1 Analysis of Over- and Underdispersed Data

Elizabeth Juarez-Colunga and C. B. Dean

1.1 Introduction

In the analysis of discrete data, for example, count data analyzed under a Poisson model, or binary data analyzed under a binomial model quite often the empirical variance exceeds the theoretical variance under the presumed model. This phenomenon is called *overdispersion*. If overdispersion is ignored, standard errors of parameter estimates will be underestimated, and therefore p-values for tests and hypotheses will be too small, leading to incorrectly declaring a predictor as significant when in fact it may not be.

The Poisson and binomial distributions are simple models but have strict assumptions. In particular, they assume a special mean-variance relationship since each of these distributions is determined by a single parameter. On the other hand, the normal distribution is determined by two parameters, the mean μ and variance σ^2, which characterize the location and the spread of the data around the mean. In both the Poisson and binomial distributions, the variance is fixed once the mean or the probability of success has been defined.

Hilbe [25] provides a very comprehensive discussion of what he calls *apparent overdispersion*, which refers to scenarios in which the data exhibit variation beyond what can be explained by the model and this lack of fit is due to several "fixable" reasons. These reasons may be omitting important predictors in the model, the presence of outliers, omitting important interactions as predictors, the need of a transformation for a predictor, and misspecifying the link function for relating the mean response to the predictors. Hilbe [25] also discusses how to recognize overdispersion, and how to adjust for it when it is present beyond apparent cases, and provides an excellent overall review of the topic.

It is important to note that if apparent overdispersion has been ruled out, in loglinear or logistic analyses, the point estimates of the covariate effects will be quite similar regardless of whether overdispersion is accounted for or not. Hence, treatment and other effects will not be aberrant or give a hint of the presence of overdispersion. As well, this suggests that adjusting for overdispersion can be handled through adjustments of variance estimates [35].

Evidence of apparent or real overdispersion exists when the Pearson or deviance residuals are too large [6]; the corresponding Pearson and deviance goodness-of-fit statistics indicate a poor fit. Several tests have been developed for overdispersion in the context of Poisson or binomial analyses [11, 12, 54], as well as in the context of zero-heavy data [30, 51, 53, 52].

1.2 Overdispersed Binomial and Count Models

1.2.1 Overdispersed Binomial Model

In the binomial context, overdispersion typically arises because the independence assumption is violated. This is commonly caused by clustering of responses; for instance, clinics or hospitals may induce a clustering effect due to differences in patient care strategies across institutions.

Let Y_i denote a binomial response for cluster i, $i = 1, \ldots, M$, which results in the sum of m_i binary outcomes Y_{ij}, that is, $Y_i = \sum_{j=1}^{m_i} Y_{ij}$, where j denotes individual j, $j = 1, \ldots, m_i$. If Y_{ij} are independent binary variables taking values 0 or 1 with probabilities $(1 - p_i)$ and p_i, respectively, then $\mathrm{E}(Y_i) = m_i p_i$ and $\mathrm{var}(Y_i) = m_i p_i(1 - p_i)$. If there exists correlation between two responses in any given cluster, with $\mathrm{corr}(Y_{ij}, Y_{ik}) = \psi > 0$, then

$$\begin{aligned} \mathrm{E}(Y_i) &= m_i p_i, \text{ and} \\ \mathrm{var}(Y_i) &= m_i p_i (1 - p_i)[1 + \psi(m_i - 1)], \end{aligned} \tag{1}$$

leading to overdispersion. Note that $\psi < 0$ leads to underdispersion. If we consider p_is as random variables with $\mathrm{E}(p_i) = \pi$ and $\mathrm{var}(p_i) = \psi\pi(1 - \pi)$, then the unconditional mean and variance also have the form of (1). And if we further assume that the p_i follow $\mathrm{Beta}(\alpha, \beta)$ distribution, the distribution of Y_i is the so called beta-binomial distribution, which has been studied extensively (see, for example, Hinde and Demétrio [26] and Molenberghs et al. [37]).

1.2.2 Overdispersed Poisson Model

Poisson and overdispersed Poisson data are examples of data from counting processes that arise when individuals experience repeated occurrence of events over time. Such data are known as recurrent event data (see, for example, Cook and Lawless [10] and Juarez-Colunga [29]). Consider M individuals each monitored for occurrence of events from a start time 0 through time τ_i, called the *termination time*, $i = 1, \ldots, M$. Let $\{N_i(t), t \geq 0\}$ be the right-continuous counting process that records the number of events for individual i over the interval $[0, t]$. The termination time is here assumed to be independent of the counting process $\{N_i(t), t \geq 0\}$. Let the intensity of the counting process be $\lambda_i(t|H(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{\Delta N_i(t) = 1 | H_i(t)\}}{\Delta t}$, where $H_i(t) = \{N_i(s) : 0 \leq s < t\}$ represents the history of the process up to time t. This intensity represents the instantaneous probability of occurrence of an event at time t. If the counting process is Poisson, given the memoryless property of the Poisson process, the intensity only depends on the history through t, $\lambda_i(t|H(t)) = \lambda_i(t)$, and the expected number of events over the entire follow-up can be written as $\mu_{i+} = \int_0^{\tau_i} \lambda(t)dt$. Let the total number of events in the entire followup be n_{i+} for individual i; then n_{i+} follows a Poisson distribution with mean $\mu_{i+} = \mathrm{E}(n_{i+}) = \mathrm{var}(n_{i+})$.

Two types of data are common in counting processes, and we will consider both here in the context of overdispersion: (*1*) individual i gives rise to n_{i+} event times recorded as $t_{i1} < t_{i2} < \cdots < t_{in_{i+}} < \tau_i$, and (*2*) only counts within specific follow-up times $0 = T_{i,0} < T_{i,1} < \ldots < T_{i,e_i} = \tau_i$ are available; these are called *panel* *counts* and are denoted $n_{ip} = N_i(T_{i,p}) - N_i(T_{i,p-1})$, $p = 1, 2 \cdots, e_i$, with the total aggregated count for individual i denoted

$$\sum_{p=1}^{e_i} n_{ip} = n_{i+}.$$

A simple way to incorporate overdispersion is through the use of an individual-specific random effect ν_i. Given ν_i, and the covariate vector $\boldsymbol{x}_i$ corresponding to the ith individual, the counting process $N_i(t)$ may be modeled as a Poisson process with intensity function

$$\lambda_i(t; \boldsymbol{x}_i) = \nu_i \rho(t; \boldsymbol{\alpha}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta}), \qquad (2)$$

where ρ is a twice-differentiable baseline intensity function, depending on the parameter $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ are the regression effects. We may take $\mathrm{E}(\nu_i) = 1$ without loss of generality, and let $\mathrm{var}(\nu_i) = \phi$. The function $\lambda(t; \boldsymbol{x})$ is now interpreted as a population average rate function among subjects with covariate vector $\boldsymbol{x}$, since $\mathrm{E}(dN(t)|\boldsymbol{x}) = \lambda(t; \boldsymbol{x})dt$. In addition to representing covariates unaccounted for, ν_i may also be a cluster effect, taking the same value for all individuals within the same cluster. This can be used to account for unknown clinic effects, for example, where individuals are patients clustered within clinics. When ν_i follows a gamma distribution, the marginal distribution of n_{i+} is negative binomial. The variance of the count of total aggregated events n_{i+} has the form $E(n_{i+}) + \phi E(n_{i+})^2_{i+}$.

Let the expected number of events over the entire follow-up $[0, \tau_i]$ be $\mu_{i+} = R_i \exp(\boldsymbol{x}_i' \boldsymbol{\beta})$, where $R_i = \int_0^{T_{e_i}} \rho(t; \boldsymbol{\alpha}) dt$ is called the cumulative baseline intensity function. Similarly, defining the cumulative baseline intensity function in panel period p as $R_{ip} = \int_{T_{i,p-1}}^{T_{i,p}} \rho(t; \boldsymbol{\alpha}) dt$, we have $\mu_{ip} = \mathrm{E}(n_{ip}) = R_{ip} \exp(\boldsymbol{x}_i' \boldsymbol{\beta})$.

The likelihood function based on continuous or panel follow-up can be expressed in the same framework as follows. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \phi)'$, and let ω_{ipl} be the time of the lth event, from the start of the study, for the ith individual in panel period p, $i = 1, \ldots, M$, $p = 1, \ldots, e_i$, $l = 1, \ldots, n_{ip}$. The likelihood based on either the full data, consisting of event times (subscripted by $d = f$), or the panel data (subscripted by $d = p$) factorizes as:

$$L_d(\boldsymbol{\theta}) = L_{\alpha,d}(\boldsymbol{\alpha}) L(\boldsymbol{\theta}), \quad d \in \{f, p\} \qquad (3)$$

where

$$L_{\alpha,f}(\boldsymbol{\alpha}) = \prod_{i=1}^{M} \prod_{p=1}^{e_i} \prod_{l=1}^{n_{ip}} \frac{\rho(\omega_{ipl}; \boldsymbol{\alpha})}{R_i}, \qquad (4)$$

and

$$L_{\alpha,p}(\boldsymbol{\alpha}) = \prod_{i=1}^{M} \left[\binom{n_{i+}}{n_{i1}, \ldots, n_{ie_i}} \prod_{p=1}^{e_i} \left(\frac{R_{ip}}{R_i} \right)^{n_{ip}} \right]; \qquad (5)$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{M} \int_0^{\infty} (\nu_i \mu_{i+})^{n_{i+}} \times e^{-\nu_i \mu_{i+}} (n_{i+}!)^{-1} G(\nu_i) d\nu_i \qquad (6)$$

is the likelihood for a mixed Poisson model based on the total counts observed for individual i. The likelihood $L(\boldsymbol{\theta})$ becomes the negative binomial if ν_i is gamma distributed (i.e., $G(\nu_i; .)$ is a gamma distribution). If there is a single panel, $L_p(\boldsymbol{\theta})$ [see Equation (3)] will reduce to the simple mixed Poisson kernel, $L(\boldsymbol{\theta})$, where the response is the total count of events in the entire follow-up time.

Overdispersed recurrent event counts are often encountered in trials where the main interest is to test whether certain treatments are effective in reducing the recurrences of events, as illustrated in the example Section 1.2.3. In this case, the βs are parametrized such that the treatment effects are measured relative to treatment 1, so that β_1 reflects the overall mean and α describes the shape of the intensity function $\rho(t, \alpha)$; common forms of $\rho(t, \alpha)$ are exponential ($\exp(\alpha t)$) and Weibull ($\alpha t^{\alpha-1}$).

1.2.3 Example

Consider a clinical trial, conducted by the Veterans Administration Co-operative Urological Research Group, that studied the effects of placebo pills, pyridoxine pills, and periodic instillation of thiotepa into the bladder on the frequency of recurrence of bladder cancer [8]. The data appear in Andrews and Herzberg [2]. All 116 patients had bladder cancer when they entered the study; the tumors were removed, and the patients were randomly assigned to one of the three treatments. Here we consider estimation of the treatment effect under both a design with continuous follow-up, as in the study, and an artificial design, for illustrative purposes, with 2 equally spaced scheduled follow-up visits over 64 months; for the panel design, we record information on event recurrences at the scheduled follow-up times and at termination times.

Table 1 reports parameter estimates and their standard errors of a Weibull baseline model for both Poisson and negative binomial analyses, under a 2-panel design as well as an analysis of the full data based on continuous follow-up. Based on both 2-panel and full data analyses there is substantial overdispersion in the data, with $\phi = 1.351$ in the analysis based on continuous follow-up. The estimate of the Weibull shape parameter α is quite close to unity, and the standard errors of the regression parameter estimates from the overdispersed model are significantly larger than those from the simple Poisson analyses. The latter leads to a significant protective effect of thiotepa treatment (β_3) based on the Poisson analysis, but not based on the overdispersed model.

1.3 Other Approaches to Account for Overdispersion

1.3.1 Generalized Linear Mixed Model

A general class of models that encompasses the incorporation of several random effects, not necessarily independent, is generalized linear mixed models. This may include an individual-specific random effect, as discussed above and also more complex structures that can accommodate dependencies in outcome variables as well as in random effects. A generalized linear mixed model specifies that

$$g(\boldsymbol{\mu}) = \boldsymbol{x}'_i\boldsymbol{\beta} + \boldsymbol{z}'_i\boldsymbol{\gamma} \qquad (7)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{x}_i$ are the mean of the response and the vector of covariates, corresponding to the ith individual, respectively; $\boldsymbol{z}_i$ is a vector of covariates determining the random effects structure, and the vector of random effects $\boldsymbol{\gamma}$ is distributed with a mean of zero and finite variance matrix; g is the link function. Conditional on $\boldsymbol{\gamma}$, the responses are assumed to have a distribution in the exponential family, for example, Poisson or binomial.

Maximum likelihood estimation involves q-dimensional integration, where q is the dimension of $\boldsymbol{\gamma}$; often random effects are assumed to be Gaussian. Tuerlinckx et al. [47] provide a review of methods used for estimation of generalized linear mixed models, discussing methods used to approximate the integral when integrating over the random effects distribution and methods that approximate the integrand of the marginal likelihood. Within the first set of methods, quadrature, Monte Carlo-based numerical methods, and expectation-maximization algorithms are reviewed; within the second, which approximate the integrand, Laplace's and quasi-likelihood methods are considered.

Table 1: Parameter estimates (Est) and their standard errors (SE), resulting from the Poisson and negative binomial (NB) likelihood fit to the bladder cancer data. The regression parameters $\beta_1, \beta_2, \beta_3$ correspond to the three treatment groups, parametrized with respect to the placebo, and α parametrizes the baseline intensity function.

	Full Data				2-Panel Data			
	Poisson		NB		Poisson		NB	
	Est	SE	Est	SE	Est	SE	Est	SE
β_1	-2.852	0.262	-2.955	0.318	-2.355	0.394	-2.483	0.474
β_2	0.008	0.170	0.132	0.332	0.016	0.170	0.114	0.328
β_3	-0.403	0.184	-0.282	0.323	-0.405	0.184	-0.299	0.320
α	0.996	0.066	1.019	0.069	0.858	0.106	0.883	0.123
ϕ			1.351	0.318			1.329	0.315

With overdispersion present, the use of the Poisson or binomial maximum likelihood equations for estimating the regression parameters in the mean is still valid. The usual likelihood equations obtained assuming a generalized linear model are unbiased, estimating equations regardless of any misspecification of the variance structure. Hence, an alternate approach to the use of generalized linear mixed models is to use the corresponding generalized linear model and adjust variance estimates. In this case, often as a final step, the variance is estimated by the sandwich estimator formula, which is an empirical estimator; this approach has become very popular in the last few decades [31, 46].

Nonparametric approaches for handling random effects have also been developed. Lindsay [32] provides a classic comprehensive source on the topic. More recently, Böhning and Seidel [3] provide a review of advances in estimation in mixture models, including nonparametric estimation, the EM algorithm, likelihood ratio tests for testing the number of components in the mixture, special mixtures such as zero-inflated Poisson models, multivariate mixtures, and testing and adjusting for heterogeneity. Groeneboom et al. [20] propose an algorithm, called the support reduction algorithm, to estimate M-estimators in mixture models through iterative unconstrained optimization. Wang [50] proposes three algorithms based on the constrained Newton method [49] to estimate semiparametric mixture models. In these, the mixture distribution G is left unspecified and a finite-dimensional parameter β is common to all mixture components. The three methods are based on (*1*) alternating estimation of parameters G and β, (*2*) profiling the likelihood, and (*3*) modifying the support set; they all use the constrained Newton method and an additional optimization algorithm for unconstrained problems.

There have been some efforts in combining models that account both for overdispersion and clustering effects, the latter perhaps arising from longitudinal measurements. Booth et al. [4] propose a negative binomial model to account for overdispersion, which incorporates random effects, in the linear predictor of the mean, to account for such clustering effects; numerical methods or the EM algorithm is proposed for estimation. Along the same lines, Molenberghs et al. [36] discuss a similar model with gamma and normal random effects to account for overdispersion and clustering effects and Molenberghs et al. [37] generalize the model to a family of generalized linear models for repeated measures with normal and conjugate random effects. Iddi and Molenberghs [27] discuss a marginalized model to account for overdispersion and longitudinal correlation.

Serial correlation may also be accommodated, in addition to overdispersion, through Gaussian time series [23]. Jowaheer and Sutradhar [28] use generalized estimating equations to account for autocorrelation structures as well as overdispersion in longitudinal counts. Parameters are estimated via a two-stage iterative procedure. Henderson and Shimakura [24] and Fiocco et al. [18] discuss a model that, conditional on a frailty, follows a Poisson distribution for counts of events and uses a gamma serially correlated process to model dependency between observations arising from the same individual. In this generalization of the individual frailty model, the random effects are first-order autocorrelated. Henderson and Shimakura [24] estimate the parameters of the model using a composite likelihood method based on pairs of time points, while Fiocco et al. [18] discuss an alternative approach using a two-stage procedure. In the two-stage procedure all parameters except the frailty correlation are estimated at the first stage while, in the second stage, the correlation of the frailties is estimated, based on pairs of observations.

1.3.2 Zero-Inflated Models

Sometimes apparent overdispersion is induced by the presence of another mode in the data, often at 0. In these cases, the remedy is to fit a model that handles the extra zeros that cannot be accounted for through the Poisson distribution [7, 40]. However, it may also occur that there is overdispersion beyond zero-inflation, in which case models accounting for both extra zeros and overdispersion have been developed, for example, the zero-inflated negative binomial [19]. There has been great interest in the last decade in accounting as well for correlation structures such as longitudinal, cluster, or spatial components. Ainsworth [1] provides a review of zero-inflated models, pointing out several references, mainly in the field of environmental statistics, that address such challenges in zero-heavy models. Hall [22] considers the challenges of simultaneously modeling within—and between—subject heterogeneity, while Dobbie and Welsh [14] consider serial correlation; both of these are framed in the context of zero-heavy count data models. Along the same lines, Wan and Chan [48] discuss a modeling approach based on a geometric process that accounts for overdispersion in zero-heavy models and, additionally, can handle serial correlation.

1.4 Underdispersion

Underdispersion is less common, but also found in count and binary data. Ridout and Besbeas [41] review methods for dealing with underdispersed counts, including (*1*) weighted Poisson models, in which weights are assigned to each probability density value [9,13]; (*2*) double Poisson models, in which the distribution has one more parameter θ than the Poisson and $\mathrm{E}(X) \approx \lambda$ and $\mathrm{var}(X) \approx \lambda/\theta$ [15]; (*3*) birth processes, which are generalizations of Poisson processes in which the birth rate at any time is a function of the number of events that have already occurred [16, 17]; [for example, Bosch and Ryan [5] propose a class of distributions $\lambda_k = \eta(k+1)^\delta$, where $\delta < 0$ corresponds to underdispersion, $\delta > 0$ to overdispersion, and $\delta = 0$ reduces to Poisson distribution]; (*4*) so-called COM-Poisson models, which are a generalization of the Poisson with one more parameter (ν) that allows it to represent under- and overdispersion with respect to Poisson [45, 44] [they can also be seen as a weighted Poisson with weights $(k!^{1-\nu})$]. Recently, Sellers et al. [43] provided a survey of the methods and applications related to the COM-Poisson models. Grunwald et al. [21] propose a birth-event process approach to model correlated over- or underdispersed data; this model can handle correlation due to clustering or serial correlation.

1.5 Software Notes

Software for incorporating overdispersion includes SAS [42], using, for instance, procedures LOGISTIC, GENMOD, GLIMMIX, and NLMIXED, and R [39] using, for example, packages glm, lmer, and lme4. Parametric mixture models can also be conducted in the MCMC framework using WinBUGS [34], OpenBUGS [33], JAGS [38], or the package mcmc in R.

References

[1] L. M. Ainsworth. *Models and Methods for Spatial Data: Detecting Outliers and Handling Zero-Inflated Counts.* PhD thesis, Simon Fraser University, 2007.

[2] D. F. Andrews and A. M. Herzberg. *Data: A Collection of Problems from Many Fields for the Student and Research Worker.* Springer-Verlag, New York, 2000.

[3] Dankmar Böhning and Wilfried Seidel. Editorial: recent developments in mixture models. *Computational Statistics & Data Analysis*, 41(3-4):349–357, January 2003.

[4] James G. Booth, George Casella, Herwig Friedl, and James P. Hobert. Negative binomial loglinear mixed models. *Statistical Modelling*, 3(3):179–191, October 2003.

[5] Ronald J. Bosch and Louise M. Ryan. Generalized poisson models arising from Markov processes. *Statistics & Probability Letters*, 39(3):205–212, August 1998.

[6] N. E. Breslow. Generalized linear models: checking assumptions and strengthening conclusions. *Statistica Applicata*, 8:23–41, 1996.

[7] Anne Buu, Runze Li, Xianming Tan, and Robert A. Zucker. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in Medicine*, 31(29):4074–4086, July 2012.

[8] D. Byar, C. Blackard, and the Veterans Administration Co-operative Urological Research Group. Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology*, 10:556–561, 1977.

[9] A. C. Cameron and P. Johansson. Count data regression using series expansions: with applications. *Journal of Applied Econometrics*, 12:203–223, 1997.

[10] R.J. Cook and J.F. Lawless. *The Statistical Analysis of Recurrent Events.* Springer, New York, 2007.

[11] C. Dean and J.F. Lawless. Tests for detecting overdispersion in Poisson regression models. *Journal of the American Statistical Association*, 84(406):467–472, June 1989.

[12] C. B. Dean. Testing for overdispersion in Poisson and binomial regression models. *Journal of the American Statistical Association*, 87(418):451–457, June 1992.

[13] Joan Del Castillo and Marta Pérez-Casany. Weighted Poisson distributions for overdispersion and underdispersion situations. *Annals of the Institute of Statistical Mathematics*, 50(3):567–585, 1998.

[14] Melissa J. Dobbie and A. H. Welsh. Modelling correlated zero-inflated count data. *Australian & New Zealand Journal of Statistics*, 43(4):431–444, December 2001.

[15] Bradley Efron. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721, September 1986.

[16] M. J. Faddy. Extended Poisson process modelling and analysis of count data. *Biometrical Journal*, 39(4):431–440, 1997.

[17] M. J. Faddy and R. J. Bosch. Likelihood-based modeling and analysis of data underdispersed relative to the Poisson distribution. *Biometrics*, 57(2):620–624, June 2001.

[18] M. Fiocco, H. Putter, and J. C. Van Houwelingen. A new serially correlated gamma-frailty process for longitudinal count data. *Biostatistics*, 10(2):245–257, April 2009.

[19] Aldo M. Garay, Elizabeth M. Hashimoto, Edwin M. M. Ortega, and Víctor H. Lachos. On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis*, 55(3):1304–1318, 2011.

[20] Piet Groeneboom, Geurt Jongbloed, and Jon A. Wellner. The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scandinavian Journal of Statistics*, 35(3):385–399, September 2008.

[21] Gary K. Grunwald, Stephanie L. Bruce, Luohua Jiang, Matthew Strand, and Nathan Rabinovitch. A statistical model for under- or overdispersed clustered and longitudinal count data. *Biometrical Journal*, 53(4):578–594, June 2011.

[22] D. B. Hall. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, 56(4):1030–1039, December 2000.

[23] J. L. Hay and A. N. Pettitt. Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics*, 2(4):433–444, December 2001.

[24] Robin Henderson and Silvia Shimakura. A serially correlated gamma frailty model for longitudinal count data. *Biometrika*, 90(2):355–366, June 2003.

[25] J. M. Hilbe. *Negative Binomial Regression*. Cambridge University Press, New York, 2nd edition, 2011.

[26] John Hinde and Clarice G. B. Demétrio. Overdispersion: Models and estimation. *Computational Statistics & Data Analysis*, 27(2):151–170, April 1998.

[27] S. Iddi and G. Molenberghs. A combined overdispersed and marginalized multilevel model. *Computational Statistics & Data Analysis*, 56(6):1944–1951, June 2012.

[28] Vandna Jowaheer and Brajendra C. Sutradhar. Analysing longitudinal count data with overdispersion. *Biometrika*, 89(2):389–399, June 2002.

[29] E. Juarez-Colunga. *Recurrent Event Studies: Efficient Panel Designs and Joint Modeling of Events and Severities*. PhD thesis, Simon Fraser University, 2011.

[30] Byoung Cheol Jung, Myoungshic Jhun, and Jae Won Lee. Bootstrap tests for overdispersion in a zero-inflated Poisson regression model. *Biometrics*, 61(2):626–628, June 2005.

[31] Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[32] Bruce G. Lindsay. Mixture models: theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, volume 5 of *Institute of Mathematical Statistics*, Hayward, 1995.

[33] David D. Lunn, David D. Spiegelhalter, Andrew A. Thomas, and Nicky N. Best. The BUGS project: Evolution, critique and future directions. *Audio, Transactions of the IRE Professional Group on*, 28(25):3049–3067, November 2009.

[34] David J. Lunn, Andrew Thomas, Nicky Best, and David Spiegelhalter. WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337, 2000.

[35] Peter McCullagh and James A. Nelder. *Generalized Linear Models*. Chapman Hall, London, 2nd edition, 1989.

[36] Geert Molenberghs, Geert Verbeke, and Clarice G. B. Demétrio. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, 13(4):513–531, December 2007.

[37] Geert Molenberghs, Geert Verbeke, Clarice G. B. Demétrio, and Afrânio M. C. Vieira. A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, 25(3):325–347, August 2010.

[38] Martyn Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, 2012.

[39] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[40] M. Ridout, C.G.B. Demetrio, and J. Hinde. Models for count data with many zeros. *Proceedings of the XIXth International Biometric Conference. Cape Town*, 1998.

[41] M. S. Ridout and P. Besbeas. An empirical model for underdispersed count data. *Statistical Modelling*, 4(1):77–89, April 2004.

[42] SAS Institute Inc. *SAS/STAT Software, Version 9.3*. Cary, NC, 2011.

[43] K. F. Sellers, S. Borle, and G. Shmueli. The COM-Poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 28:104–116, 2012.

[44] Kimberly F. Sellers and Galit Shmueli. A flexible regression model for count data. *Annals of Applied Statistics*, 4(2):943–961, November 2010.

[45] Galit Shmueli, Thomas P. Minka, Joseph B. Kadane, Sharad Borle, and Peter Boatwright. A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 54(1):127–142, January 2005.

[46] Brajendra C. Sutradhar. *Dynamic Mixed Models for Familial Longitudinal Data*. Springer, New York, 2011.

[47] Francis Tuerlinckx, Frank Rijmen, Geert Verbeke, and Paul De Boeck. Statistical inference in generalized linear mixed models: a review. *British Journal of Mathematical and Statistical Psychology*, 59(Pt 2):225–255, November 2006.

[48] Wai Yin Wan and Jennifer S.K. Chan. A new approach for handling longitudinal count data with zero-inflation and overdispersion: Poisson geometric process model. *Biometrical Journal*, 51(4):556–570, August 2009.

[49] Yong Wang. On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 69(2):185–198, January 2007.

[50] Yong Wang. Maximum likelihood computation for fitting semiparametric mixture models. *Statistics and Computing*, 20(1):75–86, March 2009.

[51] Liming Xiang, Andy H. Lee, Kelvin K. W. Yau, and Geoffrey J. McLachlan. A score test for overdispersion in zero-inflated Poisson mixed regression model. *Statistics in Medicine*, 26(7):1608–1622, 2007.

[52] F.-C. Xie, B.-C. Wei, and J.-G. Lin. Score tests for zero-inflated generalized Poisson mixed regression models. *Computational Statistics & Data Analysis*, 53(9):3478–3489, July 2009.

[53] Zhao Yang, James W Hardin, and Cheryl L. Addy. Testing overdispersion in the zero-inflated Poisson model. *Journal of Statistical Planning and Inference*, 139(9):3340–3353, September 2009.

[54] Zhao Z. Yang, James W. Hardin, Cheryl L. Addy, and Quang H. Vuong. Testing approaches for overdispersion in Poisson regression versus the generalized Poisson model. *Biometrical Journal*, 49(4):565–584, August 2007.