Nonlinear mixed-effects modeling of longitudinal count data: Bayesian inference about median counts based on the marginal zero-inflated discrete Weibull distribution

Divan Aristo Burger[1] **and Emmanuel Lesaffre**[2]

[1]*Department of Statistics, University of Pretoria, Pretoria, South Africa*
[2]*I-BioStat, KU Leuven, Leuven, Belgium*

ABSTRACT

This paper proposes a Bayesian regression model for nonlinear zero-inflated longitudinal count data that models the median count as an alternative to the mean count. The nonlinear model generalizes a recently introduced linear mixed-effects model based on the zero-inflated discrete Weibull (ZIDW) distribution. The ZIDW distribution is more robust to severe skewness in the data than conventional zero-inflated count distributions such as the zero-inflated negative binomial (ZINB) distribution. Moreover, the ZIDW distribution is attractive because of its convenience to model the median counts given its closed-form quantile function. The median is a more robust measure of central tendency than the mean when the data, for instance, zero-inflated counts, are right-skewed. In an application of the model we consider a biphasic mixed-effects model consisting of an intercept term and two slope terms. Conventionally, the ZIDW model separately specifies the predictors for the zero-inflation probability and the counting process's median count. In our application, the two latent class interpretations are not clinically plausible. Therefore, we propose a marginal ZIDW (MZIDW) model that directly models the biphasic median counts marginally. We also consider the marginal ZINB (MZINB) model to make inferences about the nonlinear mean counts over time. Our simulation study shows that the models have good properties in terms of accuracy and confidence interval coverage.

KEYWORDS: Bayesian; marginal zero-inflated discrete Weibull; marginal zero-inflated negative binomial; median counts; nonlinear mixed-effects

---

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

1 INTRODUCTION

Longitudinal data in applications such as virology, psychopathology, and neurology follow, in some cases, nonlinear trends.[1] For example, in human immunodeficiency virus (HIV) clinical trials, viral decay over time after antiviral therapy initiation is often associated with two phases.[2] Other examples include nonlinear data from depression and Alzheimer's disease studies. In depression studies, the change in depressive symptoms over time occasionally has two distinct phases.[3] An example of nonlinear data relevant to Alzheimer's disease trials is that of Li et al.[4]: These authors characterized the rate and timing of cognitive decline among Alzheimer's disease patients using a hierarchical change-point model by regressing composite cognitive measure scores against time. Therefore, nonlinear longitudinal data need to be modeled using nonlinear mixed-effects regression models rather than conventional linear mixed-effects models. Nonlinear models may include biphasic, bilinear, and change-point models as special cases. For example, biphasic and bilinear models consist of two distinct slopes over time, therefore describing the data's initial and terminal decline rates.

To model count data, log-linear models based on the Poisson and negative binomial (NB) distributions are typically used. In some applications, longitudinal count data on the log-scale are nonlinear. Hence it becomes necessary to replace the usual log-linear count model with nonlinear count models. The zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) distributions can be considered when the data contain excess zeros.[5] Applications of nonlinear modeling of zero-inflated count data can be found in econometrics and biological experiments. For example, Klein et al.[6] fitted a Bayesian generalized additive model (containing nonlinear smooth functions) based on the ZIP and ZINB distributions to analyze the number of citations of patents granted by the European Patent Office and insurance claims in Belgium during 1997. In a heritability study for tick counts on lambs, Sae-Lim et al.[7] fitted a nonlinear mixed-effects regression model to the tick counts based on the ZIP and ZINB distributions.

Longitudinal models for count data are routinely used to model the mean count, usually using the log-link function to specify the relationship between the mean counts and time. Zero-inflated count data are typically right-skewed,[8] implying that the mean count is greater than the median count. In the presence of skewness, the median count is argued to be a more appropriate characteristic of central tendency than the mean count.[9] Widely-used methods for modeling the median response are quantile regression techniques that specifically model the 50^{th} percentile of the data. Limited literature exists on quantile regression modeling of zero-inflated count data. Examples of quantile regression models for zero-inflated count data include the two-part quantile regression model of

King and Song [10] and the linear mixed-effects zero-inflated discrete Weibull (ZIDW) model of Burger et al. [11] Burger et al. [11] demonstrated in a data contamination simulation study that the ZIDW distribution better accommodates excessive skewness and outliers in the data than the ZINB distribution. Therefore, the ZIDW model is more robust to severe skewness in the data than the ZINB model and is convenient to implement since the ZIDW distribution's quantile function is available in closed form. To the best of the authors' knowledge, no literature exists on nonlinear mixed-effects quantile regression modeling of zero-inflated count data.

In this paper we propose a Bayesian regression model for nonlinear zero-inflated longitudinal count data that models the median count rather than the mean count. As a motivating example, data from a recently published extended bactericidal activity tuberculosis (TB) trial are considered. As a special case of nonlinear mixed-effects models, we consider a biphasic mixed-effects model consisting of an intercept term and two slope terms. Conventionally, ZINB models separately specify linear predictors for the zero-inflation probability and the counting process's mean count. In our application of nonlinear longitudinal modeling, the two latent class interpretations are not clinically plausible. [12] As an alternative, Preisser et al. [13] proposed a marginal ZINB (MZINB) model to model the marginal mean count directly. The current manuscript uses an approach similar to that of Preisser et al. [13] by reparameterizing the ZIDW model in such a way as to model the median count directly (i.e., obtain marginal interpretations); hence, the so-called marginal ZIDW (MZIDW) model. Furthermore, we also consider the MZINB model to make inferences about the nonlinear mean counts over time (i.e., modeling both the mean and median counts marginally) which allows us to compare the two sets of results. The application of the MZINB distribution in a nonlinear mixed-effects modeling context has also not yet been explored in previous literature.

The paper is organized as follows: Section 2 motivates our methodology using data from a longitudinal bactericidal activity TB trial. Section 3 gives an overview of the conventional ZIDW distribution, whereas Section 4 provides an overview of the MZINB and MZIDW distributions. More specifically, in Section 4, we reparameterize the conventional ZINB and ZIDW distributions to obtain the MZINB and MZIDW distributions. Section 5 introduces the nonlinear mixed-effects regression model that assumes the MZINB and MZIDW distributions for longitudinal counts. Section 6 applies the mixed-effects models to the longitudinal TB dataset. Section 7 presents a simulation study to assess the performance, identifiability and robustness of the MZINB and MZIDW models. Section 8 presents a discussion of the results and findings of the paper.

2 MOTIVATING DATA AND PREVIOUS WORK

This paper's methods are motivated by colony-forming unit (CFU) count data in extended bactericidal activity TB trials.[14] Burger and Schall[15] discuss the biological concepts involving bactericidal activity and sterilization of TB drugs, suggesting that the CFU counts on the logarithmic scale typically decline throughout treatment biphasically.[14–17] Furthermore, the CFU counts are usually zero-inflated towards the end of treatment (zero-inflation increases as the anti-TB drugs eliminate CFUs). Firstly, we introduce the TB dataset of Tweed et al.,[18] which is our motivating dataset. Secondly, Burger and Schall's[15] biphasic mixed-effects regression model, which was originally used to fit the data, is discussed. Thirdly, extensions of the previous model[15] are discussed, together with points where improvements can be made to existing models in light of the current manuscript's objectives.

2.1 TB dataset

We consider the data of Tweed et al.[18] who performed a multicentre, open-label, phase 2b bactericidal activity TB trial where a total of 180 eligible patients with newly diagnosed, smear-positive drug-sensitive pulmonary TB were randomized to one of the following three treatment regimens:

- $\text{J}_{(\text{loading dose/t.i.w.})}$PaZ: Bedaquiline 400 mg once daily (QD) on Days 1 to 14, 200 mg three times per week on Days 15 to 56; plus pretomanid 200 mg QD on Days 1 to 56; plus pyrazinamide 1500 mg QD on Days 1 to 56.
- $\text{J}_{(200\text{ mg})}$PaZ: Bedaquiline 200 mg QD on Days 1 to 56; plus pretomanid 200 mg QD on Days 1 to 56; plus pyrazinamide 1500 mg QD from Day 1 to Day 56.
- HRZE: Isoniazid 75 mg, rifampicin 150 mg, pyrazinamide 400 mg and ethambutol 275 mg on Days 1 to 56. The number of tablets the patients received depended on their weight.

Overnight sputum samples were collected before randomization (Days -2 and -1) and post-randomization (Days 3, 7, 14, 21, 28, 35, 42, 49, and 56). CFU counts on solid media were measured on each sputum sample collected. The CFU counts were calculated as:

$$\text{CFU} = \frac{1}{n}\sum_{x=1}^{n} \text{CFU}_x \times \text{factor} \times 10^{\text{dilution}} \tag{1}$$

where CFU_x is the count of culture plate x (from n replicate plates in total), and "factor $\times 10^{\text{dilution}}$" compensates for the counting process's dilution.

A total of 172 patients were included in the efficacy analysis. Supplementary Figure 1 shows profile plots of the observed $\log_{10}(\text{CFU})$ counts collected from the overnight sputum samples, by treatment group. The profile plot suggests that most profiles decline faster (on the log-scale) during the first few days than in the latter part of the curve. Summary statistics of the observed CFU counts over time are presented in Supplementary Table 1. Note that the counts' variance is considerably larger than the mean, and the percentage of zero counts significantly increases over time. These observations suggest that the CFU counts may be overdispersed relative to the Poisson distribution and, over time, zero-inflated relative to the negative binomial distribution. Furthermore, the mean counts are generally much larger than the median counts, implying that making inferences about the median counts may be more suitable than about the mean counts. Therefore, it might be preferable to use a method that models the median as a function of covariates (hence, quantile regression), while accounting for overdispersion and zero-inflation in the data.

2.2 Previous data analysis

The clinical trial's efficacy endpoints included the bactericidal activity of the three treatments quantified by overnight sputum samples. The bactericidal activity of anti-TB drugs was characterized by the rate of decline in $\log_{10}(\text{CFU})$ count.[19,20] In particular, the bactericidal activity over a certain time interval, calculated from a $\log_{10}(\text{CFU})$ vs. time profile, was expressed as follows[21]:

$$\text{BA}\,(t_1\text{–}t_2) = -\frac{\hat{f}(t_2) - \hat{f}(t_1)}{t_2 - t_1} \tag{2}$$

where $f(t)$ is the regression function for $\log_{10}(\text{CFU})$ count vs. time, and $\hat{f}(t_1)$ & $\hat{f}(t_2)$ are the corresponding fitted values at Day t_1 & Day t_2, respectively.

The regression model by Burger and Schall[15] was fitted to analyze the bactericidal activity of the three anti-TB treatments. More specifically, the following regression model was fitted to the $\log_{10}(\text{CFU})$ counts collected from overnight sputum samples, observed from Day 0 to Day 56[15,22]:

$$y_{ijk} = \alpha_{ij} - \theta_{1ij} t_{ijk} - \theta_{2ij}\gamma_j \log\left(\frac{e^{\frac{t_{ijk}-\kappa_j}{\gamma_j}} + e^{-\frac{t_{ijk}-\kappa_j}{\gamma_j}}}{e^{\frac{\kappa_j}{\gamma_j}} + e^{-\frac{\kappa_j}{\gamma_j}}}\right) + \varepsilon_{ijk} \tag{3}$$

where $\alpha_{ij} = \alpha_j + u_{0ij}$, $\theta_{1ij} = \theta_{1j} + u_{1ij}$, $\theta_{2ij} = \theta_{2j} + u_{2ij}$. y_{ijk} is the $\log_{10}(\text{CFU})$ count for patient $i = 1, \ldots, N_j$ in treatment group $j = 1, \ldots, J$ at timepoint $k = 1, \ldots, K_{ij}$, and t_{ijk} is the corresponding measurement time. u_{0ij}, u_{1ij}, and u_{2ij} denote random coefficients for patient i assigned to treatment j, and ε_{ijk} is the residual at time t_{ijk} of patient i assigned to treatment group j. α_{ij} are the random intercepts, whereas $\beta_{1ij} = \left(\theta_{1ij} - \theta_{2ij}\right)$ and $\beta_{2ij} = \left(\theta_{1ij} + \theta_{2ij}\right)$ are the random slopes respectively during the treatment period's initial and terminal phase. κ_j are the nodes at which the regression functions transition from one slope to another, whereas γ_j govern the "smoothness" or "speed" of the transition from one slope to another. The residuals and random effects were respectively assumed to follow normal and multivariate normal distributions.

Supplementary Figure 2 shows examples of the regression curve for $\log_{10}(\text{CFU})$ count over time. For this example, Equation (3) is reparameterized in terms of β_{1ij} and β_{2ij}. The regression parameters $\alpha_{ij} = 5$, $\kappa_j = 14$, and $\gamma_j = 2$ are kept fixed, but showed for different values of β_{1ij} and β_{2ij}, i.e., (i) $\beta_{1ij} = 0.05$ & $\beta_{2ij} = 0.15$, (ii) $\beta_{1ij} = 0.25$ & $\beta_{2ij} = 0.05$, and (iii) $\beta_{1ij} = 0.15$ & $\beta_{2ij} = 0.15$.

2.3 Extensions of the basic model

Burger and Schall[23] extended the previous model[15] (i.e., Equation (3)) by replacing the normal distributions for residuals and random effects with skew-t distributions to accommodate outliers and skewness in $\log_{10}(\text{CFU})$ counts (often due to data contamination).[24] The model based on the skew-t distribution makes inferences about the *mean* $\log_{10}(\text{CFU})$ count over time. However, due to skewness in the data, it would seem more appropriate to model the *median* count over time rather than the mean count.

In order to accommodate potential zero-inflation in CFU counts, Burger et al.[20] extended the previous model[15] by modeling the CFU counts on the original scale based on the ZINB distribution rather than modeling the logarithmic counts. The ZINB model of Burger et al.[20] was applied to the CFU dataset of Dawson et al.[25] However, the ZINB distribution also models the bactericidal activity of TB treatments based on the mean count. To model the median count, replacing the ZINB distribution with the ZIDW distribution can be considered.[11,26]

The ZINB model by Burger et al.[20] assumed that the mean of the "count" counterpart of the ZINB distribution, i.e., the NB counterpart, is biphasic over time. However, biologically, the *marginal* mean count is assumed to follow a biphasic trend over time.[14] In order to model the marginal mean and median, the conventional ZINB and ZIDW models can be replaced by marginal count models such as the marginal ZINB (MZINB) and marginal ZIDW (MZIDW) models.[13,27]

In summary, the data of Tweed et al.[18] motivated the fit of a biphasic regression model based on the MZINB and MZIDW distributions to make inferences about the mean and median CFU counts over time collected from overnight sputum samples.

3 CONVENTIONAL ZERO-INFLATED DISCRETE WEIBULL DISTRIBUTION

The conventional discrete Weibull distribution's key properties, including its dispersion, zero-inflation, and heavy-tail indices, can be found in Luyts et al.[8] Burger et al.[11] present an extension of the DW distribution by adding a zero-inflation parameter; hence, the so-called ZIDW distribution. The remainder of this section provides a summary of the ZIDW distribution.

If Y_i follows a ZIDW distribution, then the probability mass function (PMF) of Y_i is given by:

$$f(y_i) = \pi I(y_i = 0) + (1-\pi)\left[q^{y_i^{\phi}} - q^{(y_i+1)^{\phi}}\right]$$

where $y_i \in \{0,1,2,\ldots\}$. Here, $0 < q < 1$ and $\phi > 0$ denote the shape parameters, and $0 < \pi < 1$ is the zero-inflation probability of the ZIDW distribution. $I(a)$ denotes an indicator function taking the value 1 if condition a is true, and 0 otherwise. Supplementary Figure 3 shows examples of the ZIDW distribution's PMF for various values of q, ϕ, and π. This figure suggests that the distribution's tail becomes longer for larger values of q and smaller values of ϕ.

The mean and variance of Y_i are written as:

$$E(Y_i) = (1-\pi)\sum_{n=1}^{\infty} q^{n^{\phi}}$$

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

The cumulative distribution function of Y_i is given by:

$$F(y_i) = \sum_{x_i=0}^{y_i} f(x_i) = (\pi-1)\,q^{(y_i+1)^{\phi}} + 1$$

The τ-quantile function of Y_i is written as[11]:

$$Q(\tau) = \left(\frac{\log\left(\frac{\tau-1}{\pi-1}\right)}{\log(q)}\right)^{\frac{1}{\phi}} - 1 \tag{4}$$

The dispersion, zero-inflation, and heavy-tail indices of Luyts et al.,[8] show that the ZIDW distribution can accommodate zero-inflation, zero-deflation, over- and underdispersion, and severe skewness in the data (depending on the parameter combinations). Furthermore, the quantile function of the ZIDW distribution (Equation (4)) is available in closed form, making it convenient to model the distribution's quantiles, of which the median is a special case. Unlike the ZIDW distribution, the ZINB distribution's median cannot be written in closed form and lacks robustness to zero-deflation, underdispersion, and excessive skewness data. In this regard, the ZIDW distribution is considered more flexible than the ZINB distribution.

4 MARGINAL ZERO-INFLATED COUNT DISTRIBUTIONS

This section provides the key properties of the marginal count distribution. In particular, we reparameterize the zero-inflated negative binomial (ZINB) and zero-inflated discrete Weibull (ZIDW) distributions to model the marginal mean and median counts.

4.1 Marginal zero-inflated negative binomial distribution

The PMF of the conventional ZINB distribution for a given count y_i is written as:

$$f(y_i|\mu,\phi,\pi) = \pi I(y_i=0) + (1-\pi)\binom{y_i+\phi-1}{y_i}\left(\frac{\phi}{\mu+\phi}\right)^{\phi}\left(\frac{\mu}{\mu+\phi}\right)^{y_i}$$

Here, ϕ and π are respectively the dispersion parameter and zero-inflation probability, and μ is the mean of the conventional NB distribution. The mean of the y_i under the ZINB distribution is given by:

$$E(y_i) = \lambda = (1-\pi)\mu$$

Usually, the log-link function is used to model μ. Preisser et al.[13] considered modeling λ instead of μ. Therefore, the ZINB distribution is reparameterized in terms of λ to obtain the marginal ZINB (MZINB) distribution as follows:

$$f(y_i|\lambda,\phi,\pi) = \pi I(y_i=0) + (1-\pi)\binom{y_i+\phi-1}{y_i}\left(\frac{\phi(1-\pi)}{\lambda+\phi(1-\pi)}\right)^{\phi}\left(\frac{\lambda}{\lambda+\phi(1-\pi)}\right)^{y_i}$$

Alternatively one can use the log-link function to model the marginal mean counts as a function of covariates as follows:

$$E(y_i) = \lambda_i^* = e^{z_i'\boldsymbol{\eta}}$$

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

4.2 Marginal zero-inflated discrete Weibull distribution

The PMF of the ZIDW distribution, using the parameterization of Burger et al. [11] for a given count y_i, is written as:

$$f(y_i|\mu,\phi,\pi) = \pi I(y_i = 0) + (1-\pi)\left[\exp\left(-\log(2)\left[\frac{y_i}{\mu}\right]^{\phi}\right) - \exp\left(-\log(2)\left[\frac{y_i+1}{\mu}\right]^{\phi}\right)\right]$$

Here, ϕ and π are respectively the shape parameter and zero-inflation probability, and μ is the median of the conventional discrete Weibull distribution. The median of the y_i under the ZIDW distribution is given by:

$$M(y_i) = \lambda = \left(\frac{\log\left[\frac{0.5}{1-\pi}\right]}{\log(0.5)}\right)^{\frac{1}{\phi}}\mu$$

Recently, Burger et al. [11] suggested the use of the log-link function to model μ. Taking an approach similar to that of Preisser et al., [13] λ's modeling is suggested instead of μ. The reparameterization results in the marginal ZIDW (MZIDW) distribution:

$$f(y_i|\lambda,\phi,\pi) = \pi I(y_i = 0) + 0.5^{\left(\frac{y_i}{\lambda}\right)^{\phi}}(1-\pi)^{1-\left(\frac{y_i}{\lambda}\right)^{\phi}} - 0.5^{\left(\frac{y_i+1}{\lambda}\right)^{\phi}}(1-\pi)^{1-\left(\frac{y_i+1}{\lambda}\right)^{\phi}}$$

Again, one can alternatively use the log-link function to model the marginal median counts as a function of covariates as follows:

$$M(y_i) = \lambda_i^* = e^{z_i'\boldsymbol{\eta}}$$

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

5 BAYESIAN MIXED-EFFECTS REGRESSION MODELS

5.1 Biphasic mixed-effects regression model

Suppose that y_{ijk} is the observed count for patient $i = 1,\ldots,N_j$ in treatment group $j = 1,\ldots,J$ at timepoint $k = 1,\ldots,T_{ij}$ (or T), and $t_{ijk} \geq 0$ are the corresponding measurement times. Furthermore, assume that λ_{ijk} are the patient-specific regression functions that describe the counts' biphasic trend over time. A slightly different parameterization of the regression function by Burger and Schall [15] is chosen as:

$$\lambda_{ijk} = \exp\left(\alpha_{ij} - \frac{\beta_{1ij}+\beta_{2ij}}{2}t_{ijk} - \frac{\beta_{1ij}-\beta_{2ij}}{2}\gamma_j \log\left[\frac{e^{\frac{t_{ijk}-\kappa_j}{\gamma_j}} + e^{-\frac{t_{ijk}-\kappa_j}{\gamma_j}}}{e^{\frac{\kappa_j}{\gamma_j}} + e^{-\frac{\kappa_j}{\gamma_j}}}\right] - o_{ijk}\right) \tag{5}$$

where α_{ij} are the random intercepts, β_{1ij} & β_{2ij} the two random slopes, κ_j the change-points (or "nodes"), γ_j the smoothness parameters, and o_{ijk} are offset constants. [15,20] Like conventional Poisson and negative binomial regression, the offset constants o_{ijk} are implemented should it be more relevant to model the mean or median "rates" instead of the counts. In the context of our motivating dataset, an offset function may be necessary to account for the dilution of sputum samples that may have been required to count the bacteria (i.e., CFU counts; see Section 6 for more details).

The multivariate normal distribution is assumed to model the correlation among the random coefficients. Let $\boldsymbol{\delta}_{ij} = (\alpha_{ij}, \beta_{1ij}, \beta_{2ij})'$ and $\boldsymbol{\delta}_j = (\alpha_j, \beta_{1j}, \beta_{2j})'$ respectively represent the vectors of random and fixed intercepts and slopes. Therefore, $\boldsymbol{\delta}_{ij}$ are normally distributed as follows:

$$\boldsymbol{\delta}_{ij} \sim \text{Normal}\left(\boldsymbol{\delta}_j, \boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}\right)$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j} = \begin{bmatrix} \sigma^2_{\alpha_j} & \sigma_{\alpha_j\beta_{1j}} & \sigma_{\alpha_j\beta_{2j}} \\ \sigma_{\alpha_j\beta_{1j}} & \sigma^2_{\beta_{1j}} & \sigma_{\beta_{1j}\beta_{2j}} \\ \sigma_{\alpha_j\beta_{2j}} & \sigma_{\beta_{1j}\beta_{2j}} & \sigma^2_{\beta_{2j}} \end{bmatrix}$$

are the covariance matrices of $\boldsymbol{\delta}_{ij}$.

5.2 MZINB regression model

The PMF of the MZINB regression model for a given count y_{ijk} (i.e., for patient i, treatment j, and timepoint k) is written as:

$$f\left(y_{ijk}|\lambda_{ijk},\phi_j,\pi_{jk}\right) = \pi_{jk} I\left(y_{ijk}=0\right) + \left(1-\pi_{jk}\right)\binom{y_{ijk}+\phi_j-1}{y_{ijk}}\left(\frac{\phi_j\left(1-\pi_{jk}\right)}{\lambda_{ijk}+\phi_j\left(1-\pi_{jk}\right)}\right)^{\phi_j}\left(\frac{\lambda_{ijk}}{\lambda_{ijk}+\phi_j\left(1-\pi_{jk}\right)}\right)^{y_{ijk}}$$

where λ_{ijk} are the patient-specific biphasic regression functions in Equation (5), and ϕ_j & π_{jk} are respectively the dispersion parameter and zero-inflation probability of the MZINB distribution. The conditional mean of y_{ijk} under the MZINB regression model is given by $E\left(y_{ijk}\right) = \lambda_{ijk}$.

5.3 MZIDW regression model

The PMF of the MZIDW regression model for a given count y_{ijk} (i.e., for patient i, treatment j, and timepoint k) is written as:

$$f\left(y_{ijk}|\lambda_{ijk},\phi_j,\pi_{jk}\right) = \pi_{jk} I\left(y_{ijk}=0\right) + 0.5^{\left(\frac{y_{ijk}}{\lambda_{ijk}}\right)^{\phi_j}}\left(1-\pi_{jk}\right)^{1-\left(\frac{y_{ijk}}{\lambda_{ijk}}\right)^{\phi_j}} - 0.5^{\left(\frac{y_{ijk}+1}{\lambda_{ijk}}\right)^{\phi_j}}\left(1-\pi_{jk}\right)^{1-\left(\frac{y_{ijk}+1}{\lambda_{ijk}}\right)^{\phi_j}}$$

where λ_{ijk} are the patient-specific biphasic regression functions in Equation (5), and ϕ_j & π_{jk} are respectively the shape parameter and zero-inflation probability of the MZIDW distribution. The conditional median of y_{ijk} under the MZIDW regression model is given by $M\left(y_{ijk}\right) = \lambda_{ijk}$.

6 DATA ANALYSIS

Details on the Bayesian model specification of the MZINB and MZIDW models are presented in the supplementary material of this paper. In summary, we specified vague prior distributions for the model parameters, i.e., normal distributions for the fixed intercepts & slopes, matrix-generalized half-t distributions for the random effects covariance matrices,[28] uniform distributions for the node, smoothness & zero-inflation parameters, and gamma distributions for the dispersion & shape parameters.

6.1 Model implementation and computational issues

The MZINB and MZIDW regression models were implemented according to the model specifications discussed in Section 5. In Equation (5), $y_{ijk} = \sum_{l=1}^{n_{ijk}} \text{CFU}_{ijkl}$ is the total of n_{ijk} bacterial plate counts for patient i, treatment group j, and timepoint k. The offsets are expressed as $o_{ijk} = \log\left(c_{ijk} 10^{d_{ijk}} / n_{ijk}\right)$, where n_{ijk}, c_{ijk} & d_{ijk} are respectively n, "factor" and "dilution" as per Equation (1). The CFU counts collected before randomization (Days -2 and -1) were considered as Day 0 collections (i.e., $t_{ijk} = 0$). The posterior estimate of $\frac{\log(\lambda_{ijk}) + o_{ijk}}{\log(10)}$ can be interpreted as the "fitted" $\log_{10}(\text{CFU})$ count for patient i, treatment j, and timepoint k.

The `R` code for implementing the MZINB and MZIDW models is presented in the supplementary material of this manuscript.

The bactericidal activity of treatment group j is expressed as the daily rate of change in $\log_{10}(\text{CFU})$ count over timepoints k_1 and k_2, namely (see Equation (2)):

$$\text{BA}_j(t_{k_1}\text{–}t_{k_2}) = -\frac{\log\left[M_j(t_{k_2})\right] - \log\left[M_j(t_{k_1})\right]}{\log(10)\left(t_{k_2} - t_{k_1}\right)} \tag{6}$$

where

$$M_j(t_k) = \exp\left(\alpha_j - \frac{\beta_{1j} + \beta_{2j}}{2} t_k - \frac{\beta_{1j} - \beta_{2j}}{2} \gamma_j \log\left[\frac{e^{\frac{t_k - \kappa_j}{\gamma_j}} + e^{-\frac{t_k - \kappa_j}{\gamma_j}}}{e^{\frac{\kappa_j}{\gamma_j}} + e^{-\frac{\kappa_j}{\gamma_j}}}\right]\right) \tag{7}$$

Here, $\log_{10}\left(M_j(t_k)\right)$ derived by regression models MZINB and MZIDW is respectively the mean and median $\log_{10}(\text{CFU})$ count at time t_k of treatment group j. For our analysis, BA_j (0–56) was of primary interest.

In order to avoid numerical overflow, the regression models were fitted with the times t_{ijk} expressed in weeks instead of days. In order to avoid non-identifiability of the parameters κ_j and γ_j, the lower and upper bounds of κ_j and γ_j were set to $L_\kappa = 3$, $U_\kappa = 11$, $L_\gamma = 0.05$, and $U_\gamma = 2$ (see Section 3.1 of Burger and Schall [15]). It should be noted that the median of the MZIDW distribution exists only for $0 < \pi_{jk} < 0.5$. Therefore, the MZINB and MZIDW regression models' upper bound of π_{jk} was respectively set to 1 and 0.5.

The regression models were fitted using `JAGS` [29] via the package `runjags` [30] of the `R` project. [31] The convergence of posterior samples was confirmed using trace plots and Brooks-Gelman-Rubin statistics. [32]

Starting values for the random effects were derived by fitting the model as a linear mixed-effects regression model under the assumption that the node and smoothness parameters are respectively fixed at $\kappa_j = (L_\kappa + U_\kappa)/2 = 7$ and $\gamma_j = (L_\gamma + U_\gamma)/2 = 1.025$. The linear mixed-effects regression model was fitted to the $\log_{10}(\text{CFU})$ counts using the R project's `lmec` library,[33,34] where on the log-10 scale, the zero counts were specified as left-censored values of 1.

The `R` project was called remotely from `SAS`®,[35] and accordingly, posterior samples were exported back to `SAS`® for further computation. For each regression model, 82500 samples were simulated from the joint posterior distribution for 7 parallel chains. Among those 82500 samples (per chain), the initial 15000 samples were discarded (burn-in). High autocorrelation in the posterior samples was present. We therefore used a thinning factor of 450 to reduce autocorrelation among the samples. We ran our models on a desktop computer with a 3.00 GHz Intel® Core™ i9-10980XE processor and 64 GB installed memory (RAM). The MZINB and MZIDW models, respectively, took approximately 45 and 60 minutes to run.

We calculated the compound Laplace-Metropolis marginal likelihood (CLMML)[36,37] to discriminate between the candidate models. The CLMML compares the candidate models based on their marginal likelihood, therefore, not conditional on the random effects. Thus, the CLMML is more appropriate than the widely used deviance information criterion statistic conditional on the random effects.[38] Details on the calculation of the CLMML are presented in the supplementary material of this paper. The multidimensional integration library `cubature` of the `R` project was used to approximate the Laplace integrals.[39]

6.2 Results

Plots of the observed $\log_{10}(\text{CFU})$ counts together with fits of regression models MZINB and MZIDW are included in Figure 1 for eight randomly selected patients. The two regression models fit the data of these profiles well.

Table 1 presents the posterior estimates (PEs) and 95% highest posterior density (HPD) intervals of the bactericidal activity ($\text{BA}_j(0\text{–}56)$; see Equation (6)) and the regression model parameters. The PEs and 95% HPD intervals of the mean and median $\log_{10}(\text{CFU})$ counts (see Equation (7)) are shown in Figure 2 by treatment group and day. As expected (due to skewness in the data), the median counts are smaller than the mean counts. The PEs and 95% HPD intervals of each treatment's mean and median bactericidal activity are similar. $\text{J}_{(\text{loading dose/t.i.w.})}$PaZ shows the highest mean and median bactericidal activity, followed by $\text{J}_{(200\text{ mg})}$PaZ and HRZE; the difference between

the latter two treatments is negligible.

The log-CLMML for models MZINB and MZIDW is respectively -9906.12 and -9932.15. Hence, the CLMML favors the MZINB model over the MZIDW model. Since the two models are structured differently (i.e., modeling the mean vs. the median), one should not discriminate between models solely based on model comparison statistics (such as the CLMML); one should also consider the models' goodness of fit. As previously indicated, the profile plots in Figure 1 suggest that both models fit the data adequately.

7 SIMULATION STUDY

7.1 Model performance

We assessed the performance and the identifiability of regression models MZINB and MZIDW outlined in Section 5 in a simulation study. Datasets were simulated from regression models MZINB and MZIDW, where the model parameters were chosen to mimic the CFU count vs. time profiles of moderately and highly efficacious anti-TB drugs, respectively, each with and without zero-inflation in CFU count during the second phase of treatment.

Each fit was considered for a single treatment group separately (i.e., $j = 1$). Data on the following days were considered: $t_{ijk} \in \{0,1,3,7,14,21,28,35,42,49,56\}$. The offset, intercept, node, smoothness parameter, variance components, and dispersion/shape parameter for both models were chosen as $o_{ijk} = 5.063315552$ (i.e., $n_{ijk} = 4$, $c_{ijk} = 20$, $d_{ijk} = 1.5$), $\alpha_1 = 12.65$, $\kappa_1 = 5.25$, $\gamma_1 = 1.25$, $\sigma^2_{\alpha_1} = 0.95$, $\sigma^2_{\beta_{11}} = 0.85$, $\sigma^2_{\beta_{21}} = 0.80$, $\sigma_{\alpha_1\beta_{11}} = 0.25$, $\sigma_{\alpha_1\beta_{21}} = 0.35$, $\sigma_{\beta_{11}\beta_{21}} = -0.15$, and $\phi_1 = 0.70$. For simplicity, the zero-inflation probabilities were pooled by (i) Days 0, 1, 3, 7 ($k = 1$), and (ii) Days 14, 21, 28, 35, 42, 49, 56 ($k = 2$). The parameter scenarios were investigated for the following sets of slope terms: (i) $\beta_{11} = 3.95$, $\beta_{21} = 1.35$, and (ii) $\beta_{11} = 4.20$, $\beta_{21} = 1.80$. Both parameter scenarios were investigated for the following two sets of zero-inflation probabilities ($k = 1,2$): (i) $\pi_{11} = 0.01$, $\pi_{12} = 0.01$ (without zero-inflation), and (ii) $\pi_{11} = 0.01$, $\pi_{12} = 0.12$ (with zero-inflation). The regression models were fitted to 500 simulated datasets, each dataset consisting of 15 patient profiles ($i = 1,\ldots,15$).

The bias and root mean square error (RMSE) of the PEs were calculated, as was the average length and empirical coverage probability of the associated 95% HPD intervals. For simplicity, the regression models fitted the node and smoothness parameters as fixed values. The `autorun.jags`

function of the `runjags` package[30] was used to guarantee the successful convergence of the posterior samples for each fitted dataset.

From Table 2, we observe the following:

- Under each parameter scenario, the bias of the estimates of the fixed effects (i.e., α_1, β_{11}, and β_{21}) and dispersion/shape parameters (i.e., ϕ_1) from both regression models is small. In contrast, those of the variance components (i.e., $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}$) are somewhat large.
- The zero-inflation probability estimates of the second phase (i.e., π_{12}) are (i) biased upwards when there is no zero-inflation present in the second phase and (ii) less biased when there is zero-inflation present in the second phase. We note that our parameter scenarios mimic data where the counts during the second phase of treatment are primarily zero due to the treatment effect (as governed by the regression slopes; for example, CFU counts typically approach zero closer to the end of treatment). The latter is the case regardless of the presence of zero-inflation in the data. Therefore, it seems that adding a zero-inflation parameter for the second phase of treatment, in the absence of zero-inflation, may cause identifiability issues when the counts are inherently zero (such as "late phase" CFU counts in TB trials). We note that the zero-inflation probability estimates' bias of the first phase (i.e., π_{11}) is small under each parameter scenario.
- Under each parameter scenario, both regression models yield HPD interval coverage probabilities close to or slightly higher than the nominal value, except for the variance components' coverage of the first slope (i.e., $\sigma^2_{\beta_{11}}$, $\sigma_{\alpha_1\beta_{11}}$, and $\sigma_{\beta_{11}\beta_{21}}$) that is too conservative (equal to 100%).

We repeated the simulation study of the third parameter scenario ($\beta_{11} = 4.20$, $\beta_{21} = 1.80$, $\pi_{11} = 0.01$, $\pi_{12} = 0.01$) by simulating 100 datasets for a sample size of 50 patients ($i = 1, \ldots, 50$). The corresponding results are presented in Supplementary Table 2. We observe that the bias and HPD interval coverage under the larger and smaller sample size settings are similar. The PEs' RMSE and the associated average HPD interval length increase under the larger sample size setting.

7.2 Data contamination

We assessed the impact of the misspecification of the random effects distribution on the inferences of the fixed effects. Here, we considered the second parameter scenario in Section 7.2 ($\beta_{11} = 3.95$,

$\beta_{21} = 1.35$, $\pi_{11} = 0.01$, $\pi_{12} = 0.12$). We generated the random effects $\boldsymbol{\delta}_{ij}$ in Section 5.1 from the skew-normal distribution of Sahu et al.[40] We reparameterized the skew-normal distribution such that $E\left(\boldsymbol{\delta}_{ij}\right) = \boldsymbol{\delta}_j$ and $Var\left(\boldsymbol{\delta}_{ij}\right) = \Sigma_{\boldsymbol{\delta}_j} + \left(1 - \frac{2}{\pi}\right)\Psi^2_{\boldsymbol{\delta}_j}$. Here, $\Sigma_{\boldsymbol{\delta}_j}$ are the scale matrices of the skew-normal distribution. $\Psi_{\boldsymbol{\delta}_j}$ denote diagonal matrices with $\psi_{\alpha_j}, \psi_{\beta_{1j}}, \psi_{\beta_{2j}}$ on the diagonal, i.e., the skewness parameters of the random intercepts and two random slopes. We introduced skewness to the random intercepts only (i.e., $\psi_{\beta_{1j}} = \psi_{\beta_{2j}} = 0$). We considered the following three levels of skewness under the skew-normal distribution: $\psi_{\alpha_j} \in \{-2, -3, -4\}$. The corresponding results are presented in Supplementary Table 3 (fixed effects only). The bias and HPD interval coverage are adequate under each level of skewness. Table 3 presents the percentage difference in the RMSEs and HPD interval lengths between the contaminated and uncontaminated groups for each skewness level (ψ_{α_j}). Both models' RMSE and HPD interval length increase under higher contamination rates (skewness levels). However, the increase in these characteristics under the MZINB model is higher than that of the MZIDW model, most notably the fixed intercept term. The latter suggests that the MZIDW model is more robust to model misspecifications than the MZINB model.

8 DISCUSSION

This paper proposed a biphasic regression model for zero-inflated longitudinal counts based on the MZINB and MZIDW distributions. The current approach draws on existing literature while presenting an extension for Bayesian inference of biphasic median counts over time.

We demonstrated our methods through a reanalysis of the TB dataset of Tweed et al.[18] The nonlinear $\log_{10}(\text{CFU})$ vs. time profiles indicate that one should preferably fit nonlinear zero-inflated models to accommodate zero counts toward the end of the treatment period. Therefore, we chose to model the data using biphasic models based on the MZINB and MZIDW distributions. The biphasic model consists of an intercept term, two regression slopes, an unknown change-point (at which the slope transitions from one rate of decline to another), and a smoothness parameter governing the transition speed from one slope to another. The biphasic MZIDW model extends the conventional linear mixed-effects ZIDW model of Burger et al.[11] to describe the overall median counts over time instead of the subpopulation's latent median, that is, that of the uncured TB patients (i.e., the "at-risk" subpopulation). Similarly, the MZINB model models the biphasic mean counts on a marginal basis. In the case of modeling CFU counts in anti-TB trials such as our application, the median counts can be considered more informative than the mean count, given the data's severe skewness. For our application, the bactericidal activity of each treatment (i.e., the rate of decline in CFU counts) characterized by the mean and median counts is similar.

While the coverage of the HPD intervals for the variance components is generally very conservative, the coverage for the fixed effects is in most cases good (that is, close to the nominal coverage). We note that the bias of the zero-inflation probabilities increases over time when the counts are zero due to steep regression slopes without the presence of zero-inflation. Such instances may result in an overparameterized model, causing some parameters to become nonidentifiable. Care should, therefore, be taken to avoid such overparameterizations which potentially may yield biased estimates for the zero-inflation probabilities. The implementation of the MZINB and MZIDW models as hurdle models[41] may potentially circumvent such identifiability issues. Overall, the simulation study suggests that the proposed biphasic regression models have adequate properties in terms of accuracy and confidence interval coverage.

The "robustness" simulation study suggests that the inference about the mean counts (MZINB model) is more sensitive to model misspecifications compared to the median (MZIDW model). Extended versions of the discrete Weibull distribution may be considered for more flexibility concerning the distribution's tails.

The proposed mixed-effects MZINB and MZIDW regression models can be further extended to investigate the association between biphasic longitudinal counts and time-to-event outcomes,[42] for example, in the context of TB trials, performing biomarker analyses to assess the association between CFU counts and "time to sputum culture conversion."[25,43] The quantile function of the discrete Weibull distribution is available in closed form, making it possible to model quantiles of counts other than the median.[44]

We note that the interpretation of the models' fixed effects is not globally marginal (over all the random effects) but instead conditional on the random effects. To enable interpreting the fixed effects on a truly marginal basis, one must first integrate over the likelihood function's random effects before specifying it in the Gibbs sampling algorithm. As a subject of future research, the methods proposed by Heagerty and Zeger[45] & Geraci and Bottai[9] can be considered to perform such "marginal" analyses.

The proposed regression models treat the treatment-specific node parameters as fixed effects. As was initially done by Burger and Schall,[15] the models can treat the nodes as random effects to accommodate individual variation in the nodes (hence random change-point models[46]). The model can also be generalized to model the median counts of zero-inflated longitudinal counts based on nonlinear functions other than biphasic curves, i.e., nonlinear mixed-effect regression models in the general case.

Even though the median is a more robust measure of central tendency than the mean when the data, for instance, zero-inflated counts, are right-skewed, it may also be of clinical interest to report the mean of longitudinal count data together with the median. Therefore, we recommend fitting both the MZINB and MZIDW distributions to yield inferences about the two central tendency characteristics: the mean and median.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the TB Alliance for providing the data of Tweed et al.[18] We also thank Robert Schall at the University of the Free State for comments that improved the manuscript's quality. 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 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.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The dataset supporting this study's findings is available on request from TB Alliance via the corresponding author. The data are not publicly available due to privacy and ethics restrictions.

REFERENCES

[1] Bacon DW, Watts DG. Estimating the transition between two intersecting straight lines. *Biom*. 1971; 58: 525–534. DOI: 10.1093/biomet/58.3.525.

[2] Ding A, Wu H. Relationships between antiviral treatment effects and biphasic viral decay

rates in modeling HIV dynamics. *Math. Biosci*. 1999; 160(1): 63–82. DOI: 10.1016/S0025–5564(99)00021–8.

[3] Muggeo VMR, Atkins DC, Gallop RJ, Dimidjian S. Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study. *Stat. Model*. 2014; 14(4): 293–313. DOI: 10.1177/1471082X13504721.

[4] Li H, Benitez A, Neelon B. A Bayesian hierarchical change point model with parameter constraints. *Stat. Methods Med. Res*. 0; 0(0): DOI: 10.1177/0962280220948097.

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

[6] Klein N, Kneib T, Lang S. Bayesian generalized additive models for location, scale, and shape for zero-inflated and overdispersed count data. *J. Am. Stat. Assoc*. 2015; 110(509): 405–419. DOI: 10.1080/01621459.2014.912955.

[7] Sae-Lim P, Grøva L, Olesen I, Varona L. A comparison of nonlinear mixed models and response to selection of tick-infestation on lambs. *PLoS ONE* 2017; 12(3): e0172711. DOI: 10.1371/journal.pone.0172711.

[8] 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. DOI: 10.1177/1471082X18789992.

[9] Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostat*. 2007; 8(1): 140–154. DOI: 10.1093/biostatistics/kxj039.

[10] King C, Song JJ. A Bayesian two-part quantile regression model for count data with excess zeros. *Stat. Model*. 2019; 19(6): 653–673. DOI: 10.1177/1471082X18799919.

[11] Burger DA, Schall R, Ferreira JT, Chen DG. A robust Bayesian mixed effects approach for zero inflated and highly skewed longitudinal count data emanating from the zero inflated discrete Weibull distribution. *Stat. Med*. 2020; 39(9): 1275–1291. DOI: 10.1002/sim.8475.

[12] Long DL, Preisser JS, Herring AH, Golin CE. A marginalized zero-inflated Poisson regression model with random effects. *J. Royal Stat. Soc. Ser. C* 2015; 64(5): 815–830. DOI: 10.1111/rssc.12104.

[13] Preisser JS, Das K, Long DL, Divaris K. Marginalized zero-inflated negative binomial regression with application to dental caries. *Stat. Med*. 2016; 35(10): 1722–1735. DOI: 10.1002/sim.6804.

[14] Gillespie SH, Gosling RD, Charalambous BM. A reiterative method for calculating the early bactericidal activity of antituberculosis drugs. *Am. J. Respir. Critical Care Med*. 2002; 166(1): 31–35. DOI: 10.1164/rccm.2112077.

[15] Burger DA, Schall R. A Bayesian nonlinear mixed-effects regression model for the characterization of early bactericidal activity of tuberculosis drugs. *J. Biopharm. Stat*. 2015; 25(6): 1247–1271. DOI: 10.1080/10543406.2014.971170.

[16] Gosling RD, Heifets L, Gillespie SH. A multicentre comparison of a novel surrogate marker of determining the specific potency of anti-tuberculosis drugs. *J. Antimicrob. Chemother*. 2003; 52(3): 473–476. DOI: 10.1093/jac/dkg345.

[17] Gosling RD, Uiso LO, Sam NE, Bongard E, Kanduma EG, Nyindo M, Morris RW, Gillespie SH. The bactericidal activity of moxifloxacin in patients with pulmonary tuberculosis. *Am. J. Respir. Critical Care Med*. 2003; 168(11): 1342–1345. DOI: 10.1164/rccm.200305–682OC.

[18] Tweed CD, Dawson R, Burger DA, Conradie A, Crook AM, Mendel CM, Conradie F, Diacon AH, Ntinginya NE, Everitt DE, *et al.* Bedaquiline, moxifloxacin, pretomanid, and pyrazinamide during the first 8 weeks of treatment of patients with drug-susceptible or drug-resistant pulmonary tuberculosis: a multicentre, open-label, partially randomised, phase 2b trial. *The Lancet Respir. Med*. 2019; 7(12): 1048–1058. DOI: 10.1016/S2213–2600(19)30366–2.

[19] Diacon AH, Dawson R, Von Groote-Bidlingmaier F, Symons G, Venter A, Donald PR, Van Niekerk C, Everitt D, Winter H, Becker P, *et al.* 14-day bactericidal activity of PA-824, bedaquiline, pyrazinamide, and moxifloxacin combinations: A randomized trial. *The Lancet* 2012; 380(9846): 986–993. DOI: 10.1016/S0140–6736(12)61080–0.

[20] 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.

[21] Jindani A, Doré CJ, Mitchison DA. Bactericidal and sterilizing activities of antituberculosis drugs during the first 14 days. *Am. J. Respir. Critical Care Med*. 2003; 167: 1348–1354. DOI: 10.1164/rccm.200210–1125OC.

[22] Statistical analysis plan of clinical study protocol NC-005-(J-M-PA-Z). https://clinicaltrials.gov/ProvidedDocs/76/NCT02193776/SAP_001.pdf; 2016. Accessed: 2020-12-28.

[23] 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.

[24] Van Zyl-Smit RN, Binder A, Meldau R, Mishra H, Semple PL, Theron G, Peter J, Whitelaw A, Sharma SK, Warren R, *et al.* Comparison of quantitative techniques including Xpert MTB/RIF to evaluate mycobacterial burden. *PLoS ONE* 2011; 6(12): e28815. DOI: 10.1371/journal.pone.0028815.

[25] Dawson R, Diacon AH, Everitt D, Van Niekerk C, Donald PR, Burger DA, Schall R, Spigelman M, Conradie A, Eisenach K, *et al.* Efficiency and safety of the combination of moxifloxacin, pretomanid (PA-824), and pyrazinamide during the first 8 weeks of antituberculosis treatment: A phase 2b, open-label, partly randomised trial in patients with drug-susceptible or drug-resistant pulmonary tuberculosis. *The Lancet* 2015; 385(9979): 1738–1747. DOI: 10.1016/S0140–6736(14)62002–X.

[26] 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.

[27] Cummings TH, Hardin JW. Modeling count data with marginalized zero-inflated distributions. *The Stata J*. 2019; 19(3): 499–509. DOI: 10.1177/1536867X19874209.

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

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

[30] 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.

[31] 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/.

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

[33] Vaida F, Liu L. R lmec package: linear mixed-effects models with censored responses. R package version 3.0.3. 2012.

[34] Vaida F, Liu L. Fast implementation for normal mixed effects models with censored response. *J. Comput. Graph. Stat*. 2009; 18(4): 797–817. DOI: 10.1198/jcgs.2009.07130.

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

[36] Raftery AE. Hypothesis testing and model selection. In: Gilks WR, Richardson S, Spiegelhalter DJ, eds. Markov Chain Monte Carlo in Practice. London, UK: Chapman and Hall; 1996:163–188.

[37] 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.

[38] Quintero A, Lesaffre E. Comparing hierarchical models via the marginalized deviance information criterion. *Stat. Med*. 2018; 37(16): 2440–2454. DOI: 10.1002/sim.7649.

[39] Narasimhan B, Koller M, Johnson SG, Hahn T, Bouvier A, Kiêu K, Gaure S, Narasimhan MB. cubature: Adaptive multivariate integration over hypercubes; 2020. R package version 2.0.4.1; URL https://CRAN.R-project.org/package=cubature.

[40] Sahu SK, Dey DK, Branco MD. A new class of multivariate skew distributions with applications to Bayesian regression models. *The Can. J. Stat*. 2003; 31(2): 129–150. DOI: 10.2307/3316064.

[41] Molas M, Lesaffre E. Hurdle models for multilevel zero-inflated data via h-likelihood. *Stat. Med*. 2010; 29(30): 3294–3310. DOI: 10.1002/sim.3852.

[42] Mchunu NN, Mwambi HG, Reddy T, Yende-Zuma N, Naidoo K. Joint modelling of longitudinal and time-to-event data: an illustration using CD4 count and mortality in a cohort of patients initiated on antiretroviral therapy. *BMC Infect. Dis*. 2020; 20: 1–9. DOI: 10.1186/s12879–020–04962–3.

[43] Phillips PPJ, Mendel CM, Burger DA, Crook AM, Nunn AJ, Dawson R, Diacon AH, Gillespie SH. Limited role of culture conversion for decision-making in individual patient care and for advancing novel regimens to confirmatory clinical trials. *BMC Med*. 2016; 14: 19. DOI: 10.1186/s12916–016–0565–y.

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

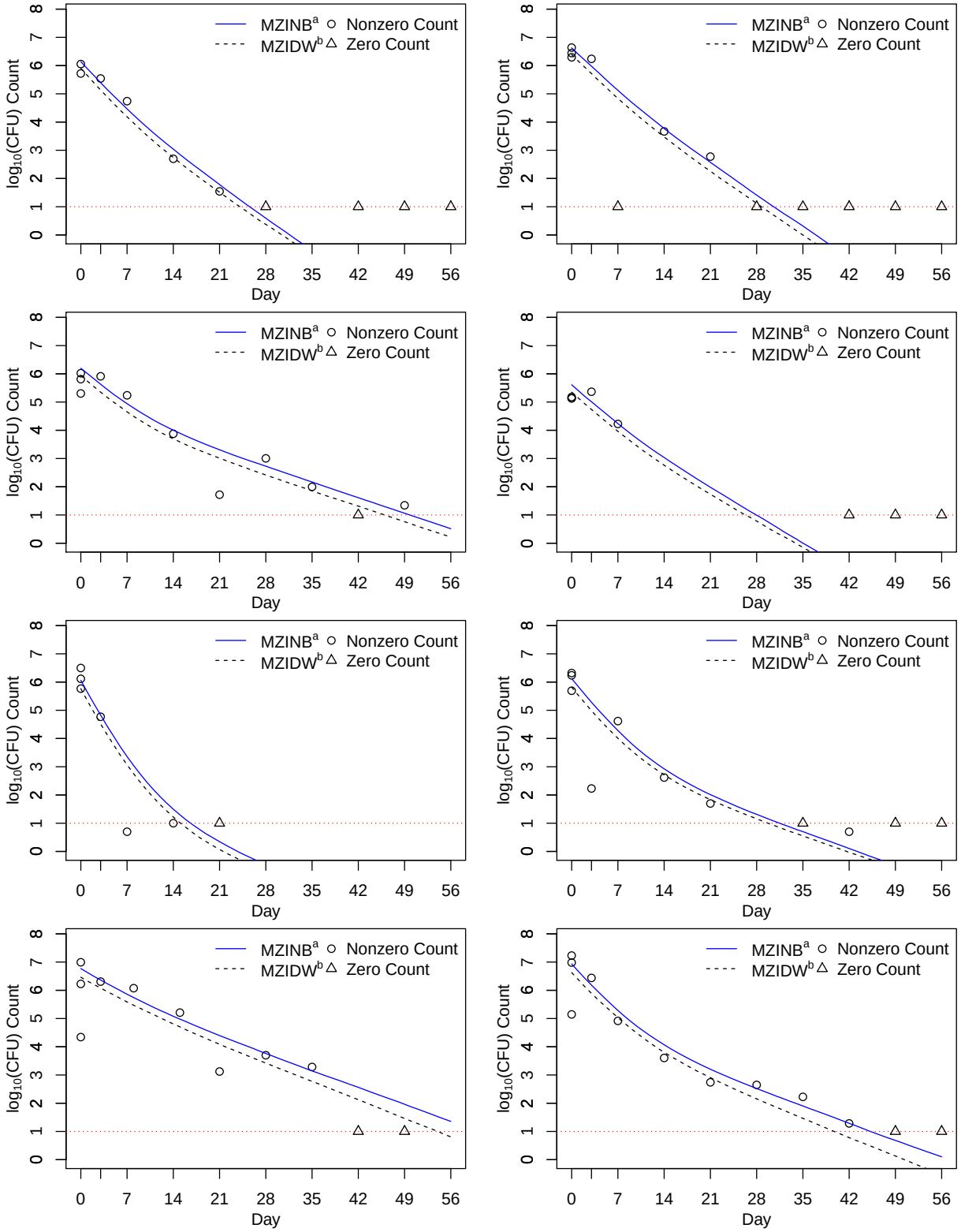
[45] Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Stat. Sci*. 2000; 15(1): 1–26. DOI: 10.1214/ss/1009212671.

[46] Dominicus A, Ripatti S, Pedersen NL, Palmgren J. *Modelling variability in longitudinal data using random change point models*. Mathematical Statistics, Stockholm University; 2006.

[47] 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.

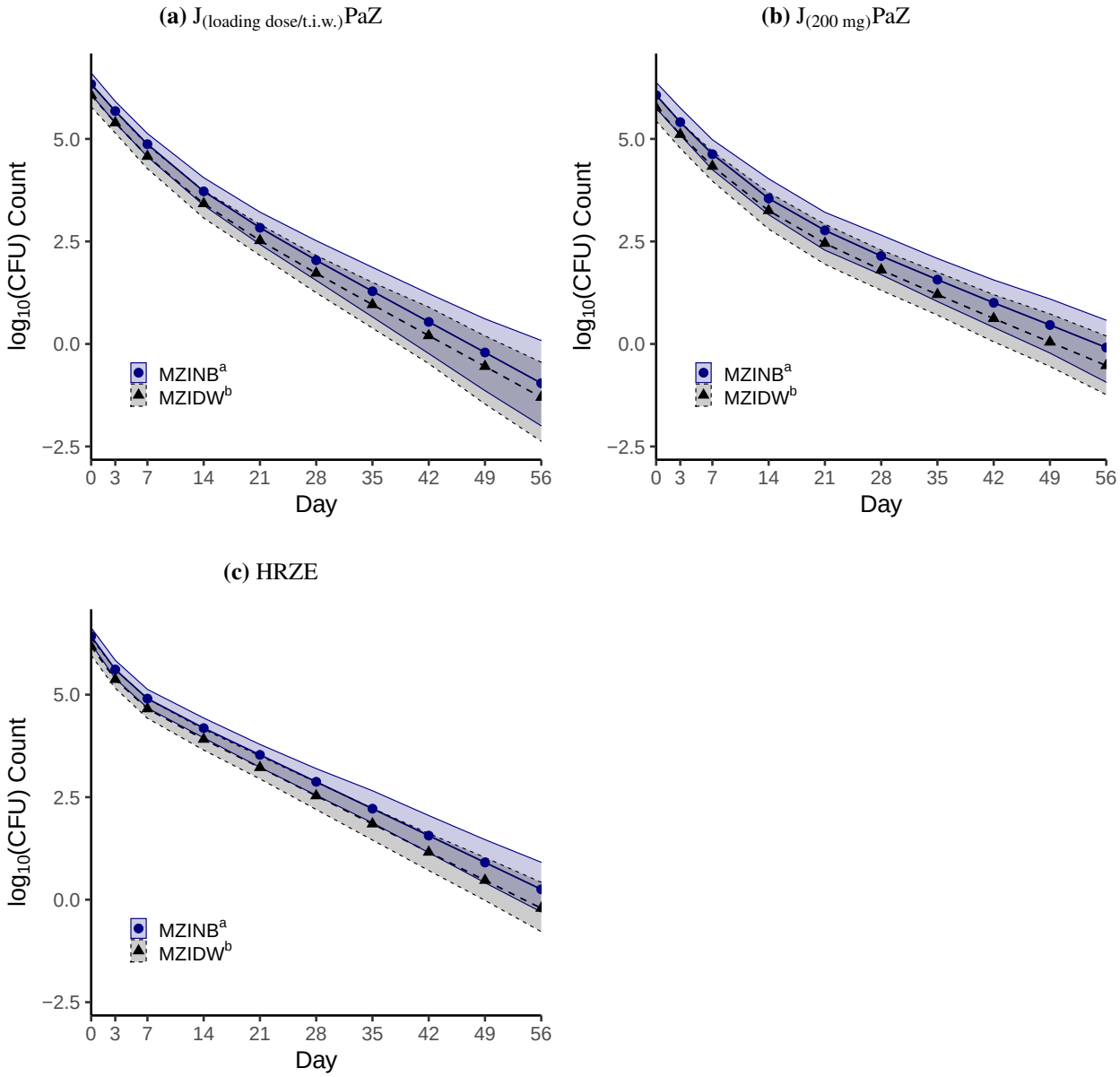
[48] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc*. 1990; 85: 398–409. DOI: 10.1080/01621459.1990.10476213.

Figure 1: TB dataset: observed and fitted $\log_{10}(\text{CFU})$ vs. time profiles



CFU: Colony-forming unit. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. [a]The log-link function is used to describe the biphasic relationship between the *mean* CFU count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* CFU count and time. CFU counts of zero are displayed on the log-10 scale as 1.

Figure 2: TB dataset: PEs and 95% HPD intervals of mean and median $\log_{10}(\text{CFU})$ count



CFU: Colony-forming unit. HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. PE: Posterior estimate. [a]The log-link function is used to describe the biphasic relationship between the *mean* CFU count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* CFU count and time.

Table 1: TB dataset: PEs and 95% HPD intervals of bactericidal activity and regression model parameters

		$J_{(\text{loading dose})}$PaZ				$J_{(200\text{ mg})}$PaZ				HRZE			
		MZINB[a]		MZIDW[b]		MZINB[a]		MZIDW[b]		MZINB[a]		MZIDW[b]	
Parameter[c]	**Day**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**	**PE**	**95% HPD**
BA_j (0–56)		0.130	[0.113; 0.148]	0.131	[0.115; 0.148]	0.110	[0.097; 0.122]	0.112	[0.101; 0.125]	0.110	[0.098; 0.120]	0.114	[0.103; 0.125]
α_j		14.59	[14.00; 15.22]	13.97	[13.31; 14.61]	13.97	[13.24; 14.69]	13.26	[12.53; 13.97]	14.79	[14.36; 15.27]	14.22	[13.72; 14.68]
β_{1j}		3.97	[3.22; 5.19]	4.05	[3.18; 5.33]	4.24	[3.33; 5.65]	4.15	[3.14; 5.46]	4.96	[3.62; 6.36]	4.94	[3.64; 6.24]
β_{2j}		1.71	[1.26; 2.08]	1.73	[1.37; 2.12]	1.26	[0.93; 1.54]	1.32	[1.05; 1.62]	1.50	[1.29; 1.71]	1.58	[1.38; 1.80]
κ_j		1.28	[0.61; 1.57]	1.25	[0.54; 1.57]	1.22	[0.57; 1.57]	1.24	[0.59; 1.57]	0.59	[0.43; 0.90]	0.58	[0.43; 0.86]
γ_j		1.53	[0.33; 2.00]	1.52	[0.35; 2.00]	1.76	[1.12; 2.00]	1.73	[1.08; 2.00]	0.44	[0.05; 0.90]	0.42	[0.05; 0.88]
$\sigma^2_{\alpha_j}$		5.09	[2.69; 7.75]	5.16	[3.28; 8.08]	6.76	[3.81; 10.38]	6.94	[3.79; 10.44]	2.74	[1.74; 4.22]	2.72	[1.43; 4.02]
$\sigma^2_{\beta_{1j}}$		1.92	[0.00; 5.33]	1.86	[0.01; 6.17]	4.43	[1.41; 9.59]	4.66	[1.29; 10.09]	2.46	[0.00; 7.06]	2.06	[0.00; 6.74]
$\sigma^2_{\beta_{2j}}$		1.06	[0.49; 1.94]	0.90	[0.38; 1.62]	0.36	[0.13; 0.74]	0.33	[0.10; 0.68]	0.36	[0.15; 0.65]	0.33	[0.15; 0.56]
$\sigma_{\alpha_j\beta_{1j}}$		0.68	[-0.90; 2.87]	0.59	[-1.13; 2.82]	-0.17	[-2.70; 2.27]	-0.11	[-2.84; 2.80]	0.35	[-0.72; 2.10]	0.25	[-0.76; 1.93]
$\sigma_{\alpha_j\beta_{2j}}$		-0.12	[-1.25; 0.76]	0.01	[-0.95; 0.92]	0.21	[-0.73; 1.03]	0.30	[-0.59; 1.18]	0.17	[-0.16; 0.54]	0.16	[-0.19; 0.49]
$\sigma_{\beta_{1j}\beta_{2j}}$		-0.04	[-1.08; 0.60]	0.02	[-1.02; 0.66]	-0.10	[-0.91; 0.59]	-0.23	[-1.10; 0.41]	0.17	[-0.25; 0.69]	0.15	[-0.33; 0.59]
ϕ_j		0.67	[0.57; 0.77]	0.75	[0.68; 0.83]	0.63	[0.55; 0.73]	0.72	[0.65; 0.81]	0.72	[0.63; 0.83]	0.78	[0.71; 0.86]
π_{jk}	0	0.01	[0.00; 0.03]	0.01	[0.00; 0.03]	0.01	[0.00; 0.04]	0.01	[0.00; 0.04]	0.01	[0.00; 0.03]	0.01	[0.00; 0.03]
	3	0.02	[0.00; 0.07]	0.02	[0.00; 0.08]	0.04	[0.00; 0.10]	0.04	[0.00; 0.10]	0.01	[0.00; 0.05]	0.01	[0.00; 0.05]
	7	0.07	[0.02; 0.15]	0.07	[0.01; 0.14]	0.01	[0.00; 0.06]	0.02	[0.00; 0.06]	0.01	[0.00; 0.04]	0.01	[0.00; 0.04]
	14	0.05	[0.00; 0.12]	0.04	[0.00; 0.10]	0.04	[0.00; 0.11]	0.03	[0.00; 0.10]	0.04	[0.00; 0.12]	0.05	[0.00; 0.13]
	21	0.04	[0.00; 0.12]	0.05	[0.00; 0.13]	0.15	[0.04; 0.27]	0.11	[0.03; 0.20]	0.09	[0.02; 0.17]	0.09	[0.02; 0.17]
	28	0.20	[0.05; 0.33]	0.17	[0.06; 0.30]	0.04	[0.00; 0.14]	0.05	[0.00; 0.16]	0.19	[0.07; 0.32]	0.22	[0.11; 0.33]
	35	0.27	[0.10; 0.47]	0.19	[0.05; 0.31]	0.14	[0.00; 0.31]	0.21	[0.04; 0.35]	0.25	[0.12; 0.41]	0.27	[0.13; 0.36]
	42	0.20	[0.05; 0.41]	0.17	[0.01; 0.31]	0.31	[0.12; 0.50]	0.28	[0.14; 0.40]	0.30	[0.12; 0.44]	0.27	[0.15; 0.38]
	49	0.22	[0.00; 0.46]	0.16	[0.00; 0.35]	0.45	[0.24; 0.66]	0.30	[0.13; 0.42]	0.32	[0.11; 0.53]	0.18	[0.02; 0.33]
	56	0.54	[0.25; 0.79]	0.42	[0.27; 0.49]	0.41	[0.09; 0.66]	0.24	[0.03; 0.42]	0.62	[0.40; 0.81]	0.37	[0.24; 0.46]

CFU: Colony-forming unit HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. PE: Posterior estimate. BA_j (0–56): Daily rate of change in $\log_{10}$ (CFU) count over Days 0 and 56. [a]The log-link function is used to describe the biphasic relationship between the *mean* CFU count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* CFU count and time. [c]For the prevention of numerical overflow, the models were fitted with time expressed in weeks. The fixed effects represent the following: α_j are the intercepts, β_{1j} & β_{2j} the two slopes, κ_j the change-points, and γ_j are the smoothness parameters.

Table 2: Simulation study: performance of regression models MZINB and MZIDW (15 patients)

						MZINB[a]				MZIDW[b]			
$\boldsymbol{\beta}_{11}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\pi}_{11}$	$\boldsymbol{\pi}_{12}$	Parameter[c]	Value	Bias	RMSE	Coverage[d]	Length[e]	Bias	RMSE	Coverage[d]	Length[e]
3.95	1.35	0.01	0.01	α_1	12.65	-0.0058	0.3550	95.4	1.4460	0.0056	0.3739	96.4	1.6001
				β_{11}	3.95	-0.0380	0.6633	96.4	2.8386	0.0236	0.7232	96.4	3.1297
				β_{21}	1.35	0.0400	0.3372	95.8	1.4371	0.0352	0.3296	98.0	1.4426
				$\sigma^2_{\alpha_1}$	0.95	0.0233	0.5837	94.8	2.5877	0.0763	0.6912	95.8	2.9061
				$\sigma^2_{\beta_{11}}$	0.85	0.3924	1.2484	100.0	6.1559	0.4614	1.2646	100.0	6.8703
				$\sigma^2_{\beta_{21}}$	0.80	0.2617	0.6942	97.8	2.8189	0.2589	0.6859	97.0	2.8269
				$\sigma_{\alpha_1\beta_{11}}$	0.25	-0.1433	0.3104	100.0	2.5898	-0.1236	0.3339	100.0	2.9237
				$\sigma_{\alpha_1\beta_{21}}$	0.35	-0.0708	0.2808	96.4	1.6878	-0.0734	0.3358	98.4	1.8046
				$\sigma_{\beta_{11}\beta_{21}}$	-0.15	-0.0134	0.3893	100.0	2.7281	-0.0086	0.3180	100.0	2.9022
				ϕ_1	0.70	0.0623	0.1289	96.8	0.4856	0.0335	0.0781	95.2	0.2885
				π_{11}	0.01	0.0165	0.0226	97.2	0.0759	0.0153	0.0210	98.2	0.0737
				π_{12}	0.01	0.0586	0.0715	94.8	0.1794	0.0515	0.0631	96.6	0.1637
3.95	1.35	0.01	0.12	α_1	12.65	-0.0061	0.3537	94.4	1.4508	-0.0055	0.3915	94.4	1.5758
				β_{11}	3.95	-0.0344	0.6619	98.0	2.8666	0.0367	0.7401	96.8	3.1692
				β_{21}	1.35	0.0268	0.3389	96.0	1.4404	0.0107	0.3337	96.4	1.4284
				$\sigma^2_{\alpha_1}$	0.95	0.0034	0.5539	96.0	2.5725	-0.0450	0.6373	94.6	2.6675
				$\sigma^2_{\beta_{11}}$	0.85	0.3086	1.1357	100.0	6.0664	0.3490	1.2183	100.0	6.5221
				$\sigma^2_{\beta_{21}}$	0.80	0.2409	0.6597	96.2	2.8170	0.2226	0.6421	97.4	2.7520
				$\sigma_{\alpha_1\beta_{11}}$	0.25	-0.1377	0.2850	100.0	2.5745	-0.1373	0.3364	100.0	2.7116
				$\sigma_{\alpha_1\beta_{21}}$	0.35	-0.0996	0.2886	96.0	1.6963	-0.1091	0.3112	96.8	1.6983
				$\sigma_{\beta_{11}\beta_{21}}$	-0.15	-0.0065	0.3699	100.0	2.7124	0.0162	0.3049	100.0	2.7801
				ϕ_1	0.70	0.0388	0.1230	97.8	0.4831	0.0179	0.0742	95.8	0.2908
				π_{11}	0.01	0.0153	0.0207	98.6	0.0745	0.0141	0.0198	97.8	0.0722
				π_{12}	0.12	-0.0165	0.0578	96.4	0.2237	-0.0178	0.0554	94.6	0.2090

HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. RMSE: Root mean square error. [a]The log-link function is used to describe the biphasic relationship between the *mean* count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* count and time. [c]For the prevention of numerical overflow, the models were fitted with time expressed in weeks. [d]95% HPD interval coverage (%). [e]95% HPD interval average length.

Table 2: Simulation study: performance of regression models MZINB and MZIDW (15 patients)

						MZINB[a]				MZIDW[b]			
$\boldsymbol{\beta}_{11}$	$\boldsymbol{\beta}_{21}$	$\boldsymbol{\pi}_{11}$	$\boldsymbol{\pi}_{12}$	Parameter[c]	Value	Bias	RMSE	Coverage[d]	Length[e]	Bias	RMSE	Coverage[d]	Length[e]
4.20	1.80	0.01	0.01	α_1	12.65	-0.0382	0.3384	95.6	1.4621	0.0028	0.3944	93.0	1.6033
				β_{11}	4.20	-0.0650	0.7565	96.6	3.2659	-0.0432	0.8662	94.6	3.6391
				β_{21}	1.80	0.0643	0.4432	95.0	1.7831	0.1138	0.4482	96.8	1.8430
				$\sigma^2_{\alpha_1}$	0.95	0.0506	0.6502	91.4	2.6332	0.0159	0.6639	94.6	2.8102
				$\sigma^2_{\beta_{11}}$	0.85	0.4846	1.4047	100.0	7.0664	0.4716	1.2207	100.0	7.5112
				$\sigma^2_{\beta_{21}}$	0.80	0.2722	0.8109	97.2	3.4511	0.3059	0.8401	94.6	3.5455
				$\sigma_{\alpha_1\beta_{11}}$	0.25	-0.1096	0.3758	100.0	2.8390	-0.1351	0.2991	100.0	3.0178
				$\sigma_{\alpha_1\beta_{21}}$	0.35	-0.1098	0.3202	94.6	1.8613	-0.1174	0.3159	94.4	1.9255
				$\sigma_{\beta_{11}\beta_{21}}$	-0.15	-0.0048	0.3416	100.0	3.0758	-0.0074	0.3533	100.0	3.2649
				ϕ_1	0.70	0.0793	0.1537	96.6	0.5394	0.0339	0.0838	95.8	0.3158
				π_{11}	0.01	0.0169	0.0229	97.4	0.0772	0.0158	0.0214	98.2	0.0751
				π_{12}	0.01	0.0836	0.0991	95.6	0.2463	0.0773	0.0928	94.6	0.2215
4.20	1.80	0.01	0.12	α_1	12.65	-0.0210	0.3361	96.6	1.4909	0.0185	0.4015	94.8	1.6164
				β_{11}	4.20	-0.1223	0.7554	96.8	3.3003	0.0052	0.8440	96.8	3.6763
				β_{21}	1.80	0.0784	0.4145	97.2	1.7771	0.0947	0.4356	97.2	1.8432
				$\sigma^2_{\alpha_1}$	0.95	0.0935	0.6907	93.8	2.7498	0.0193	0.7093	96.0	2.8393
				$\sigma^2_{\beta_{11}}$	0.85	0.4137	1.1802	100.0	6.9378	0.4750	1.3044	100.0	7.5840
				$\sigma^2_{\beta_{21}}$	0.80	0.2402	0.7560	96.4	3.3558	0.3153	0.8332	96.0	3.5511
				$\sigma_{\alpha_1\beta_{11}}$	0.25	-0.0991	0.3602	100.0	2.9094	-0.1251	0.3628	100.0	3.0431
				$\sigma_{\alpha_1\beta_{21}}$	0.35	-0.0954	0.3005	96.0	1.8745	-0.1024	0.3189	97.6	1.9484
				$\sigma_{\beta_{11}\beta_{21}}$	-0.15	0.0314	0.2954	100.0	2.9958	-0.0038	0.3683	100.0	3.2969
				ϕ_1	0.70	0.0577	0.1465	96.8	0.5306	0.0260	0.0792	97.0	0.3178
				π_{11}	0.01	0.0158	0.0214	98.4	0.0761	0.0148	0.0202	98.4	0.0737
				π_{12}	0.12	-0.0023	0.0647	98.8	0.2770	-0.0091	0.0565	98.8	0.2475

HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. RMSE: Root mean square error. [a]The log-link function is used to describe the biphasic relationship between the *mean* count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* count and time. [c]For the prevention of numerical overflow, the models were fitted with time expressed in weeks. [d]95% HPD interval coverage (%). [e]95% HPD interval average length.

Table 3: Simulation study: percentage difference in RMSEs and HPD interval lengths between "skew-normal" contamination and no contamination (15 patients)

	RMSE						**Interval Length[a]**					
	MZINB[b]			**MZIDW[c]**			**MZINB[b]**			**MZIDW[c]**		
	ψ_{α_1} **[d]**			ψ_{α_1} **[d]**			ψ_{α_1} **[d]**			ψ_{α_1} **[d]**		
Parameter[e]	**-2**	**-3**	**-4**	**-2**	**-3**	**-4**	**-2**	**-3**	**-4**	**-2**	**-3**	**-4**
α_1	36.3	61.9	102.2	25.1	49.1	86.9	39.8	76.7	115.9	35.0	66.9	102.8
β_{11}	2.0	4.4	4.0	1.3	-4.8	1.0	5.0	7.7	9.8	5.1	7.0	9.4
β_{21}	3.9	8.2	8.1	1.1	5.0	3.5	3.9	6.2	10.1	3.9	7.2	11.6

HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. RMSE: Root mean square error. [a]95% HPD interval average length. [b]The log-link function is used to describe the biphasic relationship between the *mean* count and time. [c]The log-link function is used to describe the biphasic relationship between the *median* count and time. [d]Percentage difference between "skew-normal" contamination (with skewness parameter ψ_{α_1}) and no contamination. [e]For the prevention of numerical overflow, the models were fitted with time expressed in weeks.

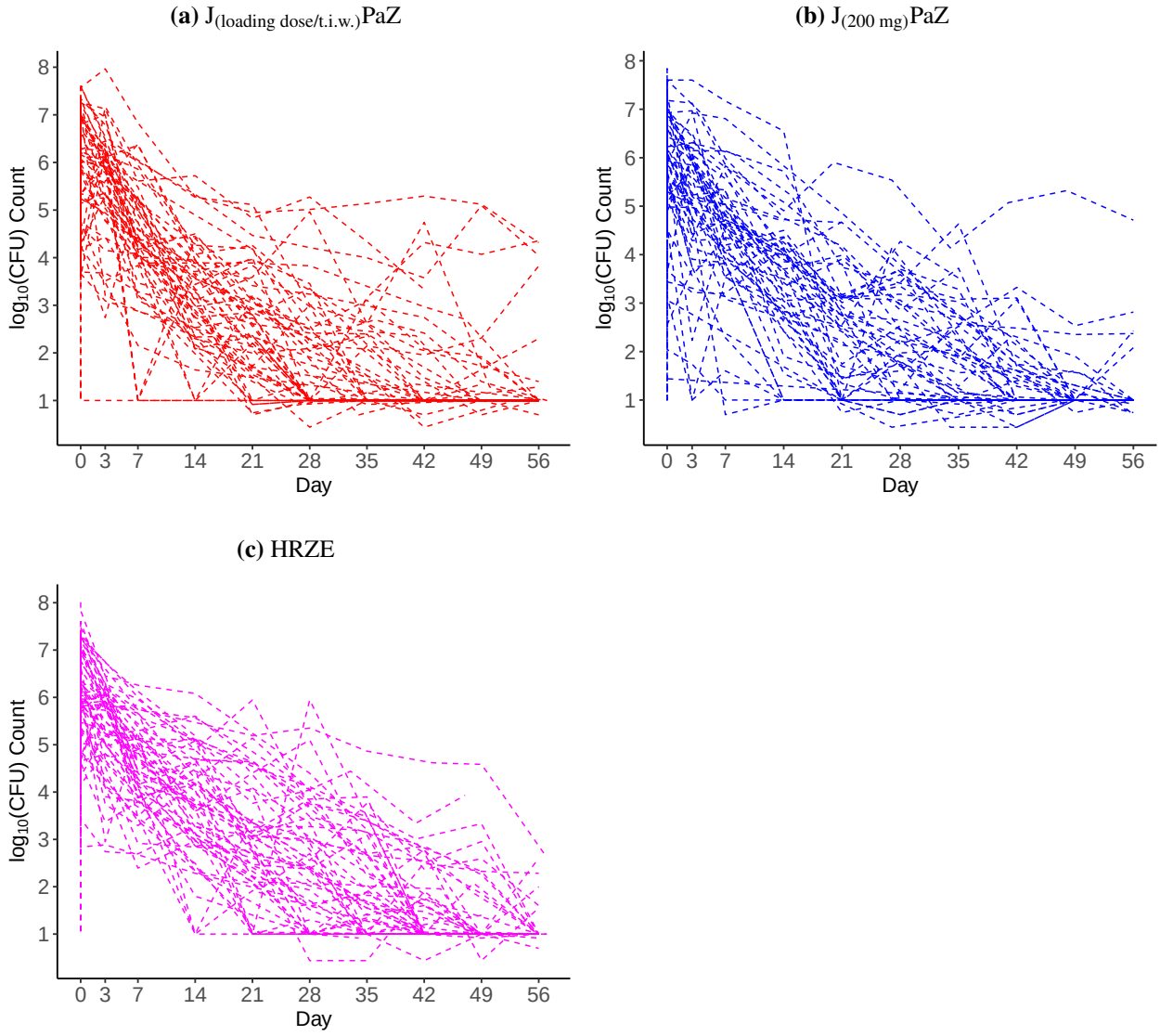
NONLINEAR MIXED-EFFECTS MODELING OF LONGITUDINAL COUNT DATA: BAYESIAN INFERENCE ABOUT MEDIAN COUNTS BASED ON THE MARGINAL ZERO-INFLATED DISCRETE WEIBULL DISTRIBUTION

Divan Aristo Burger and Emmanuel Lesaffre

SUPPLEMENTARY MATERIAL

PROFILE PLOT

Figure 1: TB dataset: observed $\log_{10}(\text{CFU})$ counts over time



CFU: Colony-forming unit. CFU counts of zero are displayed on the log-10 scale as 1.

SUMMARY STATISTICS

Table 1: Summary statistics of observed CFU count over time

Treatment	Day	n	Mean	SD	CV	Minimum	Median	Maximum	Zeros (%)
$J_{(\text{loading dose})}$PaZ	Day 0	162	5728792	9343449	163	0	1441000	40000000	1.9
	Day 3	55	3315936	12627118	381	0	679250	92400000	1.8
	Day 7	52	343078	1038131	303	0	22850	6765000	7.7
	Day 14	48	28087	87399	311	0	1788	525250	6.3
	Day 21	48	8768	24204	276	0	568	130000	6.3
	Day 28	48	8776	32969	376	0	73	188100	37.5
	Day 35	41	4257	21558	506	0	0	136950	53.7
	Day 42	49	5737	29068	507	0	0	196625	61.2
	Day 49	43	6256	27312	437	0	0	132000	72.1
	Day 56	48	1278	4682	366	0	0	22667	81.3
$J_{(200\ \text{mg})}$PaZ	Day 0	157	5127638	10585053	206	0	770000	68200000	1.9
	Day 3	52	2024968	5999341	296	0	176700	40000000	3.8
	Day 7	51	574902	2226334	387	0	42000	14750000	2.0
	Day 14	51	94971	499563	526	0	4015	3550000	9.8
	Day 21	47	22015	116787	530	0	270	800000	29.8
	Day 28	42	9833	53110	540	0	72	345000	23.8
	Day 35	41	1768	6979	395	0	15	42900	39.0
	Day 42	44	2821	17849	633	0	0	118500	54.5
	Day 49	46	4583	30960	676	0	0	210000	71.7
	Day 56	45	1174	7674	654	0	0	51500	77.8
HRZE	Day 0	167	6618662	12675684	192	0	1595000	102300000	0.6
	Day 3	55	778651	1184944	152	550	212500	5637500	0.0
	Day 7	55	193311	356931	185	248	45000	1806750	0.0
	Day 14	53	63521	179552	283	0	6600	1210000	5.7
	Day 21	51	34406	126824	369	0	1172	880000	9.8
	Day 28	52	26171	126473	483	0	312	885000	23.1
	Day 35	48	2816	11172	397	0	55	73150	33.3
	Day 42	50	997	5861	588	0	1	41525	50.0
	Day 49	46	1118	5781	517	0	0	38500	65.2
	Day 56	45	27	89	335	0	0	413	82.2

n = Number of CFU counts. CFU: Colony-forming unit. CV: Coefficient of variation. SD: Standard deviation.

REGRESSION CURVE

Figure 2: Example plot of regression curve for $\log_{10}(\text{CFU})$ count over time

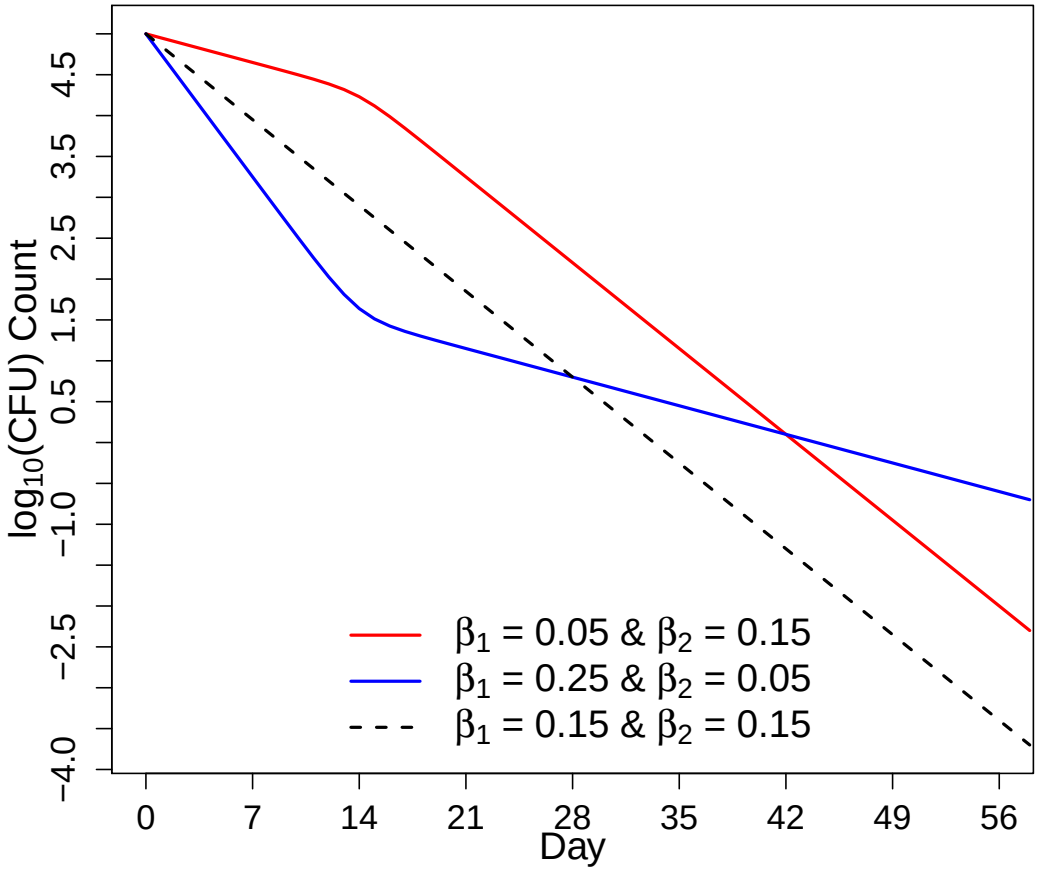
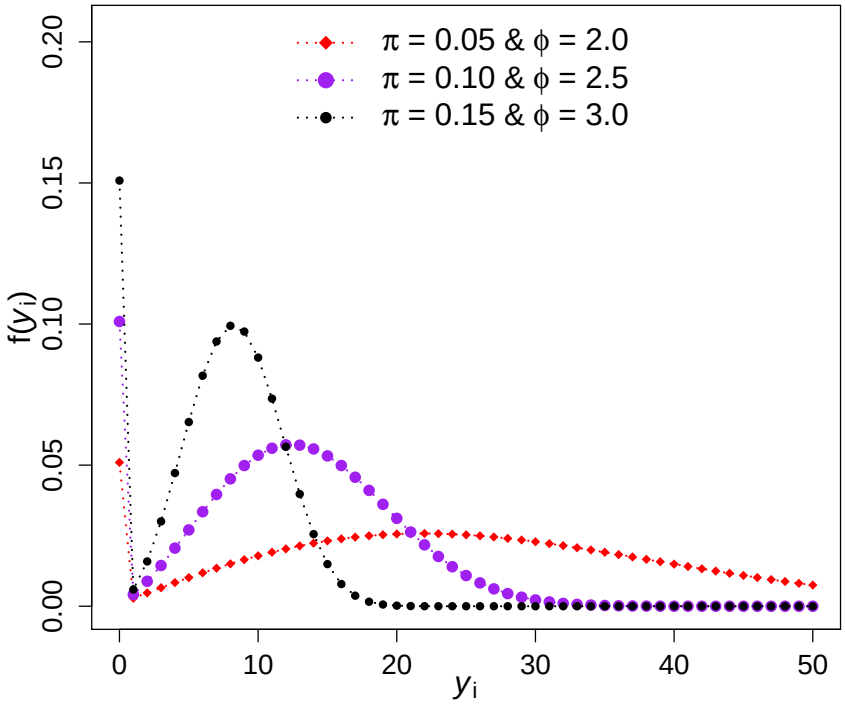
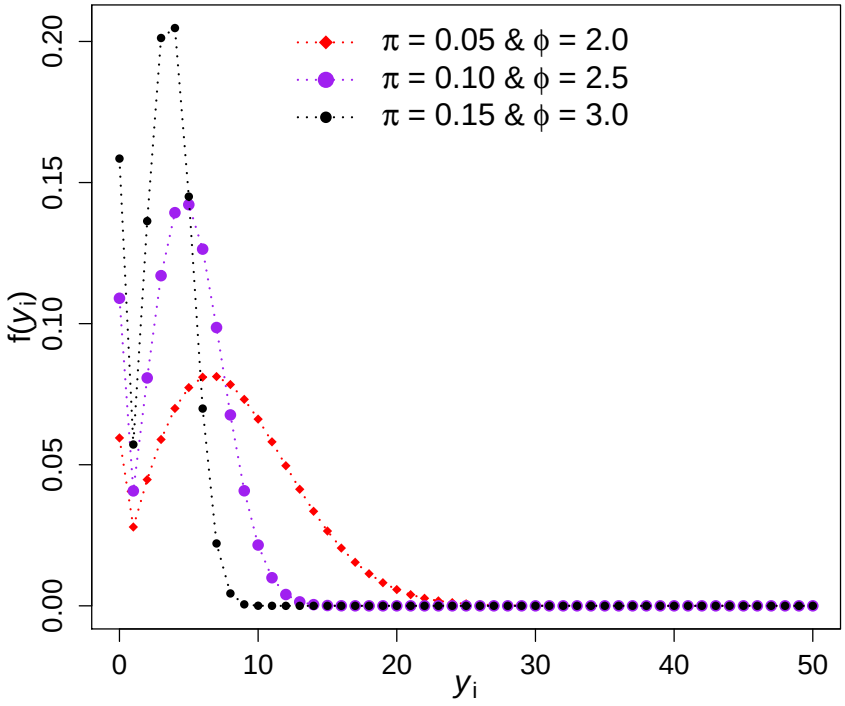


Figure 3: Probability mass function of zero-inflated discrete Weibull distribution

(a) $q = 0.999$



(b) $q = 0.99$



BAYESIAN MODEL SPECIFICATION

The Bayesian estimation procedures proposed in this manuscript are based on vague prior distributions as follows:

Multivariate normal prior distributions are specified for $\boldsymbol{\delta}_j$, namely:

$$\boldsymbol{\delta}_j \sim \text{ Normal}\left(\mathbf{0}, 10^4 \times \boldsymbol{I}\right)$$

where $\mathbf{0}$ and $\boldsymbol{I}$ respectively denote the corresponding vector of zeros and identity matrix.

The matrix-generalized half-t (MGH-t) prior distribution for the random effects covariance matrix is specified instead of the widely used Wishart prior distribution, since the latter often results in poor confidence interval coverage, especially when the variance components are small.[47] The MGH-t prior distribution for $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}$ is hierarchically specified by Wishart and gamma distributions as[28]:

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}|\boldsymbol{\Omega}_j &\sim \text{Wishart}\left(v+2, 2v\boldsymbol{\Omega}_j\right)\\ \omega_{jz} &\sim \text{Gamma}\left(0.5, 1/A^2\right)\end{aligned}$$

where $\boldsymbol{\Omega}_j$ denote diagonal matrices with ω_{jz} on the diagonal ($z = 1,2,3$). Here, $v+2$ and $2v\boldsymbol{\Omega}_j$ are respectively the degrees of freedom and inverse scale matrices of the Wishart distribution. The conditional density functions of $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}$ and ω_{jz} are written as:

$$\begin{aligned}P\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}|\boldsymbol{\Omega}_j\right) &\propto \left|\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}\right|^{\frac{v}{2}-1} \exp\left[-v\cdot\text{tr}\left(\boldsymbol{\Omega}_j\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}\right)\right];\ \boldsymbol{\Omega}_j = \text{diag}\left(\omega_{j1},\omega_{j2},\omega_{j3}\right)\\ P\left(\omega_{jz}\right) &\propto \omega_{jz}^{-\frac{1}{2}} \exp\left(-\frac{1}{A^2}\omega_{jz}\right)\end{aligned}$$

From the law of total probability, the set of nuisance parameters $\boldsymbol{\Omega}_j$ integrated out results in the MGH-t prior distribution, namely:

$$f\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}\right) \propto \left|\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}\right|^{-\frac{v+6}{2}} \prod_{z=1}^{3}\left[v\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}\right)_{zz} + 1/A^2\right]^{-\frac{v+3}{2}}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j} > \mathbf{0}$, and $\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}\right)_{zz}$ is the z^{th} diagonal entry of $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}$. This mixture representation results in the half-t prior distribution, namely half-t (v,A), for the standard deviation terms in $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}$, and the uniform prior distribution, namely $U(-1,1)$, for the correlation terms in $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}$. The corresponding quantities are set to $A = 50$ and $v = 2$. Therefore, this specification results in a weakly informative (heavy-tailed) prior for the standard deviation terms.

The parameters κ_j and γ_j are assumed to follow uniform prior distributions, namely:

$$\begin{aligned}\kappa_j &\sim \text{Uniform}\left(L_\kappa, U_\kappa\right)\\ \gamma_j &\sim \text{Uniform}\left(L_\gamma, U_\gamma\right)\end{aligned}$$

where L_κ, U_κ, L_γ, and U_γ are the pre-specified lower and upper bounds for κ_j and γ_j, respectively.

The dispersion & shape parameters ϕ_j and zero-inflation probabilities π_{jk} are assumed to follow vague gamma and uniform prior distributions, namely:

$$\begin{aligned}\phi_j &\sim \text{Gamma}\left(0.5, 0.5\right)\\ \pi_{jk} &\sim \text{Uniform}\left(0, U_\pi\right)\end{aligned} \tag{8}$$

where U_π is the upper bound for π_{jk}. Here, ϕ_j's prior variance is 2. Assuming the gamma prior distribution in Equation (8) is reasonable if the data are believed to be overdispersed (i.e., small values of ϕ_j).

The resulting joint posterior distribution of the model parameters is written as:

$$\begin{aligned}&P\left(\boldsymbol{\delta}_{ij}, \boldsymbol{\delta}_j, \boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}, \omega_{jz}, \kappa_j, \gamma_j, \phi_j, \pi_{jk}, j = 1,\ldots,J, i = 1,\ldots,N_j, k = 1,\ldots,T | \boldsymbol{y}\right)\\ &\propto \left(\prod_{j=1}^{J}\prod_{i=1}^{N_j}\prod_{k=1}^{T} f\left(y_{ijk}|\lambda_{ijk}, \phi_j, \pi_{jk}\right)\right)\left(\prod_{j=1}^{J}\prod_{i=1}^{N_j} P\left(\boldsymbol{\delta}_{ij}|\boldsymbol{\delta}_j, \boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}\right)\right)\left(\prod_{j=1}^{J} P\left(\boldsymbol{\delta}_j\right) P\left(\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}^{-1}|\boldsymbol{\Omega}_j\right)\right)\times\\ &\left(\prod_{j=1}^{J}\prod_{z=1}^{3} P\left(\omega_{jz}\right)\right)\left(\prod_{j=1}^{J} P\left(\kappa_j\right) P\left(\gamma_j\right) P\left(\phi_j\right)\right)\left(\prod_{j=1}^{J}\prod_{k=1}^{T} P\left(\pi_{jk}\right)\right)\end{aligned}$$

where $\boldsymbol{y}$ is the vector containing y_{ijk} for all $j = 1,\ldots,J$, $i = 1,\ldots,N_j$, and $k = 1,\ldots,T$. The MCMC Gibbs sampling algorithm is used to draw samples from the joint posterior distribution of the model parameters.[48] Software such as `JAGS`[29] can be employed to carry out the Gibbs sampling procedure.

PROGRAMMING CODE: MZINB MODEL

```
library(Rfast)
library(runjags)
library(parallel)
library(coda)
setwd('C:/Program Files/R/R-3.5.1/bin')
NCORES <- detectCores() - 1 ###7 Cores

DATA1 <- read.csv('C:/Progra~1/R/R-3.5.1/bin/FIVE_CFU_O_MZINB_DATA.csv')
attach(DATA1) #See below.
DATA2 <- read.csv('C:/Progra~1/R/R-3.5.1/bin/FIVE_CFU_O_MZINB_PATKVT.csv') #Dataset containing the patient IDs the treatment groups to which the patients were assigned.
attach(DATA2)

###Notes:###
#Initially, we wrote our program to accommodate multiple longitudinal outcomes. For the current manuscript, only one longitudinal outcome was used. Hence, JVINDEX = 1.
#NSUBJID: Patient ID.
#KVTRTN: Treatment group.
#TIME: Observed time.
#PROTTIMC: Protocol timepoint.
#SUMCFU: Sum of the CFU counts (all plates).
#OFFSET: Offset constant.
#PATKVT: Patient ID.

DATA <- list(JVINDEX = JVINDEX, NSUBJID = NSUBJID, KVTRTN = KVTRTN, TIME = TIME, PROTTIMC = PROTTIMC, SUMCFU = SUMCFU, OFFSET = OFFSET, PATKVT = PATKVT, ZEROVEC = rep(0, 3), IDENMAT = diag(rep(0.0001, 3)))

BAYESMODEL <- '
data {
    C <- 10000
    for (i in 1:1792) {
        ONES[i] <- 1
    }
}
model {
    for (i in 1:1792) {
        LIKE[i] <- (PI[JVINDEX[i], KVTRTN[i], PROTTIMC[i]]*equals(SUMCFU[i], 0) + (1 - PI[JVINDEX[i], KVTRTN[i], PROTTIMC[i]])*(exp(loggam(SUMCFU[i] + PHI[JVINDEX[i], KVTRTN[i]]) - loggam(SUMCFU[i] + 1) - loggam(PHI[JVINDEX[i], KVTRTN[i]]))*(PHI[JVINDEX[i], KVTRTN[i]]/(PHI[JVINDEX[i], KVTRTN[i]] + LAMBDA[i]/(1 - PI[JVINDEX[i], KVTRTN[i], PROTTIMC[i]])))^PHI[JVINDEX[i], KVTRTN[i]]*(LAMBDA[i]/(1 - PI[JVINDEX[i], KVTRTN[i], PROTTIMC[i]])/(PHI[JVINDEX[i], KVTRTN[i]] + LAMBDA[i]/(1 - PI[JVINDEX[i], KVTRTN[i], PROTTIMC[i]])))^SUMCFU[i]))/C
        ONES[i] ~ dbern(LIKE[i])
        LINK[i] <- SALPHA[JVINDEX[i], NSUBJID[i]] - (SBETA2[JVINDEX[i], NSUBJID[i]] + SBETA1[JVINDEX[i], NSUBJID[i]])/2*TIME[i]/7 - (SBETA2[JVINDEX[i], NSUBJID[i]] - SBETA1[JVINDEX[i], NSUBJID[i]])/2*MGAMMA[JVINDEX[i], KVTRTN[i]]*log((exp((TIME[i]/7 - MKAPPA[JVINDEX[i], KVTRTN[i]])/MGAMMA[JVINDEX[i], KVTRTN[i]]) + exp(-(TIME[i]/7 - MKAPPA[JVINDEX[i], KVTRTN[i]])/MGAMMA[JVINDEX[i], KVTRTN[i]]))/(exp(MKAPPA[JVINDEX[i], KVTRTN[i]]/MGAMMA[JVINDEX[i], KVTRTN[i]]) + exp(-MKAPPA[JVINDEX[i], KVTRTN[i]]/MGAMMA[JVINDEX[i], KVTRTN[i]])))
        log(LAMBDA[i]) <- LINK[i] + OFFSET[i]
    }
    for (i in 1:172) {
        SALPHA[1, i] <- SDELTA[i, 1]
        SBETA1[1, i] <- SDELTA[i, 2]
        SBETA2[1, i] <- SDELTA[i, 3]
        SDELTA[i, 1:3] ~ dmnorm(MDELTA[PATKVT[i], 1:3], SGINV[PATKVT[i], 1:3, 1:3])
    }
    for (i in 1:3) {
        MALPHA[1, i] <- MDELTA[i, 1]
        MBETA1[1, i] <- MDELTA[i, 2]
        MBETA2[1, i] <- MDELTA[i, 3]
        OMEGA[i, 1, 1] ~ dgamma(0.5, 0.0004)
        for (x in 2:3) {
            OMEGA[i, x, x] ~ dgamma(0.5, 0.0004)
```

```
            for (y in 1:(x - 1)) {
                OMEGA[i, x, y] <- 0
                OMEGA[i, y, x] <- OMEGA[i, x, y]
            }
        }
        SGINV[i, 1:3, 1:3] ~ dwish(2*2*OMEGA[i, 1:3, 1:3], 4)
        SIGMA[i, 1:3, 1:3] <- inverse(SGINV[i, 1:3, 1:3])
        ALPSIGSQ[i] <- SIGMA[i, 1, 1]
        BT1SIGSQ[i] <- SIGMA[i, 2, 2]
        BT2SIGSQ[i] <- SIGMA[i, 3, 3]
        ALPBT1SIGSQ[i] <- SIGMA[i, 1, 2]
        ALPBT2SIGSQ[i] <- SIGMA[i, 1, 3]
        BT1BT2SIGSQ[i] <- SIGMA[i, 2, 3]
        MDELTA[i, 1:3] ~ dmnorm(ZEROVEC[1:3], IDENMAT[1:3, 1:3])
        for (k in 1:1) {
            MKAPPA[k, i] ~ dunif(0.4285714286, 1.5714285714)
            MGAMMA[k, i] ~ dunif(0.05, 2)
            PHI[k, i] ~ dgamma(0.5, 0.5)
            for (j in 1:10) {
                PI[k, i, j] ~ dunif(0, 1)
            }
        }
    }
}'
OMEGA <- array(NA, dim = c(3, 3, 3))
for (k in 1:3) {
    OMEGA[k, 1, 1] <- 1
    for (i in 2:3) {
        OMEGA[k, i, i] <- 1
    }
}
ORIGPARMS <- c('SALPHA', 'SBETA1', 'SBETA2', 'MALPHA', 'MBETA1', 'MBETA2', 'MKAPPA',
'MGAMMA', 'PHI', 'PI', 'ALPSIGSQ', 'BT1SIGSQ', 'BT2SIGSQ', 'ALPBT1SIGSQ',
'ALPBT2SIGSQ', 'BT1BT2SIGSQ')
ORIGPARMS
NEWPARAMS <- c('LAMBDA')
NEWPARAMS
ALLPARAMS <- c(ORIGPARMS, NEWPARAMS)
ALLPARAMS
SDELTA =
as.matrix(read.csv('C:/Progra~1/R/R-3.5.1/bin/FIVE_CFU_O_MZINB_Initial_SDELTA.csv'))
#Dataset containing initial values.
MDELTA = t(matrix(c(colmeans(SDELTA), colmeans(SDELTA), colmeans(SDELTA)), ncol = 3))
INITIAL <- replicate(NCORES, list(list(PHI = array(rep(0.1, 3), dim = c(1, 3)), OMEGA =
OMEGA, SDELTA = SDELTA, .RNG.name = 'base::Mersenne-Twister', .RNG.seed = sample.int(n
= 100000, size = 1))))
TIME <- proc.time()
SAMPLE <- run.jags(model = BAYESMODEL, data = DATA, inits = INITIAL, monitor =
ALLPARAMS, n.chains = NCORES, burnin = 15000, thin = 450, sample = 150, summarise =
FALSE, method = 'parallel', modules = 'glm', factories = 'bugs::MNormal sampler off')
proc.time() - TIME
READ1 <- as.matrix(as.mcmc(SAMPLE, vars = c(ORIGPARMS)))
READ2 <- as.matrix(as.mcmc(SAMPLE, vars = c(NEWPARAMS)))
write.csv(READ1, 'FIVE_CFU_O_MZINB_CODA1.csv')
write.csv(t(READ2), 'FIVE_CFU_O_MZINB_CODA2.csv')
SUMMARY <- summary(SAMPLE, confidence = c(0.95))
write.csv(SUMMARY, 'FIVE_CFU_O_MZINB_LOG.csv')
pdf('FIVE_CFU_O_MZINB_DIAGN.pdf')
plot(SAMPLE, vars = c(ORIGPARMS), plot.type = c('trace'), new.window = FALSE)
plot(SAMPLE, vars = c(ORIGPARMS), plot.type = c('autocorr'), new.window = FALSE)
dev.off()
extract(SAMPLE, what = 'samplers')
```

PROGRAMMING CODE: MZIDW MODEL

```
library(Rfast)
library(runjags)
library(parallel)
library(coda)
setwd('C:/Program Files/R/R-3.5.1/bin')
NCORES <- detectCores() - 1 ###7 Cores

DATA1 <- read.csv('C:/Progra~1/R/R-3.5.1/bin/FIVE_CFU_O_MZIDW_DATA.csv')
attach(DATA1) #See below.
DATA2 <- read.csv('C:/Progra~1/R/R-3.5.1/bin/FIVE_CFU_O_MZIDW_PATKVT.csv') #Dataset containing the patient IDs the treatment groups to which the patients were assigned.
attach(DATA2)

###Notes:###
#Initially, we wrote our program to accommodate multiple longitudinal outcomes. For the current manuscript, only one longitudinal outcome was used. Hence, JVINDEX = 1.
#NSUBJID: Patient ID.
#KVTRTN: Treatment group.
#TIME: Observed time.
#PROTTIMC: Protocol timepoint.
#SUMCFU: Sum of the CFU counts (all plates).
#OFFSET: Offset constant.
#PATKVT: Patient ID.

DATA <- list(JVINDEX = JVINDEX, NSUBJID = NSUBJID, KVTRTN = KVTRTN, TIME = TIME, PROTTIMC = PROTTIMC, SUMCFU = SUMCFU, OFFSET = OFFSET, PATKVT = PATKVT, ZEROVEC = rep(0, 3), IDENMAT = diag(rep(0.0001, 3)))

BAYESMODEL <- '
data {
    C <- 10000
    for (i in 1:1792) {
        ONES[i] <- 1
    }
}
model {
    for (i in 1:1792) {
        LIKE[i] <- (PI[JVINDEX[i], KVTRTN[i], PROTTIMC[i]]*equals(SUMCFU[i], 0) + (1 - PI[JVINDEX[i], KVTRTN[i],
        PROTTIMC[i]])*(exp(log(0.5)*(SUMCFU[i]/MU[i])^PHI[JVINDEX[i], KVTRTN[i]]) -
        exp(log(0.5)*((SUMCFU[i] + 1)/MU[i])^PHI[JVINDEX[i], KVTRTN[i]])))/C
        ONES[i] ~ dbern(LIKE[i])
        MU[i] <- LAMBDA[i]*(log(0.5/(1 - PI[JVINDEX[i], KVTRTN[i],
        PROTTIMC[i]]))/log(0.5))^(-1/PHI[JVINDEX[i], KVTRTN[i]])
        LINK[i] <- SALPHA[JVINDEX[i], NSUBJID[i]] - (SBETA2[JVINDEX[i], NSUBJID[i]] +
        SBETA1[JVINDEX[i], NSUBJID[i]])/2*TIME[i]/7 - (SBETA2[JVINDEX[i], NSUBJID[i]] -
        SBETA1[JVINDEX[i], NSUBJID[i]])/2*MGAMMA[JVINDEX[i],
        KVTRTN[i]]*log((exp((TIME[i]/7 - MKAPPA[JVINDEX[i],
        KVTRTN[i]])/MGAMMA[JVINDEX[i], KVTRTN[i]]) + exp(-(TIME[i]/7 -
        MKAPPA[JVINDEX[i], KVTRTN[i]])/MGAMMA[JVINDEX[i],
        KVTRTN[i]]))/(exp(MKAPPA[JVINDEX[i], KVTRTN[i]]/MGAMMA[JVINDEX[i], KVTRTN[i]])
        + exp(-MKAPPA[JVINDEX[i], KVTRTN[i]]/MGAMMA[JVINDEX[i], KVTRTN[i]])))
        log(LAMBDA[i]) <- LINK[i] + OFFSET[i]
    }
    for (i in 1:172) {
        SALPHA[1, i] <- SDELTA[i, 1]
        SBETA1[1, i] <- SDELTA[i, 2]
        SBETA2[1, i] <- SDELTA[i, 3]
        SDELTA[i, 1:3] ~ dmnorm(MDELTA[PATKVT[i], 1:3], SGINV[PATKVT[i], 1:3, 1:3])
    }
    for (i in 1:3) {
        MALPHA[1, i] <- MDELTA[i, 1]
        MBETA1[1, i] <- MDELTA[i, 2]
        MBETA2[1, i] <- MDELTA[i, 3]
        OMEGA[i, 1, 1] ~ dgamma(0.5, 0.0004)
        for (x in 2:3) {
            OMEGA[i, x, x] ~ dgamma(0.5, 0.0004)
            for (y in 1:(x - 1)) {
                OMEGA[i, x, y] <- 0
```

```
                    OMEGA[i, y, x] <- OMEGA[i, x, y]
                }
            }
            SGINV[i, 1:3, 1:3] ~ dwish(2*2*OMEGA[i, 1:3, 1:3], 4)
            SIGMA[i, 1:3, 1:3] <- inverse(SGINV[i, 1:3, 1:3])
            ALPSIGSQ[i] <- SIGMA[i, 1, 1]
            BT1SIGSQ[i] <- SIGMA[i, 2, 2]
            BT2SIGSQ[i] <- SIGMA[i, 3, 3]
            ALPBT1SIGSQ[i] <- SIGMA[i, 1, 2]
            ALPBT2SIGSQ[i] <- SIGMA[i, 1, 3]
            BT1BT2SIGSQ[i] <- SIGMA[i, 2, 3]
            MDELTA[i, 1:3] ~ dmnorm(ZEROVEC[1:3], IDENMAT[1:3, 1:3])
            for (k in 1:1) {
                MKAPPA[k, i] ~ dunif(0.4285714286, 1.5714285714)
                MGAMMA[k, i] ~ dunif(0.05, 2)
                PHI[k, i] ~ dgamma(0.5, 0.5)
                for (j in 1:10) {
                    PI[k, i, j] ~ dunif(0, 0.5)
                }
            }
        }
    }'
    OMEGA <- array(NA, dim = c(3, 3, 3))
    for (k in 1:3) {
        OMEGA[k, 1, 1] <- 1
        for (i in 2:3) {
            OMEGA[k, i, i] <- 1
        }
    }
    ORIGPARMS <- c('SALPHA', 'SBETA1', 'SBETA2', 'MALPHA', 'MBETA1', 'MBETA2', 'MKAPPA',
    'MGAMMA', 'PHI', 'PI', 'ALPSIGSQ', 'BT1SIGSQ', 'BT2SIGSQ', 'ALPBT1SIGSQ',
    'ALPBT2SIGSQ', 'BT1BT2SIGSQ')
    ORIGPARMS
    NEWPARAMS <- c('LAMBDA')
    NEWPARAMS
    ALLPARAMS <- c(ORIGPARMS, NEWPARAMS)
    ALLPARAMS
    SDELTA =
    as.matrix(read.csv('C:/Progra~1/R/R-3.5.1/bin/FIVE_CFU_O_MZIDW_Initial_SDELTA.csv'))
    #Dataset containing initial values.
    MDELTA = t(matrix(c(colmeans(SDELTA), colmeans(SDELTA), colmeans(SDELTA)), ncol = 3))
    INITIAL <- replicate(NCORES, list(list(PHI = array(rep(0.1, 3), dim = c(1, 3)), OMEGA =
    OMEGA, SDELTA = SDELTA, .RNG.name = 'base::Mersenne-Twister', .RNG.seed = sample.int(n
    = 100000, size = 1))))
    TIME <- proc.time()
    SAMPLE <- run.jags(model = BAYESMODEL, data = DATA, inits = INITIAL, monitor =
    ALLPARAMS, n.chains = NCORES, burnin = 15000, thin = 450, sample = 150, summarise =
    FALSE, method = 'parallel', modules = 'glm', factories = 'bugs::MNormal sampler off')
    proc.time() - TIME
    READ1 <- as.matrix(as.mcmc(SAMPLE, vars = c(ORIGPARMS)))
    READ2 <- as.matrix(as.mcmc(SAMPLE, vars = c(NEWPARAMS)))
    write.csv(READ1, 'FIVE_CFU_O_MZIDW_CODA1.csv')
    write.csv(t(READ2), 'FIVE_CFU_O_MZIDW_CODA2.csv')
    SUMMARY <- summary(SAMPLE, confidence = c(0.95))
    write.csv(SUMMARY, 'FIVE_CFU_O_MZIDW_LOG.csv')
    pdf('FIVE_CFU_O_MZIDW_DIAGN.pdf')
    plot(SAMPLE, vars = c(ORIGPARMS), plot.type = c('trace'), new.window = FALSE)
    plot(SAMPLE, vars = c(ORIGPARMS), plot.type = c('autocorr'), new.window = FALSE)
    dev.off()
    extract(SAMPLE, what = 'samplers')
```

COMPOUND LAPLACE-METROPOLIS MARGINAL LIKELIHOOD

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

$$\log(f[\boldsymbol{y}|M]) = \frac{1}{2}\log(2\pi)\,pJ + \frac{1}{2}\log\left|R_{(\boldsymbol{\delta}_j,\kappa_j,\gamma_j,\phi_j,\pi_{jk},j=1,\ldots,J,k=1,\ldots,T)}\right| + s_{(\boldsymbol{\delta}_j,\kappa_j,\gamma_j,\phi_j,\pi_{jk},j=1,\ldots,J,k=1,\ldots,T)} + \sum_{j=1}^{J}\sum_{i=1}^{N_j}\left(\log\left[\int\left(\prod_{k=1}^{T} f\left(y_{ijk}|\boldsymbol{\delta}_{ij},\hat{\kappa}_j,\hat{\gamma}_j,\hat{\phi}_j,\hat{\pi}_{jk}\right)\right)P\left(\boldsymbol{\delta}_{ij}|\hat{\boldsymbol{\delta}}_j,\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}_j}\right)\mathrm{d}\boldsymbol{\delta}_{ij}\right]\right) + \sum_{j=1}^{J}\log\left[P\left(\hat{\boldsymbol{\delta}}_j\right)P\left(\hat{\kappa}_j\right)P\left(\hat{\gamma}_j\right)P\left(\hat{\phi}_j\right)P\left(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}_j}\right)\right] + \sum_{j=1}^{J}\sum_{k=1}^{T}\log\left[P\left(\hat{\pi}_{jk}\right)\right]$$

where p is the number of parameters among $\boldsymbol{\delta}_j$, κ_j, γ_j, ϕ_j, $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}$, π_{jk} of treatment group j. Here, $\hat{\boldsymbol{\delta}}_j$, $\hat{\kappa}_j$, $\hat{\gamma}_j$, $\hat{\phi}_j$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\delta}_j}$, $\hat{\pi}_{jk}$ are respectively the mean of the posterior distribution of $\boldsymbol{\delta}_j$, κ_j, γ_j, ϕ_j, $\boldsymbol{\Sigma}_{\boldsymbol{\delta}_j}$, π_{jk}. $\left|R_{(\boldsymbol{\delta}_j,\kappa_j,\gamma_j,\phi_j,\pi_{jk},j=1,\ldots,J,k=1,\ldots,T)}\right|$ and $s_{(\boldsymbol{\delta}_j,\kappa_j,\gamma_j,\phi_j,\pi_{jk},j=1,\ldots,J,k=1,\ldots,T)}$ respectively denote the determinant of the correlation matrix and the sum of the logarithm of the standard deviations of the posterior distributions of $\boldsymbol{\delta}_j$, κ_j, γ_j, ϕ_j, π_{jk}. The model with the largest CLMML is favored.

Table 2: Simulation study: performance of regression models MZINB and MZIDW (50 patients)

		MZINB[a]				MZIDW[b]			
Parameter[c]	**Value**	**Bias**	**RMSE**	**Coverage[d]**	**Length[e]**	**Bias**	**RMSE**	**Coverage[d]**	**Length[e]**
α_1	12.65	0.0107	0.1984	93.0	0.7690	0.0230	0.2045	95.0	0.8413
β_{11}	4.20	0.0493	0.4144	97.0	1.6706	0.0411	0.4832	95.0	1.8598
β_{21}	1.80	0.0136	0.2294	96.0	0.8880	0.0280	0.2116	97.0	0.9124
$\sigma^2_{\alpha_1}$	0.95	-0.0089	0.2952	97.0	1.2923	-0.0263	0.3447	96.0	1.4548
$\sigma^2_{\beta_{11}}$	0.85	-0.1258	0.7416	100.0	2.9449	-0.1335	0.6402	100.0	3.1990
$\sigma^2_{\beta_{21}}$	0.80	0.0579	0.3126	97.0	1.2932	0.0517	0.2686	99.0	1.3067
$\sigma_{\alpha_1\beta_{11}}$	0.25	-0.1490	0.2660	100.0	1.4378	-0.1680	0.2522	100.0	1.5842
$\sigma_{\alpha_1\beta_{21}}$	0.35	-0.0252	0.1552	98.0	0.8703	-0.0161	0.2074	98.0	0.9323
$\sigma_{\beta_{11}\beta_{21}}$	-0.15	0.0400	0.2506	100.0	1.3533	0.0365	0.2020	100.0	1.4497
ϕ_1	0.70	0.0349	0.0800	94.0	0.2846	0.0140	0.0445	96.0	0.1716
π_{11}	0.01	0.0071	0.0120	96.0	0.0388	0.0057	0.0102	96.0	0.0370
π_{12}	0.01	0.0422	0.0498	97.0	0.1352	0.0374	0.0439	99.0	0.1248

HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. RMSE: Root mean square error. [a]The log-link function is used to describe the biphasic relationship between the *mean* count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* count and time. [c]For the prevention of numerical overflow, the models were fitted with time expressed in weeks. [d]95% HPD interval coverage (%). [e]95% HPD interval average length.

Table 3: Simulation study: robustness of regression models MZINB and MZIDW (15 patients)

			MZINB[a]				MZIDW[b]			
ψ_{α_1}	Parameter[c]	Value	Bias	RMSE	Coverage[d]	Length[e]	Bias	RMSE	Coverage[d]	Length[e]
-2	α_1	12.65	-0.0004	0.4823	95.6	2.0276	-0.0334	0.4896	95.8	2.1268
	β_{11}	3.95	-0.0306	0.6748	97.8	3.0098	0.0037	0.7497	96.2	3.3310
	β_{21}	1.35	0.0512	0.3521	96.2	1.4967	0.0184	0.3375	96.6	1.4838
-3	α_1	12.65	0.0033	0.5726	95.4	2.5633	0.0101	0.5838	95.0	2.6299
	β_{11}	3.95	-0.0587	0.6913	96.4	3.0861	0.0197	0.7044	97.4	3.3905
	β_{21}	1.35	0.0559	0.3668	96.8	1.5293	0.0339	0.3504	96.0	1.5318
-4	α_1	12.65	-0.0420	0.7152	95.4	3.1330	-0.0051	0.7316	95.4	3.1963
	β_{11}	3.95	-0.0316	0.6887	97.4	3.1463	0.0521	0.7473	97.4	3.4657
	β_{21}	1.35	0.0410	0.3665	97.0	1.5859	0.0437	0.3455	98.6	1.5946

HPD: Highest posterior density. MZIDW: Marginal zero-inflated discrete Weibull. MZINB: Marginal zero-inflated negative binomial. RMSE: Root mean square error. [a]The log-link function is used to describe the biphasic relationship between the *mean* count and time. [b]The log-link function is used to describe the biphasic relationship between the *median* count and time. [c]For the prevention of numerical overflow, the models were fitted with time expressed in weeks. [d]95% HPD interval coverage (%). [e]95% HPD interval average length.