

A statistical model for under- or overdispersed clustered and longitudinal count data

Gary K. Grunwald^{*,1}, Stephanie L. Bruce², Luohua Jiang¹, Matthew Strand^{1,3}, and Nathan Rabinovitch⁴

¹ Department of Biostatistics and Informatics, University of Colorado Denver, USA

² Department of Mathematical Sciences, United States Air Force Academy, USA

³ Division of Biostatistics and Bioinformatics, National Jewish Health, USA

⁴ Department of Pediatrics, National Jewish Health, USA

Received 6 April 2010, revised 1 March 2011, accepted 7 March 2011

We propose a likelihood-based model for correlated count data that display under- or overdispersion within units (e.g. subjects). The model is capable of handling correlation due to clustering and/or serial correlation, in the presence of unbalanced, missing or unequally spaced data. A family of distributions based on birth-event processes is used to model within-subject underdispersion. A computational approach is given to overcome a parameterization difficulty with this family, and this allows use of common Markov Chain Monte Carlo software (e.g. WinBUGS) for estimation. Application of the model to daily counts of asthma inhaler use by children shows substantial within-subject underdispersion, between-subject heterogeneity and correlation due to both clustering of measurements within subjects and serial correlation of longitudinal measurements. The model provides a major improvement over Poisson longitudinal models, and diagnostics show that the model fits well.

Keywords: Generalized linear mixed models; Multi-level count models; Negative binomial longitudinal data; Poisson longitudinal data; Underdispersion.

Supporting Information for this article is available from the author or on the WWW under <http://dx.doi.org/10.1002/bimj.201000076>

1 Introduction

Count data occur frequently in biomedical studies and present challenges in analysis. Poisson Generalized Linear Models (GzLMs, e.g. McCullagh and Nelder, 1989; Cameron and Trivedi, 1998; Hardin and Hilbe, 2007; Winkelmann, 2008) provide a common approach, but real data often display features that make these models insufficient without extension. First, count outcomes often do not follow a Poisson distribution (conditional on covariates and other model terms), but rather are subject to over- or underdispersion relative to the Poisson distribution. Counts are overdispersed when the variance is greater than the mean, and underdispersed if the variance is less than the mean. Second, real count data are often correlated due to clustering and/or longitudinal measurement. Longitudinal studies are also subject to missing or unequally spaced data, which present an additional issue for serially correlated observations since correlation typically decreases with greater time separation. These complications require generalizations of the Poisson GzLM.

*Corresponding author: e-mail: gary.grunwald@ucdenver.edu, Phone: +1-303-724-4360, Fax: +1-303-724-4620

Overdispersion has been noted to be common in cross-sectional count data, often due to the presence of differences between subjects not captured by model covariates. The most common method for modeling overdispersion allows the mean of a Poisson distribution to be random. Common choices for the distribution of the random mean are Gamma or lognormal, as described in the above references. A Poisson distribution compounded with a Gamma-distributed random mean results in the Negative Binomial distribution, an advantage due to its closed form. This approach has been extended to further decompose components of variation in count data using generalized Waring distributions and Waring regression (e.g. Irwin, 1968; Xekalaki, 1983, 1984; Rodríguez-Avi et al., 2007, 2009), and also to allow different forms of covariate effects in Negative Binomial models (e.g. Greene, 2007). The methods and software `GAMLSS` (Rigby and Stasinopoulos, 2008, www.gamlss.com) have incorporated a number of distributions for overdispersed and zero-inflated counts, and also allow incorporation of covariates, smoothing and random effects for clustering, but they apparently do not model underdispersed counts, or allow for serial correlation. A Poisson distribution compounded with a lognormal distributed random mean does not have a closed-form distribution, but numerical methods are available and this approach allows inclusion of aspects of Gaussian models, for example normal autoregressive processes for serial correlation as described below.

Underdispersion is also seen in count data. A Poisson distribution with random mean does not provide a solution for underdispersed data, because this approach cannot remove variation from the Poisson. Ridout and Besbeas (2004) review common methods for modeling underdispersed counts. Methods include weighted Poisson distributions (Cameron and Johansson, 1997; Castillo and Pérez-Casany, 1998), Double Poisson distributions (Efron, 1986), Birth process distributions (Faddy, 1997) and COM-Poisson distributions (Conway and Maxwell, 1962, and more recently Shmueli et al., 2005 and Sellers and Shmueli, 2010).

Correlation due to clustering and/or longitudinal measurement raises further challenges. Clustering may be due to similarities among subjects (e.g. families) or, as in our example, to similarity of multiple measurements on a subject. Much has been written for Gaussian outcomes, where typical methods include multivariate models and mixed effects models (e.g. Diggle et al., 2002; Davis, 2002; Weiss, 2005; Hedeker and Gibbons, 2006). In the normal case, these approaches are equivalent in many situations, and are capable of modeling complex correlation structures. For non-normal outcomes such as counts, several approaches to modeling correlation have been proposed, some of which are mentioned in Section 3.1 below. Recently, several authors have published likelihood-based models for longitudinal count data having within-subject overdispersion that also accommodate subject heterogeneity (clustering), serial correlation and missing or unequally spaced data (Fahrmeir and Echavarria, 2006; Czado and Kolbe, 2007; Fotouhi, 2008). Overdispersion is modeled using the random Gamma or lognormal mean approach described above.

In longitudinal or clustered data, within-subject dispersion may differ from between-subject dispersion. Measurements within subjects in longitudinal studies may be subjected to repetitive, habitual or prescribed behaviors (e.g. medication use or food consumption). These similarities may give rise to underdispersion within subjects. Our example in Section 2 of asthma inhaler use by children shows such a case of longitudinal medication use, where accounting for the strong within-subject underdispersion is necessary for valid inference. There have been a few models proposed for underdispersed and correlated counts. Ridout and Besbeas (2004) used a weighted Poisson distribution to analyze several data sets including one longitudinal study. However, their models did not include serial correlation or subject heterogeneity. Toscas and Faddy (2003) incorporated underdispersion into Markov transition models (Zeger and Qaqish, 1988) for longitudinal count data and applied them to the well-known epileptic seizure data of Thall and Vail (1990). Their approach includes serial correlation but does not easily allow for subject heterogeneity, unequally spaced observations or missing data. In addition, estimates are interpreted conditionally on the previous day's count. Non-likelihood approaches have also been used, e.g. Generalized Estimating

Equations (GEE, Zeger and Liang, 1986; Liang and Zeger, 1986; Hardin and Hilbe, 2003). These methods can include under- and overdispersion and some forms of correlation and have robustness properties for estimation of covariate effects. However, direct modeling of the covariance process, probability distributions and simulations are not available from these methods because they do not provide a full likelihood model.

We were motivated by a study of the association of air pollution with daily asthma inhaler use by children during a school year. In Section 2 we present some example data. Daily inhaler use counts are small (≤ 6 in our data set), missing data are common, and simple plots show individual subjects' counts to be underdispersed. In Section 3 we propose a likelihood-based model that allows within-subject under- or overdispersion of counts, and also incorporates correlation due to clustering and/or serial correlation of longitudinal measurements. The model can accommodate missing and unequally spaced data. We also show how to estimate the model with Markov Chain Monte Carlo (MCMC) methods implemented in the standard WinBUGS software (Lunn et al., 2000). In Section 4 we analyze data from one year using several versions of the model. The results show that all of the abovementioned complications are present in the data, and that modeling within-subject underdispersion has a very substantial effect on the precision of the estimated association between air pollution and inhaler use compared with a more standard Poisson analysis. Diagnostics show that the model fits well. Section 5 gives some further discussion and suggested extensions.

2 Example

The Kunsberg School at National Jewish Health (NJH) in Denver, Colorado, is a unique day school program for children previously diagnosed with asthma by a physician and who require ongoing monitoring and medical assistance. Students ranged in age from six to thirteen years, and tend to be from lower socioeconomic backgrounds. Air pollution and health measurements were collected daily for several consecutive months during the school year. Approval for the study was obtained from the Colorado Multiple Institutional Review Board, and informed consent was obtained from all subjects. Ethical and scientific approval for the study was obtained from the National Jewish Health Institutional Review Board, and informed consent/assent was obtained from all subjects and guardians. The goal here is to model and estimate associations between daily air pollution and asthma symptoms.

Daily counts of asthma inhaler use (albuterol) were recorded electronically by a medication counter (Doser; Meditrak, Hudson, MA). Children were told to use the Doser as needed, and hence it can be considered a health outcome measure. At times they are also 'pre-treated' with the Doser, such as before exercise classes or recess at school, but this can be accounted for to some degrees as described in Section 4. We used albuterol inhaler use as measured by Doser counts during the school day (approximately from 8 am to 2:45 pm) when students attended school. Weekends, holidays and absences are treated as missing. We used data recorded during the 2002–2003 school year, 15 October 2002–22 May 2003. Our data set contained 48 children having at least 30 measurements during the year. Subjects analyzed had between 36 and 122 days of measurement, with only four subjects having fewer than 90 days. The total number of observations was 5209. Daily Doser counts are unrestricted, and in our data ranged from 0 to 6 uses (one use generally includes two puffs). Figure 1 shows daily counts for four subjects, with lines connecting measurement if they were made on consecutive days.

A graph of each child's variance versus their mean is shown in Fig. 2. The variance is always less than or equal to the mean, indicating within-subject underdispersion, likely due to habitual or prescribed patterns of inhaler use. The solid dots indicate the four subjects shown in Fig. 1, in the same order from left to right.

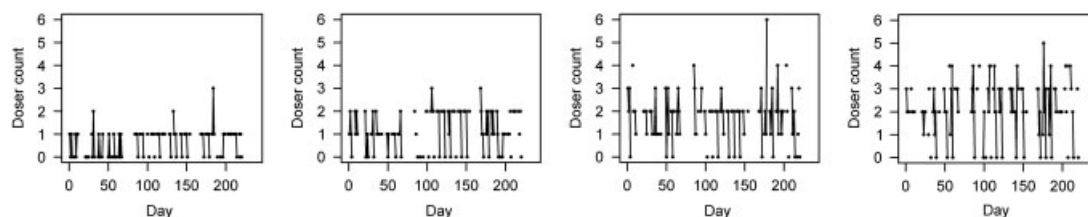


Figure 1 One year of inhaler use data for four children selected to have a range of mean counts.

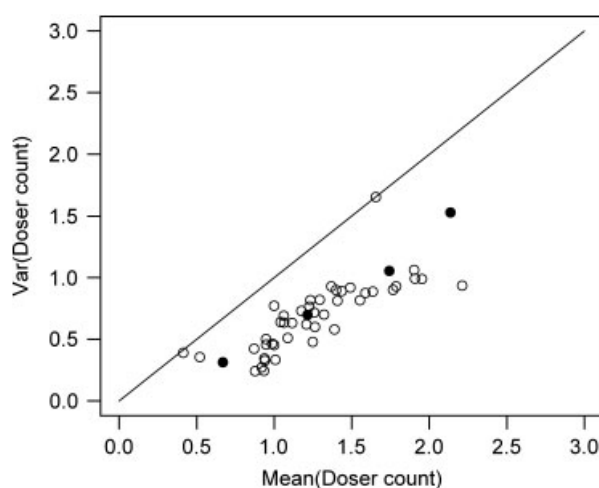


Figure 2 Variance versus mean of number of daily asthma inhaler usages during the school day for 48 children at NJH. The four solid points correspond to the subjects in Figure 1, in the same order left to right (e.g. the left most solid dot in Figure 2 corresponds to the left most panel in Figure 1). The line denotes variance equal to mean

3 Correlated count models and within-subject dispersion

3.1 Models for longitudinal count data

Models proposed for time series or longitudinal count data encompass a variety of approaches. Bockenholt (1999) generalized to the longitudinal case earlier time series work by McKenzie (1986) and Al-Osh and Alzaid (1987) based on birth-death or queueing models. Lambert (1996), Jorgensen et al. (1996), Henderson and Shimakura (2003) and Solis-Trapala and Farewell (2005) used latent Gamma processes for the mean of the Poisson distribution. Zeger and Qaqish (1988) formulated Markov transition models using functions of previous observations as regressors in a Generalized Linear Model (GzLM) to include serial correlation with non-normal outcomes. Breslow and Clayton (1993) and Wolfinger and O'Connell (1993) formulated Generalized Linear Mixed Models (GzLMM), a likelihood approach that adds random effects to the linear predictor in a GzLM. This approach has been developed more fully in the recent papers by Fahrmeir and Echavarria (2006), Xu et al. (2007), Molenberghs et al. (2007), Czado and Kolbe (2007) and Fotouhi (2008). In a somewhat different direction based on GzLM but not likelihood models, Liang and Zeger (1986) and Zeger and Liang (1986) proposed GEE, using marginal GzLMs with adjustments for correlation and over- or underdispersion. Each of these methods addresses some but not all of the

challenges mentioned above. Our approach described below is a likelihood model based on GzLMM.

While each of the approaches to longitudinal count analysis described above has appealing features, each also presents difficulties in generalizing to include all of the common data complexities mentioned in Section 1. The longitudinal count models most easily generalizable to include the complications typically seen in longitudinal count data seem to be those based on Poisson GzLMMs. A linear predictor containing observed covariates is exponentiated to give the mean of a Poisson observation. Random intercepts and a Gaussian AR(1) process can be included in the linear predictor to model subject heterogeneity and serial correlation, respectively. Overdispersion at the observation level can be included by adding a random observation-level error to the linear predictor. This will be included in the AR(1) process if it is present. Conditional on the linear predictor the observation is assumed to have a Poisson distribution. Missing or unequally spaced data can be included by using a continuous-time AR(1) process. Models like these have recently been given by Fahrmeir and Echavarria (2006), Xu et al. (2007), Czado and Kolbe (2007), and Fotouhi (2008).

Specifically, for the j -th observation Y_{ij} on subject i with covariate vector $\mathbf{x}_{ij}$, $j = 1, \dots, n_i$:

$$Y_{ij} | \mu_{ij} \sim \text{Poisson}(\mu_{ij}), \quad (1)$$

$$\mu_{ij} = \exp(\mathbf{x}_{ij}'\boldsymbol{\beta} + u_i + e_{ij}), \quad (2)$$

$$u_i \sim N(0, \sigma_u^2), \quad \text{independent for all } i, \text{ and independent of all } e_{ij}. \quad (3)$$

The random effect u_i describes subject heterogeneity. If the measurement level random errors e_{ij} are independent, they model overdispersion. If they are correlated (e.g. AR(1)) they also incorporate serial correlation.

Missing data are handled with a continuous-time AR(1) process for the measurement level errors e_{ij} that results in exponentially decreasing correlation between successive errors. For the j -th observation Y_{ij} on subject i at time t_{ij} , let $\delta_{ij} = (t_{ij} - t_{i,j-1})$ denote the time between observation $y_{i,j-1}$ and y_{ij} on subject i ($j = 2, 3, \dots, n_i$). Then following the continuous-time AR(1) formulation (e.g. Jones, 1993) independently for each subject

$$e_{ij} = \phi(\delta_{ij})e_{i,j-1} + \eta_{ij}, \quad \phi(\delta_{ij}) = \phi^{\delta_{ij}}, \quad (4)$$

$$\eta_{ij} \sim N(0, \sigma_e^2(\delta_{ij})), \quad \sigma_e^2(\delta_{ij}) = \sigma_e^2(1 - \phi^{2\delta_{ij}}). \quad (5)$$

The model is parameterized in terms of ϕ , the correlation between two errors one time unit apart, and σ_e^2 , the marginal variance of the continuous-time AR(1) process. When there is more time (e.g. missing data) between an observation and the previous one, the correlation is smaller and the error variance is larger. Each subject's process is initialized at t_{i1} by the stationary distribution, with $e_{i1} \sim N(0, \sigma_e^2)$.

3.2 Modeling dispersion

As described above, overdispersion relative to the Poisson distribution can be included in models for correlated counts using a random mean, for example lognormal or Gamma. This approach does not extend to underdispersion, because it only adds variation to the Poisson. We take a different approach to modeling dispersion, selecting a family of distributions in place of Poisson that allows both under- and overdispersion, with Poisson as a special case. Our interest is in the underdispersed case. This class of distributions, first given by Faddy (1997), includes a parameter to model a range of underdispersion and also provides a nice conceptual model for habitual event behavior. We note that there are other possible choices here, including the weighted Poisson (e.g. Ridout and Besbeas, 2004) and the COM-Poisson (e.g. Shmueli et al., 2005 and Sellers and Shmueli, 2010). We discuss these options further below.

The Faddy distribution hypothesizes a point process with an instantaneous rate parameter λ_k that is a function of the number k of previous events in the given time interval. The total number of events Y in the interval is the observed count outcome. Faddy and Smith (2005) have further extended this to a bivariate model for events and successes for binomial data.

Specifically, Faddy (1997) defined a rate parameter λ_k that is used to calculate the probabilities for counts, $k = 0, 1, 2, \dots$. The following form for dependence of the rate on the previous number of events is used:

$$\lambda_k = a(b+k)^c. \quad (6)$$

The parameter c describes dispersion. When $c = 0$, $\lambda_k = a$ is constant and the Poisson(a) distribution results; $c < 0$ gives underdispersion as the rate decreases with each new event, and $c > 0$ gives overdispersion as each new event speeds up the process; $c = 1$ gives the Negative Binomial distribution, and for theoretical reasons values of $c > 1$ are not in general considered applicable (Faddy, 1997). The parameter b ensures that the initial rate (when $k = 0$) is positive. Larger values of b correspond to starting the process with the same initial rate a , and dispersion c but after more initial events. Empirical investigations (Bruce, unpublished PhD dissertation, 2007) showed that in the range of our data the parameter b had little impact on the fit to individual subjects, and the likelihood as a function of b was very flat. We set $b = 1$ throughout our study, which makes $a > 0$ easily interpretable as the initial rate (when $k = 0$) in events per unit time. This still provides a wide range of underdispersed distributions. For example, for count distributions with mean equal to 1, as c ranges from -6 to 0 the variance ranges from 0.072 to 1 . The form (6) with $b = 1$ provides a simple model with parameters describing rate and dispersion; however, other forms are possible (e.g. Faddy and Bosch, 2001).

The probability distribution for event count Y in a unit time interval is obtained as follows. For some value K large enough to include all non-trivial probabilities for all values of a and c under consideration, the λ_k in (6) are used to define a $(K+1) \times (K+1)$ matrix

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ & & & -\lambda_{K-1} & \lambda_{K-1} \\ 0 & \cdots & & 0 & -\lambda_K \end{bmatrix}. \quad (7)$$

The vector of probabilities, $p_y = P(Y = y)$, is obtained as the solution to the Kolmogorov equations and results in the following expression involving the exponential of the matrix $\mathbf{Q}$ (e.g. Cox and Miller, 1965, Section 4.5):

$$\mathbf{p} = (p_0, p_1, \dots, p_K) = (1, 0, 0, \dots, 0) \exp(\mathbf{Q}). \quad (8)$$

Functions for computing the matrix exponential are available in most statistical software – we used `expm` from the `Matrix` library in R. We denote a count outcome Y following a Faddy distribution with mean μ and dispersion parameter c by $Y \sim Fd(\mu, c)$.

There is one difficulty in using these distributions in practice. The mean μ is not a simple closed-form function of the model parameters a , b and c in (6). Faddy (1997) gave an asymptotic approximation for μ based on diffusion theory, but in our experience with small counts it did not provide sufficient accuracy (Bruce, unpublished PhD dissertation, 2007). Even in cases when the approximation appeared quite accurate, the small errors when propagated over the long series on individual subjects (up to 122 days) resulted in substantial bias in results. Toscas and Faddy (2003) used a numerical search to find the value of a that gives a specified mean μ , for given fixed values of b and c . However, this needs to be done for each observation for each evaluation of the likelihood, which in iterative methods and large data sets becomes impractical. In Section 3.4 below we show

how this can be overcome in MCMC methods by tabling distributions over a sufficient range of parameter values. We note that other common options mentioned above for modeling under-dispersed counts, for example weighted Poisson and COM-Poisson, could be used in place of the Faddy model, but that they also do not have closed-form expressions for probabilities and moments and so would also be subject to these same difficulties.

3.3 A likelihood model for longitudinal counts with under- or overdispersion

We formulate a likelihood model including within-subject under- or overdispersion by replacing Eq. (1) with

$$Y_{ij} | \mu_{ij} \sim Fd(\mu_{ij}, c), \quad (9)$$

where this distribution is defined in Eqs. (6)–(8). Eqs. (2)–(5) remain the same to incorporate subject heterogeneity and serial correlation.

The likelihood for model (2)–(9) is calculated as a product over the independent subjects $i = 1, \dots, n$ by integrating out the unobservable random effects u_i and e_{ij} , $j = 1, \dots, n_i$. The term for the vector $\mathbf{y}_i$ of observations on subject i is

$$L(\mathbf{y}_i | \boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi) = \iint f(\mathbf{y}_i | u_i, \mathbf{e}_i, \boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi) g(u_i) h(\mathbf{e}_i) du_i d\mathbf{e}_i, \quad (10)$$

where $\mathbf{e}_i$ is the vector of autoregressive errors in (4)–(5) and g and h are defined through (3)–(5). f is the probability density function of the count outcome, defined by the Faddy distribution in (9), i.e. $f(\mathbf{y}_i | u_i, \mathbf{e}_i, \boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi) = \Pr(Y_i = \mathbf{y}_i | u_i, \mathbf{e}_i, \boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi)$.

When the dependence of observations is due to clustering or subject heterogeneity only and no serial correlation is present, the e_{ij} are independent within subjects and the likelihood for subject i factors into n_i integrals, which are in some cases numerically tractable using for example Gaussian quadrature. When there is serial correlation the dimension of the above integral is too high for such procedures. In the following section we describe estimation by MCMC methods.

3.4 Model estimation and inference using MCMC

Bayesian inference is based on the posterior distribution of the model parameters given the data,

$$p(\boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi | \mathbf{y}_i) \propto L(\mathbf{y}_i | \boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi) \cdot \pi(\boldsymbol{\beta}, c, \sigma_u^2, \sigma_e^2, \varphi)$$

with the likelihood as in (10). Vague priors were assumed for all parameters, in particular normal distributions with large standard deviations for the regression coefficients $\boldsymbol{\beta}$, Gamma distributions with large variances for precisions $\tau_u = 1/\sigma_u^2$ and $\tau_e = 1/\sigma_e^2$, and uniform priors bounded by possible ranges for parameters φ and c . All priors are assumed to be independent of each other, with $\tau_u \sim \text{Gamma}(0.001, 0.001)$, $\tau_e \sim \text{Gamma}(0.001, 0.001)$, $\varphi \sim \text{Unif}(-1, 1)$, $c \sim \text{Unif}(-4, 0)$, and $\boldsymbol{\beta} \sim N(\mathbf{0}, 10^6 \mathbf{I})$. The Gamma priors have mean 1 and variance 1000. We carried out several sensitivity studies using for example different parameters in the Gamma, and also $\text{Unif}(0, 1000)$ for the precisions, and truncated normals for restricted parameters (e.g. $\varphi \sim N(0, 10^6)U(-1, 1)$) and $c \sim N(0, 10^6)U(-5.0, 0.8)$. In no cases did results differ, indicating that the likelihood strongly dominates the priors in this example. However, with smaller data sets or less well-defined patterns, some care in selecting priors, and similar sensitivity analyses, will be warranted, e.g. Gelman (2006).

Algorithm for estimation using MCMC Metropolis–Hastings steps were used to iteratively sample from the conditional distributions in the MCMC algorithm. There are two challenging aspects to estimation: the computationally intensive evaluation of the Faddy densities and parameterization of the Faddy distributions in terms of parameters not easily related to the mean count μ_{ij} . Evaluation

of each density value requires evaluation of a large matrix exponential in (8) for each observation on each subject, at each iteration of the MCMC algorithm. Reparameterization in terms of the mean would require a numerical optimization to find the value of a in (6) that gives μ_{ij} in (2), again for each observation on each subject at each iteration. To avoid these numerically prohibitive calculations, we used the following approach, which can be done in standard MCMC software. We used WinBUGS (Lunn et al., 2000).

Part 1, in R: This step needs to be done only once.

- (i) Define a grid of values of c_r , $r = 1, \dots, R$ and μ_s , $s = 1, \dots, S$ that spans the range of interest. In this case we used $c_r = -6.0, -5.9, \dots, 0.8, 0.9$ and $\mu_s = 0.1, 0.2, \dots, 5.9, 6.0$. Let Δ_c and Δ_r denote the spacings of these grids, in this case both 0.1.
- (ii) For each r and s , use a simple one-parameter Newton–Raphson search (e.g. `nlm` in R) to find the value of a_s that gives $E(Y) = \sum_{y=0}^K y \times p_y = \mu_s$ and $\sum_{y=0}^K p_y = 1$, where K and $\{p_y\}$ are as in (8). K must be large enough to include all non-trivial probabilities for all values of c_r and μ_s in step i. Experimentation showed that $K = 49$ sufficed in our application. This step is easily accomplished in SAS PROC IML (SAS Institute Inc., Cary, NC) or with function `expm` from the `Matrix` library downloaded in R. (There are many matrix exponential functions available, our testing indicated `expm` was accurate and stable in our situation.) We used `nlm` to minimize the function $\text{abs}\left(\sum_{y=0}^K y \times p_y - \mu_s\right) + \text{abs}\left(\sum_{y=0}^K p_y - 1\right)$. To check for numerical or optimization errors, verify that for each combination of c_r and μ_s , probabilities satisfy the usual conditions $0 \leq p_y \leq 1$, $\sum_{y=0}^K p_y = 1$ and $\sum_{y=0}^K y \times p_y = \mu_s$ with ‘reasonable’ accuracy. In our experience using R, `nlm`, and `expm`, these were satisfied within tolerance 0.0001. Numerical performance in SAS IML was not as good and many more errors were noted. This step also checks that K is large enough. In our experience the tolerance was not very important and numerical errors were large and obvious, and were fixed by adjustments to the optimization (e.g. adding the second term to the function to be minimized above).
- (iii) Create a $(K+1) \times S \times R$ dimensional array in R, Entries are the probabilities $p_{y+1,r,s} = \Pr(Y = y | c = c_r, \mu = \mu_s)$. We used $K = 49$, and from step i above, $R = 70$ and $S = 60$.
- (iv) Calculate the array of *−loglikelihood* values $\text{mll}_{y+1,r,s} = -\log(p_{y+1,r,s})$.

Part 2, in MCMC software:

- (v) Use the ‘zeros method’ for defining a new distribution. This method is described in the WinBUGS online users manual (section “Specifying a new sampling distribution”), and is carried out as follows. Suppose we want the contribution to the *−loglikelihood* for an observation $Y = y$ from a distribution with $p_y = \Pr(Y = y)$, where the p_y are known. In our case these probabilities will be calculated as in (8) for the Faddy distribution and saved in the array of step iv above. If we define $Z \sim \text{Poisson}(-\log(p_y))$ we have $\Pr(Z = 0) = \exp(-1 \times -\log(p_y)) = p_y$ so the *−loglikelihood* value mll_y for a zero observation of Z is the desired probability.
- (vi) The contribution to the *−loglikelihood* for observation $Y_{ij} = y_{ij}$ is, from the array in step iv above

$$-\log[\Pr(Y_{ij} = y_{ij} | c, \mu_{ij})] = \text{mll}[\text{ind}_m, \text{ind}_c, y+1]$$

with array indices calculated as $\text{ind}_m = \text{round}((\mu_{ij} - \mu_1)/\Delta_\mu) + 1$ and $\text{ind}_c = \text{round}((c - c_1)/\Delta_c) + 1$. These indices ind_m and ind_c choose values from the array that are closest to the values of c and μ_{ij} at the current Metropolis–Hastings evaluation.

Several aspects of computing deserve comment. We carried out several small simulation studies in simple cases (e.g. two samples) to examine the estimation performance and granularity of the grids and found that the values we used showed no evidence of bias. For data simulated from Poisson distributions and estimated using Faddy distributions with unspecified c , estimated values of c were close to 0 as expected, and estimated coefficients and standard errors were similar to Poisson estimation. We do not recommend modifying the algorithm to set $c = 0$, as this restricts sampling to one dimension of the array, and we have observed some small biases in simulations. We experimented by modifying step vi with interpolation between grid values and noted no improvements in results though iteration was slower. Computing times depend on the size of the data set and complexity of the model. For our asthma data set in Section 2 with 5209 observations, the model with no autocorrelation took several hours for 110 000 iterations. The handling of unequal spacing using the continuous-time AR(1) process slows the algorithm considerably, and the full model with autocorrelation took several days. We expect that this can be improved with fine-tuning of the algorithm, in particular reducing the size of the array in step iv above; however, in the interest of maintaining more general code we did not attempt this. If using WinBUGS, save the array $mll_{y+1,r,s}$ as part of the WinBUGS data set, and change the definition of the array dimension to $.Dim = c(R, S, K+1)$ (i.e. the first and last dimensions are switched from R). We note that other choices for conditional count distributions, e.g. weighted Poisson or COM-Poisson, could be used instead of the Faddy distributions, by modifying step ii above with the chosen probabilities. Thus, this algorithm provides a flexible way of applying MCMC for two-parameter underdispersed count distributions. R and WinBUGS code are included with the Supporting Information.

3.5 Model diagnostics

A model chosen by some criterion as best among a set of candidates does not necessarily fit well. Various types of diagnostics are available for models attempting to capture complex features such as correlation and dispersion. The PIT (Probability Integral Transform) histograms of Czado et al. (2009) do very well at diagnosing under- or overdispersion in counts. We used these methods along with estimated models to study the dispersion of conditional distributions of counts relative to several candidate distributions. We have also found the parametric bootstrap approach to work well in such settings (e.g. Tsay, 1992; Grunwald et al., 2000). The basis for this very general approach is that data simulated from a model should reflect the important features of the data being modeled. In applications, a model is estimated and data are simulated from the fitted model and examined and summarized in ways that will display key features being modeled. Examples of these diagnostics are given in Section 4 below.

4 Example continued

Primary scientific interest is in assessing the association between morning air pollution as measured by ambient $PM_{2.5}$ (maximum hourly concentration in $\mu g/m^3/10$ of particulate matter less than $2.5 \mu m$ in diameter, measured by a fixed outdoor monitor between the hours of 1 am and 11 am) and asthma inhaler use. Hourly $PM_{2.5}$ concentrations were measured by a tapered element oscillating microbalance (TEOM) device located 2.7 miles west of the Kunsberg school and operated by the Colorado Department of Health Air Pollution Control Division (APCD). For this analysis we considered doser use and $PM_{2.5}$ measurements on the same day. It is common in some analyses also to include $PM_{2.5}$ one or more days prior to the health outcome measure. This was not done here but these could be included as additional covariates. We adjusted for several environmental covariates, also on the same day, including daily mean temperature (deg F/100), humidity (%/100), and barometric pressure (mmHg/1000). Temperature and relative humidity data were obtained from

APCD and barometric pressure was obtained from the National Climatic Data Center. All models included a linear time trend with time coded in units of 100 days (day/100), as well as an indicator variable for observations made on Fridays when children did not have formal exercise class at school and hence were generally not prescribed doser pre-treatment. Thus, estimates of mean as-needed doser use from the model can be made by setting Friday = 1.

4.1 Model comparison and selection

The full longitudinal model is described by Eq. (9) and Eqs. (2)–(5), where β is length 7 and with σ_u^2 , σ_e^2 , ϕ and c the model has a total of 11 parameters. We also carried out exploratory analyses, estimating several submodels to examine convergence and to study the impact of each feature: clustering of measurements within subject (subject heterogeneity), serial correlation and underdispersion. In the exploratory analyses, for each model three chains were generated, each with 4000 burn-in iterations followed by 10 000 iterations summarized for posterior estimation. Using several shorter exploratory chains allows examination of convergence with different starting values, and eliminates ill-fitting models. Initial values for parameters were chosen to be different for each chain, but fairly close to values determined from a Poisson GzLM and a pre-estimation run. Each individual chain was examined graphically for convergence, and graphs of the three individual chains were examined. In all cases we saw no evidence of non-convergence, and estimates from the three chains were similar. For each chain, the parameter estimates are the means of the posterior distributions (after burn-in). Each estimate, standard error (standard deviation of posterior values) and credible interval was estimated by an average of the same values from the three chains for each model. Deviance Information Criterion (DIC, analogous to Akaike or Bayesian Information Criteria, with smaller values indicating better fit) values were also averaged across the three chains. All covariates were standardized by subtracting the mean and dividing by the standard deviation, to improve MCMC convergence. The results are presented on the original scale, with the standardization reversed. Distributions of most estimates were fairly normal so precision is indicated by SEs. Credible intervals are also given for the pollution parameter of interest. For the final model with all components, more precise estimation and formal model diagnostics were done using two chains, each with 10 000 burn-in iterations followed by 100 000 estimation iterations.

The results for the exploratory analyses are shown in the top portion of Table 1. According to DIC the full model fits best (DIC = 10311). The average of the dispersion parameter estimates was $\hat{c} = -2.38$, individual values from three chains were -2.47 , -2.30 and -2.37 , and the average 95% credible interval was $(-2.63, -2.17)$. These negative values indicate substantial within-subject underdispersion since $c = 0$ corresponds to the Poisson distribution. The average estimated pollution parameter was 0.0133, with 95% credible interval $(-0.0033, 0.030)$.

The same model assuming a Poisson observation distribution in Eq. (1) rather than the Faddy distribution in Eq. (9) was much worse, with DIC = 12486. The estimate of the pollution parameter, $\hat{\beta}_{PM25} = 0.0136$, is similar in magnitude to that from the Faddy model but the 95% credible interval $(-0.0123, 0.0390)$ is more than 50% wider, indicating a much too conservative result. The autocorrelation parameter $\hat{\phi} = 0.05$ is near 0 and non-significant, because the observation model is so poor. The observation error standard deviation $\hat{\sigma}_e$ was near zero because this term contributes overdispersion and the Poisson distribution already contains too much dispersion for the data.

Underdispersed models omitting serial correlation (DIC = 10686) and omitting subject heterogeneity (DIC = 10493) also did not fit as well. Estimates and SEs in Table 1 are not too different from those of the full Faddy model; however, models without these terms would be subject to question. For example, it is well known that ignoring serial correlation of errors in normal regression tends to artificially reduce standard errors (e.g. Anderson, 1971), and the slightly lower SEs observed for our model omitting serial correlation may be due to the same effect. Other regression parameters remained stable across models, except for the decreases in precision for the Poisson

Table 1 Model estimation results, mean (SE).

Model	β_0	β_{time}	β_{fri}	c	φ	σ_e	σ_u	β_{hum}	β_{pres}	β_{temp}	β_{PM25}	DIC
<i>Exploratory analyses</i>												
Full	0.174 (1.23)	0.0223 (0.019)	-1.235 (0.039)	-2.38 (0.12)	0.737 (0.048)	0.241 (0.0154)	0.316 (0.035)	-0.296 (0.063)	0.688 (1.963)	-0.466 (0.107)	0.0133 (0.0083)	10 311
Poisson	0.579 (1.71)	0.0192 (0.024)	-1.227 (0.048)		0.050 (0.429)	0.0334 (0.0069)	0.317 (0.036)	-0.200 (0.089)	-0.065 (2.735)	-0.335 (0.150)	0.0136 (0.0135)	12 486
No autocorrelation	1.212 (1.07)	0.0211 (0.015)	-1.245 (0.034)	-1.69 (0.05)			0.308 (0.034)	-0.263 (0.058)	-0.983 (1.720)	-0.438 (0.094)	0.0136 (0.0079)	10 686
No subject heterogeneity	-0.005 (1.10)	-0.034 (0.029)	-1.247 (0.037)	-1.88 (0.06)	0.995 (0.001)	0.349 (0.029)		-0.233 (0.055)	0.903 (1.757)	-0.307 (0.107)	0.0096 (0.0083)	10 493
<i>Final estimation</i>												
Faddy full	0.126 (1.21)	0.019 (0.018)	-1.239 (0.039)	-2.33 (0.13)	0.760 (0.056)	0.237 (0.017)	0.320 (0.036)	-0.287 (0.060)	0.749 (1.932)	-0.455 (0.108)	0.0128 (0.0083)	10 310
Faddy full	0.150 (1.18)	0.021 (0.019)	-1.239 (0.039)	$c_{FR} = -2.08$ (0.23)	0.755 (0.055)	0.239 (0.016)	0.319 (0.036)	-0.292 (0.062)	0.734 (1.891)	-0.467 (0.108)	0.0130 (0.0081)	10 333
Different c^a				$c_{NF} = -2.27$ (0.14)								

^{a)} Different values of c for Friday (c_{FR}) and Mon–Thurs (c_{NF}).

model noted above and for the barometric pressure coefficient which was poorly determined by the data.

4.2 Model results

The results for the full model combining the two chains of 100 000 estimation iterations each are shown in the bottom section of Table 1. The results from the two chains were very similar, and similar to the exploratory results, and model diagnostics indicated convergence. For example, the Gelman–Rubin R value was < 1.02 for all parameters, and Geweke diagnostic z -values were less than 2.2 for all parameters using data splits of 0.20 and 0.50. The parameter of interest, $\beta_{\text{PM}_{2.5}}$, passed all diagnostics, and the Raftery–Lewis measure indicated sufficient precision to estimate the 0.025 and 0.975 quantiles to within tolerance of 0.0025 with probability 0.90. The distribution of chain values was approximately normal. The estimate of the parameter of interest was $\hat{\beta}_{\text{PM}_{2.5}} = 0.0128 \pm 0.0083$ with 95% credible interval $(-0.0034, 0.0279)$, just containing 0. For a $10 \mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ (about one SD of variation in $\text{PM}_{2.5}$), the expected number of daily doser uses was about 1.3% higher, with 95% credible interval $(-0.3\%, 2.8\%)$.

Figure 3 shows conditional probability distributions of doser counts from Faddy and Poisson models (rows 5 and 2 from Table 1) for an average subject ($u_i = 0$) for several covariate patterns. The underdispersion of the estimated Faddy probabilities is evident, especially on weekdays (left two panels). The Friday distributions (right panel) are quite different from weekdays, but there is little difference between weekdays with low pollution ($\text{PM}_{2.5}$ at its 10th percentile) and high pollution ($\text{PM}_{2.5}$ at its 90th percentile). All these are consistent with the results in Table 1, and provide detail on the estimated probability distributions.

4.3 Diagnostics

Figure 4 shows two parametric bootstrap diagnostics for the full model, based on the estimates in row 5 of Table 1. The left plots show variance versus mean and histogram plots for observed inhaler use, and the three plots to the right show three realizations of the estimated full model. The histograms indicate simulated marginal count ranges and distributions very similar to the observed data, and the simulated variance versus mean plots also show very similar patterns to the observed data.

Figure 5 shows further diagnostics based on the PIT histogram. Estimated mean counts for each observation were calculated using (2) with regression parameters as estimated by the full model (row

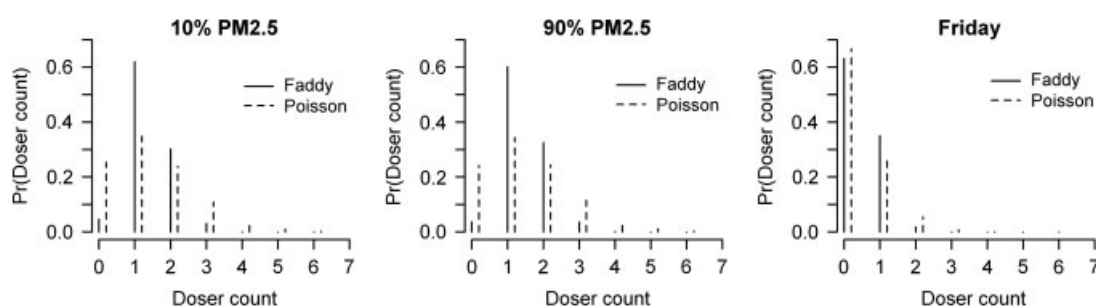


Figure 3 Estimated conditional probability distributions of doser count from Faddy and Poisson models (rows 5 and 2 from Table 1) at the final day of the study for an average subject ($u_i = 0$) and weekdays with low pollution ($\text{PM}_{2.5}$ at its 10th percentile, left panel) and high pollution ($\text{PM}_{2.5}$ at its 90th percentile, middle panel), and for Fridays with median pollution (right panel). All other covariates were fixed at their median values

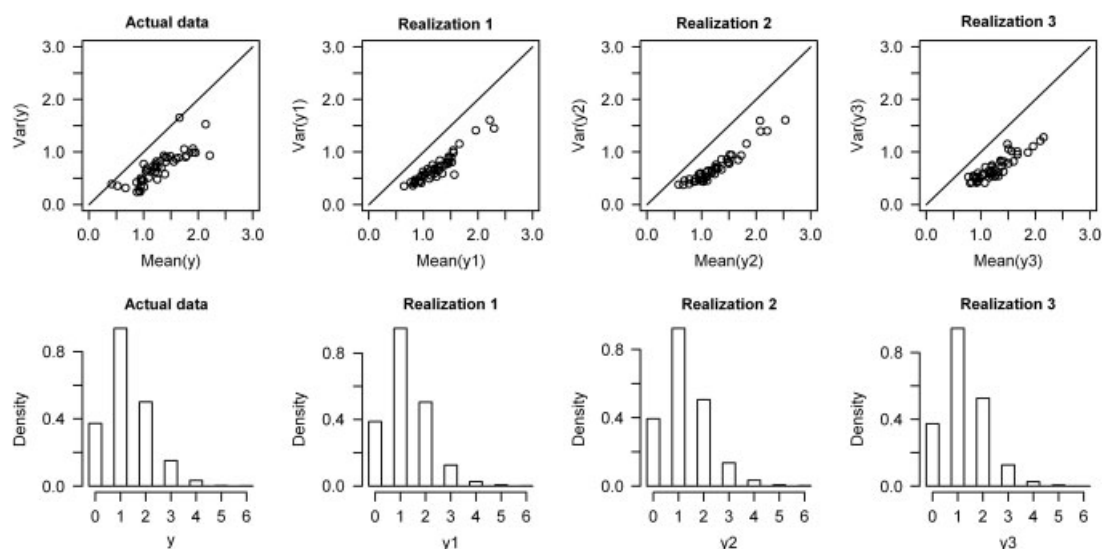


Figure 4 Parametric bootstrap diagnostics for the full model. The left plots show daily asthma inhaler use y (the top left panel is repeated from Figure 2), and the right three plots are for three realizations y_1 - y_3 of the estimated full model in Table 1

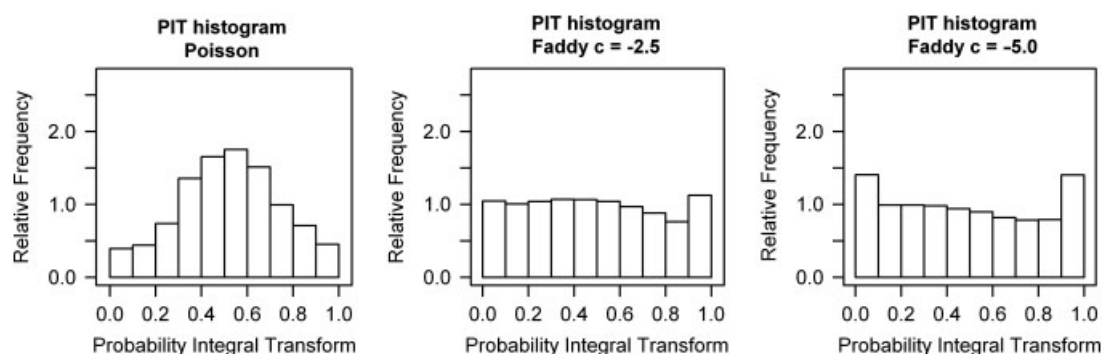


Figure 5 Diagnostics using PIT (Probability Integral Transform) histograms (Czado, Gneiting, Held, 2009). The mean of each observation was calculated using (2) with regression parameters as estimated by the full model (row 5 of Table 1), and with $e_{ij} = 0$ (i.e. an average day). This shows the Poisson (dispersion parameter $c = 0$) is overdispersed for these data, $c = -5.0$ is underdispersed, and $c = -2.5$ is about right

5 in Table 1), and with $e_{ij} = 0$ (i.e. an average day). We used in-sample predictions rather than out-of-sample predictions at the final study day, since only 7 of the 48 final-day values were non-zero. The principle is the same, and these models are based on so much data we do not consider this a problem in this example. The histograms again show that the Poisson (dispersion parameter $c = 0$) is strongly overdispersed for these data, a Faddy distribution with $c = -5.0$ is underdispersed, and $c = -2.5$ is about right, again corresponding with the results in Table 1. The diagnostics in Figs. 4 and 5 are very general and applicable to complex count models.

4.4 Further modeling

One advantage of a likelihood model is to allow formulation and exploration of further patterns in the data. We estimated one further model, allowing different values of c for weekdays and Fridays, a reasonable possibility due to the lack of prescribed exercise on Fridays and the large and highly significant Friday parameter. The results for one chain of 10 000 burn-in iterations and 100 000 estimation iterations are shown in Table 1. The model was not better than the model with a single value of c , and the estimated dispersion values ($\hat{c}_{FR} = -2.08 \pm 0.23$ and $\hat{c}_{NF} = -2.27 \pm 0.14$) were similar. This also indicates that the cause of the underdispersion is not due entirely to the prescribed doses for some children on exercise days.

5 Discussion

It is often noted that overdispersion relative to the Poisson is the norm for count data. In studies with count outcomes that are correlated due to clustering or longitudinal measurement, it is important to distinguish within- and between-subject or -cluster dispersion, which may be quite different. Our example shows clear evidence of underdispersion within subjects as indicated by the variance to mean graph (Fig. 2), PIT histograms (Fig. 5), and the estimated dispersion parameter $c \ll 0$. This contrasts with the observed overdispersion between subjects as indicated by the highly significant between-subject variance σ_u^2 . Accounting for within-subject underdispersion greatly affects the precision of estimated parameters of interest. The example also illustrates the ability of our model to incorporate complexities of longitudinal count data including subject heterogeneity, serial correlation and unequally spaced or missing data, and to provide information on estimated probabilities.

There is an extensive literature on the analysis of count outcomes, with several areas related to our work including other underdispersed distributions, decompositions of variation based on Waring and related distributions, models for overdispersed longitudinal counts and non-likelihood methods. We now relate some of this work to ours.

In previous work less attention has been given to underdispersion, probably because it is seen less frequently in cross-sectional data, and also because it cannot easily be modeled using a Poisson distribution with a random mean. The family of probability distributions we have used to model underdispersion conditional on covariates and other random effects, due to Faddy (1997), is based on an event process with a rate that slows (for underdispersion) each time an event occurs. This seems to be a reasonable general stochastic model for habitual behaviors such as medication use or intake of servings of certain foods, where intake beyond some number of doses or servings becomes increasingly unlikely. Other two-parameter distributions such as weighted Poisson (Ridout and Besbeas, 2004) or COM-Poisson (Conway and Maxwell, 1962; Shmueli et al., 2005; Sellers and Shmueli, 2010), could be used in (9) in place of the Faddy model, by calculating an array of probabilities based on those distributions. In that case, our methods would give a flexible means for incorporating covariates and correlation into other underdispersed distributions. These other distributions also seem to have similar challenges to the Faddy distribution, that is a lack of closed-form expressions for exact probabilities and moments.

There have been other decompositions of variability in count data proposed. For example, methods based on Waring distributions and regression (e.g. Irwin, 1968; Xekalaki, 1983, 1984; Rodríguez-Avi et al., 2007, 2009) use parametric distributions based on Poisson, Gamma and Beta distributions to decompose variation into several components: randomness and between-subject heterogeneity explainable by covariates (liability) or intrinsic to individuals (proneity). Our models could be viewed in terms of decompositions using random effects, into components between subjects described by u_i and any subject-level covariates (e.g. sex or age at start of study, which could easily be included in our models) and components within-subjects described by day-level covariates (as all of ours are) and randomness described by the Faddy distribution. These approaches are quite different in spirit, and also in application. The former methods are better suited for quantification of

variance components, using the parametric distributions involved. The methods in our paper seem better suited for more complex correlated and multilevel data. Also, Waring and related distributions are based on compounding Poisson distributions and as such do not appear capable of modeling underdispersed data, without further development.

The Faddy family of distributions also includes Poisson and overdispersed distributions (e.g. Negative Binomial) so our methods apply there as well, though we have not fully investigated our methods in those cases, and other methods exist there (e.g. Fahrmeir and Echavarria, 2006; Czado and Kolbe, 2007; Fotouhi, 2008).

Another method commonly used with correlated non-normal outcomes is GEE, first proposed by Zeger and Liang (1986) and Liang and Zeger (1986), with more recent developments described for example in Hardin and Hilbe (2003). GEE does not use a likelihood model but rather a method of inference that adjusts estimates and standard errors for dispersion and correlation. Likelihood models and the resulting probability distributions have several uses and advantages. The probability distributions can be used in diagnostics, as in Figs. 4 and 5, and probability distributions for various covariate combinations are available, as in Fig. 3. The direct model description of the data can be extended to describe particular features or patterns in the data, for example the model in Section 4.4 allowing for different dispersions for Monday–Thursday and Friday. Data can be simulated from the model, which is useful for model diagnostics as well as for evaluating and comparing performance of various models.

We have given a method for estimating parameters in the distribution using MCMC with standard software (e.g. WinBUGS) that overcomes the problems of the Faddy distribution not having exact closed-form expressions for probabilities or moments. This method applies in other settings, and can be used to extend other count models to include both under- and overdispersion and correlation. While simulations indicate good estimation properties, the algorithm has several limitations. It is versatile in the models it can use, but is currently limited to two-parameter models for underdispersed distributions, due to the limit of three dimensions for arrays in WinBUGS. It requires specification of the range of the parameters and counts in a particular application, though only at the initial creation of the array of probabilities in steps i–iv of the algorithm. Finally, it can be slow for complex models and large data sets, as described in the computing notes following the algorithm. Obviously, this can be improved with faster computers, and further optimization of the algorithm may be possible (e.g. further limiting the range of parameter values to reduce the size of the probability array). Of course, the MCMC iterations could be programmed directly in C, though the complications previously described make this challenging.

There is room for further work in this area. We used example data from one year, but data are available from several years. Exploratory analyses indicate a meta-analytic approach based on separate fits each year will be more useful than a single model including interactions with year. We are proceeding with these analyses. There is also room for further model development. A likelihood model makes available sophisticated modeling of dispersion. The model with different dispersion parameters for Monday–Thursday versus Friday is a simple example. We have also observed within-subject underdispersion in longitudinal measurement of weekly count of servings of particular foods in nutrition studies. In those data, simulations such as those in Fig. 4 indicate a diversity of values of c across subjects, indicating habitual behavior by some but not all subjects. Our model might be extended to allow subject-specific values of c . In another direction, measurement intervals in our situation were school days and so were all of roughly equal length. In models based on Poisson distributions, unequal measurement intervals are easily handled by including an offset in the model (e.g. McCullagh and Nelder, 1989). This assumes a constant event rate throughout the measurement interval, and so would not apply when using within-subject distributions such as those in (9). The underdispersion models given by Faddy (1997) include an adjustment for the length of the measurement interval that involves replacing $\exp(\mathbf{Q})$ in (8) with $\exp(\mathbf{Q}t)$, where t is the length of the measurement interval. However, some thought would be needed about whether the resulting event process gives a sensible model for the stochastic behavior of the system being modeled.

Acknowledgements The authors are grateful to Dr Stan Xu and Prof. Richard Jones for very helpful discussions, and to two reviewers and the Associate Editor for many comments that greatly improved the manuscript. This work was supported by the United States Air Force (S. L. B.); the Environmental Protection Agency (R825702 to M. S.); the Thrasher Research Fund (02816 to M. S.); and the Colorado Tobacco Research Program (2R-020 to M. S.).

Conflict of interest

The authors have declared no conflict of interest.

References

- Al-Osh, M. and Alzaid, A. (1987). First-order inter-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis* **8**, 261–275.
- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.
- Bockenholt, U. (1999). Mixed INAR(1) Poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data. *Journal of Econometrics* **89**, 317–338.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association* **88**, 9–25.
- Cameron, A. C. and Johansson, P. (1997). Count data regression using series expansions: with applications. *Journal of Applied Econometrics* **12**, 203–223.
- Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.
- Castillo, J. and Pérez-Casany, M. (1998). Weighted Poisson distributions for overdispersion and underdispersion situations. *Annals of the Institute of Statistical Mathematics* **50**, 567–585.
- Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering* **12**, 132–136.
- Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Processes*. Chapman & Hall, New York.
- Czado, C. and Kolbe, A. (2007). Model-based quantification of the volatility of options at transaction level with extended count regression models. *Applied Stochastic Models in Business and Industry* **23**, 1–21.
- Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data. *Biometrics* **65**, 1254–1261.
- Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York.
- Diggle, P., Heagerty, P., Liang, K. and Zeger, S. (2002). *Analysis of Longitudinal Data* (2nd edn.) Oxford University Press, Oxford.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**, 709–721.
- Faddy, M. J. (1997). Extended Poisson process modeling and analysis of count data. *Biometrical Journal* **39**, 431–440.
- Faddy, M. J. and Bosch, R. J. (2001). Likelihood-based modeling and analysis of data underdispersed relative to the Poisson distribution. *Biometrics* **57**, 620–624.
- Faddy, M. J. and Smith, D. M. (2005). Modeling the dependence between the number of trials and the success probability in binary trials. *Biometrics* **61**, 1112–1114.
- Fahrmeir, L. and Echavarria, L. (2006). Structured additive regression for overdispersed and zero-inflated count data. *Applied Stochastic Models in Business and Industry* **22**, 351–369.
- Fotouhi, A. (2008). Modelling overdispersion in longitudinal count data in clinical trials with application to epileptic data. *Contemporary Clinical Trials* **29**, 547–554.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.
- Greene, W. (2007). Functional forms for the negative binomial model for count data. *Econometrics Letters* **99**, 585–590.
- Grunwald, G. K., Hyndman, R. J., Tedesco, L. and Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics* **42**, 479–495.
- Hardin, J. W. and Hilbe, J. M. (2003). *Generalized Estimating Equations*. Chapman & Hall/CRC, Florida.
- Hardin, J. W. and Hilbe, J. M. (2007). *Generalized Linear Models and Extensions* (2nd edn.) Stata Press, TX.

- Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Wiley, NJ
- Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika* **90**, 355–366.
- Irwin, J. O. (1968). The generalized Waring distribution applied to accident theory. *Journal of the Royal Statistical Society Series A* **131**, 205–225.
- Jones, R. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*. Chapman & Hall, London.
- Jorgensen, B., Lundbye-Christensen, S., Song, P. X.-K. and Sun, L. (1996). State-space models for multivariate longitudinal data of mixed types. *Canadian Journal of Statistics* **24**, 385–402.
- Lambert, P. (1996). Modelling of repeated series of count data measured at unequally spaced times. *Applied Statistics* **45**, 31–38.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.
- McCullagh, P., and Nelder, J. (1989). *Generalized Linear Models* (2nd edn.) Chapman & Hall, London.
- Mckenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric marginal distributions. *Advances in Applied Probability* **18**, 679–705.
- Molenberghs, G., Verbeke, G. and Demétrio, C. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis* **13**, 513–531.
- Ridout, M. and Besbeas, P. (2004). An empirical model for underdispersed count data. *Statistical Modelling* **4**, 77–89.
- Rigby, R. and Stasinopoulos, M. (2008). A flexible regression approach using GAMLSS in R.
- Rodríguez-Avi, J., Conde-Sánchez, A., Sáiz-Castillo, A. J. and Olmo-Jiménez, M. J. (2007). A new generalization of the Waring distribution. *Computational Statistics and Data Analysis* **51**, 6138–6150.
- Rodríguez-Avi, J., Conde-Sánchez, A., Sáiz-Castillo, A. J., Olmo-Jiménez, M. J. and Martínez-Rodríguez, A. M. (2009). A generalized Waring regression model for count data. *Computational Statistics and Data Analysis* **53**, 3717–3725.
- Sellers, K. F. and Shmueli, G. (2010). A flexible regression model for count data. *Annals of Applied Statistics* **4**, 943–961.
- Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S. and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the COM-Poisson. *Journal of the Royal Statistical Society Series C* **54**, 127–142.
- Solis-Trapala, I. and Farewell, V. (2005). Regression analysis of overdispersed correlated count data with subject specific covariates. *Statistics in Medicine* **24**, 2557–2575.
- Thall, P. and Vail, S. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–671.
- Toscas, P. and Faddy, M. J. (2003). Likelihood-based analysis of longitudinal count data using a generalized Poisson model. *Statistical Modelling* **3**, 99–108.
- Tsay, R. S. (1992). Model checking via parametric bootstraps in time series analysis. *Applied Statistics* **41**, 1–15.
- Weiss, R. E. (2005). *Modeling Longitudinal Data*. Springer, New York.
- Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Springer, Berlin.
- Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48**, 233–243.
- Xekalaki, E. (1983). The univariate generalized Waring distribution in relation to accident theory: proneness, spells or contagion. *Biometrics* **39**, 887–895.
- Xekalaki, E. (1984). The bivariate generalized Waring distribution and its application to accident theory. *Journal of the Royal Statistical Society Series A* **147**, 488–498.
- Xu, S., Jones, R. H. and Grunwald, G. K. (2007). Analysis of longitudinal count data with serial correlation. *Biometrical Journal* **49**, 1–12.
- Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130.
- Zeger, S. L. and Qaqish, B. (1988). Markov regression models for time series: a quasi-likelihood approach. *Biometrics* **44**, 1019–1031.