerably larger compared to the MGH-t prior distribution when σ_2^2 is small. In general, the SE and RMSE of the variance component estimates differ considerably between the two prior specifications. The coverage probabilities of the 95% HPD intervals of the fixed effects under the specification of the MGH-t prior distribution are close to the nominal value. However, the coverage probabilities of β_2 and σ_2^2 under the specification of the Wishart prior distribution are respectively quite conservative (close to 100%) and extremely small (close to 0%), whereas those of σ_2^2 under the MGH-t prior distribution are close to 100%.

6 DISCUSSION

This paper proposed a Bayesian mixed effects ZIDW regression model for longitudinal count data as an alternative to mixed effects regression models that are based on the usual NB, ZINB and conventional DW distribution. The mixed effects ZIDW regression model is an extension of the method proposed by Luyts et al.[11], and uses the log-link function to specify the relationship between the linear predictors and the median counts, therefore offering a robust characteristic of central tendency, compared to the mean count, when the distribution of the data is skew.

An advantage of the Bayesian implementation of the mixed effects DW regression model is that it does not involve asymptotic approximations, unlike the frequentist method by Luyts et al.[11] (i.e. the `SAS`® procedure `NLMIXED`[12]), and therefore may be more suitable for small sample problems. Moreover, the implementation of the ZIDW regression model (using `JAGS`) is user friendly and competitive with the ZINB regression model in terms of computational speed and model convergence.

The DIC statistics obtained from a reanalysis of the dataset of Leppik et al.[28] suggest that the mixed effects ZIDW regression model is more suitable than the competing mixed effects regression models (that is, regression models NB, ZINB, and DW). The CLMMLs, however, indicate no distinct preference between any of the mixed effects regression models. The results for the mixed effects NB and ZINB regression models suggest that the treatment effect is statistically significant in favor of progabide (lower seizure counts), whereas for the mixed effects DW and ZIDW regression models the treatment effect is not statistically significant.

In a data contamination simulation study we demonstrated that, when the data are contaminated (excess zeros and outliers), the mixed effects ZIDW regression model provides better fits than the competing mixed effects regression models; furthermore, the mixed effects ZIDW regression model is competitive (gives similar fits to its competitors) when the data are uncontaminated. Even though the estimated treatment effect of progabide under the zero inflated regression models (ZINB and ZIDW) was similar to those under the non-zero inflated regression models (NB and DW), the data contamination study points to the ZIDW regression model as the most suitable count model for this type of data (compared to regression models NB, ZINB, and DW).

For our primary model (that is, the mixed effects ZIDW regression model) we carried out a simulation study to illustrate the impact of the choice of prior distributions for the random effects covariance matrix on the estimation and inference of fixed effects and variance components. The data were generated to mimic the seizure count dataset of Leppik et al.[28]. The MGH-t and Wishart prior distribution were studied under four simulation scenarios. The simulation study suggests that estimates of the fixed effects are not particularly sensitive to the choice of prior distributions for the random effects covariance matrix. On the other hand, the variance component estimates differed considerably between the two different prior specifications, especially when the true variability between random slopes (over time) is small. In these cases the HPD intervals under the MGH-t prior for the random slope variance component are extremely conservative, whereas those under the Wishart prior have zero coverage. Although we found that the HPD intervals under the MGH-t prior for the fixed effects were suitable, HPD intervals under the Wishart prior for the fixed effects counterpart of near-zero variance components (that is, the fixed effects slope term) are inappropriate. No major concerns were therefore raised under the MGH-t prior (unlike the Wishart prior) about the estimation and inference of the fixed effects. Evidently, care should be taken in the selection of prior distributions for the random effects covariance matrix, and sensitivity analyses should be performed to confirm findings from the analysis of longitudinal count data. We further note that the conditional conjugacy property associated with the mixture representation of the MGH-t prior distribution reduces computational burden in `JAGS`, and is regarded as a routine implementation.

In conclusion, the proposed mixed effects ZIDW regression model for longitudinal or clustered count data provides better fits than its competitors when the data are skewed (contain outliers) and contain excess zeros. The model can conveniently be specified on the median scale; in contrast, model specification in terms of the conditional mean is difficult given its infinite series expression.

DATA ACCESSIBILITY

The programming code supporting this paper is included as supplementary material to the paper.

ACKNOWLEDGEMENTS

We thank Judy Kleyn and Christelle Catharina Pieterse for discussions that improved the quality of the manuscript. This work is based upon research supported by the South Africa National Research Foundation and South Africa Medical Research Council (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant number 114613); and the Research Development Programme 219/2018 & 296/2019 at the University of Pretoria, South Africa. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

REFERENCES

[1] Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychol. Addict. Behav*. 2013; 27(1): 166–177. DOI: 10.1037/a0029508.

[2] Zhang X, Pei YF, Zhang L, Guo B, Pendegraft A, Zhuang W, Yi N. Negative binomial mixed models for analyzing longitudinal microbiome data. *Front. Microbiol*. 2018; 9: 1683. DOI: 10.3389%2Ffmicb.2018.01683.

[3] Booth JG, Casella G, Friedl H, Hobert JP. Negative binomial loglinear mixed models. *Stat. Model*. 2003; 3(3): 179–191. DOI: 10.1191/1471082X03st058oa.

[4] Molenberghs G, Verbeke G, Demétrio CGB, Vieira AMC. A family of generalized linear models for repeated measures with normal and conjugate random effects. *Stat. Sci*. 2010; 25(3): 325–347. DOI: 10.1214/10–STS328.

[5] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34: 1–14. DOI: 10.2307/1269547.

[6] Gupta R, Marino BS, Cnota JF, Ittenbach RF. Finding the right distribution for highly skewed zero-inflated clinical data. *Epidemiol. Biostat. Public Heal*. 2013; 10(1): DOI: 10.2427/8732.

[7] Xie FC, Wei BC, Lin JG. Score tests for zero-inflated generalized Poisson mixed regression models. *Comput. Stat. & Data Anal.* 2009; 53(9): 3478–3489. DOI: 10.1016/j.csda.2009.02.017.

[8] Joe H, Zhu R. Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biom. J.* 2005; 47(2): 219–229.

[9] Shmueli G, Minka TP, Kadane JB, Borle S, Boatwright P. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *J. Royal Stat. Soc. Ser. C* 2005; 54(1): 127–142. DOI: 10.1111/j.1467–9876.2005.00474.x.

[10] Zha L, Lord D, Zou Y. The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data. *J. Transp. Saf. & Secur.* 2016; 8(1): 18–35. DOI: 10.1080/19439962.2014.977502.

[11] Luyts M, Molenberghs G, Verbeke G, Matthijs K, Ribeiro Jr EE, Demétrio CGB, Hinde J. A Weibull-count approach for handling under-and overdispersed longitudinal/clustered data structures. *Stat. Model.* 2019; 19(5): 569–589.

[12] SAS Institute. SAS/IML user's guide, Version 9.4. 2013.

[13] Klakattawi HS, Vinciotti V, Yu K. A simple and adaptive dispersion regression model for count data. *Entropy* 2018; 20(2): 142. DOI: 10.3390/e20020142.

[14] Haselimashhadi H, Vinciotti V, Yu K. A novel Bayesian regression model for counts with an application to health data. *J. Appl. Stat.* 2018; 45(6): 1085–1105. DOI: 10.1080/02664763.2017.1342782.

[15] Puig P, Valero J. Count data distributions: some characterizations with applications. *J. Am. Stat. Assoc.* 2006; 101(473): 332–340.

[16] Kalktawi HS. Discrete Weibull regression model for count data. Ph.D. thesis; Brunel University London; 2017.

[17] Fortin M, DeBlois J. Modeling tree recruitment with zero-inflated models: the example of hardwood stands in southern Québec, Canada. *For. Sci.* 2007; 53(4): 529–539. DOI: 10.1093/forestscience/53.4.529.

[18] Peluso A. Novel regression models for discrete response. Ph.D. thesis; Brunel University London; 2017.

[19] Hurtado Rua SM, Mazumdar M, Strawderman RL. The choice of prior distribution for a covariance matrix in multivariate meta-analysis: a simulation study. *Stat. Med.* 2015; 34(30): 4083–4104. DOI: 10.1002/sim.6631.

[20] Schuurman NK, Grasman RPPP, Hamaker EL. A comparison of inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models. *Multivar. Behav. Res.* 2016; 51(2-3): 185–206. DOI: 10.1080/00273171.2015.1065398.

[21] Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* 2006; 1(3): 515–533. DOI: 10.1214/06–BA117A.

[22] Kass RE, Natarajan R. A default conjugate prior for variance components in generalized linear mixed models (comments on article by Browne and Draper). *Bayesian Anal.* 2006; 1(3): 535–542. DOI: 10.1214/06–BA117B.

[23] Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat. Med.* 2005; 24: 2401–2428. DOI: 10.1002/sim.2112.

[24] Huang A, Wand MP. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.* 2013; 8(2): 439–452. DOI: 10.1214/13–BA815.

[25] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis, Texts in Statistical Science*. 3rd ed.; London: Chapman & Hall/CRC; 2013.

[26] Goldenholz DM, Goldenholz SR, Moss R, French J, Lowenstein D, Kuzniecky R, Haut S, Cristofaro S, Detyniecki K, Hixson J. Is seizure frequency variance a predictable quantity? *Annals Clin. Transl. Neurol.* 2018; 5(2): 201–207. DOI: 10.1002/acn3.519.

[27] Perumean-Chaney SE, Morgan C, McDowall D, Aban I. Zero-inflated and overdispersed: what's one to do? *J. Stat. Comput. Simul.* 2013; 83(9): 1671–1683. DOI: 10.1080/00949655.2012.668550.

[28] Leppik IE, Dreifuss FE, Bowman T, Santilli N, Jacobs M, Crosby C, Cloyd J, Stockman J, Graves N, Sutula T. A double-blind crossover evaluation of progabide in partial seizures. *Neurol.* 1985; 35(4): 285.

[29] Thall PF, Vail SC. Some covariance models for longitudinal count data with overdispersion. *Biom.* 1990; 46: 657–671.

[30] Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D. MASS: Support functions and datasets for Venables and Ripley's MASS; 2019. R package version 7.3-51.4; URL https://CRAN.R-project.org/package=MASS.

[31] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc*. 1990; 85: 398–409.

[32] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *J. Royal Stat. Soc. Ser. B* 2002; 64(4): 583–639. DOI: 10.1111/1467–9868.00353.

[33] Lewis SM, Raftery AE. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *J. Am. Stat. Assoc*. 1997; 92(438): 648–655. DOI: 10.2307/2965712.

[34] Burger DA, Schall R, Jacobs R, Chen DG. A generalized Bayesian nonlinear mixed-effects regression model for zero-inflated longitudinal count data in tuberculosis trials. *Pharm. Stat*. 2019; 18(4): 420–432. DOI: 10.1002/pst.1933.

[35] Burger DA, Schall R. Robust fit of Bayesian mixed effects regression models with application to colony forming unit count in tuberculosis research. *Stat. Med*. 2018; 37(4): 544–556. DOI: 10.1002/sim.7529.

[36] Plummer, M. JAGS Version 4.3.0 user manual. 2017. URL http://mcmc-jags.sourceforge.net/.

[37] Denwood MJ. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *J. Stat. Softw*. 2016; 71(9): 1–25. DOI: 10.18637/jss.v071.i09.

[38] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria; 2018. URL https://www.R-project.org/.

[39] Ntzoufras I. *Bayesian Modeling Using WinBUGS*. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2009.

[40] Narasimhan B, Johnson SG, Hahn T, Bouvier A, Kiêu K. cubature: Adaptive multivariate integration over hypercubes; 2018. R package version 2.0.3; URL https://CRAN.R-project.org/package=cubature.

[41] Bayes CL, Bazán JL, García C. A new robust regression model for proportions. *Bayesian Anal*. 2012; 7(4): 841–866. DOI: 10.1214/12–BA728.

APPENDIX

Zero inflated discrete Weibull distribution

If Y follows a ZIDW distribution, then the probability mass function of Y is given by:

$$f(y) = \pi I(y=0) + (1-\pi)\left[q^{y^{\rho}} - q^{(y+1)^{\rho}}\right] \tag{2}$$

where $y \in \{0,1,2,\ldots\}$. Here, $0 < q < 1$ and $\rho > 0$ denote the shape parameters, and $0 < \pi < 1$ the zero inflation probability, of the ZIDW distribution. $I(a)$ denotes an indicator function taking the value 1 if a is true, and 0 otherwise.

The mean and variance of Y are written as:

$$E(Y) = (1-\pi)\sum_{n=1}^{\infty} q^{n^{\rho}}$$
$$Var(Y) = (1-\pi)\left(2\sum_{n=1}^{\infty} nq^{n^{\rho}} - \sum_{n=1}^{\infty} q^{n^{\rho}}\right) - (1-\pi)^2\left(\sum_{n=1}^{\infty} q^{n^{\rho}}\right)^2$$

The cumulative distribution function (CDF) of Y is given by:

$$F(y) = \sum_{x=0}^{y} f(x) = (\pi-1)\,q^{(y+1)^{\rho}} + 1$$

The τ-quantile function $Q(\tau)$ is obtained from the inverse CDF, i.e. $F^{-1}(\tau)$, as follows:

$$(\pi-1)\,q^{[Q(\tau)+1]^{\rho}} + 1 = \tau \tag{3}$$

Solving $Q(\tau)$ in Equation (3) gives:

$$Q(\tau) = \left(\frac{\log\left(\frac{\tau-1}{\pi-1}\right)}{\log(q)}\right)^{\frac{1}{\rho}} - 1$$

The median of Y can therefore written as:

$$M(Y) = Q(0.5) = \left(\frac{\log\left(\frac{0.5}{1-\pi}\right)}{\log(q)}\right)^{\frac{1}{\rho}} - 1$$

However, since $\lim_{q \to 0} M(Y) = -1$, we consider the following expression for the median of Y as an alternative:

$$M(Y) + 1 = \left(\frac{\log\left(\frac{0.5}{1-\pi}\right)}{\log(q)} \right)^{\frac{1}{\rho}} \tag{4}$$

Conventional discrete Weibull distribution

Considering the conventional DW distribution ($\pi = 0$) the median in Equation (4) reduces to:

$$M(Y) + 1 = \left(\frac{\log(0.5)}{\log(q)} \right)^{\frac{1}{\rho}} \tag{5}$$

By solving q in Equation (5), we consider the reparameterization of the DW distribution in terms of its median.[16] That is:

$$q = \exp\left(-\frac{\log(2)}{[M(Y)+1]^{\rho}} \right) \tag{6}$$

Therefore, substituting Equation (6) in Equation (2):

$$f(y) = \exp\left(-\log(2) \left[\frac{y}{M(Y)+1} \right]^{\rho} \right) - \exp\left(-\log(2) \left[\frac{y+1}{M(Y)+1} \right]^{\rho} \right)$$

This reparameterization therefore allows for the direct regression modeling of the median counts. The log-link function is used to model the median counts as follows:

$$M(Y) + 1 = e^{\boldsymbol{x}'\boldsymbol{\alpha}} \tag{7}$$

where $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ are respectively a set of covariates and regression coefficients.

Finally, substituting Equations (7) and (6) in Equation (4) results in the median of the ZIDW distribution as follows:

$$M(Y) + 1 = \left(\frac{\log\left[\frac{0.5}{1-\pi}\right]}{\log(0.5)} \right)^{1/\rho} e^{\boldsymbol{x}'\boldsymbol{\alpha}} \tag{8}$$

Figure 1: Epilepsy dataset: Observed seizure counts over time

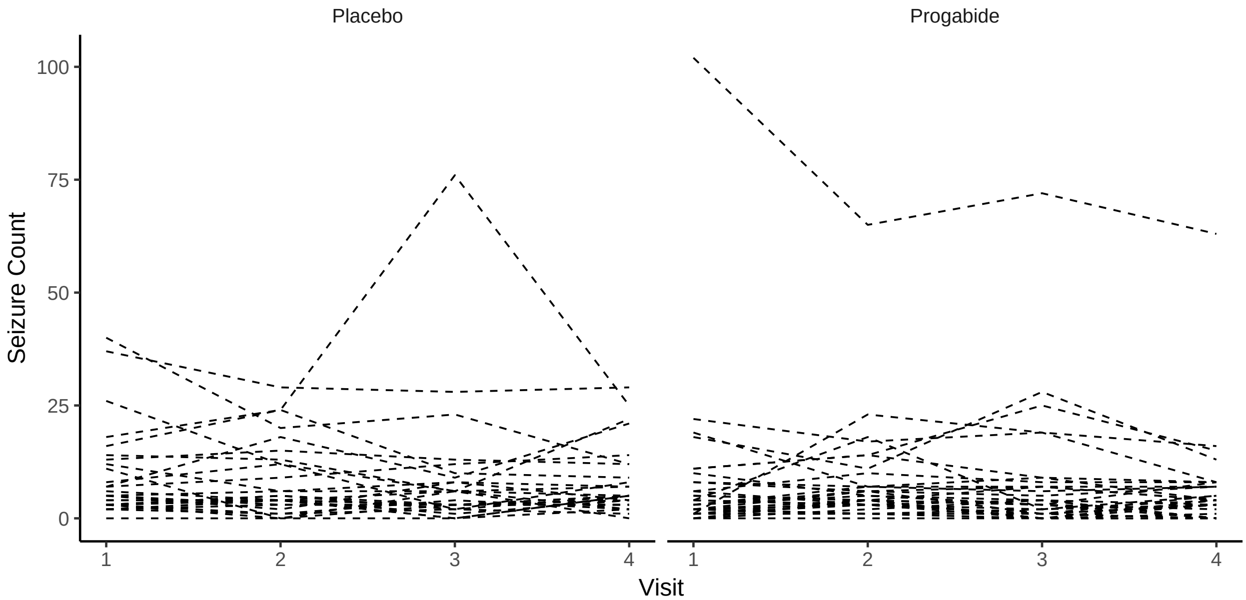
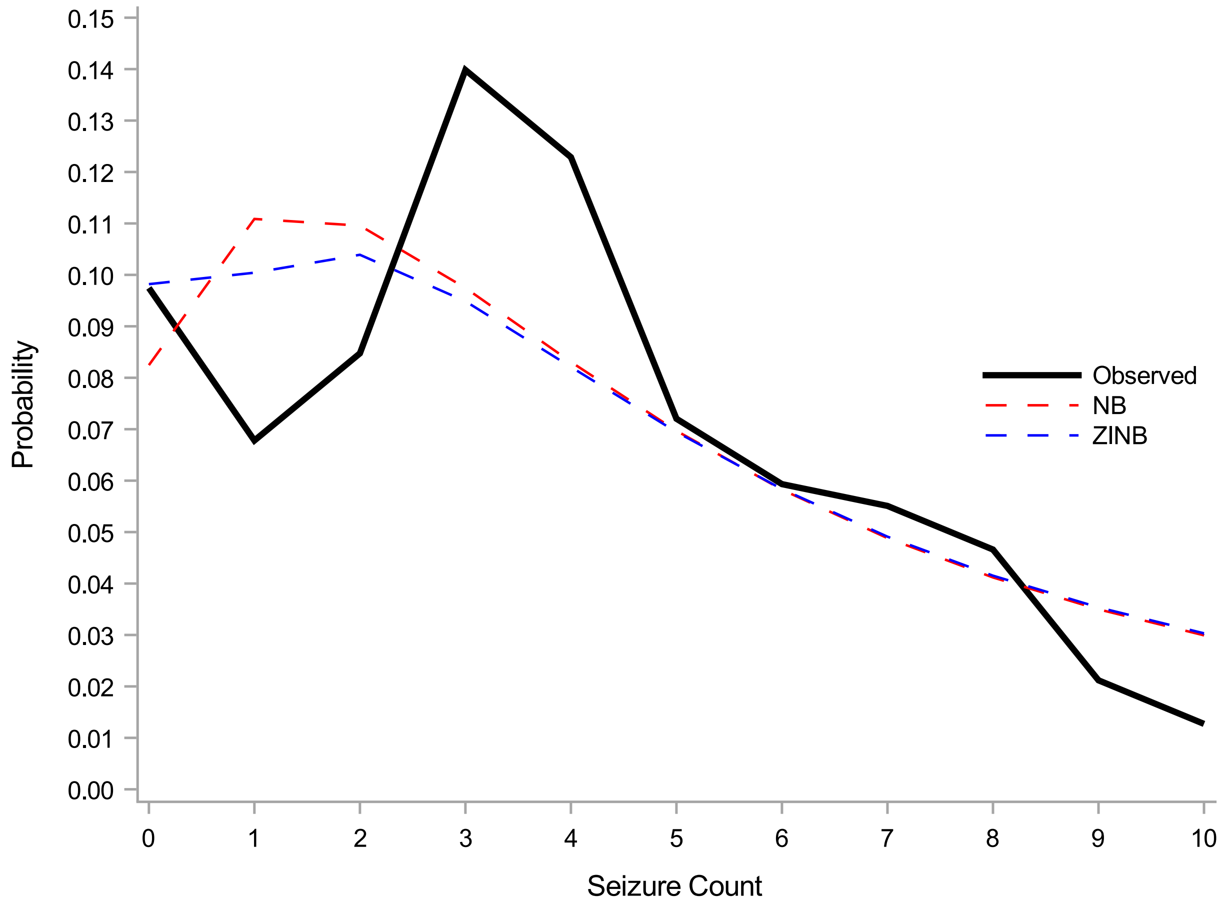
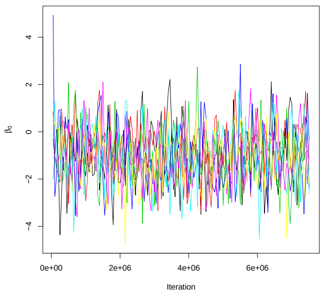


Figure 2: Epilepsy dataset: Average predicted probabilities of seizure counts

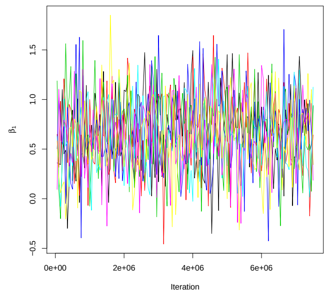


NB: Negative binomial. ZINB: Zero inflated negative binomial.

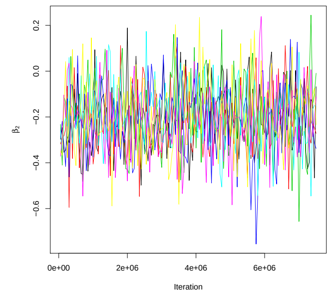
Figure 3: Epilepsy dataset: Iteration plots of posterior samples of regression model ZIDW



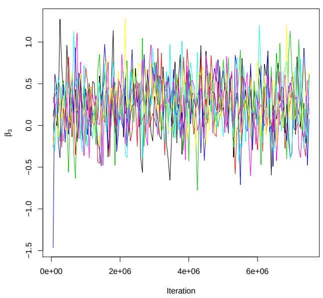
(a) β_0



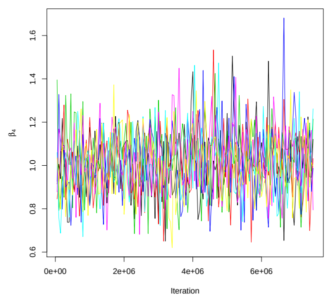
(b) β_1



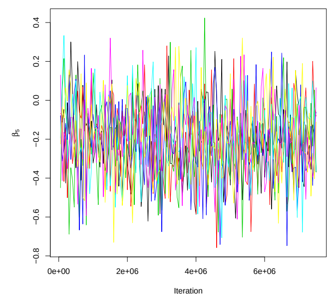
(c) β_2



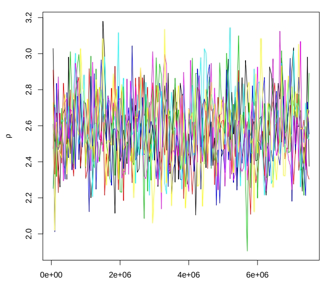
(d) β_3



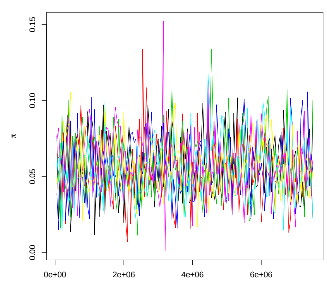
(e) β_4



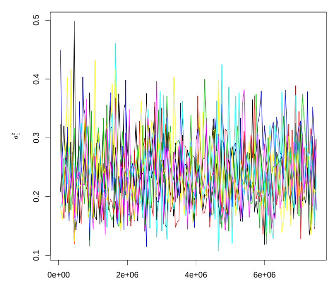
(f) β_5



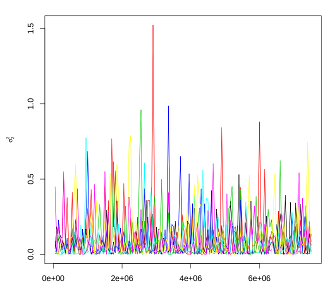
(g) ρ



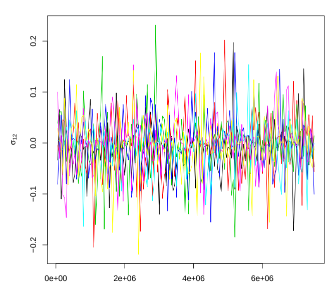
(h) π



(i) σ_1^2



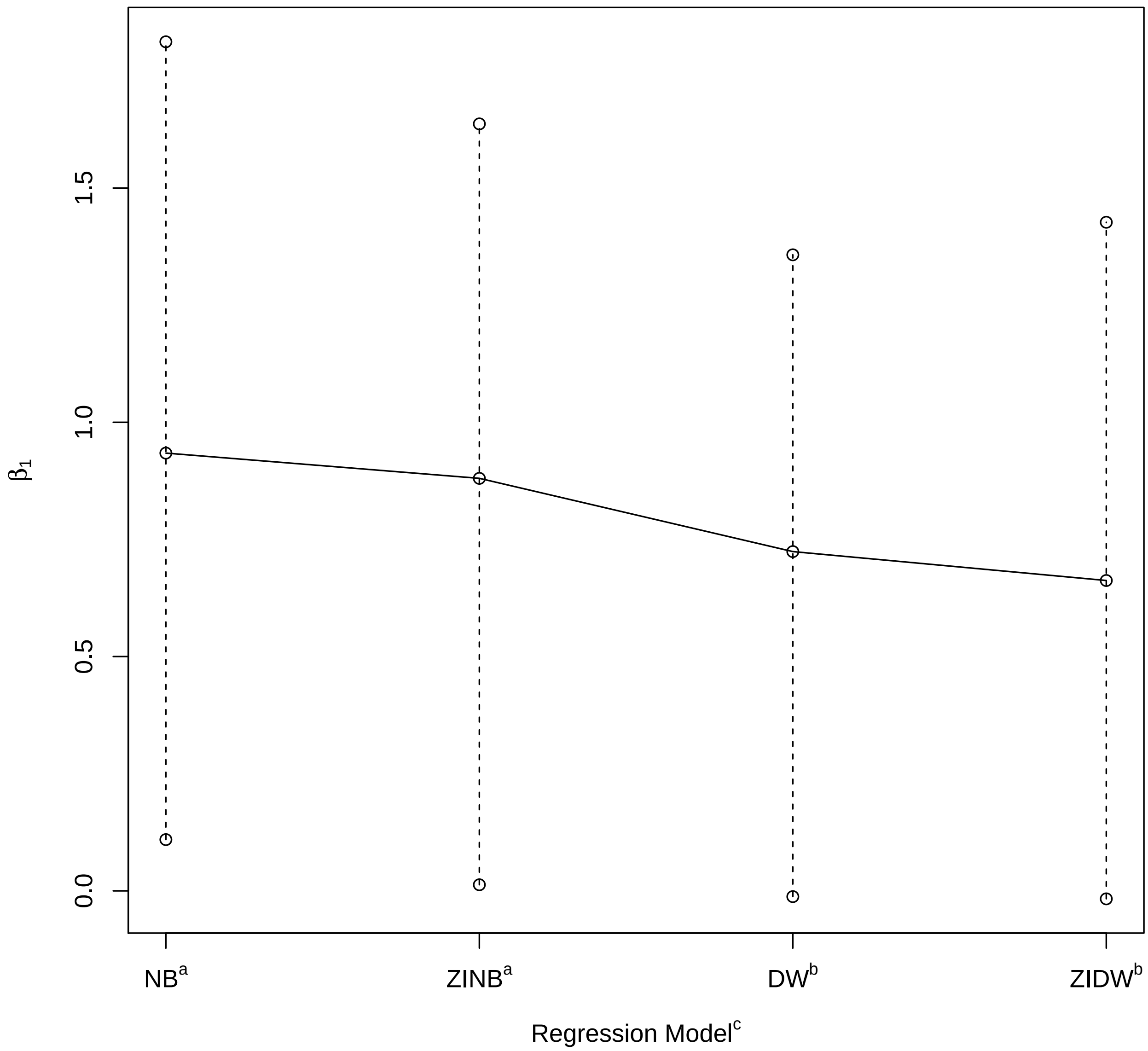
(j) σ_2^2



(k) σ_{12}

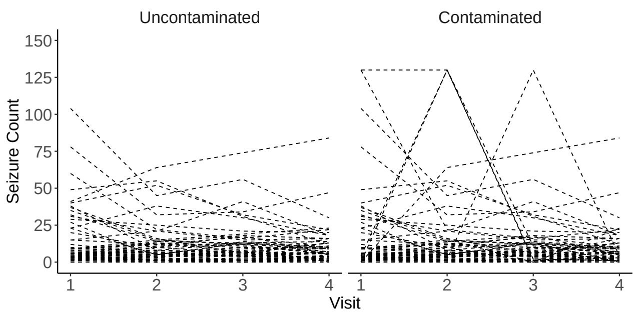
ZIDW: Zero inflated discrete Weibull.

Figure 4: Epilepsy dataset: Posterior estimates and 95% HPD intervals of treatment effects

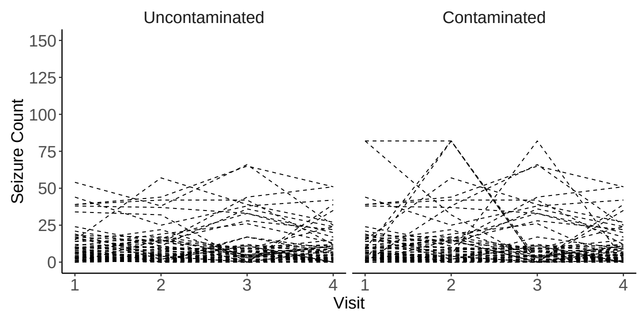


NB: Negative binomial. ZINB: Zero inflated negative binomial. DW: Discrete Weibull. ZIDW: Zero inflated discrete Weibull. HPD: Highest posterior density. [a]The log-link function is used to describe the relationship between the linear predictors and the *mean* of the distribution function. [b]The log-link function is used to describe the relationship between the linear predictors and the *median* of the distribution function. [c]The MGH-*t* prior distribution is specified for the random effects covariance matrix.

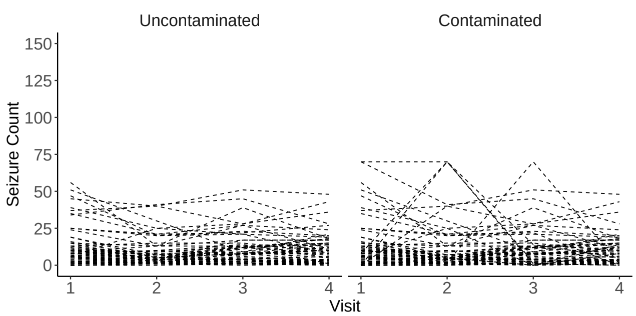
Figure 5: Simulation study: Simulated seizure counts over time – uncontaminated and contaminated datasets



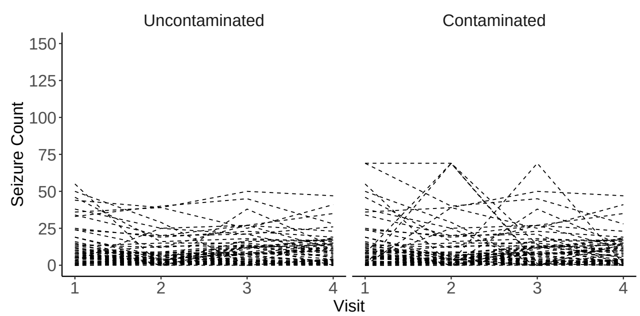
(a) NB



(b) ZINB



(c) DW



(d) ZIDW

NB: Negative binomial. ZINB: Zero inflated negative binomial. DW: Discrete Weibull. ZIDW: Zero inflated discrete Weibull.

Table 1: Epilepsy dataset: Sample and estimated skewness measure based on the NB distribution

	Placebo		Progabide	
Visit	**Sample**	**Estimate**	**Sample**	**Estimate**
1	1.764	1.745	4.325	2.471
2	1.056	1.922	3.533	1.849
3	3.493	2.240	3.251	2.582
4	1.400	1.591	4.007	2.304

NB: Negative binomial.

Table 2: Epilepsy dataset: Posterior estimates and 95% HPD intervals of regression model parameters

		Regression Model									
		MGH-*t* Prior[a]								Wishart Prior[b]	
		NB[c]		ZINB[c]		DW[d]		ZIDW[d]		ZIDW[d]	
Parameter	**Predictor**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**
β_0		-2.241	[-4.595; 0.570]	-2.097	[-4.699; 0.522]	-1.189	[-3.695; 1.133]	-0.886	[-3.185; 1.560]	-0.951	[-3.482; 1.468]
β_1	Trt	0.934	[0.109; 1.812]	0.880	[0.013; 1.637]	0.724	[-0.012; 1.358]	0.662	[-0.017; 1.427]	0.674	[0.040; 1.488]
β_2	Visit	-0.267	[-0.629; 0.065]	-0.236	[-0.576; 0.120]	-0.244	[-0.536; 0.098]	-0.201	[-0.479; 0.072]	-0.196	[-0.509; 0.101]
β_3	Age	0.467	[-0.294; 1.145]	0.455	[-0.315; 1.145]	0.270	[-0.435; 0.898]	0.217	[-0.481; 0.872]	0.236	[-0.511; 0.907]
β_4	Base	1.218	[0.892; 1.544]	1.191	[0.892; 1.528]	1.044	[0.758; 1.323]	1.021	[0.746; 1.303]	1.031	[0.751; 1.336]
β_5	Base$\times$Trt	-0.328	[-0.800; 0.071]	-0.298	[-0.718; 0.106]	-0.233	[-0.573; 0.190]	-0.201	[-0.553; 0.166]	-0.209	[-0.604; 0.173]
ρ		6.518	[4.232; 9.616]	7.772	[4.919; 11.870]	2.101	[1.815; 2.388]	2.584	[2.179; 2.966]	2.625	[2.195; 3.068]
π				0.033	[0.004; 0.068]			0.058	[0.023; 0.096]	0.056	[0.021; 0.094]
σ_1^2		0.255	[0.132; 0.430]	0.235	[0.122; 0.393]	0.233	[0.126; 0.360]	0.234	[0.134; 0.362]	0.248	[0.151; 0.387]
σ_2^2		0.058	[0.000; 0.462]	0.044	[0.000; 0.371]	0.048	[0.000; 0.388]	0.051	[0.000; 0.377]	0.249	[0.082; 0.590]
σ_{12}		0.000	[-0.152; 0.142]	0.001	[-0.129; 0.120]	-0.002	[-0.139; 0.109]	-0.001	[-0.136; 0.099]	-0.006	[-0.164; 0.151]
DIC[e]		1227.86		1224.90		1222.26		1200.21		1205.78	
CLMML[f]		-655.22		-654.42		-657.68		-652.80		-663.20	

NB: Negative binomial. ZINB: Zero inflated negative binomial. DW: Discrete Weibull. ZIDW: Zero inflated discrete Weibull. PE: Posterior estimate. HPD: Highest posterior density. DIC: Deviance information criterion. CLMML: Compound Laplace-Metropolis marginal likelihood. [a]The MGH-*t* prior distribution is specified for the random effects covariance matrix. [b]The Wishart prior distribution is specified for the random effects covariance matrix. [c]The log-link function is used to describe the relationship between the linear predictors and the *mean* of the distribution function. [d]The log-link function is used to describe the relationship between the linear predictors and the *median* of the distribution function. [e]Models with small DICs are favored. [f]Models with large CLMMLs are favored.

Table 3: Simulation study: Comparison of regression models fitted to uncontaminated and contaminated datasets

	Regression Model: DIC Statistic[a]							
	Uncontaminated				**Contaminated**			
Simulated Dataset	**NB**	**ZINB**	**DW**	**ZIDW**	**NB**	**ZINB**	**DW**	**ZIDW**
NB	1343.00	1344.81	1327.13	1341.02	1588.56	1571.74	1569.74	1542.96
ZINB	1431.46	1380.22	1423.24	1367.30	1564.89	1536.53	1558.61	1515.85
DW	1292.23	1295.35	1253.75	1258.52	1524.76	1494.88	1493.95	1447.91
ZIDW	1387.57	1315.65	1357.91	1280.01	1540.28	1483.50	1526.73	1440.09

NB: Negative binomial. ZINB: Zero inflated negative binomial. DW: Discrete Weibull. ZIDW: Zero inflated discrete Weibull. DIC: Deviance information criterion. [a]Models with small DIC statistics are favored.

Table 4: Simulation study: Performance of regression model ZIDW under specification of MGH-t and Wishart prior distributions

				MGH-t Prior[a]				Wishart Prior[b]			
Scenario	**Parameter**	**Predictor**	**Value**	**Bias**	**SE**	**RMSE**	**Coverage[c]**	**Bias**	**SE**	**RMSE**	**Coverage[c]**
$\pi = 0.05$ & $\sigma_2^2 = 0.05$	β_0		1.66	-0.0275	0.2855	0.2867	95.0	-0.0122	0.2798	0.2800	95.0
	β_1	Trt	0.28	0.0184	0.3981	0.3983	95.1	-0.0031	0.3929	0.3927	94.2
	β_2	Visit	-0.20	-0.0131	0.2165	0.2168	97.6	-0.0101	0.2198	0.2199	97.8
	ρ		2.85	0.0823	0.3729	0.3817	96.4	0.1368	0.3879	0.4111	96.6
	π		0.05	0.0063	0.0267	0.0274	95.8	0.0069	0.0269	0.0278	96.2
	σ_1^2		0.72	0.0989	0.3113	0.3265	95.7	0.0508	0.2616	0.2663	97.3
	σ_2^2		0.05	0.1765	0.2479	0.3043	99.8	0.3644	0.1231	0.3847	0.0
	σ_{12}		-0.01	0.0128	0.0886	0.0895	100.0	0.0047	0.1584	0.1584	99.9
$\pi = 0.05$ & $\sigma_2^2 = 0.75$	β_0		1.66	-0.0156	0.2791	0.2794	96.2	-0.0175	0.2762	0.2767	95.2
	β_1	Trt	0.28	-0.0033	0.4056	0.4054	94.4	0.0143	0.3923	0.3923	95.0
	β_2	Visit	-0.20	0.0127	0.2836	0.2837	95.9	0.0052	0.2959	0.2958	95.6
	ρ		2.85	-0.0677	0.4062	0.4116	95.2	-0.0446	0.3852	0.3876	95.6
	π		0.05	0.0067	0.0267	0.0275	96.5	0.0069	0.0267	0.0276	96.7
	σ_1^2		0.72	0.0890	0.3029	0.3156	96.1	0.0412	0.2693	0.2723	95.9
	σ_2^2		0.75	0.0445	0.6142	0.6155	96.2	-0.0177	0.3591	0.3594	99.6
	σ_{12}		-0.01	0.0034	0.1774	0.1774	99.5	0.0106	0.2275	0.2276	98.8

ZIDW: Zero inflated discrete Weibull. SE: Standard error. RMSE: Root mean square error. [a]The MGH-t prior distribution is specified for the random effects covariance matrix. [b]The Wishart prior distribution is specified for the random effects covariance matrix. [c]95% highest posterior density (HPD) interval coverage (%).

Table 4: Simulation study: Performance of regression model ZIDW under specification of MGH-t and Wishart prior distributions

				MGH-t Prior[a]				Wishart Prior[b]			
Scenario	**Parameter**	**Predictor**	**Value**	**Bias**	**SE**	**RMSE**	**Coverage[c]**	**Bias**	**SE**	**RMSE**	**Coverage[c]**
$\pi = 0.10$ & $\sigma_2^2 = 0.05$	β_0		1.66	-0.0174	0.2819	0.2823	94.9	-0.0138	0.2779	0.2781	94.9
	β_1	Trt	0.28	0.0014	0.3908	0.3906	95.8	0.0001	0.4063	0.4061	93.7
	β_2	Visit	-0.20	0.0079	0.2326	0.2326	97.3	-0.0030	0.2238	0.2237	99.1
	ρ		2.85	0.0410	0.4163	0.4181	95.2	0.1078	0.4427	0.4554	94.9
	π		0.10	0.0032	0.0364	0.0366	95.9	0.0024	0.0362	0.0363	93.9
	σ_1^2		0.72	0.0829	0.2907	0.3022	96.7	0.0341	0.2512	0.2534	97.0
	σ_2^2		0.05	0.1998	0.2532	0.3224	100.0	0.3864	0.1409	0.4113	0.0
	σ_{12}		-0.01	0.0066	0.1061	0.1063	100.0	-0.0006	0.1647	0.1646	99.8
$\pi = 0.10$ & $\sigma_2^2 = 0.75$	β_0		1.66	-0.0071	0.2795	0.2794	94.7	-0.0105	0.2796	0.2796	95.9
	β_1	Trt	0.28	-0.0097	0.4063	0.4062	94.7	-0.0100	0.3993	0.3992	94.9
	β_2	Visit	-0.20	-0.0098	0.3079	0.3079	96.8	-0.0084	0.3115	0.3114	95.7
	ρ		2.85	-0.0447	0.4272	0.4293	96.3	-0.0740	0.4142	0.4205	94.6
	π		0.10	0.0025	0.0365	0.0366	94.5	0.0018	0.0363	0.0363	94.8
	σ_1^2		0.72	0.0751	0.2957	0.3050	96.2	0.0331	0.2469	0.2490	96.8
	σ_2^2		0.75	0.0797	0.6541	0.6586	96.7	-0.0110	0.3656	0.3656	99.9
	σ_{12}		-0.01	0.0068	0.1700	0.1700	100.0	0.0070	0.2266	0.2266	99.5

ZIDW: Zero inflated discrete Weibull. SE: Standard error. RMSE: Root mean square error. [a]The MGH-t prior distribution is specified for the random effects covariance matrix. [b]The Wishart prior distribution is specified for the random effects covariance matrix. [c]95% highest posterior density (HPD) interval coverage (%).