




# Bayesian joint modeling of patient-reported longitudinal data on frequency and duration of migraine

Gül İnan 

*Department of Mathematics, Istanbul Technical University, Istanbul, 34469, Turkey*

## Abstract

In this methodological study, we address the joint modeling of longitudinal data on the frequency and duration migraine attacks collected from patients in a clinical study in which patients were repeatedly asked at each hospital visit to report the number of days of migraine attacks they had in the last 30 days and the corresponding average duration of attacks. In our motivating data set, the migraine frequency outcome is a count variable inflated at multiples of 5 and 10 days, whereas the migraine duration outcome is reported entirely in discrete hours, including 0 for non-migraine days and inflated at multiples of 12 hours. In our study, we propose a joint modeling approach that models each migraine outcome by a multiple inflated negative binomial model with random effects and assumes a bivariate normal distribution for the random effects. We estimate the model parameters under Bayesian inference. We examine the performance of the proposed joint model using a Monte Carlo simulation study and compare its performance with a separate modeling approach in which each longitudinal count outcome is modeled separately. Finally, we present the results of the analysis of migraine data.

**Mathematics Subject Classification (2020).** 62H99, 62P10, 62F15

**Keywords.** Count outcomes, migraine days, migraine duration, multiple inflation, self-reported outcomes

## 1. Introduction

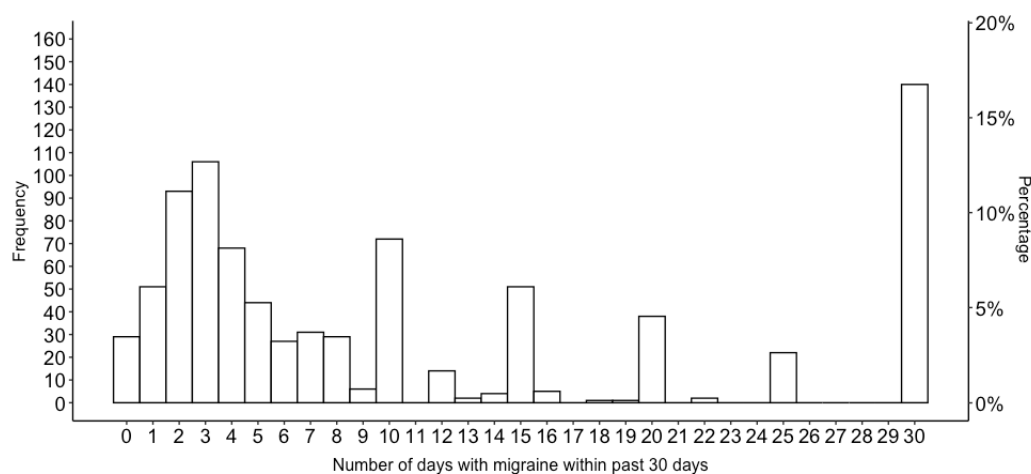
This study addresses a methodological approach to analyse longitudinal data on the frequency and duration of migraine attacks collected from migraine patients in a clinical study at the Department of Neurology, Faculty of Medicine, Mersin University, Turkey. In the clinical study, migraine specialists were interested in maintaining an electronic migraine database in the hospital to study the mechanisms of migraine in detail [25]. With this in mind, migraine patients, who visited the hospital at least once between the years 2004 and 2010, were asked to report the number of days with migraine attacks they had within the last 30 days and the average duration of these migraine attacks during each hospital visit. The result of this migraine study is an electronic database consisting of patient-reported migraine days and duration and other demographic and clinical information collected

from  $N = 179$  sufferers over several months/years. Thus, the migraine data retrieved from the database is in a form of longitudinal multivariate data with two count outcomes: i) migraine frequency (days): an outcome representing the number of days with migraine within the last 30 days and ii) migraine duration: an outcome augmented with zeros representing the corresponding average duration of migraine in discrete hours. In this case, the duration outcome is completely reported in discrete hours including 0 (i.e., if the frequency outcome is 0, the duration takes the value 0).

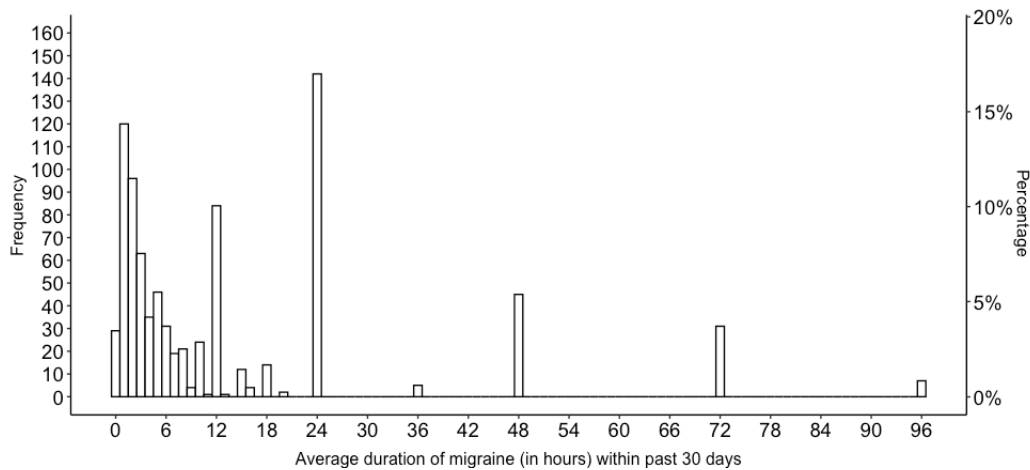
In the migraine study, migraine specialists were mainly interested in the co-evolution of migraine frequency and duration over time, as these two outcomes are biologically associated [25]. This medical research question led us to consider joint modeling of these two longitudinal outcomes, as it is known that joint models provide better insights in the analysis of longitudinal multivariate data with increased efficiency due to information exchange between outcomes and allow estimation of the association between outcomes [3, 10–13]. In this sense, following the novel papers of [8] and [16], a joint model can be constructed as follows: First, separate generalized linear mixed models (GLMMs) can be used to model each longitudinal outcome under an appropriate distribution from the exponential family, and then a bivariate GLMM can be constructed by imposing a bivariate normal distribution on random effects to jointly analyze the longitudinal migraine data with two count outcomes.

For exploratory analysis, we plotted the distribution of migraine days and migration duration reported by patients in the migraine study and presented them in Figure 1 and Figure 2, respectively. However, a closer look at Figure 1 and Figure 2 shows that migraine days are inflated for numbers that are multiples of 5 and 10, whereas migraine duration is inflated for numbers that are multiples of 12. More specifically, we found that the frequency (percentage) of days with migraine that are inflated at 10, 15, 20, 25, and 30 are 72 (9%), 51 (6%), 38 (5%), 22 (3%), and 140 (17%), respectively. Similarly, we found that the frequency of the average migraine duration that are inflated at 12, 24, 48, 72, and 96 are 84 (10%), 142 (17%), 45 (5%), 31 (4%), and 7 (1%), respectively.

The Figures 1 and 2 apparently show that migraine patients tend to report the frequency and duration of attacks by rounding up or down to a nearby number since they were asked to give precise information retrospectively, but, they could not remember exactly how many days with migraine they had in the last 30 days and how long the migraine lasted on average.



**Figure 1.** The distribution of number of days with migraine reported by patients in the migraine study.



**Figure 2.** The distribution of average duration of migraine (in hours) reported by patients in the migraine study.

In the statistical literature, rounding a numerical value to a nearby number is referred to by various terms, such as heaping, coarsening, or misreporting, although we would prefer the word “heaping” in this paper. Heaping is a very common phenomenon in applied studies where self-reported data are collected retrospectively. For example, heaping may occur in count data when reporting the number of cigarettes consumed [22, 23], the number of sexual partners [5], the number of depressed days [15], the number of work disability [2], and the number of unprotected sexual relationships [14]. On the other hand, heaping in positive continuous data may occur when age [9], unemployment duration [17, 21], income [6, 26], birth weight [4], etc. are reported. Heitjan and Rubin [9], Wang and Heitjan [22], Wang et al. [23], Allen et al. [1] have shown that in self-reported retrospective studies, there is a loss of precision in the data due to recall errors in the true responses, which in turn affects the true distribution of the data. These authors have shown that statistical conclusions based on heaped data can be misleading if the heaping in the data has not been properly accounted for in the data modeling and analysis. For this reason, they suggest that heaping in data should be seriously considered in data analysis in order to draw reliable statistical conclusions.

Heitjan and Rubin [9] was the first to discuss heaping in detail and to propose multiple imputation inference to deal with heaping at reporting age. Then, Wang and Heitjan [22] used a proportional odds regression modeling approach to model different rounding behaviors in the self-reported number of cigarettes consumed per day and Wang et al. [23] extended the same approach to the longitudinal analysis of self-reported cigarette consumption. Li et al. [14] proposed a multiple inflated Poisson regression model to analyze cross-sectional count data with multiple inflated values without assuming that all inflated values are rounded values, which is a more flexible assumption than [9] and [22] and also results in a less complicated and more interpretable model.

In this study, we would like to propose a bivariate GLMM for analysis of longitudinal data on the frequency and duration of migraine outcomes with inflated values. Following [14], our proposed approach models each migraine outcome by a multiple inflated negative binomial model with random effects and then models both outcomes jointly by imposing a bivariate normal distribution on the random effects. To our knowledge, this study is the first work to account for inflation in both outcomes of a bivariate longitudinal data model, motivated by a real-world problem.

The remainder of the paper is organized as follows. In Section 2, we first give a general overview of a multiple inflated negative binomial model with random effects. We then

extend this model to propose a joint model for analyzing longitudinal data on migraine frequency and duration outcomes with inflated values, and then explain the Bayesian estimation of the proposed model. Section 3 presents the application of the proposed joint model to the motivating migraine data. Section 4 provides the results of a simulation study conducted to investigate the performance of the proposed model and compare it with alternatives. Finally, Section 5 provides some concluding remarks.

## 2. Statistical methods

### 2.1. Multiple inflated negative binomial model with random effects

Consider a longitudinal study with  $N$  patients ( $i = 1, \dots, N$ ) repeatedly observed at  $n_i$  number of visits ( $j = 1, \dots, n_i$ ). Let  $Y_{ij} \in \mathbb{N}$  denote the  $j$ th response of  $i$ th patient and  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$  be the longitudinal count response vector for  $i$ th patient. Assume further that a total of  $K$  positive integer values such as  $\{r_0, r_1, \dots, r_{K-1}\}$  occur more frequently in the complete data than those expected under standard count data distributions. Following [14], we assume that the  $j$ th response of  $i$ th patient,  $Y_{ij}$  takes an integer value of  $r_k$  ( $k = 0, \dots, K - 1$ ) with probability  $\pi_k$  and any other integer value  $\in \mathbb{N} - \{r_0, \dots, r_{(K-1)}\}$  with probability  $\pi_K$ . Then we assume a mixture of  $K$  degenerate distributions with a negative binomial distribution to model  $Y_{ij}$  as follows:

$$Y_{ij} \sim \begin{cases} r_0, & \text{with probability } \pi_0 \\ r_1, & \text{with probability } \pi_1 \\ \dots & \dots \\ r_k, & \text{with probability } \pi_k \\ \dots & \dots \\ r_{K-1}, & \text{with probability } \pi_{K-1} \\ \text{NegBin}(\mu_{ij}, \phi), & \text{with probability } \pi_K, \end{cases} \quad (2.1)$$

where  $\pi_k$  is the probability that  $Y_{ij}$  arises from a degenerate distribution at  $r_k$  for  $k = 0, \dots, K - 1$ , and  $\pi_K$  is the probability that  $Y_{ij}$  arises from a negative binomial distribution, with the restriction that  $\sum_{k=0}^K \pi_k = 1$ . The parameters  $\mu_{ij}$  ( $\mu_{ij} > 0$ ) and  $\phi$  ( $\phi > 0$ ) in Equation (2.1) characterize the mean and the over-dispersion of the negative binomial distribution, respectively. Hence, the formulation in the Equation (2.1) accommodates more values from the elements of the set  $\{r_0, r_1, \dots, r_{K-1}\}$  than expected for the negative binomial distribution.

Thus, the probability mass function of the multiple inflated negative binomial (MINB) model can be written as follows:

$$Pr(Y_{ij} = r_k) = \pi_k + \pi_K \frac{\Gamma(r_k + \frac{1}{\phi})}{r_k! \Gamma(\frac{1}{\phi})} (1 + \phi \mu_{ij})^{-\frac{1}{\phi}} \left(1 + \frac{1}{\phi \mu_{ij}}\right)^{-r_k},$$

for  $k = 0, \dots, (K - 1)$  and

$$Pr(Y_{ij} = r) = \pi_K \frac{\Gamma(r + \frac{1}{\phi})}{r! \Gamma(\frac{1}{\phi})} (1 + \phi \mu_{ij})^{-\frac{1}{\phi}} \left(1 + \frac{1}{\phi \mu_{ij}}\right)^{-r},$$

for  $r \in \mathbb{N} - \{r_0, \dots, r_{(K-1)}\}$ .

In longitudinal regression modeling, the mean of the  $j$ th response of the  $i$ th patient under the negative binomial distribution is associated with a number of covariates and subject-specific random effects as given below:

$$g(\mu_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\beta} + b_i, \quad (2.2)$$

where  $g(\cdot)$  is a link function that maps the interval  $(0, \infty)$  to  $(-\infty, \infty)$  such as  $\log(\cdot)$  link function,  $\mathbf{X}_{ij}$  is a  $p \times 1$  vector of fixed effects covariates,  $\boldsymbol{\beta}$  is the corresponding  $p \times 1$  vector of fixed effects regression coefficients. The term  $b_i$  is the random intercept at the subject-level representing heterogeneity between subjects and is assumed to follow a normal distribution with a mean of 0 and a variance of  $\sigma_b^2$ . For simplicity, only models with random intercepts are considered in this paper.

Furthermore, we assume conditional independence, i.e., given the subject-specific random intercepts, repeated measurements within a subject are independent of each other, and measurements from different subjects are also independent of each other.

Let  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \pi_0, \pi_1, \dots, \pi_{(K-1)}, \pi_K, \phi, \sigma_b^2)$  be the vector of unknown parameters of the model. Then the contribution of the  $i$ th subject to the marginal likelihood of the observed data involves integration over the distribution of random intercepts as follows:

$$L(\boldsymbol{\theta}|\mathbf{Y}_i) = \int \prod_{j=1}^{n_i} \left[ \prod_{k=0}^{(K-1)} \left( \pi_k + \pi_K \frac{\Gamma(r_k + \frac{1}{\phi})}{r_k! \Gamma(\frac{1}{\phi})} (1 + \phi \mu_{ij})^{-\frac{1}{\phi}} \left( 1 + \frac{1}{\phi \mu_{ij}} \right)^{-r_k} \right)^{\delta_{ijk}} \right] \times \left[ \left( \pi_K \frac{\Gamma(y_{ij} + \frac{1}{\phi})}{y_{ij}! \Gamma(\frac{1}{\phi})} (1 + \phi \mu_{ij})^{-\frac{1}{\phi}} \left( 1 + \frac{1}{\phi \mu_{ij}} \right)^{-y_{ij}} \right)^{\delta_{ijK}} \right] f(b_i) db_i,$$

where  $\delta_{ijk} = \mathbb{1}_{\{y_{ij}=r_k\}}$  for  $k = 0, \dots, (K - 1)$  and  $\delta_{ijK} = \mathbb{1}_{\{y_{ij} \in \mathbb{N} - \{r_0, \dots, r_{(K-1)}\}\}}$  are indicator functions,  $f(b_i)$  denotes the normal distribution density function of the random intercept  $b_i$ , and the regression parameter vector  $\boldsymbol{\beta}$  enters the marginal likelihood of the data through the Equation (2.2).

### 2.2. Proposed joint model

In this section, we propose a joint MINB regression model which extends MINB regression model in Section 2.1 to longitudinal data with bivariate count outcomes.

Following the notation introduced in Section 2.1, let  $\mathbf{Y}_i = (\mathbf{Y}_{i1}^T, \mathbf{Y}_{i2}^T)^T$  be the bivariate response vector for the  $i$ th patient ( $i = 1, \dots, N$ ). In particular, here  $\mathbf{Y}_{i1} = (Y_{i11}, \dots, Y_{i1n_i})^T$  is the  $n_i \times 1$  response vector of the  $i$ th patient for the number of days with migraine over  $n_i$  visits and  $\mathbf{Y}_{i2} = (Y_{i21}, \dots, Y_{i2n_i})^T$  is the corresponding  $n_i \times 1$  response vector for the average duration of the migraine. Suppose further that a total of  $K_1$  integer values  $\{r_{10}, r_{11}, \dots, r_{1K_1-1}\}$  and a total of  $K_2$  integer values  $\{r_{20}, r_{21}, \dots, r_{2K_2-1}\}$  are more frequent in migraine frequency and duration, respectively, than expected according to the negative binomial distribution.

Then, a joint MINB model for the analysis of longitudinal data on the frequency and duration of migraine outcomes with inflation at  $\{r_{10}, r_{11}, \dots, r_{1K_1-1}\}$  and  $\{r_{20}, r_{21}, \dots, r_{2K_2-1}\}$ , respectively, can be specified as follows:

$$Y_{i1j} \sim \begin{cases} r_{1k_1}, & \text{with probability } \pi_{1k_1} \text{ for } k_1 = 0, \dots, (K_1 - 1) \\ NegBin(\mu_{i1j}, \phi_1), & \text{with probability } \pi_{1K_1}, \end{cases}$$

$$Y_{i2j} \sim \begin{cases} r_{2k_2}, & \text{with probability } \pi_{2k_2} \text{ for } k_2 = 0, \dots, (K_2 - 1) \\ NegBin(\mu_{i2j}, \phi_2), & \text{with probability } \pi_{2K_2}, \end{cases}$$

and

$$\begin{aligned} \log(\mu_{i1j}) &= \mathbf{X}_{i1j}^T \boldsymbol{\beta}_1 + b_{i1}, \\ \log(\mu_{i2j}) &= \mathbf{X}_{i2j}^T \boldsymbol{\beta}_2 + b_{i2}, \\ \mathbf{b}_i &= (b_{i1}, b_{i2}) \sim MVN_2(\mathbf{0}, \boldsymbol{\Sigma}_b), \end{aligned} \tag{2.3}$$

where  $i = 1, \dots, N$  and  $j = 1, \dots, n_i$  stand for patients and visits, respectively. The parameter  $\mu_{i1j}$  is the expected number of days with migraine within the last 30 days for the  $i$ th patient at the  $j$ th visit conditional on response and subject-level random intercept term  $b_{i1}$ . It is associated with  $p_1 \times 1$  vector of covariates  $\mathbf{X}_{i1j}^T$  by the  $\log(\cdot)$  link function. Similarly,  $\mu_{i2j}$  is the expected average duration of migraine within the last 30 days for the  $i$ th patient at the  $j$ th visit conditional on the response and subject-level random intercept  $b_{i2}$ . It is associated with  $p_2 \times 1$  vector of covariates  $\mathbf{X}_{i2j}^T$  by the  $\log(\cdot)$  link function. The parameters  $\beta_1$  and  $\beta_2$  are the corresponding response-specific vectors of the regression coefficients. The parameters  $\phi_1$  and  $\phi_2$  denote response-specific over-dispersion parameters. We further assume that the vector of random intercepts for the  $i$ th patient,  $\mathbf{b}_i = (b_{i1}, b_{i2})$ , has a bivariate normal distribution with mean vector  $\mathbf{0}$  and variance-covariance matrix:

$$\Sigma_b = \begin{bmatrix} \sigma_{b_1}^2 & \rho_b \sigma_{b_1} \sigma_{b_2} \\ \rho_b \sigma_{b_1} \sigma_{b_2} & \sigma_{b_2}^2 \end{bmatrix}, \quad (2.4)$$

where  $\Sigma_b$  represents the association between two negative binomial outcomes of the patient. We extend the conditional independence assumptions given in Section 2.1 and assume that the observations in  $\mathbf{Y}_{i1j}$  and  $\mathbf{Y}_{i2j}$  are independent given  $\mathbf{b}_i$ .

Let  $\theta = (\beta_1, \beta_2, \pi_{10}, \pi_{11}, \dots, \pi_{1(K_1-1)}, \pi_{1K_1}, \pi_{20}, \pi_{21}, \dots, \pi_{2(K_2-1)}, \pi_{2K_2}, \phi_1, \phi_2, \Sigma_b)$  denote the vector of unknown parameters in the proposed joint model. Then, the contribution of the  $i$ th subject to the marginal likelihood of the observed data can be expressed as follows:

$$\begin{aligned} L(\theta | \mathbf{Y}_i) &= \int \prod_{j=1}^{n_i} \left[ \prod_{k_1=0}^{(K_1-1)} \left( \pi_{1k_1} + \pi_{1K_1} \frac{\Gamma(r_{1k_1} + \frac{1}{\phi_1})}{r_{1k_1}! \Gamma(\frac{1}{\phi_1})} (1 + \phi_1 \mu_{i1j})^{-\frac{1}{\phi_1}} \left(1 + \frac{1}{\phi_1 \mu_{i1j}}\right)^{-r_{1k_1}} \right)^{\delta_{ij k_1}} \right] \\ &\quad \times \left[ \left( \pi_{1K_1} \frac{\Gamma(y_{i1j} + \frac{1}{\phi_1})}{y_{i1j}! \Gamma(\frac{1}{\phi_1})} (1 + \phi_1 \mu_{i1j})^{-\frac{1}{\phi_1}} \left(1 + \frac{1}{\phi_1 \mu_{i1j}}\right)^{-y_{i1j}} \right)^{\delta_{ij K_1}} \right] \\ &\quad \times \left[ \prod_{k_2=0}^{(K_2-1)} \left( \pi_{2k_2} + \pi_{2K_2} \frac{\Gamma(r_{2k_2} + \frac{1}{\phi_2})}{r_{2k_2}! \Gamma(\frac{1}{\phi_2})} (1 + \phi_2 \mu_{i2j})^{-\frac{1}{\phi_2}} \left(1 + \frac{1}{\phi_2 \mu_{i2j}}\right)^{-r_{2k_2}} \right)^{\delta_{ij k_2}} \right] \\ &\quad \times \left[ \left( \pi_{2K_2} \frac{\Gamma(y_{i2j} + \frac{1}{\phi_2})}{y_{i2j}! \Gamma(\frac{1}{\phi_2})} (1 + \phi_2 \mu_{i2j})^{-\frac{1}{\phi_2}} \left(1 + \frac{1}{\phi_2 \mu_{i2j}}\right)^{-y_{i2j}} \right)^{\delta_{ij K_2}} \right] f(\mathbf{b}_i) d\mathbf{b}_i, \end{aligned}$$

where  $\delta_{ij k_1} = \mathbb{1}_{\{y_{i1j}=r_{1k_1}\}}$  for  $k_1 = 0, \dots, (K_1 - 1)$ ,  $\delta_{ij K_1} = \mathbb{1}_{\{y_{i1j} \in \mathbb{N} - \{r_{10}, \dots, r_{1(K_1-1)}\}\}}$ ,  $\delta_{ij k_2} = \mathbb{1}_{\{y_{i2j}=r_{2k_2}\}}$  for  $k_2 = 0, \dots, (K_2 - 1)$ , and  $\delta_{ij K_2} = \mathbb{1}_{\{y_{i2j} \in \mathbb{N} - \{r_{20}, \dots, r_{2(K_2-1)}\}\}}$  are indicator functions, respectively. The vector of parameters  $\beta_1$  and  $\beta_2$  enter the model through Equation (2.3). The function  $f(\mathbf{b}_i)$  denotes the bivariate normal distribution density for  $\mathbf{b}_i$  given by Equation (2.4).

### 2.3. Bayesian inference

We used Bayesian inference to deal with the complexity of the proposed model due to the large number of random effects involved in the model. The Markov Chain Monte Carlo (MCMC) algorithm (specifically Gibbs sampling) is used to sample from the posterior distribution of parameters via JAGS (version 4.3.0) and the R package rjags [18, 19]. As prior distribution, we assigned an independent normal distribution with mean 0 and large variance 1000 for each element in the vector of regression coefficients  $\beta_1$  and  $\beta_2$ , Dirichlet distribution for degenerate distribution probabilities  $(\pi_{10}, \pi_{11}, \dots, \pi_{1(K_1-1)})$  and  $(\pi_{20}, \pi_{21}, \dots, \pi_{2(K_2-1)})$ , uniform distribution in interval (0.001, 5) for the over-dispersion parameters  $\phi_1$  and  $\phi_2$ , and Wishart distribution with  $\mathbf{I}$  scale matrix and 3 degrees of freedom for the inverse covariance matrix of the random effects  $\Sigma_b^{-1}$ .

### 3. Analysis of migraine data

As mentioned in Section 1, our motivating data set comes from a migraine study conducted at the Neurology Department of Mersin University, one of the leading migraine research centres in Turkey. The data set consists of longitudinal information on the number of days with migraine within the last 30 days and the average duration of migraine (in discrete hours) reported by  $N = 179$  patients who visited the hospital, after the first visit at least once between 2004 and 2010. As shown in Table 1, of 179 patients, 151 (84%) are women and 28 (16%) are men. The mean age at baseline is 38.27 years (standard deviation (SD) 13.69), with the youngest patient 10 years old and the oldest 84 years old. The total follow-up time per patient ranges from 1 to 57 months with a mean of 13.20 months (SD 11.16) and the total number of visits per patient ranges from 2 to 15 with a mean of 4.68 (SD 2.12). In addition, the mean of frequency outcome for  $N = 179$  patients is 10.91 days (SD 10.30) and that of duration outcome is 14.18 hours (SD 18.29).

**Table 1.** Summary statistics for  $N = 179$  patients in the migraine study data.

Variable	Range	Mean (SD)	Frequency
Gender			
Females	-	-	151 (84%)
Males	-	-	28 (16%)
Age at baseline (in years)	10 – 84	38.27 (13.69)	-
Total follow-up time per patient (in months)	1 – 56	13.20 (11.16)	-
Total number of visits per patient	2 – 15	4.68 (2.12)	-
Migraine frequency (in days)	0 – 30	10.91 (10.30)	-
Migraine duration (in hours)	0 – 96	14.18 (18.29)	-

We are interested in modeling patient-reported longitudinal data on migraine frequency and duration together by accounting for inflated values in both outcomes. Following the Figures 1 and 2, we assumed that a total of  $K_1 = 5$  integer values  $\{10, 15, 20, 25, 30\}$  in migraine frequency and a total of  $K_2 = 6$  integer values  $\{12, 24, 36, 48, 72, 96\}$  in migraine duration are observed more frequently than would be expected according to the negative binomial distribution. In addition, we assumed the following explanatory variables: the time of hospital visit (in months) since the first visit, gender (male = 0, female = 1), and the interaction of time and gender. Thus, under the proposed joint model, the expected number of days with migraine within the last 30 days and the expected average duration of migraine for the  $i$ th patient at the  $j$ th visit are associated with the above covariates by the  $\log(\cdot)$  link function, respectively, as follows:

$$\begin{aligned} \log(\mu_{i1j}) &= \beta_{10} + \beta_{11}time_{ij} + \beta_{12}gender_i + \beta_{13}(time_{ij} * gender_i) + b_{i1}, \\ \log(\mu_{i2j}) &= \beta_{20} + \beta_{21}time_{ij} + \beta_{22}gender_i + \beta_{23}(time_{ij} * gender_i) + b_{i2}, \end{aligned} \tag{3.1}$$

where  $i = 1, \dots, 179$ ,  $j = 1, \dots, n_i$ , the  $time_{ij}$  covariate is standardized over all patients and time points, and  $\mathbf{b}_i = (b_{i1}, b_{i2})$  is the vector of subject-level random intercepts coming from  $MVN_2(\mathbf{0}, \Sigma_b)$  (see Equation (2.4)).

For model fitting under Bayesian framework, the first 15,000 iterations are considered as the burn-in period. After the burn-in period, 50,000 iterations are drawn from 2 different chains with a thinning value of 50. The convergence of MCMC chains is checked using trace and autocorrelation plots and the Geweke Z-score (Z-score  $< |1.96|$  for all

parameters) obtained from the R package `coda` [20]. The initial values for  $\beta_1$ ,  $\beta_2$ ,  $\phi_1$ ,  $\phi_2$ ,  $\sigma_{b_1}^2$ , and  $\sigma_{b_2}^2$  are obtained from the SAS NLMIXED procedure by fitting separate models to the data, and the initial values for the degenerate probabilities ( $\pi_{10}, \pi_{11}, \dots, \pi_{15}$ ) and ( $\pi_{20}, \pi_{21}, \dots, \pi_{26}$ ) are obtained from the migraine data. The posterior mean estimates, standard deviations, and 95% credible intervals (CrIs) obtained from the analysis of the migraine data by the proposed joint model are given in Table 2.

**Table 2.** Posterior mean estimates, standard deviations, and 95% credible intervals (CrIs) obtained through the analysis of the migraine data with the proposed joint model.

Parameter	Est.	SD	2.5%	97.5%
Migraine frequency outcome				
$\beta_{10}$ (Int.)	2.0922	0.1394	1.8235	2.3680
$\beta_{11}$ (time)	-0.3549	0.0839	-0.5261	-0.1936
$\beta_{12}$ (gender)	0.1900	0.1490	-0.0986	0.4719
$\beta_{13}$ (time*gender)	0.2191	0.0923	0.0406	0.4011
$\phi_1$	0.5155	0.0344	0.4489	0.5837
$\sigma_{b_1}^2$	0.3751	0.0616	0.2737	0.5091
Migraine duration outcome				
$\beta_{20}$ (Int.)	1.9477	0.1981	1.5482	2.3347
$\beta_{21}$ (time)	-0.2405	0.0968	-0.4338	-0.0586
$\beta_{22}$ (gender)	0.4022	0.2121	0.0067	0.8192
$\beta_{23}$ (time*gender)	0.0898	0.1050	-0.1209	0.2919
$\phi_2$	0.7289	0.0446	0.6472	0.8181
$\sigma_{b_2}^2$	0.8368	0.1202	0.6142	1.1045
$\rho_b$	-0.0257	0.1000	-0.2131	0.1805

The results in Table 2 first show that in the migraine frequency model, the interaction between time and gender is statistically significant at the 5% level (95% CrI : [0.0406, 0.4011]). When  $b_{i1}$  is fixed, for females, the mean number of days with migraine decreases by 0.87 fold (e.g.,  $\exp(\hat{\beta}_{11} + \exp(\hat{\beta}_{13})) = \exp(-0.3549 + 0.2191) = 0.87$ ) when the time since the first visit increases by one unit. When  $b_{i1}$  is fixed, the mean number of days with migraine in males gradually decreases by 0.70 fold (e.g.,  $\exp(\hat{\beta}_{11}) = \exp(-0.3549) = 0.70$ ) when the time since the first visit increases by one unit. These results suggest that patients, regardless of gender, are more likely to report more migraine attacks when the time to the next visit increases after the first visit. For the migraine duration model, the interaction between time and gender is not statistically significant at the 5% level (95% CrI : [-0.1209, 0.2919]). For a given  $b_{i2}$  and time, the mean migraine duration with gender factor is not statistically significant at the 5% level (95% CrI : [-0.0067, 0.8192]). Similarly, for a given  $b_{i2}$  and gender, expected migraine duration gradually decreases by a factor of 0.79 fold (e.g.,  $\exp(\hat{\beta}_{21}) = \exp(-0.2405) = 0.79$ ) as time since first visit increases by one unit. Parameters  $\phi_1$  and  $\phi_2$  are estimated to be 0.5155 and 0.7289, respectively, indicating over-dispersion in the outcomes. The parameters ( $\pi_{10}, \pi_{11}, \pi_{12}, \pi_{13}, \pi_{14}$ ) and ( $\pi_{20}, \pi_{21}, \pi_{22}, \pi_{23}, \pi_{24}, \pi_{25}$ ) are estimated as (0.0867, 0.0611, 0.0460, 0.0267, 0.1685) and (0.1008, 0.1710, 0.0060, 0.0540, 0.0371, 0.0085).

In our analysis, the correlation between the random-intercept of the migraine frequency outcome and the migraine duration outcome  $\rho_b$  is estimated to be -0.0257. These results suggest that the average duration of migraine decreases as the number of patient-reported



migraine frequencies increases. Because the responses were self-reported, patients may have had only a rough recollection of the number of migraine attacks in the past 30 days and little recollection of the exact average duration. Therefore, this behaviour of the patients may have resulted in an inverse relationship between migraine frequency and migraine duration. However, as shown in Table 2, the correlation estimate is at the 5% level (95% CrI : [-0.2131, 0.1805]) turns out not to be statistically significant, indicating that migraine frequency and duration outcomes are not correlated. In this sense, we also fitted the separate models mentioned in Section 2.1 to each migraine outcome. Posterior mean estimates, standard deviations, and 95% credible intervals are presented in Table 3. It should be noted here that in the separate modeling, the gamma distribution (i.e., Gamma(0.001, 0.001)) is assumed to be the prior distribution for the inverse of the  $\sigma_{b_1}^2$  and  $\sigma_{b_2}^2$  parameters. The results in Table 2 are consistent with those in Table 3, with slightly lower standard deviations for the parameter estimates of the joint modeling approach. To compare the proposed joint model with the separate models, we also used the widely applicable information criterion (WAIC) [7, 24] in the Appendix. The smaller the WAIC value, the better the model. The WAIC value of the proposed joint model is 11945.89 and the sum of the WAIC values of the separate models is 12173.48, indicating that the joint modeling is better than the separate modeling for this data.

**Table 3.** Posterior mean estimates, standard deviations, and 95% credible intervals (CrIs) obtained through the analysis of the migraine data with separate models.

Parameter	Est.	SD	2.5%	97.5%
Migraine frequency outcome				
$\beta_{10}$ (Int.)	2.0903	0.1431	1.7902	2.3776
$\beta_{11}$ (time)	-0.3598	0.0863	-0.5266	-0.2017
$\beta_{12}$ (gender)	0.1927	0.1554	-0.1044	0.4977
$\beta_{13}$ (time*gender)	0.2214	0.0928	0.0452	0.4042
$\phi_1$	0.5138	0.0350	0.4471	0.5817
$\sigma_{b_1}^2$	0.3784	0.0586	0.2749	0.5062
Migraine duration outcome				
$\beta_{20}$ (Int.)	1.9416	0.1930	1.5675	2.3183
$\beta_{21}$ (time)	-0.2438	0.0999	-0.4502	-0.0421
$\beta_{22}$ (gender)	0.4068	0.2108	0.0057	0.8292
$\beta_{23}$ (time*gender)	0.0958	0.1058	-0.1040	0.3035
$\phi_2$	0.7283	0.0459	0.6417	0.8234
$\sigma_{b_2}^2$	0.8251	0.1201	0.6148	1.0876

#### 4. Simulation study

In this section, we conducted a Monte Carlo simulation study to evaluate the performance of the proposed joint model and compare its performance with separate models under two different correlation structures such as weak and strong.

In the simulation study, we closely followed the structure of motivating migraine data in Table 1. We assumed that the simulation study consisted of  $N = 200$  patients. The number of repeated measurements  $n_i$  per patient was independently generated from a discrete uniform distribution between 3 and 6. The variable  $time_{ij}$  was randomly drawn from the possible values  $\{1, 2, \dots, 56\}$  at each time point of each patient in ascending order.

The variable  $gender_i$  was generated independently for each patient from  $Ber(0.84)$ . The vector of random intercepts for the  $i$ th patient,  $\mathbf{b}_i = (b_{i1}, b_{i2})$ , was generated from bivariate normal distribution with mean vector  $\mathbf{0}$  and variance-covariance matrix:

$$\Sigma_b = \begin{bmatrix} \sigma_{b_1}^2 = 0.3792 & \rho_b \sigma_{b_1} \sigma_{b_2} \\ \rho_b \sigma_{b_1} \sigma_{b_2} & \sigma_{b_2}^2 = 0.8324 \end{bmatrix},$$

with  $\rho_b = 0.1$  (weak correlation) and 0.80 (strong correlation). Longitudinal bivariate data with count outcomes  $Y_{i1j}$  and  $Y_{i2j}$  were generated according to the proposed joint model, using the parameter estimates obtained from the analysis of migraine data in Table 2 as the true value for the data generation mechanism. Then, inflation of the  $Y_{i1j}$  and  $Y_{i2j}$  outcomes was performed in a similar manner to the analysis of the migraine data. After generating 100 longitudinal bivariate count data sets with inflated values, we fitted the proposed joint model and separate models to the generated data sets with similar MCMC settings as in the migraine data analysis.

While the simulation results for the regression parameters  $\beta_1$  and  $\beta_2$  of the joint modeling approach and those of the separate modeling approach when correlation was weak were presented in Table 4, simulation results when correlation was strong were presented in Table 5. To measure performance, for each individual regression coefficient, the mean of the posterior estimates of the regression parameters over 100 Monte Carlo runs (Mean), the absolute bias (abs(Bias): absolute value of the difference between mean and true value), the sample standard deviation of the posterior mean of the parameters over 100 Monte Carlo runs (SD), and the mean of the posterior standard deviation of the parameters over 100 Monte Carlo runs (MSD) were calculated. Finally, the efficiency gain was calculated as the mean of the posterior variance of the parameter estimates from the separate modeling over 100 Monte Carlo runs divided by the mean of the posterior variance of the parameter estimates from the joint modeling over 100 Monte Carlo runs. Thus, the basic objective of our simulation study is to examine the efficiency gain of the regression parameters from the proposed joint model over the separate model when the inherent correlation between the outcomes is strong.

**Table 4.** Summary of the performance measures of regression parameters under weak correlation scenario.

Parameter	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_{13}$	$\beta_{20}$	$\beta_{21}$	$\beta_{22}$	$\beta_{23}$
<b>True value</b>	2.10	-0.35	0.19	0.21	1.90	-0.25	0.40	0.10
Joint model								
Mean	2.1152	-0.3554	0.1773	0.2186	1.9072	-0.2466	0.4109	0.0997
abs(Bias)	0.0152	0.0054	0.0127	0.0086	0.0072	0.0034	0.0109	0.0003
SD	0.1238	0.0746	0.1280	0.0773	0.1814	0.1033	0.2061	0.1072
MSD	0.1310	0.0761	0.1424	0.0827	0.1862	0.0948	0.2028	0.1027
Separate model								
Mean	2.1148	-0.3551	0.1773	0.2184	1.9054	-0.2463	0.4126	0.0993
abs(Bias)	0.0148	0.0051	0.0127	0.0084	0.0054	0.0037	0.0126	0.0007
SD	0.1221	0.0741	0.1265	0.0766	0.1833	0.1032	0.2081	0.1070
MSD	0.1316	0.0763	0.1432	0.0826	0.1865	0.0950	0.2029	0.1030
Efficiency	1.0116	1.0000	1.0098	1.0000	1.0057	1.0120	1.0024	1.0094

The results in Table 4 show that with weak correlation, there is no significant difference between the joint and separate modeling approaches in the estimates of the regression parameters in terms of bias. The bias is much smaller for the parameter estimates of the

duration outcome because the range of the duration outcome is larger than the range of the migraine frequency outcome. The SD and the MSD of the regression parameter estimates are slightly smaller for the joint modeling approach than for the separate modeling approach. On the other hand, the results in Table 5 show that when the correlation is strong, the bias of the parameter estimates in the joint modeling approach decreases significantly, leading to larger biases in the separate modeling approach. Moreover, an increase in correlation leads to higher efficiency of the parameter estimates of the joint modeling approach compared to the parameter estimates of the separate modeling approach.

**Table 5.** Summary of the performance measures of regression parameters under strong correlation scenario.

Parameter	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_{13}$	$\beta_{20}$	$\beta_{21}$	$\beta_{22}$	$\beta_{23}$
<b>True value</b>	2.10	-0.35	0.19	0.21	1.90	-0.25	0.40	0.10
Joint model								
Mean	2.1150	-0.3470	0.1859	0.2091	1.9039	-0.2329	0.4074	0.0809
abs(Bias)	0.0150	0.0030	0.0041	0.0009	0.0039	0.0171	0.0074	0.0191
SD	0.1336	0.0711	0.1421	0.0776	0.1976	0.0976	0.1943	0.1063
MSD	0.1335	0.0721	0.1425	0.0832	0.1877	0.0943	0.2040	0.1021
Separate model								
Mean	2.1127	-0.3462	0.1877	0.2079	1.9030	-0.2341	0.4088	0.0821
abs(Bias)	0.0127	0.0038	0.0023	0.0021	0.0030	0.0159	0.0088	0.0179
SD	0.1348	0.0755	0.1469	0.0784	0.1983	0.0997	0.1944	0.1080
MSD	0.1347	0.0774	0.1461	0.0838	0.1866	0.0954	0.2024	0.1031
Efficiency	1.1222	1.1167	1.0142	1.1143	1.0888	1.1232	1.0857	1.1190

### 5. Conclusion

According to the Migraine Research Foundation (MRF), migraine is the 3rd most common and the 6th most disabling disease in the world, affecting 39 million people in the United States and 1 billion worldwide. The MRF also believes that migraine is a public health problem, such that the annual health care and lost productivity costs associated with migraine in the U.S. are estimated to be approximately 36 billion. These descriptive statistics highlight the importance of developing new statistical models for migraine studies to provide reliable statistical inference for better medical and health decisions.

In this paper, we proposed a joint modeling strategy to analyze longitudinal patient-reported data on frequency and duration of migraine with inflated values under a Bayesian estimation framework. Our simulation study showed that when the correlation between outcomes are strong, the joint modeling of the outcomes results in efficiency gain over the separate modeling of each outcome. Broadly speaking, note that the elements of the set  $\{r_0, r_1, \dots, r_{K-1}\}$  can be treated as ordinal variables and their cumulative probabilities  $Pr(Y_{ij} \leq r_k) = \sum_0^k Pr(Y_{ij} = r_k) = \sum_0^k \pi_{ij_k}$  can be associated with a set of covariates, which can be a follow-up study.

**Acknowledgment.** This study was supported by Istanbul Technical University with a grant ID 41881. The authors thank Prof. Dr. Aynur Ozge from Department of Neurology, Faculty of Medicine, Mersin University, Turkey and Dr. Osman Ozgur Yalin from Department of Neurology, Istanbul Education and Research Hospital, Turkey for permission to use the migraine data.

## References

- [1] C.M. Allen, S.D. Griffith, S. Shiffman and D.F. Heitjan, *Proximity and gravity: Modelling heaped self-reports*, Stat. Med. **36** (20), 3200–3215, 2017.
- [2] L. Bermúdez, D. Karlis and M. Santolino, *A finite mixture of multiple discrete distributions for modelling heaped count data*, Comput. Statist. Data Anal. **112**, 14–23, 2017.
- [3] E. Buta, S.S. O’Malley and R. Gueorguieva, *Bayesian joint modelling of longitudinal data on abstinence, frequency and intensity of drinking in alcoholism trials*, J. Roy. Statist. Soc. Ser. A **81** (3), 869–888, 2018.
- [4] C.G. Camarda, P.H. Eilers and J. Gampe, *Modelling trends in digit preference patterns*, J. R. Stat. Soc. Ser. C. Appl. Stat. **66** (5), 893–918, 2017.
- [5] F.W. Crawford, R.E. Weiss and M.A. Suchard, *Sex, lies and self-reported counts: Bayesian mixture models for heaping in longitudinal count data via birth-death processes*, Ann. Appl. Stat. **9** (2), 572–596, 2015.
- [6] J. Drechsler and H. Kiesel, *Beat the heap: An imputation strategy for valid inferences from rounded income data*, J. Surv. Stat. Methodol. **4** (1), 22–42, 2015.
- [7] A. Gelman, J. Hwang and A. Vehtari, *Understanding predictive information criteria for Bayesian models*, Stat. Comput. **24** (6), 997–1016, 2014.
- [8] R. Gueorguieva, *A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family*, Stat. Model. **1** (3), 177–193, 2001.
- [9] D.F. Heitjan and D.B. Rubin, *Inference from coarse data via multiple imputation with application to age heaping*, J. Amer. Statist. Assoc. **85** (410), 304–314, 1990.
- [10] E. Juárez-Colunga, G.L. Silva and C.B. Dean, *Joint modeling of zero-inflated panel count and severity outcomes*, Biometrics **73** (4), 1413–1423, 2017.
- [11] W. Kassahun, T. Neyens, G. Molenberghs, C. Faes and G. Verbeke, *A joint model for hierarchical continuous and zero-inflated overdispersed count data*, J. Stat. Comput. Simul. **85** (3), 552–571, 2015.
- [12] H. Li, J. Staudenmayer, T. Wang, S.K. Keadle and R.J. Carroll, *Three-part joint modeling methods for complex functional data mixed with zero-and-one-inflated proportions and zero-inflated continuous outcomes with skewness*, Stat. Med. **37** (4), 611–626, 2018.
- [13] Q. Li, J. Pan and J. Belcher, *Bayesian inference for joint modelling of longitudinal continuous, binary and ordinal events*, Stat. Methods Med. Res. **25** (6), 2521–2540, 2016.
- [14] Q. Li, G.K. Tso, Y. Qin, T.I. Lovejoy, T.G. Heckman and Y. Li, *Penalized multiple inflated values selection method with application to SAFER data*, Stat. Methods Med. Res. **28** (10-11), 3205–3225, 2019.
- [15] B.E. Magnus and D. Thissen, *Item response modeling of multivariate count data with zero inflation, maximum inflation, and heaping*, J. Educ. Behav. Stat. **42** (5), 531–558, 2017.
- [16] C. McCulloch, *Joint modelling of mixed outcome types using latent variables*, Stat. Methods Med. Res. **17** (1), 53–73, 2008.
- [17] F.E. Messlaki, *Making use of multiple imputation to analyze heaped data*, Master’s thesis, Utrecht University, 2010.
- [18] M. Plummer, *JAGS: Just another Gibbs sampler*, <http://mcmc-jags.sourceforge.net/>, 2017.
- [19] M. Plummer, A. Stukalov and M. Denwood, Package “rjags: Bayesian graphical models using MCMC”, R package version: 4-13, 2022.
- [20] M. Plummer, N. Best, K. Cowles and K. Vines, Package “CODA: Convergence diagnosis and output analysis for MCMC”, R package version: 0.19-4, 2022.

- [21] J. Van der Laan and L. Kuijvenhoven, *Imputation of rounded data*, Technical report, Statistics Netherlands, 2011.
- [22] H. Wang and D.F. Heitjan, *Modeling heaping in self-reported cigarette counts*, *Stat. Med.* **27** (19), 3789–3804, 2008.
- [23] H. Wang, S. Shiffman, S.D. Griffith and D.F. Heitjan, *Truth and memory: Linking instantaneous and retrospective self-reported cigarette consumption*, *Ann. Appl. Stat.* **6** (4), 1689–1706, 2012.
- [24] S. Watanabe, *Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory*, *J. Mach. Learn. Res.* **11**, 3571–3594, 2010.
- [25] O.O. Yalin, A. Ozge, M. Turkegun, B. Tasdelen and D. Uluduz, *Course of migraine with aura: A follow-up study*, *J. Neurol. Sci-Turk.* **33** (2), 254–263, 2016.
- [26] S. Zinn and A. Würbach, *A statistical approach to address the problem of heaping in self-reported income data*, *J. Appl. Stat.* **43** (4), 682–703, 2016.

## Appendix

The widely available (or Watanabe-Akaike) information criterion (WAIC) is calculated as follows:

$$WAIC = -2(lppd - pWAIC),$$

where “lppd” is the log-pointwise predictive density and is computed as:

$$lppd = \sum_{i=1}^N \sum_{j=1}^{n_i} \log \left[ \frac{1}{S} \sum_{s=1}^S f(y_{ij}; \boldsymbol{\theta}^s) \right],$$

with  $\boldsymbol{\theta}^s$  denoting the sth ( $s = 1, \dots, 1000$ ) sample value from the posterior predictive distribution. The effective number of parameters is computed as:

$$pWAIC = \sum_{i=1}^N \sum_{j=1}^{n_i} V_{s=1}^S (\log(f(y_{ij}; \boldsymbol{\theta}^s))).$$