Statistical Modelling 2018; **xx(x)**: 1–21

A Weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures

Martial Luyts[1], Geert Molenberghs[1], Geert Verbeke[1], Koen Matthijs[2], Eduardo E Ribeiro Jr[3], Clarice GB Demétrio[3] and John Hinde[4]
[1]Interuniversity Institute for Biostatistics and Statistical Bioinformatics, KU Leuven and Universiteit Hasselt, Leuven, Belgium.
[2]Family and Population Studies, KU Leuven, Leuven, Belgium.
[3]ESALQ, Piracicaba, University of São Paulo, São Paulo, Brazil.
[4]School of Mathematics, Statistics and Applied Mathematics, NUI Galway, Galway, Ireland.

Abstract: A Weibull-model-based approach is examined to handle under- and overdispersed discrete data in a hierarchical framework. This methodology was first introduced by Nakagawa and Osaki (1975, *IEEE Transactions on Reliability*, **24**, 300–301), and later examined for under- and overdispersion by Klakattawi et al. (2018, *Entropy*, **20**, 142) in the univariate case. Extensions to hierarchical approaches with under- and overdispersion were left unnoted, even though they can be obtained in a simple manner. This is of particular interest when analysing clustered/longitudinal data structures, where the underlying correlation structure is often more complex compared to cross-sectional studies. In this article, a random-effects extension of the Weibull-count model is proposed and applied to two motivating case studies, originating from the clinical and sociological research fields. A goodness-of-fit evaluation of the model is provided through a comparison of some well-known count models, that is, the negative binomial, Conway–Maxwell–Poisson and double Poisson models. Empirical results show that the proposed extension flexibly fits the data, more specifically, for heavy-tailed, zero-inflated, overdispersed and correlated count data. Discrete left-skewed time-to-event data structures are also flexibly modelled using the approach, with the ability to derive direct interpretations on the median scale, provided the complementary log–log link is used. Finally, a large simulated set of data is created to examine other characteristics such as computational ease and orthogonality properties of the model, with the conclusion that the approach behaves best for highly overdispersed cases.

Key words: longitudinal profiles, clustering, dispersion, random effects, weibull-count approach

Received October 2017; **revised** April 2018; **accepted** June 2018

1 Introduction

The analysis of count data has received considerable attention in the literature, with practical applications in public health, and social and behavioural sciences. Since the

Address for correspondence: Martial Luyts, L-Biostat, KU Leuven, Kapucijnenvoer 35, B-3000 Leuven, Belgium.
E-mail: martial.luyts@kuleuven.be

 10.1177/1471082X18789992

introduction of generalized linear models (GLM's) by Nelder and Wedderburn (1972), a GLM based on the Poisson distribution, a well-known member of the exponential family, is a commonly applied statistical model for count data analysis. In spite of its many advantages, for example, the ability of fitting skewed non-negative data, the model possesses a too restricted mean–variance relationship (equidispersion), a characteristic that is often violated in the data. In particular, two situations can occur, (a) the variability in the data is larger than the theoretical variance implied by the model (overdispersion), and (b) the variability in the data is smaller than the theoretical variance (underdispersion). For these and other reasons, for example, zero-inflation (Iddi and Molenberghs, 2013) and heavy-tailed profiles (Zhu and Joe, 2009), many alternative and extended models have been proposed in the literature.

These models can often be classified as exponential dispersion models (EDM's), introduced by Jørgensen (1987), which include the GLM families as a special case. More specifically, EDM's increase the range of univariate/multivariate variance functions for which generalized linear type models exist. Kokonendji et al. (2004), for example, investigated two classes of EDM's for count data that is overdispersed compared to the Poisson distribution, that is, the Poisson–Tweedie and Hinde–Demétrio classes. Efron (1986), on the other hand, proposed a different class of regression families, by introducing a second parameter in the exponential family that controls the dispersion independently of the mean while still carrying out the usual regression analysis in a GLM context. These are the so-called double-exponential families because they enjoy exponential family properties simultaneously for the mean and dispersion parameters. A popular member is the double Poisson (DP) model (Appendix A4). A general overview of some popular models is given in Appendix A for subsequent comparison (Section 4).

While most of these models find their origin back in the Poisson GLM framework, alternative approaches for modelling count data based on time-to-event distributions have recently been developed. These approaches are mainly built upon the direct relationship between the Poisson and exponential distributions (Cooper, 2005). Zeviani et al. (2014), for example, focused on a discrete version of the Gamma distribution to model count data by following the two-step approach of Winkelmann (1995): (a) define the Poisson process as a sequence of iid exponentially distributed waiting times (Cox, 1962); and (b) replace the exponential distribution with a less-restrictive (extended) non-negative distribution such as the Gamma distribution. For the Weibull distribution, Morais and Barreto-Souza (2011) constructed count versions, that is, the generalized Weibull power series (GWPS) class of distributions. Another, simple discrete approach based on the Weibull distribution, is that of Nakagawa and Osaki (1975). In particular, Klakattawi et al. (2018) recently pointed out that the corresponding regression model can model over- and underdispersed count data. Moreover, they showed that the model is able to adequately fit highly skewed count data with excessive zeros, without the need for introducing zero-inflated or hurdle components, in contrast to other existing methods, for example, the zero-inflated Conway–Maxwell–Poisson (ZICOM) model (Sellers and Raim, 2016). A further generalization of this approach was introduced by Nekoukhou and Bidram (2015), where the exponentiated discrete Weibull (EDW) distribution is defined.

Apart from the presence of extra-dispersion, extended structures such as longitudinally collected data, where subjects/patients are repeatedly measured over time, and hierarchical structures, originating from hierarchical designs such as multicentre trials, can also be present. For the GLM framework, the generalized linear mixed model (GLMM), discussed by Engel and Keen (1994), Breslow and Clayton (1993) and Wolfinger and O'Connell (1993), has been suggested, and became a popular framework for taking into account hierarchical data structures. In these models, random effects are introduced to capture the association structure and to some extent dispersion. Molenberghs et al. (2007) extended this approach by introducing the so-called combined modelling (CM) framework, that was mainly developed to encompass both aspects: (a) overdispersion and (b) hierarchical/longitudinal structures, simultaneously, by adding an extra random effect into the GLMM framework.

In this article, we examine the existing (univariate) discrete Weibull-based approach of Nakagawa and Osaki (1975), and extend it with random effects to accommodate more complex data structures. This approach assumes that extra dispersion is captured in the pre-specified distribution, and differs from that in Molenberghs et al. (2007) where the extra dispersion is captured by an additional random effect. In addition, various settings (heavy tails, zero-inflation) in combination with dispersion and correlation are examined, and compared with other well-known count models (Appendix A). Conclusions are supported with some characteristics of the model.

The remainder of our article is organized as follows. In Section 2, two motivating case studies are presented, stemming from patients with epileptic seizures, and historic data on household members from a Belgian town. An overview of the discrete Weibull version of Nakagawa and Osaki (1975) is sketched in Section 3, alongside its extended version and characteristics. Section 4 is devoted to the analysis of our case studies, where a comparison is made between this approach and other count models (Appendix A). A simulation study is reported in Section 5 to investigate other characteristics of the framework, and concluding remarks are given in Section 6.

2 Case studies

2.1 Epilepsy dataset

The epilepsy dataset comes from a randomized, double-blinded, parallel group multi-centre study aimed at comparing placebo with a new anti-epileptic drug (AED), in combination with one or two other AEDs. In total, 45 patients were assigned to the placebo group, and 44 to the active (new) treatment group. Patients were then followed for several weeks—during which the number of epileptic seizures experienced in the last week—were counted, that is, since the last time the outcome was measured. The main research question is whether or not the new treatment reduces the number of epileptic seizures. A full description of the epilepsy dataset is provided in Faught et al. (2011). Figure 1 (top) shows the individual profiles with corresponding mean and median profiles of the seizure counts for every study week,

and Figure 1 (bottom) shows the observed mean and variance of the seizure counts per patient ID, categorized for both treatment groups. The figure shows highly variable longitudinal count data with the presence of extreme values, zero-inflation and very few observations available at some of the time-points, especially past week 20.

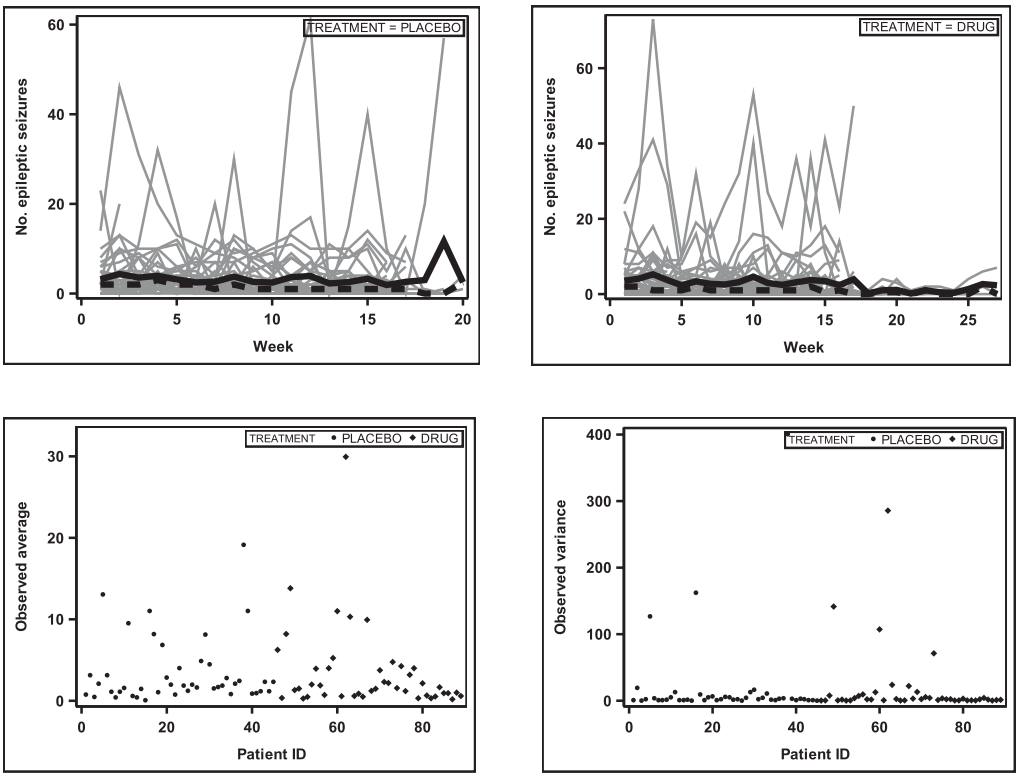


Figure 1 Epilepsy data (Faught et al., 2011). Subject-specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the number of epileptic attacks for every visit (top), and observed mean and variance of the seizure counts per patient ID (bottom), categorized for both treatments

2.2 Moerzeke dataset

The second dataset comes from a demographic, historical database of Moerzeke, a small village in the centre of Flanders (the Dutch-speaking part of Belgium) within the province of East Flanders. Information in the database is drawn from church and civil registers, which can be taken as high quality and appropriate for population studies, and includes all individuals who were born, married or died in Moerzeke.

In this study, a sample of 457 families was taken—by selecting all fathers born between 1750 and 1830, and then forming a family by also including their first born children and the children's mothers. To avoid overlap, children already selected are not included again, as either father or mother of new families. For the group under study, the mean age at death for those who were born and deceased in Moerzeke

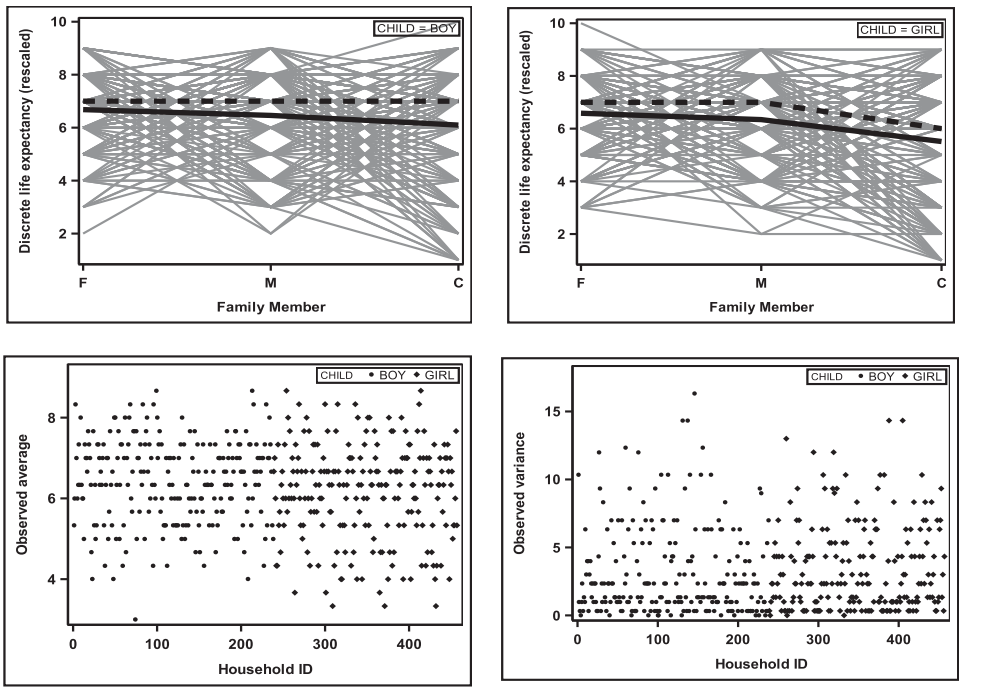


Figure 2 Moerzeke data. Household-specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the (discrete) life expectancy (rescaled) for every household member (top), and observed mean and variance of the (discrete) life expectancy (rescaled) per household ID (bottom), categorized for the gender of the first born child. The indexes F, M and C refer to father, mother and the first-born child, respectively

was 71.9 years for men and 71.7 for women, respectively. The main interest lies in the exploration of different social and/or household characteristics (e.g., gender of first born child) on the (discrete) life expectancy of family members. Figure 2 (top) shows the household profiles with corresponding average and median profiles of the (discrete) life expectancy, and Figure 2 (bottom) shows the observed mean and variance of the (discrete) life expectancy per household ID, categorized for the gender of the first-born child. On the average and median scales, a higher life expectancy of first-born male children is observed compared to first-born female children.

3 The Weibull-count approach

Due to the reproductive property of the Gamma distribution, that is, the sum of two Gamma distributed random variables again follows a Gamma distribution. Winkelmann (1995) pointed out that the Gamma distribution is a useful choice for his two-step approach. Unfortunately, this reproductive property does not hold for the Weibull distribution. As an alternative, the discrete approach of Nakagawa and Osaki (1975), which is here referred to as the discrete Weibull (DW) model, can be

used instead and gives a simple and adaptable alternative for the Weibull case. In what follows, we will give a general overview of the DW approach of Nakagawa and Osaki (1975).

Let Y_i, $i = 1, \ldots, n$, be (type 1) DW distributed (Nakagawa and Osaki, 1975) with parameters $0 < q < 1$ and $\rho > 0$. The probability mass function, cumulative distribution function and hazard function are given by

$$P(Y_i = y_i) = q^{y_i^\rho} - q^{(y_i+1)^\rho}, \qquad F(y_i) = 1 - q^{(y_i+1)^\rho},$$
$$h(y_i) = q^{y_i^\rho - (y_i+1)^\rho} - 1,$$

respectively. Special cases result from this. When $\rho = 1$ and $q = 1 - p$, the geometric distribution follows. Particularly, when $\rho = 1$ and $q = e^{-\lambda}$, the discrete exponential (DE) distribution results (Sato et al., 1999), which is overdispersed relative to the standard Poisson distribution (Appendix B). In addition, when $\rho = 2$ and $q = \theta$, the discrete Rayleigh (DR) distribution of Roy (2004) obtains. If $\rho \to +\infty$, the DW approaches a Bernoulli distribution with probability q. When q is small, an excessive zero case occurs (Klakattawi et al., 2018).

The mean and variance of the DW are given by

$$\mathrm{E}(Y_i) = \mu = \sum_{n=1}^{+\infty} q^{n^\rho},$$
$$\mathrm{Var}(Y_i) = 2 \cdot \sum_{n=1}^{+\infty} n \cdot q^{n^\rho} - \mu - \mu^2.$$

It can easily be shown that both of these infinite series converge (Appendix C). Based on the integral test, general approximations can be found consisting of incomplete gamma functions, for example, Englehardt and Li (2011).

To explore the characteristics of the DW model, we compute indexes for dispersion (DI), zero-inflation (ZI) and heavy-tail (HT), which are respectively given by

$$\mathrm{DI} = \frac{\mathrm{Var}(Y_i)}{\mathrm{E}(Y_i)}, \qquad \mathrm{ZI} = 1 + \frac{\log P(Y_i = 0)}{\mathrm{E}(Y_i)}, \qquad \mathrm{HT} = \frac{P(Y_i = y_i + 1)}{P(Y_i = y_i)}, \text{ for } y_i \to \infty.$$

Note that these indices are defined in relation to the Poisson distribution. Thus, the dispersion index indicates over-, under- and equidispersion for, respectively, $\mathrm{DI} > 1$, $\mathrm{DI} < 1$ and $\mathrm{DI} = 1$. The zero-inflation index indicates zero-inflation for $\mathrm{ZI} > 0$, zero-deflation for $\mathrm{ZI} < 0$ and no excess of zeros for $\mathrm{ZI} = 0$. Finally, the heavy-tail index indicates a heavy-tail distribution for $\mathrm{HT} \to 1$ when $y \to \infty$. Figure 3 shows that the DW framework is able to model highly overdispersed, zero-inflated and heavy-tailed data. The approach also allows the fit of low overdispersed, zero-deflated data and even some amount of underdispersion.

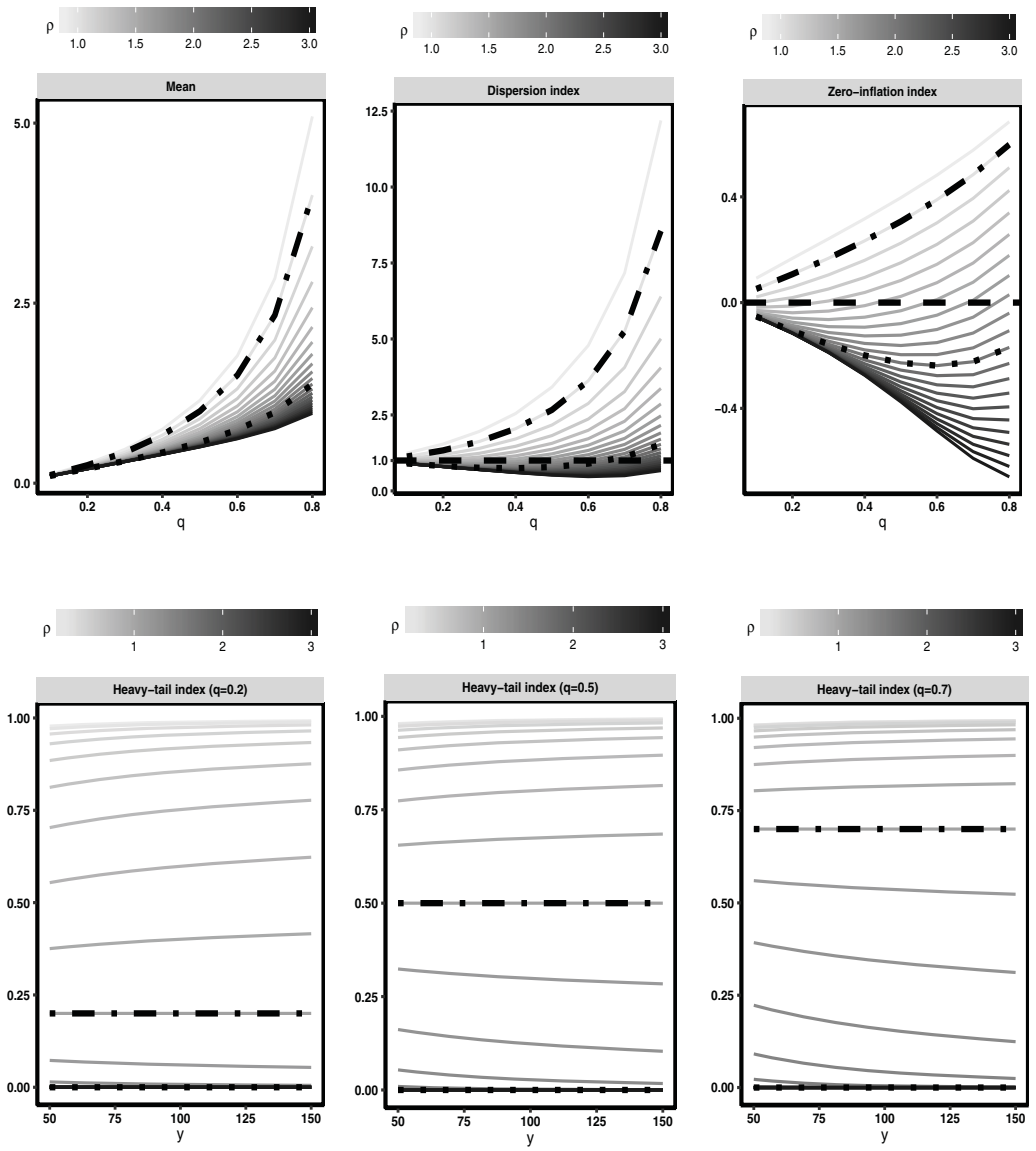


Figure 3 Characteristic indexes related to the Poisson distribution. Dashed, dot dashed and dotted lines represent the Poisson, DE and DR distribution, respectively

In a regression framework, Klakattawi et al. (2018) assumed that the response Y_i has a DW distribution, where a subject-specific parameter q_i is related to a p-dimensional vector of covariates $\mathbf{x}_i$ for ith observation through the complementary log–log link function

$$\ln[-\ln(q_i)] = \mathbf{x}_i^{'} \cdot \boldsymbol{\beta}$$
$$\Leftrightarrow$$
$$q_i = \mathrm{e}^{-\mathrm{e}^{\mathbf{x}_i^{'} \cdot \boldsymbol{\beta}}} (= \mathrm{e}^{-\lambda_i}).$$

Note that the complementary log–log link for q_i corresponds to a log link for λ_i. In addition, $\boldsymbol{\beta}$ represents the associated regression parameter vector, which can directly be interpreted in terms of the logarithm of the (closed-form) median. This is of particular interest when modelling, for example, highly skewed data, which often occurs in count data. Particularly, by splitting the regression parameters $\boldsymbol{\beta}$ into $\{\beta_0\} \cup \{\beta_l \mid l = 1, \ldots, p\}$, it can easily be shown, thanks to the use of the complementary log–log link function (Klakattawi et al., 2018), that $\{\ln[\ln(2)] - \beta_0\}/\rho$ is related to the conditional median when all covariates are set to zero, whereas $-\beta_l/\rho$ $(l = 1, \ldots, p)$ can be related to the change in the median of the response corresponding to a one unit change of $\mathbf{x}_{li}$, keeping all other covariates constant.

In terms of estimation procedures, Klakattawi et al. (2018) and Kulasekera (1994) used maximum likelihood for parameter estimation, while Haselimashhadi et al. (2017) proposed a Bayesian approach for estimating the parameters.

3.1 The extended hierarchical Weibull-count approach

If the discrete data are hierarchically structured, with Y_{ij} denoting the jth discrete outcome measured for cluster (subject) i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, univariate models are often not appropriate to take into account the underlying correlation structure of the data. Therefore, mixed-effects models are often proposed where, in addition to fixed effects, random effects are added to the model to allow for the correlation structure of the data. These approaches have been studied extensively in the GLM framework, for example, LMM and GLMM (Molenberghs and Verbeke, 2005), while little research has focused on dispersion models outside of this framework. In our context, where the focus is on the DW approach, a dispersion model extension with random effects can simply be achieved as follows:

$$\ln[-\ln(q_{ij})] = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\beta} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i,$$

where $\mathbf{z}_{ij}$ represents a q-dimensional vector of known covariate values corresponding to the q-dimensional random effects vector $\mathbf{b}_i$ following a multivariate normal distribution with mean vector $\mathbf{0}$ and variance–covariance matrix D.

In the following, we will analyse the epilepsy and Moerzeke datasets introduced in Section 2. Maximum likelihood principles are used to obtain parameter estimates. The SAS procedure NLMIXED is used for the computations (Appendix D).

4 Analysis of case studies

4.1 Epilepsy dataset

The epilepsy data of Section 2 will be analysed with the DW and its nested DE model (Section 3), and compared with some conventional models from Appendix A, that is, the classical Poisson log-linear (P), negative binomial (NB), Conway–Maxwell–Poisson (COM) and DP models. Previous work on this dataset was reported by Molenberghs and Verbeke (2005) and Molenberghs et al. (2007) in the context of generalized estimating equations (Liang and Zeger, 1986) and the CM framework, respectively.

Let Y_{ij} be the number of epileptic seizures that patient i experiences during week j of the follow-up period, and let t_{ij} be the time-point at which outcome Y_{ij} has been measured, that is, $t_{ij} = 1, 2, \ldots$, until at most 27. The following specific choice is made for the linear predictor:

$$\eta_{ij} = \beta_0 + b_i + \beta_0^{'} \cdot T_i + (\beta_1 + \beta_1^{'} \cdot T_i) \cdot t_{ij},$$

where $T_i = 1$ if patient i receives the treatment, and 0 for placebo. Here, $\beta_0^{'}$ and $\beta_1^{'}$ represent differences between treatment and placebo in terms of intercept and slope, respectively. The link functions are $\eta_{ij} = \exp(\lambda_{ij})$ for the P, DE, NB, COM and DP models, and $\eta_{ij} = \ln[-\ln(q_{ij})]$ for the DW model. The random intercept b_i is assumed to be normally distributed with mean 0 and variance σ^2, reflecting the between-patient variability within the data. Maximum likelihood estimates and corresponding standard errors of the parameters are reported in Table 1 (for the univariate case, that is, without the subject random effect) and Table 2 (for the clustered case, that is, with the subject random effect).

In the univariate case, that is, where clustering is ignored (Table 1), we clearly observe very large improvements in the DE, NB, DP and DW models, in terms of the likelihood, relative to the classical Poisson model. This, of course, is to be expected since the Poisson model assumes equidispersion while the parameters α, ϕ and ρ (see Appendix A for details) provide significant evidence of overdispersion. Furthermore, when a comparison is made between the DW and the conventional models, for example, NB and DP, w.r.t the Poisson model, we could consider the DW model as the better one in terms of log-likelihood. Indeed, similar to the NB approach (Appendix E1), the DW model is able to capture highly overdispersed, zero-inflated and heavy-tailed data (Figure 3), characteristics that are definitely present within the epilepsy dataset. Bar charts of the fitted univariate models are given in Figure 4.

Furthermore, we should mention that 'illegal' estimates were obtained for the COM model, implying that no valid conclusions can be made from it. Indeed, when looking at the fitted dispersion parameter τ, a negative estimate (-0.1188) is observed which is outside the parameter space (Appendix A3). This, of course, can easily be explained by the fact that the COM distribution limits itself in flexibility towards underdispersed data with narrow flexibility to zero-inflation (Appendix E2).

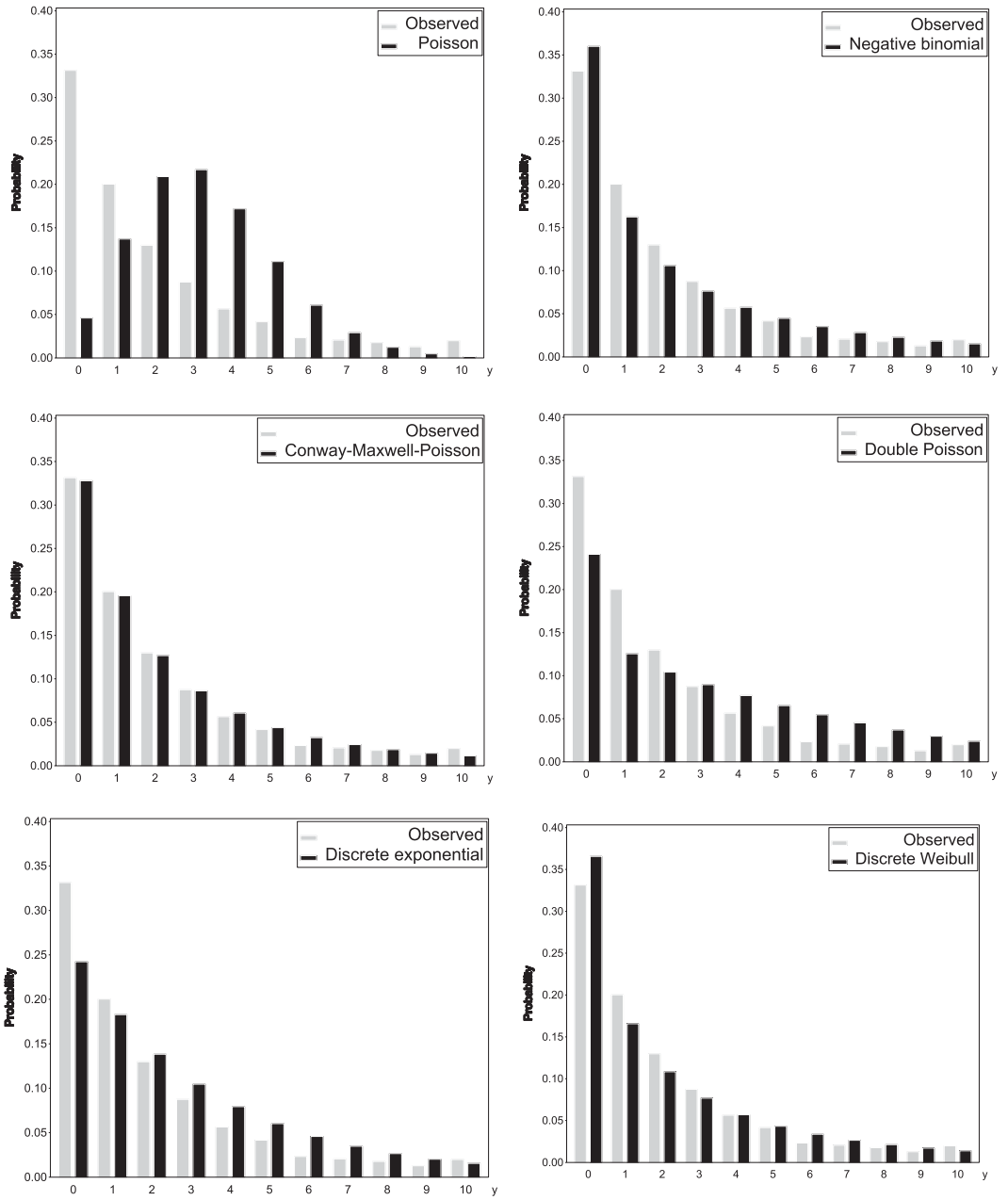


Figure 4 Bar charts of fitted univariate models

Table 1 Epilepsy dataset. Parameter estimates and standard errors for the (a) Poisson (P) model, (b) discrete exponential (DE) approach, (c) negative binomial (NB) model, (d) Conway–Maxwell–Poisson (COM) model, (e) double Poisson (DP) model and (f) the discrete Weibull (DW) model

Effect	Par.	**P** Est. (s.e.)	**DE** Est. (s.e.)	**NB** Est. (s.e.)
Intercept placebo	β_0	1.2662 (0.0424)	1.2601 (0.0864)	1.2594 (0.1119)
Difference in intercepts	β_0'	0.1869 (0.0571)	0.2115 (0.1202)	0.2156 (0.1564)
Slope placebo	β_1	−0.0134 (0.0043)	−0.0126 (0.0086)	−0.0126 (0.0111)
Difference in slopes	β_1'	−0.0195 (0.0058)	−0.0222 (0.0116)	−0.0227 (0.0150)
Ratio of slopes	$1+\frac{\beta_1'}{\beta_1}$	2.4576 (0.8480)	2.7586 (1.9721)	2.8081 (2.6066)
	α	–	–	1.8961 (0.0918)
	τ	–	–	–
	ϕ	–	–	–
	ρ	–	–	–
−2 loglik		11 590.0	6 502.5	6 326.1
AIC		11 598.0	6 510.5	6 336.1

Effect	Par.	**COM** Est. (s.e.)	**DP** Est. (s.e.)	**DW** Est. (s.e.)
Intercept placebo	β_0	−0.5054 (0.0189)	1.2662 (0.1054)	0.7341 (0.1002)
Difference in intercepts	β_0'	0.0131 (0.0144)	0.1869 (0.1421)	0.0936 (0.1307)
Slope placebo	β_1	−0.0011 (0.0012)	−0.0134 (0.0108)	−0.0174 (0.0095)
Difference in slopes	β_1'	−0.0017 (0.0017)	−0.0195 (0.0144)	−0.0143 (0.0127)
Ratio of slopes	$1+\frac{\beta_1'}{\beta_1}$	2.5663 (3.1297)	2.4576 (2.1093)	1.8189 (1.1027)
	α	–	–	–
	τ	−0.1188 (0.0051)	–	–
	ϕ	–	0.1616 (0.0061)	–
	ρ	–	–	0.7383 (0.0172)
−2 loglik		6 256.2	6 815.6	6 291.3
AIC		6 266.2	6 825.6	6 301.3

For the clustered case, that is, where a subject-specific random intercept is added to account for correlation (Table 2), we find that the DWN is considerably better in terms of likelihood. Moreover, point and precision estimates of such key parameters as the slope difference and the slope ratio are strongly affected when a random effect is added to the models. This remark was also made by Molenberghs et al. (2010), who noted an impact on hypothesis testing. Surprisingly, a valid interpretation on the extended COM approach can now be given, while this was not possible in the univariate case. To explain this phenomenon, attention should be directed towards the limited flexibility of COM in terms of overdispersion and the multiplicity effect of the random effects. In particular, a limited number of highly overdispersed regions can be modelled with the COM approach (Appendix E2). By adding a random effect to the model, extra flexibility has been given towards capturing overdispersed regions. Indeed, since random effects are mainly used to capture the underlying correlation structure of the data, they are also able to seize a certain amount of dispersion. Therefore, more flexibility has been gained with the inclusion of random effects towards the modelling of overdispersed data. In addition, a much lower parameter

Table 2 Epilepsy dataset. Parameter estimates and standard errors for the (a) Poisson-normal (PN) model, (b) discrete exponential-normal (DEN) approach, (c) combined (CM) model, (d) Conway–Maxwell–Poisson-normal (COMN) model, (e) double Poisson-normal (DPN) model and (f) the discrete Weibull-normal (DWN) model

		PN	**DEN**	**CM**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	0.8177 (0.1677)	0.9443 (0.1843)	0.9112 (0.1755)
Difference in intercepts	$\beta_0^{'}$	−0.1705 (0.2387)	−0.2670 (0.2620)	−0.2556 (0.2500)
Slope placebo	β_1	−0.0143 (0.0044)	−0.0271 (0.0101)	−0.0248 (0.0077)
Difference in slopes	$\beta_1^{'}$	0.0023 (0.0062)	0.0145 (0.0140)	0.0130 (0.0107)
Ratio of slopes	$1+\frac{\beta_1^{'}}{\beta_1}$	0.8398 (0.3979)	0.4663 (0.3953)	0.4751 (0.3345)
Std. dev. random effect	σ	1.0755 (0.0857)	1.0436 (0.0888)	1.0626 (0.0871)
	α	–	–	0.4059 (0.0348)
	τ	–	–	–
	ϕ	–	–	–
	ρ	–	–	–
−2 loglik		6 271.9	5 543.9	5 417.0
AIC		6 281.9	5 553.9	5 429.0
		COMN	**DPN**	**DWN**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	−0.2384 (0.0779)	0.8314 (0.1721)	1.4319 (0.2183)
Difference in intercepts	$\beta_0^{'}$	−0.0947 (0.1042)	−0.1582 (0.2451)	−0.2970 (0.3005)
Slope placebo	β_1	−0.0040 (0.0023)	−0.0146 (0.0067)	−0.0297 (0.0098)
Difference in slopes	$\beta_1^{'}$	0.0005 (0.0032)	0.0018 (0.0093)	0.0180 (0.0135)
Ratio of slopes	$1+\frac{\beta_1^{'}}{\beta_1}$	0.8646 (0.7451)	0.8778 (0.5980)	0.3947 (0.3382)
Std. dev. random effect	σ	0.4475 (0.0433)	1.0458 (0.0875)	1.2658 (0.1063)
	α	–	–	–
	τ	0.1563 (0.0196)	–	–
	ϕ	–	0.4355 (0.0169)	–
	ρ	–	–	1.3074 (0.0340)
−2 loglik		5 473.8	5 652.2	5 451.1
AIC		5 485.8	5 664.2	5 463.1

estimate for σ was obtained for the COMN case, compared to all other models. This directly results from the main disadvantage of the COM regression model, that is, its location parameter does not correspond to the expectation, which complicates the interpretation of regression models towards the mean specified using this distribution (Sellers and Shmueli, 2010). Even though the CM model is a more viable candidate in terms of likelihood (related to the Poisson model), one should be aware of the restricted mean scale interpretation in this framework, especially when dealing with skewed data. In this setting, right-skewed data (Figure 1) is observed, making the inferences less attractive from an interpretational point of view (similar to the DPN approach). The DW model avoids this problem by allowing inferences directly on the median scale (Section 3), making the approach more interesting here.

Finally, we expand our analysis with random slopes in the DWN model, that is, considering two random effects instead of a single one to reflect the between- and within-patient variability of the data. The linear predictor becomes:

Table 3 Epilepsy dataset. Parameter estimates and standard errors for the discrete Weibull-normal (DWN) model with (a) random intercept, (b) random slope with uncorrelated random effects (IND) and (c) random slope with correlated random effects (UN)

		Random intercept	**Random slope (IND)**	**Random slope (UN)**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	1.4319 (0.2183)	1.4973 (0.2183)	1.4947 (0.2287)
Difference in intercepts	$\beta_0^{'}$	−0.2970 (0.3005)	−0.2909 (0.2996)	−0.2984 (0.3150)
Slope placebo	β_1	−0.0297 (0.0098)	−0.0339 (0.0120)	−0.0327 (0.0126)
Difference in slopes	$\beta_1^{'}$	0.0180 (0.0135)	0.0169 (0.0168)	0.0167 (0.0176)
Ratio of slopes	$1+\frac{\beta_1^{'}}{\beta_1}$	0.3947 (0.3382)	0.5016 (0.3920)	0.4884 (0.4219)
Std. dev. random intercept	σ_1	1.2658 (0.1063)	1.2553 (0.1114)	1.3333 (0.1302)
Std. dev. random slope	σ_2	– (–)	0.0417 (0.0092)	0.0474 (0.0099)
Cov. between random effects	σ_{12}	– (–)	– (–)	−0.0177 (0.0142)
	ρ	1.3074 (0.0340)	1.3393 (0.0362)	1.3463 (0.0366)
−2 loglik		5 451.1	5 439.6	5 437.7
AIC		5 463.1	5 453.6	5 453.7

$$\eta_{ij} = \beta_0 + b_{1i} + \beta_0^{'} \cdot T_i + (\beta_1 + \beta_1^{'} \cdot T_i + b_{2i}) \cdot t_{ij},$$

where the random effects vector $\mathbf{b}_i = (b_{1i}, b_{2i})^{'}$ is assumed to be multivariate normally distributed with mean vector $\mathbf{0}$ and variance–covariance matrix

$$D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

A comparison with the random-intercept model will be made in two ways, that is, (a) a random-slopes model with uncorrelated random effects ($\sigma_{12} = 0$; IND) and (b) a random slopes model with correlated random effects ($\sigma_{12} \neq 0$; UN). Maximum likelihood estimates and corresponding standard errors of the parameters are reported in Table 3.

A significant improvement in likelihood is observed when adding a random slope to the model (likelihood ratio test $p = 0.0007$). However, there are no qualitative changes in the results of hypothesis testing for the main effects of interest. Furthermore, by comparing the independent random effects (IND) with correlated random effects (UN), no significant improvements were obtained (likelihood ratio test $p = 0.1692$). This extension at the same time illustrates the ease with which more than one random effect can be included.

4.2 Moerzeke dataset

While previous work on the Moerzeke data was provided by Matthijs et al. (2002) for the examination of historical mortality in terms of sociological and biological components, there has been no consideration of dispersion aspects. To this end, the

Table 4 Moerzeke dataset. Parameter estimates and standard errors for the (a) Poisson (P) model, (b) discrete exponential (DE) approach, (c) Conway–Maxwell–Poisson (COM) model, (d) double Poisson (DP) model and (e) discrete Weibull (DW) model

		P	**DE**	**COM**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0288)	1.7068 (0.0735)	3.0527 (0.1294)
Intercept father	$\beta_0^{'}$	1.8473 (0.0268)	1.8473 (0.0727)	3.2891 (0.1373)
Intercept mother	$\beta_0^{''}$	1.8847 (0.0263)	1.8847 (0.0725)	3.3522 (0.1395)
Gender effect on first born child	β_1	0.1009 (0.0390)	0.1009 (0.1014)	0.1697 (0.0509)
Gender effect on father	$\beta_1^{'}$	0.0187 (0.0370)	0.0187 (0.1007)	0.0316 (0.0481)
Gender effect on mother	$\beta_1^{''}$	0.0145 (0.0364)	0.0145 (0.1005)	0.0247 (0.0473)
	τ	–	–	1.7484 (0.0690)
	ϕ	–	–	–
	ρ	–	–	–
−2 loglik		5 834.3	7 985.1	5 669.3
AIC		5 846.3	7 997.1	5 683.3

		DP	**DW**
Effect	Par.	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0225)	8.9228 (0.2301)
Intercept father	$\beta_0^{'}$	1.8473 (0.0210)	9.0796 (0.2293)
Intercept mother	$\beta_0^{''}$	1.8847 (0.0206)	9.1660 (0.2301)
Gender effect on first born child	β_1	0.1009 (0.0305)	0.1699 (0.0957)
Gender effect on father	$\beta_1^{'}$	0.0187 (0.0290)	0.0831 (0.0955)
Gender effect on mother	$\beta_1^{''}$	0.0145 (0.0285)	0.0350 (0.0954)
	τ	–	–
	ϕ	1.6333 (0.0624)	–
	ρ	–	4.5377 (0.1055)
−2 loglik		5 693.3	5 512.3
AIC		5 707.3	5 526.3

DW and its nested DE models are considered in the analysis of the Moerzeke dataset (Section 2) and compared with the count models from Appendix A.

Let Y_{ij} represent the (discrete) life expectancy of the mother, father and first-born child ($j = 1, 2, 3$) in household $i = 1, \ldots, 457$. We assume the following predictor:

$$\eta_{ij} = \beta_0 \cdot I_{C\ddot{i}j} + \beta_0^{'} \cdot I_{M\ddot{i}j} + \beta_0^{''} \cdot I_{F\ddot{i}j} + b_i + (\beta_1 \cdot I_{C\ddot{i}j} + \beta_1^{'} \cdot I_{M\ddot{i}j} + \beta_1^{''} \cdot I_{F\ddot{i}j}) \cdot G_i,$$

where $I_{C\ddot{i}j}$, $I_{M\ddot{i}j}$ and $I_{F\ddot{i}j}$ are dummy variables for first-born child, mother and father, respectively, and G_i is the binary indicator for the gender of the first-born child, that is, 1 for male and 0 for female. Similar to the epilepsy analysis, the link functions are $\eta_{ij} = \exp(\lambda_{ij})$ for the P, DE, NB, COM and DP models, and $\eta_{ij} = \ln[-\ln(q_{ij})]$ for the DW model. The random intercept b_i is used to capture between-household variability, which here is assumed normally distributed with mean 0 and variance σ^2. Maximum likelihood estimates and corresponding standard errors of the parameters are reported in Table 4 (for the univariate case without the random effect) and Table 5 (for the clustered case, including the random effect).

Table 5 Moerzeke dataset. Parameter estimates and standard errors for the (a) Poisson-normal (PN) model, (b) discrete exponential-normal (DEN) approach, (c) Conway–Maxwell–Poisson-normal (COMN) model, (d) double Poisson-normal (DPN) model and (e) discrete Weibull-normal (DWN) model

		PN	**DEN**	**COMN**
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0288)	1.7068 (0.0735)	3.0529 (0.1294)
Intercept father	β_0'	1.8473 (0.0268)	1.8472 (0.0727)	3.2895 (0.1373)
Intercept mother	β_0''	1.8847 (0.0263)	1.8847 (0.0727)	3.3527 (0.1395)
Gender effect on first born child	β_1	0.1009 (0.0390)	0.1009 (0.1014)	0.1698 (0.0509)
Gender effect on father	β_1'	0.0187 (0.0370)	0.0187 (0.1007)	0.0317 (0.0481)
Gender effect on mother	β_1''	0.0145 (0.0364)	0.0145 (0.1005)	0.0245 (0.0473)
Std. dev. random effect	σ	$1.16E-4$ (0.0119)	$1.68E-4$ (0.0215)	$7.72E-4$ (0.1039)
	τ	–	–	1.7486 (0.0690)
	ϕ	–	–	–
	ρ	–	–	–
−2 loglik		5 834.3	7 985.1	5 669.3
AIC		5 848.3	7 999.1	5 685.3

		DPN	**DWN**
Effect	Par.	Est. (s.e.)	Est. (s.e.)
Intercept first born child	β_0	1.7068 (0.0225)	8.9228 (0.2301)
Intercept father	β_0'	1.8473 (0.0210)	9.0795 (0.2293)
Intercept mother	β_0''	1.8846 (0.0206)	9.1660 (0.2301)
Gender effect on first born child	β_1	0.1010 (0.0305)	0.1699 (0.0957)
Gender effect on father	β_1'	0.0187 (0.0290)	0.0831 (0.0955)
Gender effect on mother	β_1''	0.0145 (0.0285)	0.0350 (0.0954)
Std. dev. random effect	σ	$1.85E-4$ (0.0293)	$2.33E-4$ (0.0420)
	τ	–	–
	ϕ	1.6333 (0.0624)	–
	ρ	–	4.5376 (0.1055)
−2 loglik		5 693.3	5 512.3
AIC		5 709.3	5 528.3

In the univariate case (Table 4), the COM, DP and DW models significantly improved the model fit, compared to the classical Poisson model, while, in terms of likelihood, a worse fit is observed for the DE case. Indeed, when considering the dispersion parameters (τ, ϕ and ρ), we observe the clear presence of underdispersion within the data. While the COM, DP and DW models are able to capture this phenomenon (Figures 3, Appendix E1 and Appendix E2), this is not the case for the DE (Appendix B) and Poisson models. Therefore, it is fair to say that the DE model is completely wrong, not just in terms of underdispersion but also in the fact that it fails to capture the unimodal shape, as expected from a geometric distribution. The underdispersion result can be explained by the fact that Moerzeke has characteristics of a geographically isolated area as it is almost completely surrounded by a meander in the river Scheldt and by the river Durme. This was an important geographical limitation within the time bracket at which data were collected, and led to more genetic homogeneity than in the typical town. We observe that the DW model indicates the best fit, relative to the COM and DP models, in terms of likelihood

compared to the Poisson model. A possible reason for this result is the presence of left-skewed discrete time-to-event data, which can flexibly be modelled with the DW approach due to the underlying Weibull connection. Bar charts of the fitted univariate models are given in Figure 5.

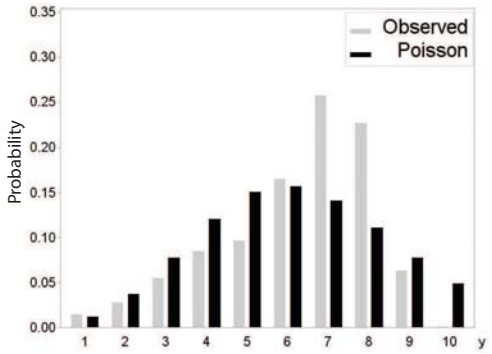
In the clustered case (Table 5), noteworthy results were obtained for the estimated variance component σ^2. In all clustered models, the estimated component is very close to 0, leaving the standard error estimates unchanged relative to the univariate cases. This phenomenon, while strange at first sight, is reasonably well understood in the literature. More specifically, partial-marginalization is used here, in agreement with Molenberghs et al. (2010), where adaptive Gaussian quadrature principles are used to approximate the marginal likelihood obtained from integrating over the normal random effects. This automatically adopts a hierarchical perspective, implying the restriction that no negative estimates of σ^2 can be achieved, even though this could be present for several reasons (e.g., negative intra-class correlation, underdispersion, etc.). Molenberghs and Verbeke (2011) and Verbeke and Molenberghs (2003), for example, discussed this phenomenon in the context of linear mixed models. Pryseley et al. (2011) extended this discussion to non-Gaussian outcomes, while Oliveira et al. (2017) illustrated how such negative variance components play a natural role in modelling both the correlation between repeated measures on the same experimental unit and over- or underdispersion from a CM perspective. While a zero variance component could in principle also point to the absence of correlation, this is not something one would expect in view of these data.

To conclude, we should mention that, even though the DW model fits the data quite well in the context of underdispersed data for the univariate case, there is still scope for further research in the context of underdispersed clustered data. Even though it is not our scope to fully encounter this problem here, boundary issues are suggested for the variance component. Also note that the random-effects variability is very different between the epilepsy and Moerzeke studies, underscoring that a large range of situations can be handled. Of course, this does not preclude further research towards underdispersion.

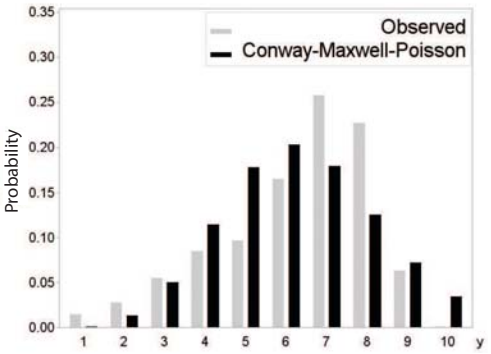
5 A large simulated set of data

To further explore the DW approach with dispersed count data, a large simulated set of data is obtained to examine the deviance surface under different dispersion situations. This highlights some other characteristics of the model such as the orthogonality and computational ease of estimating the parameters (q, ρ).

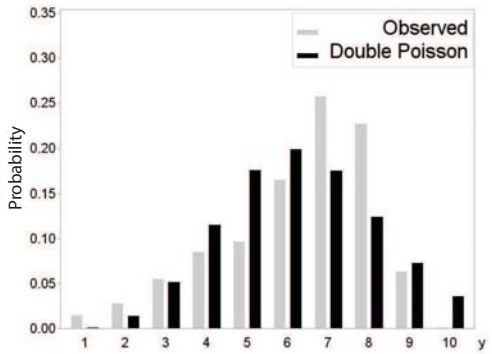
Figure 6 presents contour plots of the deviance surfaces for five different simulated DW data of size 1 000, with expectation fixed at 1 and dispersion indices at 0.25 (very strong underdispersion), 0.5 (strong underdispersion), 1 (equidispersion), 5 (strong overdispersion) and 10 (very strong overdispersion). As a result, the figure indicates that the parameters are highly intra-related in the likelihood function, consequently the maximum likelihood estimators for ρ and q are correlated. More specifically, a decreasing trend in the correlation seems to correspond with an increasing dispersion



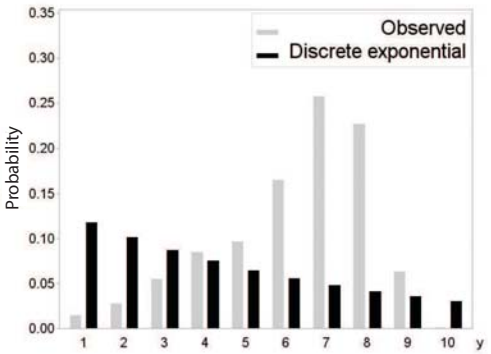
(a) Poisson model



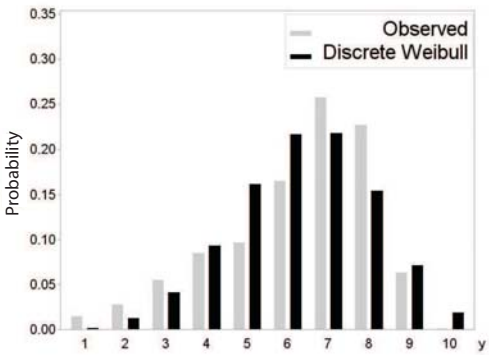
(b) Conway–Maxwell– Poisson model



(c) Double Poisson model



(d) Discrete exponential model



(e) Discrete Weibull model

Figure 5 Bar charts of fitted univariate models

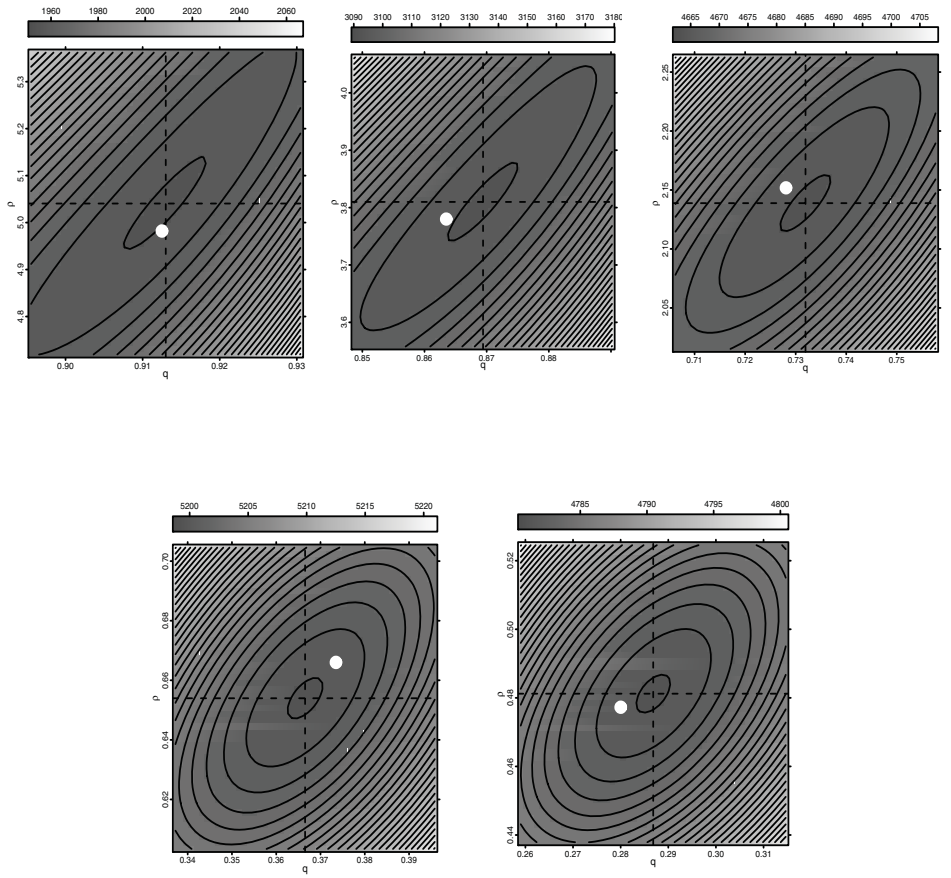


Figure 6 Deviance surfaces for discrete Weibull model fitted to five simulated data with expectation 1 and dispersion 0.25, 0.5, 1, 5 and 10. Dotted lines are the maximum likelihood estimates, and white points are the parameters used in the simulation

index. Based on the deviance surface, computational ease is combined with the ability to perform asymptotic (normally based) inferences in the regions with high dispersion, that is, $DI \to \infty$. Note that this is not a genuine simulation study; such will be the topic of future research.

6 Concluding remarks

Starting from an existing univariate framework, we have proposed an extended version that can handle both under- and overdispersed, and hierarchical data structures. In both case studies, we showed that the model fits the data well, for both under- and overdispersed situations. More specifically, the approach used is able to flexibly model highly overdispersed, zero-inflated, heavy-tailed and correlated data,

similar to the CM approach. In addition, the approach is capable of modelling some low overdispersed regions with zero-deflation (e.g., the DR approach for small values of q) and even underdispersed data, regions that cannot be captured within the CM framework. Due to the presence of a closed-form median expression, interpretations of the parameters can directly be related to the median profile, which is of particular interest when modelling skewed data. Finally, orthogonality properties are examined through a large simulated set of data. The resulting outcome indicates the presence of correlation between maximum likelihood estimators, related to the dispersion index.

Acknowledgements

We thank Mr R. Bijl, a Flemish genealogist and member of the V.V.F.-Dendermonde ('Vlaamse Vereniging voor Familiekunde', Dendermonde), for granting us access to his demographic data on Moerzeke.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. This work was partially supported by CNPq, a Brazilian science funding agency. The research leading to these results has also received funding from KU Leuven GOA project: 'New approaches to the social dynamics of long term fertility change'.

Supplementary material

For accessing the appendices, please visit http://www.statmod.org/smij/archive.html

References

Breslow NE (1984) Extra-Poisson variation in log-linear models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **33**, 38–44.

Breslow NE and Clayton DG (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Cameron AC and Trivedi PK (1986) Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, **1**, 29–53.

Conway RW and Maxwell WL (1962) A queuing model with state dependent service rates. *Journal of Industrial Engineering*, **12**, 132–36.

Cooper JCB (2005) A simple approach for the analysis of generalized linear mixed models. *Mathematical Spectrum*, **37**, 123–25.

Cox DR (1962) *Renewal Theory*. New York, NY: John Wiley & Sons.

Duchateau L and Janssen P (2007) *The Frailty Model*. New York, NY: Springer Science & Business Media.

Efron B (1986) Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, **81**, 709–21.

Engel B and Keen A (1994) A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, **48**, 1–22.

Englehardt JD and Li R (2011) The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water. *Risk Analysis*, **31**, 370–81.

Faught E, Wilder BJ, Ramsay RE, Reife RA, Kramer LD, Pledger, GW and Karim RM (2011) Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400- and 600-mg daily dosages. *Neurology*, **46**, 1684–90.

Haselimashhadi H, Vinciotti V and Yu K (2017) A novel Bayesian regression model for counts with an application to health data. *Journal of Applied Statistics*, **45**, 1–21.

Hilbe JM (2011) *Negative Binomial Regression*. Cambridge: Cambridge University Press.

Hinde J and Demétrio CGB (1998) Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*, **27**, 151–70.

Iddi S and Molenberghs G (2013) A marginalized model for zero-inflated, overdispersed and correlated count data. *Electronic Journal of Applied Statistical Analysis*, **6**, 149–65.

Jørgensen B (1987) Exponential dispersion models. *Journal of the Royal Statistical Society, Series B(Methodological)*, **49**, 127–62.

Kadane JB, Shmueli G, Minka TP, Borle S and Boatwright P (2006) Exponential dispersion models. *Bayesian Analysis*, **1**, 363–74.

Klakattawi HS, Vinciotti V and Yu K (2018) A simple and adaptive dispersion regression model for count data. *Entropy*, **20**, 142.

Knopp K (1951) *Theory and Application of Infinite Series*. Dover: Dover Publications.

Kokonendji CC, Dossou-Gbété S and Demétrio CGB (2004) Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *Statistics and Operations Research Transactions*, **28**, 201–13.

Kulasekera KB (1994) Approximate MLE's of the parameters of a discrete Weibull distribution with type 1 censored data. *Microelectronics Reliability*, **34**, 1185–88.

Lawless JF (1987) Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, **15**, 209–25.

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

Matthijs K, Van de Putte B and Vlietinck R (2002) The inheritance of longevity in a Flemish village (18th–20th century). *European Journal of Population/Revue europ´eenne de D´emographie*, **18**, 59–81.

Molenberghs G and Verbeke G (2005). *Models for Discrete Longitudinal Data*. New York, NY: Springer-Verlag.

——— (2011) A note on a hierarchical interpretation for negative variance components. *Statistical Modelling*, **11**, 389–408.

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

Molenberghs G, Verbeke G, Demétrio CGB and Vieira AMC (2010) A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–47.

Morais AL and Barreto-Souza W (2011) A compound class of Weibull and power series distributions. *Computational Statistics & Data Analysis*, **55**, 1410–25.

Nakagawa T and Osaki S (1975) The discrete Weibull distribution. *IEEE Transactions on Reliability*, **24**, 300–01.

Nekoukhou V and Bidram H (2015) The exponentiated discrete Weibull distribution. *Statistics and Operations Research Transactions*, **39**, 127–46.

Nelder JA and Wedderburn RWM (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, **135**, 370–84.

Oliveira IRC, Molenberghs G, Verbeke G, Demétrio CGB and Dias, CTS (2017) Negative variance components for non-negative hierarchical data with correlation, over- and/or underdispersion. *Journal of Applied Statistics*, **44**, 1047–63.

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 Statistical Data Analysis*, **55**, 1071–85.

Roy D (2004) Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, **53**, 255–60.

Sato H, Ikota M, Sugimoto A and Masuda H (1999) A new defect distribution metrology with a consistent discrete exponential formula and its applications. *IEEE Transactions on Semiconductor Manufacturing*, **12**, 409–18.

Sellers KF and Raim A (2016) A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis*, **99**, 68–80.

Sellers KF and Shmueli G (2010) A flexible regression model for count data. *The Annals of Applied Statistics*, **4**, 943–61.

Shmueli G, Minka TP, Kadane JB, Borle S and Boatwright P (2005) A useful distribution for fitting discrete data: Revival of the Conwa–Maxwell–Poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 127–42.

Verbeke G and Molenberghs G (2003) The use of score tests for inference on variance components. *Biometrics*, **59**, 254–62.

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

Winkelmann, R. (1995) Duration dependence and dispersion in count-data models. *Journal of Business & Economic Statistics*, **13**, 467–74.

——— (2008) *Econometric Analysis of Count Data*. New York, NY: Springer Science & Business Media.

Wolfinger R and O'Connell M (1993) Generalized linear mixed models a pseudo likelihood approach. *Journal of Statistical Computation & Simulation*, **48**, 233–43.

Zeviani WM, Ribeiro PJ, Bonat WH, Shimakura SE and Muniz JA (2014) The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics*, **41**, 2616–26.

Zhu R and Joe H (2009) Modelling heavy-tailed count data using a generalised Poisson-inverse Gaussian family. *Statistics & Probability Letters*, **79**, 1695–1703.

Zou Y, Geedipally SR and Lord D (2013) Evaluating the double Poisson generalized linear model. *Accident Analysis & Prevention*, **59**, 497–505.