A hierarchical Bayesian approach for the analysis of longitudinal count data with overdispersion: A simulation study

Peer-reviewed author version

A Hierarchical Bayesian Approach for the Analysis of Longitudinal Count Data with Overdispersion: A Simulation Study

Mehreteab Aregay[a], Ziv Shkedy[b], Geert Molenberghs[b,a]

[a]*I-BioStat, Katholieke Universiteit Leuven, Leuven, Belgium*

[b]*I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium*

Abstract

In sets of count data, the sample variance is often considerably larger or smaller than the sample mean, known as a problem of over- or underdispersion. The focus is on hierarchical Bayesian modeling of such longitudinal count data. Two different models are considered. The first one assumes a Poisson distribution for the count data and includes a subject-specific intercept, which is assumed to follow a normal distribution, to account for subject heterogeneity. However, such a model does not fully address the potential problem of extra-Poisson dispersion. The second model, therefore, includes also random subject and time dependent parameters, assumed to be gamma distributed for reasons of conjugacy. To compare the performance of the two models, a simulation study is conducted in which the mean squared error, relative bias, and variance of the posterior means are compared.

Keywords: Deviance information criteria, Hierarchical Poisson-Normal model (HPN), Hierarchical Poisson-Normal overdispersed model (HPNOD), Overdispersion.

1. Introduction

In medical research, data are often collected in the form of counts, e.g., corresponding to the number of times that a particular event of interest occurs. A common model for count data is the Poisson model, which is rather restrictive, given that variance and mean are equal. Often, in observed count data, the sample variance is considerably larger (smaller) than the sample mean — a phenomenon called overdispersion (underdispersion). Generically, this is referred to as extra-(Poisson)-dispersion (Iddi and Molenberghs 2012). If not appropriately accounted for, extra-dispersion may cause serious flaws in precision estimation, and inferences based there upon (Breslow, 1990). However, such excess variation has little effect on estimation of the regression coefficients of primary interest (Cox, 1983).

One of the approaches to this problem is to assume a specific, flexible parametric distribution for the Poisson means associated with each observed count. Margolin *et al.* (1981) assumed a gamma mixing distribution for the Poisson means which leads to the negative binomial model. The advantage of this parametric approach is that parameter estimates may be obtained by maximum

---

Email address: `Correspondence: I-BioStat, Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium; geert.molenberghs@uhasselt.be` (Geert Molenberghs)

likelihood, leading to estimates that are asymptotically normal, consistent, and efficient if the parametric assumptions are accurate (Cramér, 1946; Wald, 1949).

Under conditions discussed by Cox (1983), maximum likelihood methods maintain high efficiency for modest amounts of extra-dispersion, even when not explicitly accounted for in the parametric model. Pocock *et al.* (1981) proposed an intermediate solution, via maximum likelihood, to the problem of fitting regression models to tables of frequencies when the residual variation is substantially larger than would be expected from assumptions. Williams (1982) proposed a moment method for logistic linear models, and Breslow (1984) used the method proposed by Pocock *et al.* (1981) and Williams (1982) for log-linear models. Furthermore, the quasi-likelihood method, which can be considered a moment method, was applied for overdispersion by McCullagh and Nelder (1989) and Wedderburn (1974). The asymptotic properties of all these moment method for extra-binomial and extra-Poisson variation were studied by Moore (1986).

For modeling longitudinal count data with overdispersion, similarly to Zeger (1988), Thall and Vail (1990) developed a mixed-effects approach in which the regression coefficients are estimated by generalized estimating equation and the variance component is estimated using method of moments. This may be viewed as an extension of Liang and Zeger's (1986) model for longitudinal count data. Variance components are generally of broad interest (Pryseley *et al.* 2011).

Besides, Booth *et al.* (2003) and Molenberghs *et al.* (2007) brought together both modeling strands and allowed at the same time correlation between repeated measures and overdispersion in the counts. This work was extended by Molenberghs *et al.* (2010) to data types different from counts. Molenberghs *et al.* (2007) termed their model the *combined model*. All of these authors conducted parameter estimation and inferences using a likelihood paradigm. In contrast, this paper takes a likelihood perspective. In particular, two versions of a hierarchical Poisson model for longitudinal count data are studied. The first one includes subject-specific random effects to account for subject heterogeneity (a conventional generalized linear mixed model) and the second one includes an additional parameter accounting for overdispersion, generated through an additional gamma distributed random effect (a combined model). The two models are applied to real longitudinal count data and compared using a simulation study.

This paper proceeds as follows. In Section 2, the motivating study is described, which comprises a set of data on epileptic patients. The statistical methodologies is laid out in Section 3. In Section 4, the dataset is analyzed, followed by a simulation study in Section 5.

2. The Epilepsy Data

The data set used in this study is obtained from 89 epileptic patients that are randomized into either placebo or novel anti-epileptic drug (AED), in combination with one or two other AED's after a 12-week run-in period. 45 patients were assigned to the placebo group, the rest to AED. This is a double-blind, parallel group multi-center study. Patients were measured weekly and followed during 16 weeks. That said, some patients were measured up to 27 weeks. The aim of the study was to compare between the groups, the number of seizures experienced during the last week. Note that there are relatively few observations from 20 weeks onwards. Table 1 shows the number of measurements at a selection of time-points. These data were used as one of the three illustrating examples in Booth *et al.* (2003) who also considered models for longitudinally observed counts that accommodate, at the same time, overdispersion and correlation between repeated measures; for a more elaborate discussion regarding the data, refer

2

Table 1: *Epilepsy Data. Number of measurements available at a selection of time points, for both treatment groups separately.*

	# Observations		
Week	Placebo	Treatment	Total
1	45	44	89
5	42	42	84
10	41	40	81
15	40	38	78
16	40	37	77
17	18	17	35
20	2	8	10
27	0	3	3

Table 2: *Sample mean (Sample variance) at a selection of time-points, for both treatment groups separately.*

	mean (variance)	
Week	Placebo	Treatment
1	3.17 (17.19)	3.55 (26.39)
3	3.56 (27.87)	5.25 (157.45)
5	3.14 (14.86)	2.33 (8.66)
10	2.44 (8.30)	4.63 (109.37)
15	3.30 (47.49)	3.47 (55.28)
16	1.90 (6.55)	2.38 (22.63)
17	2.61 (14.84)	3.94 (143.56)
19	11.60 (644.30)	1.00 (2.00)
20	2.50 (4.50)	1.13 (2.41)
27	-	2.33 (16.33)

to Faught *et al.* (1996) and Molenberghs *et al.* (2007). The individual profile curves for both arms is shown in Figure 1 and reveals substantial variability between subjects; the graphs also show the presence of rather extreme values. It was noticed that there was up and down behavior in the mean evolution. Specifically, on average, there was a substantially higher number of epileptic seizures on week 19 in the placebo group than in the treatment group (Figure 2). The observed variances at each week are shown in Figure 2. Notice that there is very high variability in week 19 in the placebo group.

To gain insight into the extent of overdispersion, the sample mean and sample variance at each week for the treatment and placebo group was calculated (Table 2). Clearly, the sample variance is much larger than the sample mean, underscoring the presence of overdispersion in the data. This effect is evident as well from the scale of the mean evolution and variance structure in Figure 2.
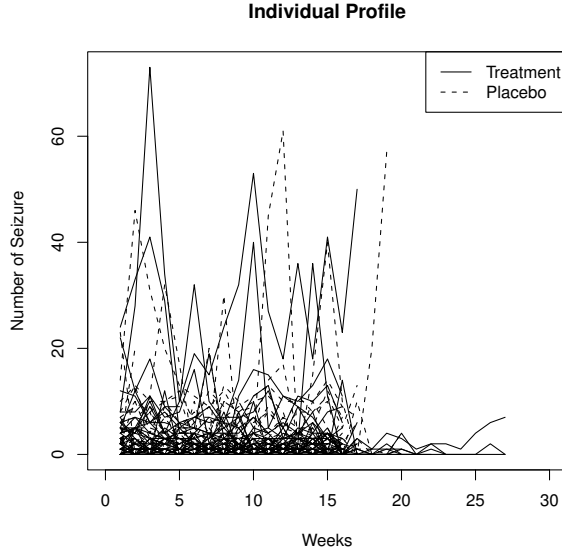
Individual Profile

Figure 1: *Individual profiles of the epilepsy data for both treatment groups*

3. A Hierarchical Poisson Normal Model with Extra-dispersion

Let Y_{ij} represent the number of epileptic seizures that patient i experiences during week j, $i = 1, 2, \ldots, 89$ and $j = 1, 2, \ldots, n_i$, where n_i is the number of repeated measurements for patient i. There are 1419 measurements available in total. We assume the following hierarchical Poisson-Normal model (HPN): $Y_{ij} \sim \text{Poisson}(\lambda_{ij}|b_i)$ with

$$\eta_{ij} = \log(\lambda_{ij}) = \beta_{00} \times I_i + \beta_{01} \times (1 - I_i) + \beta_{10} \times I_i \times t_{ij} + \beta_{11} \times (1 - I_i) \times t_{ij} + b_i, \quad (1)$$

where I_i is an indicator variable which takes value 1 for a treated subject and 0 for a placebo subject. Hence, the mean response is given by

$$E(Y_{ij}|b_i) = \begin{cases} \beta_{00} + \beta_{10} \times t_{ij} + b_i, & \text{if active,} \\ \beta_{01} + \beta_{11} \times t_{ij} + b_i, & \text{if placebo.} \end{cases}$$

Here, $\boldsymbol{\beta} = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})'$ is the parameter vector of the fixed effects and b_i is the subject-specific parameter. Independent, non-informative prior distributions for both $\boldsymbol{\beta}$ and b_i are used: $b_i \sim N(0, \sigma_b^2)$ and $\beta_{lk} \sim N(\mu_{lk}, \sigma_{\beta_{lk}}^2)$, for $l = 0, 1$, $k = 0$ (treatment) or 1 (placebo). To complete the specification of the hierarchical model, assume the following hyper-prior distributions: $\sigma_b^{-2} \sim G(0.01, 0.01)$ and $\sigma_{\boldsymbol{\beta}_{lk}}^{-2} \sim G(0.01, 0.01)$ Non-informative independent normal priors $\mu_{lk} \sim N(0; 100, 000)$ were specified for the prior means of the components of $\boldsymbol{\beta}$ (Gelman 2006).

The hierarchical model specified above assumes that the sources of variability associated with the response variable is related to the Poisson distribution and the subject heterogeneity. For Poisson regression, this is often not the case and extra variability is called for. A number of early extensions of the HPN models have been proposed by Breslow (1984) and Lawless (1987). A commonly encountered step is to allow an overdispersion parameter $\phi \neq 1$, so that $\text{Var}(Y) = \phi\mu$.
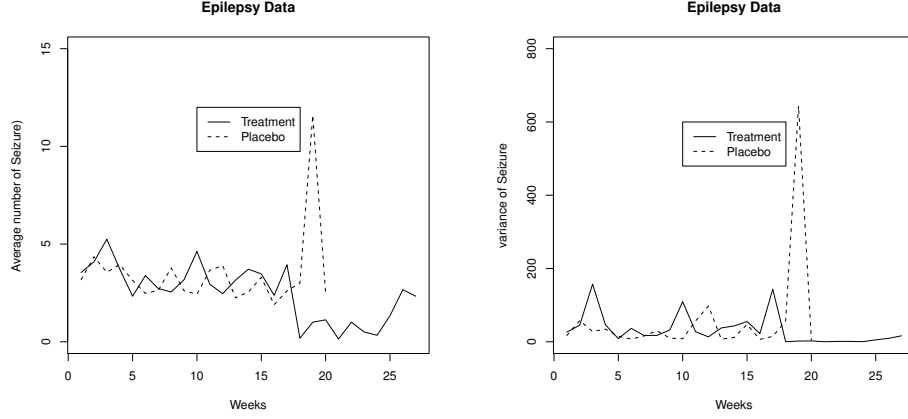
4

Figure 2: *Mean evolution over time (left panel) and variance structure (right panel) of the epilepsy data for both treatment groups*

This is similar to the moment-based approach, though such moments may arise from a random sum of Poisson variables, a point made by Hinde and Demétrio (1998a) and Hinde and Demétrio (1998b). Molenberghs *et al.* (2007) as well discussed an extension to the HPN model that account for extra-dispersed count data (HPNOD model). This is effectuated by adding a multiplicative dispersion parameter to the mean structure, i.e.,

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij}\theta_{ij}|b_i, \theta_{ij}), \tag{2}$$
$$\eta_{ij} = \log(\lambda_{ij}). \tag{3}$$

Hence, the Poisson parameter in (2) has two components. The first is identical to the linear predictor in (1) and is expressed as (3) and captures covariate dependence, while the second one captures overdispersion. It is assumed that $\theta_{ij} \sim \text{Gamma}(\alpha, 1/\alpha)$. Thus, $E(\theta_{ij}) = \alpha^2$ and $\text{Var}(\theta_{ij}) = \alpha^3$. Molenberghs *et al.* (2007) assumed that the components θ_{ij} of the vector $\theta_i = (\theta_{i1}, \theta_{i2}, \ldots, \theta_{in_i})'$ are independent. Note that for $1/\alpha \to \infty$, $\text{Var}(\theta_{ij}) \to 0$, and the above HPNOD model is reduced to the HPN model. The same prior distributions for the parameters as discussed in Section 3 were used. In addition, a uniform distribution $U(0, 100)$ was considered as the prior distribution of α.

Breslow and Clayton (1993) analyzed the epilepsy data set from Thall and Vail (1990) by considering the covariates: logarithm of baseline seizure count, treatment, logarithm of age, visit, and the treatment by log(base) interaction. Similarly, the model specified in (1) is extended with these effects. The resulting HPN is:

$$
\begin{aligned}
Y_{ij} &\sim \text{Poisson}(\lambda_{ij}|b_i), \\
\eta_{ij} &= \log(\lambda_{ij}) = \beta_0 + \beta_1 \times I_i + \beta_2 \times \log(\text{base}_i) + \beta_3 \times t_{ij} + \beta_4 \times \log(\text{age}_i) \\
&\quad + \beta_5 \times I_i \times t_{ij} + \beta_6 \times I_i \times \log(\text{base}_i) + b_i.
\end{aligned} \tag{4}
$$

Likewise, the HPNOD becomes:

$$
\begin{aligned}
Y_{ij} &\sim \text{Poisson}(\lambda_{ij}\theta_{ij}|b_i, \theta_{ij}), \\
\eta_{ij} &= \beta_0 + \beta_1 \times I_i + \beta_2 \times \log(\text{base}_i) + \beta_3 \times t_{ij} + \beta_4 \times \log(\text{age}_i) \\
&\quad + \beta_5 \times I_i \times t_{ij} + \beta_6 \times I_i \times \log(\text{base}_i) + b_i.
\end{aligned} \tag{5}
$$

5

To aid convergence when fitting the HPN and HPNOD models (4) and (5), respectively, the covariates $\log(\text{base})$, $\log(\text{age})$, and $I_i \times \log(\text{base}_i)$ were centered about their mean. The same prior distribution was considered for the parameters as in (1) and (2).

4. Analysis of the Epilepsy Data Set

The model discussed in Section 3 was fitted using the R2Winbugs package. A MCMC simulation of 100,000 iterations from which the first 10,000 were considered to be the burn-in period and discarded from analysis, was used to estimate the model parameters. For both models, convergence was assessed using trace plots, estimated potential scale reduction factor, $\hat{R}$, and Brooks, Gelman and Rubin's (BGR) plot (Gelman and Rubin, 1992). Model selection was done using the Deviance Information Criteria; DIC (Gelman *et al.*, 2004). Diagnostic plots for all parameters considered were studied and indicated convergence for all model parameters. The BGR plot indicates convergence for all model parameters. Moreover, the estimated potential scale reduction factor $\hat{R}$ values for all the parameters were close to one, which indicates convergence for all model parameters.

The posterior means for parameters and DIC values are presented in Table 3. The DIC values of the HPNOD model (4830.54) is smaller than that of the HPN model (6047.67), which indicates that the first model is to be preferred.

The posterior mean for the variance of the random effects obtained for the HPNOD model is equal to 1.152 (credible interval [0.829;1.598]) slightly smaller than the posterior mean obtained for the HPN model, which is 1.213 (credible interval [0.882;1.657]). Figure 3 shows the scatter plot of the estimated posterior means $\bar{b}_i$, obtained from the HPN and HPNOD models. The agreement between both sets is striking but not unexpected. Because the overdispersion random effect enters the linear predictor as an additional component in the intercept, it does not distort the interpretation of the random effecs b_i, a point on which will be elaborated in Section 5.2.3.

The posterior mean for α is equal to 2.48, which implies that $\text{Var}(\theta_{ij}) = 2.48^3 = 15.3$ and $E(\theta_{ij}) = 2.48^2 = 6.21$. Figure 4 shows the posterior distribution of $\bar{\theta}_{ij}$, the posterior mean of the overdispersion parameter for the ith subject at the jth occasion. it is noticeable that the empirical mean is

$$E(\bar{\theta}_{ij}) = \frac{1}{1419} \sum_{i=1}^{89} \sum_{j=1}^{n_i} \bar{\theta}_{ij} = 6.21,$$

which equals 2.48^2, as reported in Table 3, and as it should. Further, only 5 from $\bar{\theta}_{ij}$ (out of 1419) are smaller than 1. This indicates overdispersion. It is observed that there are differences in the parameter estimates between the HPN and HPNOD models. The estimated posterior means for the intercept for both treatment groups obtained for the HPN models are positive, while the sign is reversed for the intercept obtained for the HPNOD model. However, the slope parameters obtained from both models have similar magnitude and sign. On the other hand, both models produce non-significant values for the difference and ratio in slopes.

The posterior summary statistics for the model specified by (4) and (5) are shown in Table 4. Similar to the model in (1) and (2), the DIC value for the HPNOD (4833.79) is smaller than the DIC value for the HPN (6049.19) model, which establishes that the first model fits better. The posterior mean of α was equal to 2.495, indicating substantial overdispersion in the data. Note that the variance of the random effect for both HPN and HPNOD is approximately 0.4. Given

6

Table 3: *Epilepsy Data. Posterior summary statistics for the HPN and HPNOD*

Parameter	Mean		SD		MC error		95% Credible interval		Rhat	
	HPN	Comb	HPN	Comb	HPN	Comb	HPN	Comb	HPN	Comb
β_{00}	0.6441	-1.184	0.1686	0.2494	0.0033	0.0063	$(0.306, 0.971)$	$(-1.676, -0.702)$	1.00	1.00
β_{01}	0.8243	-0.8717	0.1709	0.2493	0.0033	0.0063	$(0.494, 1.164)$	$(-1.353, -0.376)$	1.00	1.00
β_{10}	-0.0120	-0.0114	0.0043	0.0074	2.17E-05	6.86E-05	$(-0.021, -0.004)$	$(-0.023, 0.003)$	1.00	1.00
β_{11}	-0.0143	-0.0252	0.0044	0.0072	2.22E-05	7.12E-05	$(-0.023, -0.006)$	$(-0.039, -0.009)$	1.00	1.00
σ^2	1.213	1.152	0.1993	0.1929	6.44E-04	7.09E-04	$(0.882, 1.657)$	$(0.829, 1.598)$	1.00	1.00
β_{10}- β_{11}	0.0023	0.0128	0.0062	0.0107	3.06E-05	9.84E-05	$(-0.009, 0.014)$	$(-0.008, 0.034)$	1.00	1.00
β_{10}/β_{11}	0.9544	0.5398	4.272	3.883	0.0086	0.0089	$(0.227, 2.367)$	$(-0.126, 1.648)$	1.08	1.05
α		2.48		0.2138		0.0060		$(2.096, 2.938)$	1.00	1.00
DIC	6047.57	4830.54								

the strong overdispersion in the data (Table 2), the advantage of using HPNOD over HPN is that the overdispersion not captured by the normal random effect can be accommodated. In line with expectation, both models produced similar results for the slopes. The difference in intercepts is not worrisome; this point is addressed in the next section.

Further, both models show that there is an effect of baseline seizure rate and time on the number of epileptic seizures. However, it was found that the treatment and its interaction with the baseline seizure count and time do not have a significant effect on the response. These models produce also non-significant value for age.

Note that the Monte Carlo error quantifies the efficiency of the posterior sample mean for the corresponding population estimand. As a rule of thumb, a Monte Carlo error $< 1 - 5\%$ of the posterior standard deviation is requested. It is clear from Tables 3 and 4 that the Monte Carlo error for all parameters satisfies this rule of thumb.

Table 4: *Epilepsy Study. posterior summary statistics using HPN and HPNOD model with covarites Age, base, time, treatment by base interaction and treatment by time interaction.*

Parameters	Mean		Sd		MC error		95% Credible interval		Rhat	
	HPN	HPNOD	HPN	HPNOD	HPN	HPNOD	HPN	HPNOD	HPN	HPNOD
β_0	0.9691	-0.7611	0.2488	0.3099	0.0075	0.0095	$(0.4867, 1.464)$	$(-1.355, -0.1556)$	1.01	1.02
β_1	-0.4984	-0.5664	0.4612	0.4815	0.0143	0.0150	$(-1.414, 0.3857)$	$(-1.502, 0.3538)$	1.01	1.01
β_2	0.9518	0.9566	0.1239	0.131	0.0033	0.0036	$(0.7023, 1.194)$	$(0.6969, 1.209)$	1.01	1.00
β_3	-0.0139	-0.0236	0.0044	0.0076	2.22E-05	1.06E-04	$(-0.0226, -0.0052)$	$(-0.0385, -0.0086)$	1.00	1.00
β_4	-0.0186	-0.0292	0.2672	0.268	0.0036	0.0036	$(-0.5478, 0.5061)$	$(-0.5477, 0.5082)$	1.00	1.00
β_5	0.0007	0.0086	0.0062	0.0107	4.99E-04	1.49E-04	$(-0.0114, 0.0128)$	$(-0.0123, 0.0295)$	1.00	1.00
β_6	0.0392	0.0353	0.1706	0.1756	0.0053	0.0055	$(-0.287, 0.3786)$	$(-0.2987, 0.3738)$	1.00	1.00
σ^2	0.4111	0.3868	0.0078	0.0790	4.46E-04	4.81E-04	$(0.2826, 0.5865)$	$(0.2565, 0.565)$	1.00	1.00
α		2.495		0.2083		0.0058		$(2.116, 2.935)$	1.01	1.00
DIC	6049.19	4833.79								

5. Simulation Study

In this section, a simulation study is presented, conducted to evaluate the performance of the models discussed above with and without adjustment for extra-dispersion.
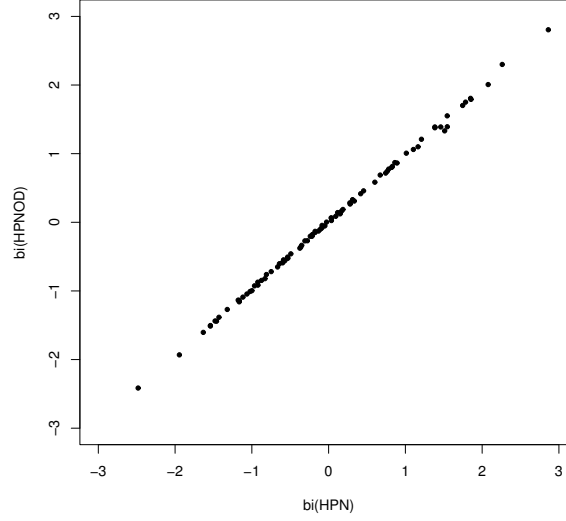
Figure 3: *A plot of the random-effect estimates obtained from HPN (x-axis) versus the random effect obtained from the HPNOD (y-axis), for the analysis of epilepsy data set*

5.1. Setup of the Simulation Study

Different settings were considered in this simulation study. The steps for the first setting are as follows: (1) Data are generated from a Poisson distribution with mean $\theta_{ij}\lambda_{ij}$. For the extra-dispersion parameter, $\theta_{ij} \sim \text{Gamma}(\alpha, 1/\alpha)$ was assumed. High, moderate, and low overdispersion levels were induced by setting $\alpha = 0.25; 1; 25$, respectively. Additionally, data without overdispersion were generated as well, i.e., $\theta_{ij} = 1$. Note that the mean structure for λ_{ij} was the same as defined in (4), with true values $(\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}) = (2, -2, 0.05, 0.2)$. These values are chosen to have induce covariate effects and opposite sign for the intercept. To check for robustness, different true values for the regression coefficients in the second and third settings were used. The main focus was be on the other model parameters, as well as on sample and cluster sizes. The intra-cluster correlation was varied by specifying different values for the standard deviation parameter of the random-effects term, i.e., equal to either 0.1 or 0.5. Two covariates were used; treatment and time. Each subject forms a cluster. Initially, all clusters were equal in size, with cluster sizes 2, 5, 10, and 20. Sample sizes in this setting were equal to 30, 60, and 120 subjects. Half of the individuals were placed in the treatment arm, with the other half in the placebo group. This setting led to $4 \times 4 \times 2 \times 3 = 96$ different scenarios. (2) Both the HPN and HPNOD models are fitted, using Bayesian methods; (3) The first and second steps were repeated 100 times for each scenario. (4) The bias was calculated, relative bias, variance, and MSE of the parameters, and compared both models.

The second setting was similar to the aforementioned one except for $\beta_{10} = -0.2$, which allows opposite signs for the slopes. The true values of the third setting were taken from the results of the epilepsy data set, that is, $\beta_{00} = 0.655$, $\beta_{01} = 0.9112$, $\beta_{10} = -0.0118$, and $\beta_{11} = -0.0248$. This leads to a total of $4 \times 2 \times 2 = 16$ scenarios for α, σ^2 and $\boldsymbol{\beta}$. In this setting, $n = 60$ and the cluster size is equal to 10. Evidently, the algorithmic steps in these settings are the same as in
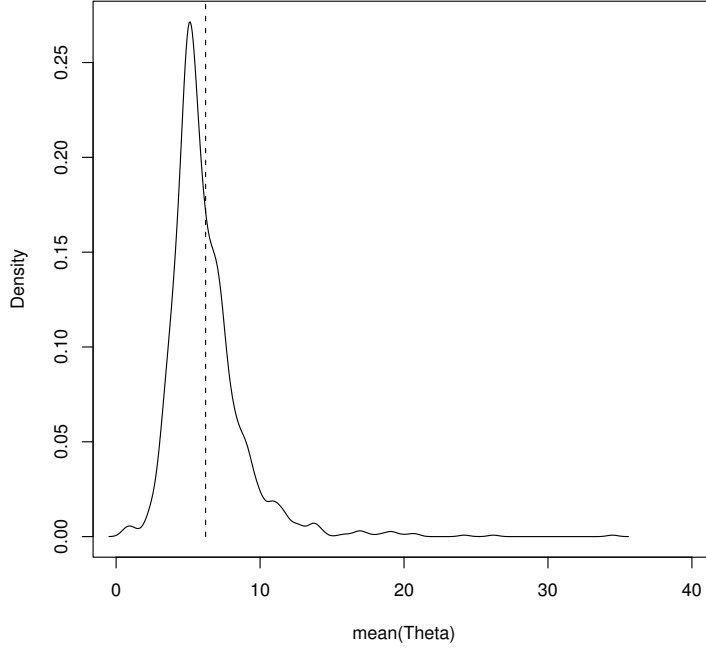
Figure 4: *Kernel density for the posterior distribution of $\bar{\theta}_{ij}$ for the analysis of epilepsy data set. The vertical dashed line is the mean of $\bar{\theta}_{ij}$*

the first setting.

For each simulated dataset, the model was estimated using 60,000 MCMC iterations from which the first 30,000 were considered burn in and discarded from analysis. The number of chains was 1.

5.2. Simulation Results

5.2.1. Effect on α and σ

For the first simulation study with $n = 60$, for all settings with overdispersion, the MSE for σ for the HPNOD model is smaller or equal to that for the HPN model, as can be clearly seen from Figures 5 and 6. Besides, for all settings, the MSE of the overdispersion parameter α increases when the overdispersion level increases. There is more bias in α when there is low overdispersion. For the high overdispersion scenario, the MSE of the HPN model for β_{00} and β_{01} is smaller than the MSE obtained for the HPNOD model. Note that, when the level of overdispersion is moderate ($\alpha = 1$), the MSE of β_{00} and β_{01} for the HPN model is slightly smaller compared with the MSE obtained for the HPNOD model. For the low ($\alpha = 25$) and no ($\alpha \to \infty$) overdispersion scenarios, the MSE for β_{00} and β_{01} obtained for the HPN model is much smaller than the MSE obtained for the HPNOD model. However, for all levels of overdispersion, the MSE for the slope parameters β_{10} and β_{11} is of the same magnitude for the two models (Table 5; see Supplementary Appendix). The findings for HPN and HPNOD were similar for

9

$\sigma = 0.1$ and $\sigma = 0.5$, except that the estimates were slightly more precise and unbiased for $\sigma = 0.5$ (Table 5, see Supplementary Appendix).

5.2.2. Effect of Sample Size and Cluster Size

The results obtained for both models indicate that, as expected, the MSE, variance, and bias decrease as the sample size increase (Figures 5, 6, and 7, see Supplementary Appendix). Note that for the low overdispersion scenario, for a cluster size of 2, the HPN model introduces high variance and MSE for β_{01} and β_{11}. On the other hand, if the data are generated without overdispersion, for clusters of size 2, the HPNOD model results in high variability and MSE value for β_{01} and β_{11}. This suggests caution, for both models, with small cluster size

5.2.3. Effect of β

To check the robustness of the simulation result with regard to the true values of β, two additional scenarios were considered. The findings (see Supplementary Appendix) were similar to the previous findings for the first scenario in Section 5.2.1.

Furthermore, data were generated according to the model specified in (1) and (2) with unbalanced time points (unequal cluster size), similar to the epilepsy data set as discussed in Section 2. In this setting, 60 individuals were considered over variable numbers of periods of time up to a maximum of 27 time points. Half of the individuals were assigned to the treatment group. The true values for the regression coefficients were also here $\beta = (2, -2, 0.05, 0.2)$, while the standard deviation of the random effects was 0.1 and 0.5. The results are shown in Table 6. One notices that the bias, relative bias, variance, and MSE were similar to that of the first scenario (Section 5.2.1). Also 30 and 120 individuals with unbalanced time points were considered, and the results (details not shown) were similar to the balanced time points case with the same number of individuals. Additionally, to gauge the effect of the standard deviation of the random effect, $\sigma = 2$ was employed. The findings were very close to these of $\sigma = 0.5$ (details not shown).

To assess the effect of assigning patients randomly to the treatment group, a completely randomized design was used. Each patients has equal probability of being assigned either to the treatment or to the placebo group. Data were generated according to (1) and (2) and both the HPN and the HPNOD model were fitted for each simulated data set. The true values for the regression coefficients were also here $\beta = (2, -2, 0.05, 0.2)$, while the standard deviation of the random effects was 0.1 and 0.5. Initially, a sample size of 60 subjects was used, together with clusters of size 10. The results are shown in Table 7. It can be clearly seen that the results are similar to the previous setting with the exception that, for the data generated with low overdispersion level ($\alpha = 25$), the bias and MSE of α in this setting were smaller than the previous one. However, the conclusions which can be drawn from these results are similar to these for the earlier scenarios. Note that the sample size was varied to 60 and 120 subjects, while the cluster sizes were set to 2, 5, and 20. The results, omitted for brevity, are similar.

In general, when there is low overdispersion and when the data are generated without overdispersion, the HPN revealed more precise and unbiased estimates than HPNOD for the intercepts, while there was similar precision and unbiasedness in the slopes. This is not unexpected, because there are three contributions to the intercept in a Poisson model with normal and overdispersion random effects. Indeed, it can be shown (details omitted), that the marginal expectation of the
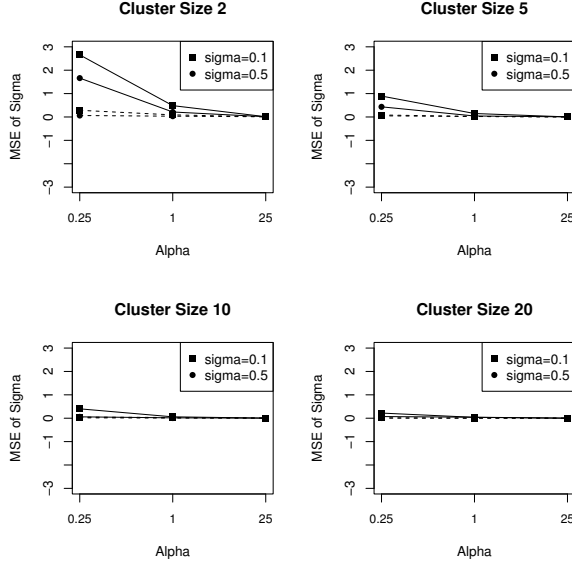
Figure 5: *Simulation Study. Comparison of the MSE of the standard deviation of the random effects for different cluster sizes, using the HPN model (solid line) and HPNOD (dashed line). The x-axis represents the value of α which shows the amount of overdispersion amount and the y-axis represents the MSE. $n = 60$*

count in the HPNOD equals

$$E(Y_{ij}) = e^{\log E(\theta_{ij}) + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\sigma^2}.$$

Hence, $\log E(\theta_{ij}) + \beta_0 + 0.5\sigma^2$ form the intercept. When comparing the HPN and HPNOD, not simply β_0, but rather $\beta_0 + \log E(\theta_{ij})$ ought to be compared. With this in mind, the bias in the intercept becomes an apparent bias. No such phenomenon plays for the other covariate effects. On the other hand, the HPNOD revealed less biased and precise estimates than the HPN for σ. Besides, for all simulation settings, except for clusters of size 2, the MSE and bias of all the parameters obtained from HPN increases when the overdispersion level increases. This indicates that HPN is not performing well when there is high overdispersion.

The aforementioned issue with the intercept also shows through the correlation structure of the various contributions. For example, when there is low overdispersion and when the data are generated without overdispersion, the correlation between the intercept and overdispersion parameter α was calculated to be more than 0.94 (Table 8).

To select the best model, the DIC for the HPN and HPNOD was calculated for each data set. When α is equal to 0.25, the mean DIC value of the 100 data set for the HPN was 7314 which is too large when compared with the mean DIC value of the HPNOD model: 1629. Similarly, the DIC value of the HPN model when $\alpha = 1$ was higher than the DIC value for the HPNOD. Furthermore, when $\alpha = 25$ the DIC value for the HPN is higher than for the HPNOD. This suggests that the HPNOD is performing better when there is overdispersion. On the other hand, when the data are generated without overdispersion, the HPNOD has slightly smaller DIC value than the HPN (Table 5 and Figure 8). This indicates that when the data are generated without overdispersion, both models perform similarly.

11

Table 5: *Summary of the simulation result of $n = 60$, $t = 10$ for over dispersed data with different α values and without over dispersion data with HPN and HPNOD*

σ	Parameter		α							
			0.25		1		25		$\alpha\to\infty$	
			GLMM	HPNOD	GLMM	HPNOD	GLMM	HPNOD	GLMM	HPNOD
0.1	β_{00}	Bias	-0.2116	2.6697	-0.0343	-0.104	0.0031	-7.0672	-0.0032	-8.7611
		Rel.Bias	-0.1058	1.3348	-0.0172	-0.052	0.0016	-3.531	-0.0016	-4.3806
		var	0.0757	0.1186	0.0191	0.0535	0.0032	0.6291	0.0028	0.0511
		MSE	0.1205	7.2457	0.0203	0.0643	0.0032	50.511	0.0028	76.8
	β_{01}	Bias	-0.2395	2.7100	-0.0559	-0.0937	-0.0617	-7.0687	- 0.0309	-8.8116
		Rel.Bias	0.1197	-1.355	0.0279	0.0468	0.0308	3.534	0.0155	4.4058
		var	0.1426	0.1789	0.0849	0.1142	0.0691	0.6546	0.0497	0.1060
		MSE	0.1999	7.5233	0.088	0.1231	0.0731	50.621	0.0506	77.8
	β_{10}	Bias	-0.0023	0.0048	-0.002	0.0014	-0.0015	0.0004	0.0007	-0.0014
		Rel.Bias	-0.0466	0.0961	-0.0406	0.0272	-0.029	0.0086	0.0138	-0.0284
		var	0.0021	0.0022	0.0005	0.0004	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0021	0.0022	0.0005	0.0004	<0.0001	<0.0001	<0.0001	<0.0001
	β_{11}	Bias	-0.0051	-0.0043	-0.0016	-0.0055	0.0046	-0.0022	0.0028	0.0037
		Rel.Bias	-0.0253	-0.0215	-0.0078	-0.0276	0.0232	-0.0011	0.0138	0.0187
		var	0.0031	0.0034	0.0018	0.0012	0.0011	0.0009	0.0008	0.0011
		MSE	0.0031	0.0034	0.0018	0.0012	0.0011	0.0009	0.0008	0.0011
	σ	Bias	0.6268	0.2106	0.2396	0.1339	0.0609	0.0606	0.0102	0.0122
		Rel.Bias	6.2679	2.1061	2.396	1.3885	0.6089	0.6055	0.1021	0.1215
		var	0.0073	0.0025	0.0017	0.0131	0.0003	0.0004	0.0003	0.0003
		MSE	0.4001	0.0468	0.0592	0.031	0.0039	0.0078	0.0004	0.0005
	α	Bias		0.0073		0.0527		12.915		
		Rel.Bias		0.0293		0.0527		0.5166		
		Var		0.0004		0.0091		254.35		
		MSE		0.0004		0.0119		421.135		
	DIC		7313.68	1628.53	4131.24	2226.50	2154.22	2105.32	2062.11	2051.63
0.5	β_{00}	Bias	-0.2116	2.8153	-0.0095	-0.0329	-0.0123	-6.9979	-0.0032	-8.8303
		Rel.Bias	-0.1058	1.4708	-0.0476	-0.0165	-0.0062	-3.4989	0.0016	-4.4152
		var	0.0758	0.1314	0.0369	0.0532	0.0126	0.851	0.0088	0.0438
		MSE	0.1205	8.0576	0.046	0.0543	0.0128	49.8	0.0088	78
	β_{01}	Bias	-0.2395	2.7409	-0.0869	-0.0287	-0.0319	-7.0473	-0.0681	-8.8047
		Rel.Bias	0.1197	-1.3705	0.0434	0.0143	0.0159	3.5236	0.0340	4.4024
		var	0.1426	0.2374	0.0908	0.1253	0.0525	0.821	0.0667	0.1059
		MSE	0.1999	7.7503	0.0983	0.1261	0.0535	50.5	0.0714	77.6
	β_{10}	Bias	-0.0023	-0.0040	0.0057	0.0024	<0.0001	-0.0003	<0.0001	-0.0002
		Rel.Bias	-0.0467	-0.0805	0.1141	0.0472	<0.0001	-0.0059	-0.0011	-0.0032
		var	0.0021	0.0018	0.0007	0.0005	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0021	0.0018	0.0007	0.0005	<0.0001	<0.0001	<0.0001	<0.0001
	β_{11}	Bias	-0.0051	0.0026	0.0035	0.006	0.0011	0.0034	0.0066	-0.0028
		Rel.Bias	-0.0253	0.0128	0.0174	0.0300	0.0056	-0.0059	0.0328	-0.0140
		var	0.0031	0.0035	0.0015	0.0016	0.0007	0.0010	0.0009	0.0010
		MSE	0.0031	0.0035	0.0015	0.0016	0.0007	0.0011	0.0009	0.0010
	σ	Bias	0.2268	-0.0265	0.1043	-0.0124	0.0106	-0.0073	0.0020	-0.0044
		Rel.Bias	0.4536	-0.0531	0.2087	-0.0247	0.0213	-0.0146	0.0040	-0.0087
		var	0.0073	0.0168	0.0056	0.0069	0.0041	0.0033	0.0031	0.0035
		MSE	0.0587	0.0175	0.0165	0.0071	0.0042	0.0034	0.0031	0.0035
	α	Bias		0.0021		0.0087		12.5088		
		Rel.Bias		0.0085		0.0087		0.5044		
		Var		0.0006		0.0081		335		
		MSE		0.0006		0.0082		492		
	DIC		7935.78	1625.49	4374.21	2240.69	2207.83	2152.35	2097.15	2090.63

6. Concluding Remarks

A Bayesian inferential route was proposed for the HPNOD (and the HPN), and compared the performance of the HPN and HPNOD models on data generated with and without overdispersion. A Bayesian approach was adopted. When the data are generated with high overdispersion levels, the HPN model leads to higher bias and less precise estimates for the variance of the random effect (σ^2) than the HPNOD. HPN and HPNOD produce similar results for the slopes.

HPNOD and HPN provide similar bias and precision for the slopes and for the random effects variance σ. To check the problem with the intercept estimates using the HPNOD model, the correlation between the parameters was calculated.

The intercepts between the two models cannot be directly compared, but only indirectly, given that it takes the form $\log E(\theta_{ij}) + \beta_0 + 0.5\sigma^2$ in the HPNOD and $\beta_0 + 0.5\sigma^2$ in the HPN.

A Deviance Information Criterion (DIC) was applied to check the overall performance of both

Table 6: *Summary of the simulation result of the data generated from HPNOD and HPN model with n = 60 individuals, and with unbalanced time points in a way similar to the epilepsy data*

			α							
			0.25		1		25		$\alpha\to\infty$	
σ	Parameter		HPN	HPNOD	HPN	HPNOD	HPN	HPNOD	HPN	HPNOD
0.1	β_{00}	Bias	-0.2742	2.7185	-0.057	-0.0529	-0.0025	-6.7331	-0.0023	-8.8627
		Rel.Bias	-0.1371	1.3593	-0.0285	-0.0265	-0.0012	-3.3665	-0.0012	-4.4313
		var	0.0417	0.0559	0.0132	0.0287	0.0016	0.3048	0.0009	0.0648
		MSE	0.1169	7.4464	0.0165	0.0315	0.0016	45.639	0.0009	78.61
	β_{01}	Bias	-0.2201	2.6775	0.0039	-0.0528	-0.0143	-6.744	-0.0189	-8.8863
		Rel.Bias	0.1101	-1.3387	-0.0019	0.0264	0.0072	3.3721	0.0095	4.4431
		var	0.1538	0.0699	0.0512	0.0495	0.0159	0.3159	0.0129	0.0740
		MSE	0.2022	7.2391	0.0513	0.0523	0.0162	45.800	0.0133	79.04
	β_{10}	Bias	0.0013	0.0013	0.0004	-0.0005	<0.0001	0.0001	<0.0001	0.0001
		Rel.Bias	0.0263	0.0264	0.0074	-0.0094	0.0006	0.0029	<0.0001	0.0024
		var	0.0004	0.0003	0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0004	0.0003	0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	β_{11}	Bias	0.0017	0.0051	-0.0039	-0.0002	0.0004	0.0008	0.0008	0.0012
		Rel.Bias	0.0084	0.0254	-0.0195	-0.0007	0.0018	0.0043	0.0039	0.0062
		var	0.0011	0.0004	0.0003	0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0011	0.0004	0.0003	0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	σ	Bias	0.6264	0.1261	0.2439	0.0584	0.0241	0.0109	0.0104	0.0082
		Rel.Bias	6.2639	1.2608	2.4398	0.5841	0.2414	0.1090	0.1042	0.0819
		var	0.0101	0.0027	0.0022	0.001	0.0004	0.0003	0.0002	0.0004
		MSE	0.4025	0.0186	0.0617	0.0044	0.0010	0.0003	0.0004	0.0004
	α	Bias		0.0035		0.0298		5.644		
		Rel.Bias		0.0138		0.0298		0.2258		
		Var		0.0003		0.0048		92.475		
		MSE		0.0003		0.0057		124.351		
0.5	β_{00}	Bias	-0.2972	2.7655	-0.0679	-0.0306	-0.0137	-6.5554	-0.0085	-8.8618
		Rel.Bias	-0.1486	1.3827	-0.0339	-0.0153	-0.0069	-3.2777	-0.0042	-4.4309
		var	0.0562	0.0583	0.0167	0.0418	0.0118	0.2643	0.0100	0.0676
		MSE	0.1445	7.7061	0.0213	0.0427	0.0119	43.237	0.0101	78.599
	β_{01}	Bias	-0.2978	2.7757	-0.0592	-0.0345	-0.0151	-6.564	-0.0043	-8.8713
		Rel.Bias	0.1489	-1.3878	0.0296	0.0172	0.0075	3.2819	0.0021	4.4356
		var	0.2927	0.1167	0.0819	0.0609	0.0268	0.2874	0.0204	0.0959
		MSE	0.3814	7.8209	0.0854	0.0621	0.0269	43.371	0.0204	78.7959
	β_{10}	Bias	0.0024	-0.0003	-0.0008	0.0007	<0.0001	-0.0002	0.0004	0.0003
		Rel.Bias	0.0484	-0.0052	-0.0166	0.0145	0.0015	-0.0039	0.0086	0.0064
		var	0.0005	0.0004	0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0005	0.0004	0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	β_{11}	Bias	0.0058	0.0013	-0.0005	0.001	0.0173	0.0002	-0.0012	-0.0005
		Rel.Bias	0.0289	0.0065	-0.0027	0.0052	0.0346	0.0008	-0.0060	-0.0026
		var	0.0019	0.0005	0.0004	0.0002	0.0033	<0.0001	<0.0001	<0.0001
		MSE	0.0019	0.0005	0.0004	0.0002	0.0036	<0.0001	<0.0001	<0.0001
	σ	Bias	0.3992	-0.0107	0.1057	0.017	0.0173	0.0172	0.0159	0.0075
		Rel.Bias	0.7983	-0.0215	0.2115	0.0340	0.0346	0.0345	0.0319	0.0149
		var	0.0171	0.0185	0.0055	0.00056	0.0033	0.0031	0.0036	0.0033
		MSE	0.1764	0.0186	0.01657	0.0056	0.0036	0.0034	0.0039	0.0034
	α	Bias		0.0022		0.0157		2.9087		
		Rel.Bias		0.0089		0.0157		0.1163		
		Var		0.0003		0.0057		65.3326		
		MSE		0.0003		0.0059		73.793		

Table 7: *Summary of the simulation result of $n = 60$, $t = 10$ for over dispersed data with different α values and without over dispersion data with HPN and HPNOD using random allocation of patients in to either Treatment or Placebo group*

σ	Parameter		α							
			\multicolumn{2}{c}{0.25}	\multicolumn{2}{c}{1}	\multicolumn{2}{c}{25}	\multicolumn{2}{c}{$\alpha\to\infty$}
			HPN	HPNOD	HPN	HPNOD	HPN	HPNOD	HPN	HPNOD
0.1	β_{00}	Bias	-0.1802	2.7192	-0.0556	-0.0349	-0.0074	-6.9011	-0.005	-8.7534
		Rel.Bias	-0.0901	1.3596	-0.0278	-0.0174	-0.0037	-3.4506	-0.003	-4.3767
		var	0.0604	0.1065	0.0175	0.0442	0.0025	0.5803	0.002	0.0913
		MSE	0.0929	7.5007	0.0205	0.0454	0.0025	48.206	0.002	76.71
	β_{01}	Bias	-0.2637	2.6582	-0.0943	-0.0842	-0.0375	-6.9305	-0.0007	-8.7487
		Rel.Bias	0.1318	-1.3291	0.0472	0.0421	0.0188	3.4665	0.0004	4.3767
		var	0.1708	0.1579	0.0827	0.1019	0.0647	0.6739	0.054	0.1651
		MSE	0.2403	7.2238	0.0916	0.1089	0.0661	48.741	0.054	76.71
	β_{10}	Bias	-0.0049	-0.0006	-0.0009	-0.0011	0.0002	0.0003	0.0009	0.0010
		Rel.Bias	-0.0990	-0.0111	-0.0179	-0.0209	0.0029	0.0053	0.019	0.0202
		var	0.0017	0.0018	0.0005	0.0005	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0017	0.0018	0.0005	0.0005	<0.0001	<0.0001	<0.0001	<0.0001
	β_{11}	Bias	-0.0033	0.0062	0.0022	0.0039	0.0021	0.0024	-0.003	-0.0031
		Rel.Bias	-0.0165	0.0312	0.0108	0.0195	0.0106	0.0122	-0.015	-0.0155
		var	0.0036	0.0029	0.0016	0.0016	0.0011	0.0011	0.001	0.0011
		MSE	0.0036	0.0029	0.0016	0.0016	0.0011	0.0011	0.001	0.0011
	σ	Bias	0.6288	0.1649	0.2581	0.0871	0.0251	0.0154	0.015	0.0111
		Rel.Bias	6.2885	1.6496	2.5808	0.8709	0.2511	0.1543	0.145	0.1113
		var	0.0078	0.0050	0.0027	0.0016	0.0004	0.0003	0.0005	0.00005
		MSE	0.4033	0.0322	0.0693	0.0092	0.001	0.0006	0.0007	0.00006
	α	Bias		0.0057		0.0223		9.6462		
		Rel.Bias		0.0227		0.0223		0.3858		
		Var		0.0005		0.0075		189.53		
		MSE		0.0006		0.0080		282.58		
0.5	β_{00}	Bias	-0.1795	2.7272	-0.0730	0.0005	-0.0095	-6.7143	0.0024	-8.8129
		Rel.Bias	-0.0898	1.3636	-0.0365	0.0003	-0.0047	-3.3572	0.0012	-4.4065
		var	0.0795	0.0984	0.0354	0.0724	0.0106	0.3633	0.0092	0.0467
		MSE	0.1117	7.5363	0.0408	0.0724	0.0107	45.445	0.0092	77.71
	β_{01}	Bias	-0.3174	2.7003	-0.0725	0.0005	-0.0297	-6.7362	-0.007	-8.8225
		Rel.Bias	0.1587	-1.3501	0.0362	-0.0003	0.0149	3.3681	0.0004	4.4112
		var	0.1637	0.1716	0.0997	0.1152	0.0639	0.4184	0.0607	0.1119
		MSE	0.2644	7.4629	0.1049	0.1152	0.0648	45.794	0.0608	77.95
	β_{10}	Bias	-0.0070	0.0004	0.0033	0.0039	0.0008	0.0007	-0.0005	-0.0005
		Rel.Bias	-0.1404	0.0071	0.0652	0.0789	0.0166	0.0139	-0.0096	-0.0105
		var	0.0020	0.0017	0.0006	0.0006	<0.0001	<0.0001	<0.0001	<0.0001
		MSE	0.0020	0.0017	0.0007	0.0006	<0.0001	<0.0001	<0.0001	<0.0001
	β_{11}	Bias	0.0066	0.0007	-0.0002	0.0015	0.0001	0.0003	-0.0033	-0.0034
		Rel.Bias	0.0329	0.0034	-0.0008	0.0073	0.0007	0.0015	-0.0163	-0.0169
		var	0.0029	0.0030	0.0015	0.0014	0.0009	0.0009	0.0008	0.0008
		MSE	0.0030	0.0030	0.0015	0.0014	0.0009	0.0009	0.0008	0.0008
	σ	Bias	0.3807	-0.0039	0.1165	0.0048	0.0259	0.0216	0.0218	0.0204
		Rel.Bias	0.7614	-0.0078	0.2330	0.0095	0.0518	0.0432	0.0437	0.0409
		var	0.0114	0.0241	0.0050	0.0077	0.0043	0.0044	0.0048	0.0048
		MSE	0.1563	0.0241	0.0186	0.0077	0.005	0.0048	0.0052	0.0052
	α	Bias		-0.003		-0.0008		5.564		
		Rel.Bias		-0.0102		-0.0008		0.2226		
		Var		0.0005		0.0089		89.665		
		MSE		0.0005		0.0089		120.623		

models. The DIC result seems to imply that the HPNOD is much better than the HPN model for data with high, moderate, and low overdispersion. Nevertheless, the HPNOD model has slightly smaller DIC values than the HPN for data without overdispersion.

The results of the simulation study also show that there is an effect of cluster size and sample size. The bias and the MSE decrease when the cluster size increase and there is a slight decrease of the bias and the MSE when the sample size increases. To investigate the robustness of the simulation study, three different true values for $\boldsymbol{\beta}$ were chosen. The results obtained were similar under these three different true values of $\boldsymbol{\beta}$ which shows the robustness of the simulation study.

Most of our findings for the analysis of the epilepsy data set are in agreement with the findings reported in Molenberghs *et al.* (2007). In both studies, there was a difference in the estimates of the intercepts and also on the inference of the slopes using both models. The HPNOD model shows also that there is no significant change in the number of epileptic seizures over time for the patients who received the treatment while the HPN models does. This underscores the importance of careful extra-dispersion modeling. Further, both models produce non-significant values

Table 8: *Summary of the correlation between α and the parameters for $n = 60$ and $t = 10$ for the HPNOD model*

	$\alpha = 0.25$	$\alpha = 1$	$\alpha = 25$	$\alpha \to \infty$
β_{00}	-0.58953	-0.8739	-0.96311	-0.97212
β_{01}	-0.55579	-0.53827	-0.94476	-0.94223
β_{10}	-0.08627	-0.0351	0.067976	0.005266
β_{11}	0.103553	-0.24591	0.005057	0.230678
σ	0.103553	-0.24591	0.005057	0.230678
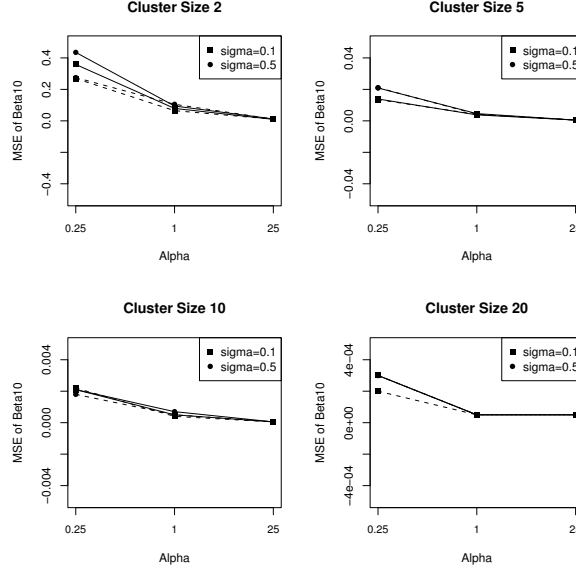
Figure 6: *Simulation Study. Comparison of the MSE of β_{10} for different cluster sizes using the HPN model (solid line) and HPNOD (dashed line). The x-axis represents the value of α which shows the overdispersion amount and the y-axis represents the MSE. $n = 60$*

for the difference and ratio in slopes. However, the study done by Molenberghs *et al.* (2007) shows that there is significant difference in the slopes using the HPN. In both studies, the HPNOD model fits better than the HPN model. Note that our findings are different form the ones reported in Thall and Vail (1990) and in Lindsey (1993). This should not come as a surprise, because these authors consider a different set of data, studying different compounds.

To conclude, the HPNOD model performs better than the HPN model for data featuring high, moderate and low overdispersion level. However, both models perform similarly for data without overdispersion. Using the HPN model, the bias and MSE of all parameters increases when the overdispersion level increases. The HPN model results in bias and inefficient estimates for all parameters, especially for σ and for data with high overdispersion ($0 < \alpha <= 0.25$). This may be due to the excess variability resulting from overdispersion not taken into account with the HPN model. This underscores that we should accommodate the extra-model variability. Further investigation is needed to answer the question why the HPNOD model is providing unbiased estimate of the intercepts when the data are generated with moderate overdispersion level but not when there is high overdispersion, low overdispersion, and no overdispersion.
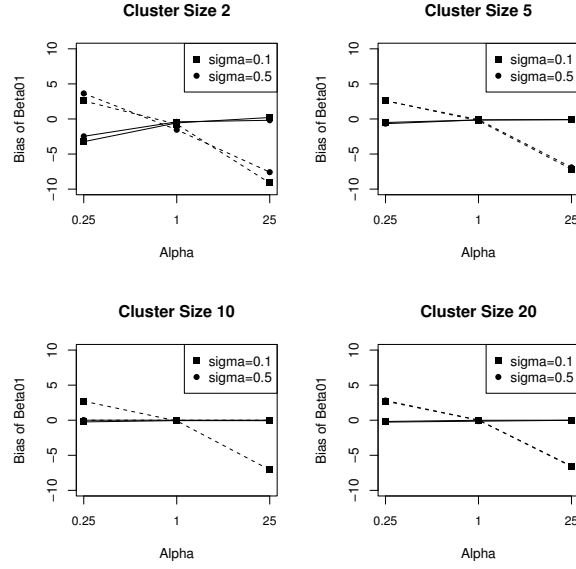
Figure 7: *Simulation Study. Comparison of the bias of β_{01} for different cluster sizes using the HPN model (solid line) and HPNOD (dashed line). The x-axis represents the value of α which shows the overdispersion amount and the y-axis represents the Bias. $n = 60$*

Supplementary Materials

The web-based Supplementary Materials presents the SAS programs for both the epileptic patients data analysis as well as for the onychomycosis trial.

Acknowledgment

References

Booth, J.G., Casella, G., Friedl, H., and Hobert, J.P. (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, **3**, 179–181.

Breslow, N. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, **33**, 38–44.

Breslow, N. (1990). Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models. *Journal of the American Statistical Association*, **85**, 565–571.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
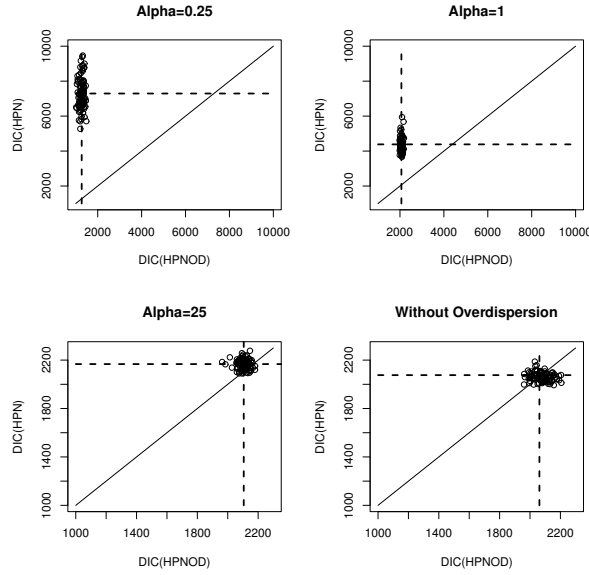
Figure 8: *Simulation Study. Comparison of the DIC of HPN and HPNOD for 100 dataset that are generated from $\beta_{00} = 2$, $\beta_{01} = -2$, $\beta_{10} = 0.05$, $\beta_{11} = 0.2$, and $\sigma = 0.1$. The x-axis represents the value of DIC for the HPNOD model and the y-axis represents the DIC for HPN and the dashed vertical and horizontal line shows the mean of the DIC for the 100 data sets.*

Cox, D.R. (1983). Some remarks on overdispersion. *Biometrika*, **70**, 269–74.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

Faught, E., Wilder, B.J., Ramsay, R.E., Reife, R.A., Kramer, L.D., Pledger, G.W., and Karim, R.M. (1996). Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. *Neurology*, **46**, 1684–1690.

Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, **3**, 515-533.

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457-511.

Gelman, A., Carlin, J.B., Stern, H.S, and Rubin, D.B. (2004). *Bayesian Data Analysis.* New York: Chapman and Hall.

Hinde, J. and Demétrio, C.G.B. (1998a). Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.

Hinde, J. and Demétrio, C.G.B. (1998b). *Overdispersion: Models and Estimation.* São Paulo: XIII Sinape.

Iddi, S., and Molenberghs, G. (2012). A combined overdispersed and marginalized multilevel model. *Computational Statistics and Data Analysis*, **56**, 1944–1951.

Lawless, J. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, **15**, 209–225.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lindsey, J.K. (1993). *Models for Repeated Measures.* Oxford: Oxford University Press.

Margolin, B. H., Kaplan, N., and Zeiger, E. (1981) Statistical Analysis of the Ames Salmonella Microsome Test.*Proceedings of the National Academy of Sciences*, **76**, 3779–3783.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* London: Chapman & Hall.

Molenberghs, G., Verbeke, G., and Demétrio, C.G. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–31

Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

Moore, D.F. (1986). Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika*, **73**, 583–588.

Pocock, S.J., Cook, D.G. and Beresford, S.A. (1981). Regression of area mortality rates on explanatory variables: what weighting is appropriate? *Applied Statistics*, **30**, 286–295.

Pryseley, A., Tchonlafi, C., Verbeke, G., and Molenberghs, G. (2011). Estimating negative variance components from Gaussian and non-Gaussian data: a mixed models approach. *Computational Statistics and Data Analysis*, **55**, 1071–1085.

Thall, P.F. and Vail, S.C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–671.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, **15**, 358-372.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss- Newton method. *Biometrika*, **61**, 439–47.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, **30**, 144-8.

Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika*, **75**, 621–9.