Made available by Hasselt University Library in https://documentserver.uhasselt.be

Second-order generalized estimating equations for correlated count data

Peer-reviewed author version

KALEMA, George; MOLENBERGHS, Geert & KASSAHUN, Wondwosen (2015)
Second-order generalized estimating equations for correlated count data. In:
COMPUTATIONAL STATISTICS 31(2), p. 749-770.

DOI: 10.1007/s00180-015-0599-1
Handle: http://hdl.handle.net/1942/20859

Computational Statistics manuscript No.
(will be inserted by the editor)

Second-order generalized estimating equations for correlated count data

George KALEMA · Geert MOLENBERGHS · Wondwosen KASSAHUN

Received: date / Accepted: date

Abstract Generalized estimating equations have been widely used in the analysis of correlated count data. Solving these equations yields consistent parameter estimates while the variance of the estimates is obtained from a sandwich estimator, thereby ensuring that, even with misspecification of the so-called working correlation matrix, one can draw valid inferences on the marginal mean parameters. That they allow misspecification of the working correlation structure, though, implies a limitation of these equations should scientific interest also be in the covariance or correlation structure. We propose herein an extension of these estimating equations such that, by incorporating the bivariate Poisson distribution, the variance-covariance matrix of the response vector can be properly modeled, which would permit inference thereon. A sandwich estimator is used for the standard errors, ensuring sound inference on the parameters estimated. Two applications are presented.

Keywords Bivariate Poisson distribution · first four moments · generalized linear models · longitudinal count data · sandwich estimator · time varying covariates

1 Introduction

Count data, as the name suggests, arises as a result of a counting process in a given interval of time and therefore takes on non-negative integer values. Examples may

George KALEMA (✉) · Geert MOLENBERGHS · Wondwosen KASSAHUN
I-Biostat, Universiteit Hasselt, Martelarenlaan 42, 3500 Hasselt, Belgium
Tel.: +32 11 26 82 84
E-mail: george.kalema@uhasselt.be

George KALEMA
School of Statistics and Applied Economics, Makerere University, P.O. Box 7062, Kampala, Uganda

Geert MOLENBERGHS
I-Biostat, KU Leuven, Kapucijnenvoer 35, B3000 Leuven, Belgium

Wondwosen KASSAHUN
Department of Epidemiology and Biostatistics, Jimma University, Jimma, Ethiopia

include: number of doctor visits, number of epileptic seizures, number of accidents, etc. To draw inferences from such data, a Poisson distribution is usually assumed as the data generating mechanism and a log-likelihood function is constructed which, when maximized, yields parameters of scientific interest. The standard Poisson model implies that the mean and variance are equal (McCullagh and Nelder (1989). However, in practice, this implication is usually restrictive because count data samples often have the mean either greater than the variance (so-called underdispersion) or less than the variance (also known as overdispersion). Therefore, using the Poisson model in its basic form would not account for this feature correctly. To account for overdispersion, the negative-binomial (NB; Breslow (1984); Lawless (1987)) model is an option. Also, count data regularly has an incidence of zero counts greater than expected from the Poisson model. The zero-inflated Poisson (ZIP; Lambert (1992)) or zero-inflated negative binomial (ZINB; Ridout et al (2001)) model account for the extra zeros.

Further, count data is often collected repeatedly over time. Such studies aim at describing, for example, the evolution of the subjects' condition over time, given certain characteristics of interest. This repetition in the observation of the patients or cluster or subjects induces the aspect of correlation because responses from the same subject will be more alike than those between different subjects. Also here, extensions from cross-sectional or univariate data to correlated data have been proposed in the literature and implemented in statistical software packages. Some of these include generalized estimating equations (GEE1; Liang and Zeger (1986)), the Poisson-normal model, which belongs to the generalized linear mixed model family (GLMM; Breslow and Clayton (1993); Wolfinger and O'Connell (1993)) or more generally the combined model (Molenberghs et al (2007), Molenberghs et al (2010)), the multivariate negative binomial model (Solis-Trapala and Farewell (2005); Winkelmann (2008)), etc. In this paper, we shall generically refer to all estimating equations as GEE while GEE1 denotes the method put forward by Liang and Zeger (1986) in which the correlation structure is calculated using the method of moments. Furthermore, GEE1.5 denotes the extension of GEE1 by replacing the moment-based estimation of the working correlation parameters with a second set of estimating equations (Prentice (1988); Kim and Shults (2010) and, Lipsitz et al (1991) are some examples). In GEE1.5, the two sets of estimating equations for the marginal mean and correlation structure are assumed orthogonal or independent, which simplifies the computational burden that would be encountered otherwise. Because these methods aim at obtaining marginal mean parameters that are consistent and asymptotically normally distributed, they permit inferences on the marginal mean regression parameters and standard errors even when the correlation structure is not correctly specified. As a result, and akin to GEE1, no scientific inference can be made on the correlation structures in GEE1.5 given that these association structures are allowed to be misspecified. On the other hand, allowing the two sets of estimating equations to be correlated results in GEE2 (Liang et al 1992; Zhao and Prentice 1990; Prentice and Zhao 1991). This implies that the first and second moments are then fully modeled while making working assumptions about the third- and higher-order moments. Research in marginal models for hierarchical or correlated count data is certainly ongoing. Iddi and Molenberghs (2013) also contributed to this area of correlated and overdispersed

count data by proposing a marginalized model for zero-inflated, overdispersed, and correlated count data. We refer interested readers in the topic of generalized estimating equations to Molenberghs and Verbeke (2005), Hardin and Hilbe (2003), Diggle et al (2002), Fitzmaurice et al (2004), and Ziegler (2011). Modeling of the covariance structure simultaneously with the mean is a field of ongoing research and many other researchers, for example, Ye and Pan (2006), Leng et al (2010), Fan et al (2007), etc., have also made contributions to the subject.

Generally, when analyzing correlated data, 3 different modeling frameworks can be chosen from, depending on the objective of the study. In Section 5.3 of Molenberghs and Verbeke (2005), these frameworks or model families (marginal models, conditionally specified models, and subject-specific models) are presented and characterized. Note that Lee and Nelder (2004) regard the subject-specific models also as conditional models. For completeness, we in turn briefly define these 3 families but refer to Molenberghs and Verbeke (2005), Fahrmeir and Tutz (1994), Fahrmeir and Tutz (2001), Diggle et al (2002), and Lee and Nelder (2004) for detailed discussions of these families. A marginal model is one where the marginal distribution of the response of interest is modeled as a function of covariates. One models the expectation of the response variable conditioning only on the covariates. An example is a comparison of males and females in terms of the mean response or a contrast in the average number of epileptic seizures between patients who received a treatment versus patients in a control (placebo) group. A random-effects or subject-specific model further conditions on unobserved or latent subject-specific random effects in addition to the covariates of interest. For a conditionally specified model, typically an autoregressive or transition model, the expectation of the response variable is modeled while conditioning on part or all of the remaining set of responses for a subject as well as covariates; in a transition model, conditioning is on past measurements.

In this paper, marginal models are of interest, especially motivated by two datasets presented in Section 3 and analyzed in Section 4. We hence-forth limit our discussion to the marginal-models framework for correlated count data, highlighting the (relevant) limitations and then discuss our proposed solution. GEE1 is a common tool used when modeling correlated count data. Since its introduction by Liang and Zeger (1986), it has been extensively studied, implemented in statistical software packages, and applied in research. Its strengths and limitations are very thoroughly described in the literature. Extensions of GEE1 to allow for the simultaneous estimation of both the marginal mean and correlation parameters assuming independence between them have been proposed in literature and are herein referred to as GEE1.5. One strength of GEE1 and GEE1.5 is that the parameter estimates are consistent as long as the mean structure is correctly specified, even if the working covariance or correlation structure is misspecified. They are also computationally easier (faster) to evaluate than their GEE2 counterparts. A major limitation suffered by both methods is that, because the estimating equations allow for the misspecification of the correlation/covariance structure, GEE1 and GEE1.5 fall short if scientific interest is not only in the marginal mean parameters but also in the association structure. Further detail about the different GEE methods is presented in Section 2.2, while Section 2.3 presents our extension of the estimating equations to model the covariance structure via covariates simultaneously with the marginal mean parameters by incorporating

the bivariate Poisson distribution into the estimating equations thereby permitting inference on both the marginal mean and covariance parameters.

2 Methodology

2.1 Notation

Henceforth, the term *subject* will be used to mean the independently replicated entity within which the repetition occurs; for example, patient, subject, cluster, or unit. We use the random variable Y_{ij} to denote the j-th observation of subject i, $i = 1,\ldots,K$ and $j = 1,\ldots,n_i$. Because the responses for each subject i are repeatedly recorded, subject i has an $n_i \times 1$ vector $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^\top$ of measurements. Further, let $\mathbf{X}_{ij}$ denote a $p \times 1$ vector of covariates, thus $\mathbf{X}_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijp})^\top$, that are to be investigated for possible association with the response variable Y_{ij}. In matrix notation,

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{X}_{i1}^\top \\ \mathbf{X}_{i2}^\top \\ \vdots \\ \mathbf{X}_{in_i}^\top \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \ldots & X_{i1p} \\ X_{i21} & X_{i22} & \ldots & X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{in_i1} & X_{in_i2} & \ldots & X_{in_ip} \end{pmatrix}.$$

It is important to mention that the covariates contained in $\mathbf{X}_i$ may be either changing over time j in which case they would be referred to as time-varying covariates, or otherwise time-stationary. In Section 2.3, we show how our proposal adapts to both time-stationary and time-varying covariates.

2.2 Generalized Estimating Equations

Following from the theory of generalized linear models (GLM; Agresti (2002), Nelder and Wedderburn (1972), McCullagh and Nelder (1989)), the first two moments derived (Molenberghs and Verbeke (2005), ch. 3) from a distribution that belongs to the exponential family of distributions are the mean and variance, expressed as,

$$\mathrm{E}(Y_{ij} \mid \mathbf{X}_i) = \mu_{ij}, \tag{1a}$$

$$\mathrm{Var}(Y_{ij} \mid \mathbf{X}_i) = V_{ij} = \phi\,\upsilon(\mu_{ij}), \tag{1b}$$

respectively, where ϕ is a scale parameter for the variance and $\upsilon(\cdot)$ is the variance function, which describes the dependency of the variance on the mean. Mean (1a) is related to covariates in $\mathbf{X}_{ij}$ via a known link function $g(\cdot)$ (for example, log link for counts/Poisson data, logit or probit link for binary/binomial data) as $g(\mu_{ij}) = \mathbf{X}_{ij}^\top\beta$, where β is a $p \times 1$ vector of unknown regression parameters. If we let $\mathrm{Corr}(Y_{ij}, Y_{ik} \mid \mathbf{X}_i) = \rho_{ijk}$, then $\mathrm{Cov}(\mathbf{Y}_i \mid \mathbf{X}_i) = V_i(\beta,\phi,\alpha) = \phi C_i(\beta)^{\frac{1}{2}} R_i(\alpha) C_i(\beta)^{\frac{1}{2}}$, where R_i is a correlation matrix, $C_i = \mathrm{diag}(\upsilon(\mu_{ij}))$ is a diagonal matrix of variances and α is a vector of correlation parameters. Specific to count data, Y_{ij} is assumed to follow a Poisson distribution with mean μ_{ij}, thus $Y_{ij} \sim \mathrm{Poisson}(\mu_{ij})$. The Poisson

density can be expressed as belonging to the exponential family by letting $\log \mu_{ij}$ to be the natural parameter, $\phi = 1$ and $\upsilon(\mu_{ij}) = \mu_{ij}$. The marginal mean (1a) is then modeled in terms of covariates as $\log(\mu_{ij}) = \mathbf{X}_{ij}^{\top}\beta$, therefore referred to as a log-linear or Poisson regression model and $\mathrm{Var}(Y_{ij} \mid \mathbf{X}_i) = \mu_{ij}$. This model, therefore, specifically implies that the mean is equal to the variance, a phenomenon usually termed equi-dispersion. In practice though, deviations from this are common, so that the mean is greater than the variance (under-dispersion) or that the mean is less than the variance (over-dispersion). While it is not this paper's goal to fully address over(under)-dispersion, we refer interested readers to, for example, Molenberghs et al (2007), Molenberghs et al (2010) and references therein for further details. Indeed, our intention is to model correlated count data such that inference is allowed on both the marginal parameters β and the covariance structure $\mathrm{Cov}(\mathbf{Y}_i \mid \mathbf{X}_i)$.

Given (ϕ, α), Liang and Zeger (1986) iteratively solve the generalized estimating equation given by

$$\sum_{i=1}^{K} U_i(\beta) = \sum_{i=1}^{K} \frac{\partial \mu_i}{\partial \beta}^{\top} V_i^{-1} (\mathbf{Y}_i - \mu_i) = 0 \tag{2}$$

to obtain the estimates for β $(\widehat{\beta})$. The iterative algorithm is as follows:

1. Obtain the starting values of $\widehat{\beta}$ from fitting a GLM (thus assuming independence).
2. Given $\widehat{\beta}$ or $\widehat{\beta}^{(l)}$, calculate $(\hat{\phi}, \widehat{\alpha})$ and therefore $\hat{V}_i = \hat{\phi} C_i^{\frac{1}{2}} \left(\widehat{\beta}\right) R_i(\widehat{\alpha}) C_i^{\frac{1}{2}} \left(\widehat{\beta}\right)$ using the method of moments (see Chapter 8 of Molenberghs and Verbeke (2005)).
3. Given $\hat{\phi}, \widehat{\alpha}$ and $\hat{V}_i$, update $\widehat{\beta}$ by using modified Fisher's scoring algorithm:

$$\widehat{\beta}^{(l+1)} = \widehat{\beta}^{(l)} - \left[\sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)^{\top} V_i^{-1} \left(\frac{\partial \mu_i}{\partial \beta}\right)\right]^{-1} \left[\sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)^{\top} V_i^{-1} (\mathbf{Y}_i - \mu_i)\right].$$

The solution is obtained by iterating between steps 2 and 3 above until convergence meaning that the change in the parameter estimates satisfies (e.g., is less than) a pre-specified criterion. Assuming the marginal mean (μ_i) is correctly specified, consistent and asymptotically normally distributed parameter estimates $\widehat{\beta}$ with mean β and variance-covariance matrix

$$\mathrm{Var}(\widehat{\beta}) = I_0^{-1} I_1 I_0^{-1}, \tag{3}$$

where

$$I_0 = \sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)^{\top} V_i^{-1} \left(\frac{\partial \mu_i}{\partial \beta}\right) \text{ and}$$

$$I_1 = \sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)^{\top} V_i^{-1} \mathrm{Var}(\mathbf{Y}_i) V_i^{-1} \left(\frac{\partial \mu_i}{\partial \beta}\right),$$

are obtained. The variance estimator in (3) is commonly referred to as the sandwich estimator and results in the so-called empirically corrected standard errors. The parameter estimates and standard errors are asymptotically correct whether or not the working correlation structure is correctly specified.

Much as GEE1 has been found appealing to many data analysts and researchers, it has quite a number of issues associated with it. It is not our intention to exhaustively list them herein but refer to, for example, Lee and Nelder (2004), Lindsey and Lambert (1998), Crowder (1995), Sun et al (2009), Wang and Carey (2004), among others, for further discussion of these issues. Specifically, GEE1 allowing the misspecification of the working correlation structure, thereby rendering it a nuisance, implies that the response vector $(\mathbf{Y}_i)$ is given an arbitrary distribution and hampers checking assumptions about the covariance structure (Lee and Nelder (2004)). Specifying a covariance structure based on a model straightforwardly allows for inference on this covariance structure. Further and more importantly, although consistent parameter estimates and standard errors can be obtained even with a misspecification of the working correlation assumption, careful estimation of the covariance/correlation is needed since it may affect the iterative updating of β and α, leading to a breakdown of the iterative procedure (Sun et al (2009)). As an alternative estimation approach to the method of moments used by Liang and Zeger (1986) for the correlation structure, Kim and Shults (2010) use a two-stage approach to estimate the regression parameters $\widehat{\beta}$ and the correlation parameters $\widehat{\alpha}$. At stage 1, they iterate between (2), with $V_i^{-1} = C_i^{-\frac{1}{2}}(\beta)R_i^{-1}(\alpha)C_i^{-\frac{1}{2}}(\beta)$, and the estimating equation for α, namely:

$$\frac{\partial}{\partial\alpha}\left\{\sum_{i=1}^{K}\mathbf{Z}_i^{\top}(\beta)R_i^{-1}\mathbf{Z}_i(\beta)\right\} = 0, \tag{5}$$

where $\mathbf{Z}_i(\beta) = (z_{i1},\ldots,z_{in_i})^{\top}$ are the j-th Pearson residuals for subject i given by $z_{ij} = (y_{ij} - \hat{\mu}_{ij})/\sqrt{\upsilon(\hat{\mu}_{ij})}$ and evaluated at the current $\widehat{\beta}$, until convergence. At stage 2, they plug $\widehat{\alpha}$ from stage 1 into

$$\sum_{i=1}^{K}\text{trace}\left\{\frac{\partial R_i^{-1}(\delta)}{\partial\alpha}R_i(\alpha)\right\}\Bigg|_{\delta=\widehat{\alpha}} = 0, \tag{6}$$

and update $\widehat{\alpha}$. The final $\widehat{\beta}$ is obtained by solving (2) at the final $\widehat{\alpha}$ from (6).

Yet another alternative to the method of moments for the correlation parameters is the proposal by Prentice (1988) in which estimating equations (2) are simultaneously solved with those of pairwise correlations (α) given by;

$$\sum_{i=1}^{K}\left(\frac{\partial\zeta_i}{\partial\alpha}\right)^{\top}H_i^{-1}(W_i - \zeta_i) = 0, \tag{7}$$

where $W_i = (z_{i1}z_{i2}, z_{i1}z_{i3},\ldots,z_{i,n_i-1}z_{i,n_i}, z_{i1}^2, z_{i2}^2,\ldots,z_{in_i}^2)^{\top}$ contains the products of subject i's pairs and squares of Pearson residuals $z_{is}z_{it}$ where $1 \le s < t \le n_i$, $H_i = \text{Var}(W_i)$ and $\zeta_i = \text{E}(W_i)$. It is common for binary responses that the last n_i components of W_i, i.e., the squared residuals, are left out due to the mean-variance relationship. Calculating $\text{Var}(W_i) = \text{Var}(z_{is}z_{it}) = \text{E}(\{z_{is}z_{it}\}^2) - \text{E}(z_{is}z_{it})^2$ requires

$$\begin{aligned}\text{E}(\{z_{is}z_{it}\}^2) = \big[&\text{E}(Y_{is}^2Y_{it}^2) - 2\hat{\mu}_{it}\text{E}(Y_{is}^2Y_{it}) + \hat{\mu}_{it}^2\text{E}(Y_{is}^2) - 2\hat{\mu}_{is}\text{E}(Y_{is}Y_{it}^2) + \\ &4\hat{\mu}_{is}\hat{\mu}_{it}\text{E}(Y_{is}Y_{it}) - 3\hat{\mu}_{is}^2\hat{\mu}_{it}^2 + \hat{\mu}_{is}^2\text{E}(Y_{it}^2)\big]\left[\upsilon(\hat{\mu}_{is})\upsilon(\hat{\mu}_{it})\right]^{-1}.\end{aligned} \tag{8}$$

For binary response data, for example, and unlike the counts case, (8) simplifies (as $Y_{ij}^2 = Y_{ij}$) such that

$$\mathrm{Var}(z_{is}z_{it}) = 1 + (1-2\hat{\mu}_{is})(1-2\hat{\mu}_{it})\left(\upsilon(\hat{\mu}_{is})\upsilon(\hat{\mu}_{it})\right)^{-1/2}\psi_{ist} - \psi_{ist}^2,$$

where $\psi_{ist} = \mathrm{E}(z_{is}z_{it})$ are entries in ζ_i. The binary response case, thus, turns out to be special since ζ_i and H_i are then fully determined by the mean and correlation models without necessitating additional assumptions about higher-order moments. Generally though, obtaining matrix H_i involves the third and fourth moments of $\mathbf{Y}_i$, which are usually assumed to be equal to zero. Alternatives to this assumption may be sought depending on the type of response variable under consideration. In the context of binary response data, for example, Diggle et al (2002) suggest $H_i = \mathrm{diag}\,[\mathrm{Var}(z_{i1}z_{i2}), \ldots, \mathrm{Var}(z_{i,n_i-1}, z_{i,n_i})]$, which only depends on $\widehat{\alpha}$ and $\widehat{\beta}$ while they propose the use of the identity matrix for count responses. By using the identity matrix for the counts, there is a loss of efficiency in estimating α. They, however, argue that this efficiency loss has very little impact, in practice, on the estimation of β and yet simplifies computation by avoiding the estimation of additional higher-order parameters. Note that while H_i is a working variance-covariance matrix (meaning that it contains working assumptions usually being that the third- and fourth-order correlations are equal to zero, matters not whether it is correctly specified or not, only aides the estimation of the regression parameters β and cannot be used for formal inferences) for W_i; V_i in (2), on the other hand, is not a working covariance matrix because the second moments are specified by (7).

Note that Prentice (1988) assumes independence between (2) and (7). Again, this assumption implies a loss of efficiency but is defendable because the consistency and asymptotic normality of the marginal mean regression parameters is not hampered by the misspecification of the correlation structure. Also important to mention is that the sets of parameters α and β both come with precision estimates and formal inference can be made on these parameters as long as the equations can be believed to have been correctly specified (Molenberghs and Verbeke (2005)). It may be desirable, however, to relax the independence assumption between (2) and (7). This may be the case if interest lies in the efficient estimation of both β and α. One may then be interested in minimizing the loss of efficiency accruing to the orthogonality assumption in GEE1.5. This leads us to the so-called second-order GEE (GEE2). Zhao and Prentice (1990) proposed an alternative to GEE1 or GEE1.5 in terms of correlations while Liang et al (1992) used odds ratios, with both proposals aimed at modeling mutivariate binary responses. Prentice and Zhao (1991) extended the equations of Zhao and Prentice (1990) to the general case of discrete or continuous response vectors. They combine the response vector $\mathbf{Y}_i$ and the pairwise crossproducts W_i into one outcome vector $\mathbf{T}_i^\top = (\mathbf{Y}_i^\top, W_i^\top)$ and solve the equations:

$$\begin{aligned}\sum_{i=1}^{K} U_i(\Theta) &= \sum_{i=1}^{K} \mathbf{D}_i^\top(\Theta)\boldsymbol{\Sigma}_i^{-1}(\Theta)\mathbf{f}_i(\Theta) \\ &= \begin{pmatrix} \frac{\partial \mu_i}{\partial \beta} & \frac{\partial \zeta_i}{\partial \beta} \\ \mathbf{0} & \frac{\partial \zeta_i}{\partial \alpha} \end{pmatrix} \begin{pmatrix} \mathrm{Var}(\mathbf{Y}_i) & \mathrm{Cov}(\mathbf{Y}_i, W_i) \\ \mathrm{Cov}(W_i, \mathbf{Y}_i) & \mathrm{Var}(W_i) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_i - \mu_i \\ W_i - \zeta_i \end{pmatrix} = 0,\end{aligned} \tag{9}$$

where $\Theta = (\beta^\top, \alpha^\top)^\top$. Since solving (9) is computationally unattractive, Prentice and Zhao (1991) suggested specifying working variance matrices in Σ_i such that the third and fourth moments are expressed as functions of μ_i and ζ_i. We show in the next section how we incorporate the bivariate Poisson distribution, rather than the multivariate Poisson distribution, into score equation (2) to allow scientific inference on the covariance as well as the mean parameters while modeling the covariance of $\mathbf{Y}_i$ at pair level. Use of the multivariate Poisson distribution would result in a full likelihood approach, which would maximize efficiency. However, it is hampered by the complexity of the probability function due to summations which may increase the computational burden with increase in the number of measurements per subject and/or sample size (Karlis (2003)), the very reason estimating equations are being sought after. By using the bivariate Poisson distribution, closed form expressions for the third and fourth moments in (8) are easily obtainable and one could go ahead with the suggestion of Prentice (1988) but for correlated count data. A somewhat different route is taken here in that estimating equations are proposed at the level of subject i's pair of responses (Y_{is}, Y_{it}). The use of pairs rather than the whole vector of responses $(\mathbf{Y}_i)$ would lead to loss of efficiency in estimating the parameters of interest but would simplify the computational unattractiveness of having to obtain the third and fourth moments that is evident as long as scientific interest lies in both the marginal mean parameters (β) and the association structure.

2.3 Extension of GEE Using the Bivariate Poisson Distribution

To put matters into perspective, consider the following bivariate Poisson distribution which is derived using the trivariate reduction method (Kocherlakota and Kocherlakota (1992), Kocherlakota and Kocherlakota (2001)) based on a convolution of independent Poisson variables. Note that there are several derivations of the bivariate Poisson distribution in the literature. For example, Lakshminarayana et al (1999) derive their bivariate Poisson distribution based on a polynomial factor. Assume that W_{ic} are independent Poisson random variables such that $\mathrm{E}(W_{ic}) = \eta_{ic}$, $c = s, t$ or st. The random variables $Y_{is} = (W_{is} + W_{ist})$ and $Y_{it} = (W_{it} + W_{ist})$ then follow a bivariate Poisson distribution. Thus, $(Y_{is}, Y_{it}) \sim BP(\eta_{is}, \eta_{it}, \eta_{ist})$ characterized by

$$f(y_{is}, y_{it}) = e^{-(\eta_{is}+\eta_{it}+\eta_{ist})} \frac{\eta_{is}^{y_{is}}}{y_{is}!} \frac{\eta_{it}^{y_{it}}}{y_{it}!} \sum_{l=0}^{\min(y_{is}, y_{it})} \binom{y_{is}}{l} \binom{y_{it}}{l} l! \left(\frac{\eta_{ist}}{\eta_{is}\eta_{it}} \right)^l. \tag{10}$$

Marginally, $\mathrm{E}(Y_{is}) = \eta_{is} + \eta_{ist}$, $\mathrm{E}(Y_{it}) = \eta_{it} + \eta_{ist}$ and $\mathrm{Cov}(Y_{is}, Y_{it}) = \eta_{ist}$. We propose the score equation $(U_{i,st})$ to be computed at each pair $\{s, t\}$ of responses from subject i such that the estimates for the β regression parameters are obtained by solving

$$\sum_i^K U_i = \sum_i^K \sum_{1 \le s < t \le n_i} U_{i,st}(\beta) = \sum_i^K \sum_{s<t} \frac{\partial \mu_{i,st}}{\partial \beta^\top} V_{i,st}^{-1} \left(\mathbf{Y}_{i,st} - \mu_{i,st} \right) = 0, \tag{11}$$

where

$\mathbf{Y}_{i,st} = \begin{pmatrix} Y_{is} \\ Y_{it} \\ Y_{is}Y_{it} \end{pmatrix}$, $\mu_{i,st} = \begin{pmatrix} \mathrm{E}(Y_{is}) \\ \mathrm{E}(Y_{it}) \\ \mathrm{E}(Y_{is}Y_{it}) \end{pmatrix}$ and

$$V_{i,st} = \begin{pmatrix} \mathrm{Var}(Y_{is}) & \mathrm{Cov}(Y_{is},Y_{it}) & \mathrm{Cov}(Y_{is},Y_{is}Y_{it}) \\ \mathrm{Cov}(Y_{it},Y_{is}) & \mathrm{Var}(Y_{it}) & \mathrm{Cov}(Y_{it},Y_{is}Y_{it}) \\ \mathrm{Cov}(Y_{is}Y_{it},Y_{is}) & \mathrm{Cov}(Y_{is}Y_{it},Y_{it}) & \mathrm{Var}(Y_{is}Y_{it}) \end{pmatrix},$$

with $\mathrm{E}(Y_{is}Y_{it}) = \mathrm{E}(Y_{is})\mathrm{E}(Y_{it}) + \mathrm{Cov}(Y_{is},Y_{it})$, $\mathrm{Var}(Y_{is}) = \mathrm{E}(Y_{is})$ and $\mathrm{Var}(Y_{it}) = \mathrm{E}(Y_{it})$.

To derive the covariance terms $\mathrm{Cov}(Y_{is},Y_{is}Y_{it})$, $\mathrm{Cov}(Y_{it},Y_{is}Y_{it})$ and $\mathrm{Var}(Y_{is}Y_{it})$ in V_{ist} in (11), we need to calculate the following four moments of the Poisson distribution that turn out to be essential. If $\tilde{X} \sim \mathrm{Poisson}(\lambda)$ with probability mass function $f(\tilde{X};\lambda) = \dfrac{\lambda^{\tilde{x}}e^{-\lambda}}{\tilde{x}!}$ where $\lambda > 0$ and $\tilde{x} = 0,1,2,\ldots$, then the nth moment $\mathrm{E}(\tilde{X}^n)$, $n = 1,2,3,4$ is as follows;

$$\mathrm{E}(\tilde{X}) = e^{-\lambda}\sum_{\tilde{x}=0}^{\infty}\tilde{x}\frac{\lambda^{\tilde{x}}}{x!} = \lambda e^{-\lambda}\sum_{\tilde{x}=1}^{\infty}\frac{\lambda^{\tilde{x}-1}}{(\tilde{x}-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda, \tag{12a}$$

$$\begin{aligned}\mathrm{E}(\tilde{X}^2) &= e^{-\lambda}\sum_{\tilde{x}=0}^{\infty}\tilde{x}^2\frac{\lambda^{\tilde{x}}}{\tilde{x}!} = \lambda e^{-\lambda}\sum_{\tilde{x}=1}^{\infty}\tilde{x}\frac{\lambda^{\tilde{x}-1}}{(\tilde{x}-1)!} = \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left[\lambda\sum_{\tilde{x}=1}^{\infty}\frac{\lambda^{\tilde{x}-1}}{(\tilde{x}-1)!}\right] \\ &= \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left(\lambda e^{\lambda}\right) = \lambda e^{-\lambda}(1+\lambda)e^{\lambda} = \lambda(1+\lambda),\end{aligned} \tag{12b}$$

$$\begin{aligned}\mathrm{E}(\tilde{X}^3) &= e^{-\lambda}\sum_{\tilde{x}=0}^{\infty}\tilde{x}^3\frac{\lambda^{\tilde{x}}}{\tilde{x}!} = \lambda e^{-\lambda}\sum_{\tilde{x}=1}^{\infty}\tilde{x}^2\frac{\lambda^{\tilde{x}-1}}{(\tilde{x}-1)!} = \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left[\lambda\sum_{\tilde{x}=1}^{\infty}\frac{\tilde{x}\lambda^{\tilde{x}-1}}{(\tilde{x}-1)!}\right] \\ &= \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left[\lambda\frac{\partial}{\partial\lambda}\left(\lambda\sum_{\tilde{x}=1}^{\infty}\frac{\lambda^{\tilde{x}-1}}{(\tilde{x}-1)!}\right)\right] = \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left[\lambda\frac{\partial}{\partial\lambda}\left(\lambda e^{\lambda}\right)\right] \\ &= \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left(\lambda(1+\lambda)e^{\lambda}\right) = \lambda e^{-\lambda}(\lambda(1+\lambda)+1+2\lambda)e^{\lambda}) \\ &= \lambda(1+3\lambda+\lambda^2), \text{ and}\end{aligned} \tag{12c}$$

$$\begin{aligned}\mathrm{E}(\tilde{X}^4) &= \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left[\lambda(1+3\lambda+\lambda^2)e^{\lambda}\right] \\ &= \lambda\left[\lambda(1+3\lambda+\lambda^2)+1+6\lambda+3\lambda^2\right] = \lambda(\lambda^3+6\lambda^2+7\lambda+1).\end{aligned} \tag{12d}$$

Generally, if $\mathrm{E}(\tilde{X}^n) = f_n(\lambda)$, then

$$\begin{aligned}f_{n+1}(\lambda) &= \lambda e^{-\lambda}\frac{\partial}{\partial\lambda}\left[f_n(\lambda)e^{\lambda}\right] \\ &= \lambda e^{-\lambda}\left[f_n'(\lambda)+f_n(\lambda)\right]e^{\lambda} = \lambda\left[f_n'(\lambda)+f_n(\lambda)\right],\end{aligned} \tag{13}$$

where $f_n'(\cdot)$ is the first derivative of $f_n(\cdot)$. Now, from $\mathrm{Var}\left(\tilde{X}\right) = \mathrm{E}\left(\tilde{X}^2\right) - \left[\mathrm{E}\left(\tilde{X}\right)\right]^2$, it follows that $\mathrm{Var}(Y_{is}Y_{it}) = \mathrm{E}\left[(Y_{is}Y_{it})^2\right] - [\mathrm{E}(Y_{is}Y_{it})]^2$ such that $\mathrm{E}\left[(Y_{is}Y_{it})^2\right]$ or $\mathrm{E}(Y_{is}^2Y_{it}^2)$ is to be replaced with

$$\begin{aligned}\mathrm{E}(Y_{is}^2Y_{it}^2) &= \mathrm{E}\left[(W_{is}+W_{ist})^2(W_{it}+W_{ist})^2\right] \\ &= \mathrm{E}\left[\left(W_{is}^2+2W_{is}W_{ist}+W_{ist}^2\right)\left(W_{it}^2+2W_{it}W_{ist}+W_{ist}^2\right)\right] \\ &= \mathrm{E}\begin{bmatrix} W_{is}^2W_{it}^2+2W_{is}^2W_{it}W_{ist}+W_{is}^2W_{ist}^2+2W_{is}W_{it}^2W_{ist}+ \\ 4W_{is}W_{it}W_{ist}^2+2W_{is}W_{ist}^3+W_{it}^2W_{ist}^2+2W_{it}W_{ist}^3+W_{ist}^4 \end{bmatrix},\end{aligned} \tag{14}$$

where further simplification is possible by applying the expectation to the independent Poisson variables W_s, W_t, W_{st} and using the moments in (12). This leads to the solution

$$\begin{aligned} \mathrm{E}(Y_{is}^2 Y_{it}^2) = \mathrm{E}(Y_{is})^2\mathrm{E}(Y_{it})^2 + \mathrm{E}(Y_{it})\mathrm{E}(Y_{is})^2) + \mathrm{E}(Y_{is})\mathrm{E}(Y_{it})^2 + 2\eta_{ist}^2 + \\ \mathrm{E}(Y_{is})\mathrm{E}(Y_{it})(1+4\eta_{ist}) + 2\eta_{ist}(\mathrm{E}(Y_{is}) + \mathrm{E}(Y_{it})) + \eta_{ist}. \end{aligned} \tag{15}$$

The covariances $\mathrm{Cov}(Y_{is}, Y_{is}Y_{it})$ and $\mathrm{Cov}(Y_{is}, Y_{is}Y_{it})$ are calculated as

$$\mathrm{Cov}(Y_{is}, Y_{is}Y_{it}) = \mathrm{E}(Y_{is}Y_{is}Y_{it}) - \mathrm{E}(Y_{is})\mathrm{E}(Y_{is}Y_{it}) = \mathrm{E}(Y_{is}^2 Y_{it}) - \mathrm{E}(Y_{is})\mathrm{E}(Y_{is}Y_{it})$$

and

$$\mathrm{Cov}(Y_{it}, Y_{is}Y_{it}) = \mathrm{E}(Y_{it}Y_{is}Y_{it}) - \mathrm{E}(Y_{it})\mathrm{E}(Y_{is}Y_{it}) = \mathrm{E}(Y_{is}Y_{it}^2) - \mathrm{E}(Y_{it})\mathrm{E}(Y_{is}Y_{it}),$$

respectively, where similar algebra as in (14) leads to the following quantities:

$$\begin{aligned} \mathrm{E}(Y_{is}^2) &= \mathrm{E}(Y_{is}) + \mathrm{E}(Y_{is})^2, \\ \mathrm{E}(Y_{it}^2) &= \mathrm{E}(Y_{it}) + \mathrm{E}(Y_{it})^2, \\ \mathrm{E}(Y_{is}^2 Y_{it}) &= \mathrm{E}(Y_{is}^2)\mathrm{E}(Y_{it}) + 2\eta_{ist}\mathrm{E}(Y_{is}) + \eta_{ist}, \\ \mathrm{E}(Y_{is}Y_{it}^2) &= \mathrm{E}(Y_{is})\mathrm{E}(Y_{it}^2) + 2\eta_{ist}\mathrm{E}(Y_{it}) + \eta_{ist}. \end{aligned}$$

The means $\mathrm{E}(W_{ic})$, are related to covariates as $\log[\mathrm{E}(W_{ic})] = \mathbf{X}_{ic}^\top \boldsymbol{\beta}$, where

$$\begin{pmatrix} \mathbf{X}_{is}^\top \\ \mathbf{X}_{it}^\top \\ \mathbf{X}_{ist}^\top \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & X_{is1} & X_{is2} & \dots & X_{isp} \\ 0 & 1 & 0 & X_{it1} & X_{it2} & \dots & X_{itp} \\ 0 & 0 & 1 & X_{ist1} & X_{ist2} & \dots & X_{istp} \end{pmatrix}. \tag{16}$$

The vector of unknown regression parameters $\boldsymbol{\beta} = (\beta_{0s}, \beta_{0t}, \beta_{0st}, \beta_1, \beta_2, \dots, \beta_p)^\top$, allowing for an intercept β_0 specific to time point s, t and their product st in addition to the regression parameters $(\beta_1, \dots, \beta_p)$ shared for the rest of the variables. Furthermore, the model-based standard errors are obtained as the square root of the diagonal entries of

$$U^* = \left(\sum_i^K \sum_{s<t} \frac{\partial \mu_{i,st}}{\partial \beta^\top} V_{i,st}^{-1} \frac{\partial \mu_{i,st}}{\partial \beta} \right)^{-1},$$

while the sandwich standard errors are calculated as the square root of the diagonal of

$$U^{**} = U^* \cdot I^* \cdot U^* = U^* \cdot \sum_i^K \sum_{s<t} U_{i,st} U_{i,st}^\top \cdot U^*.$$

As mentioned in Section 2.1, covariates under consideration in (16) may be either time-stationary or time-varying. When time-stationary, each of the columns $1, 2, \dots, p$ would contain the same values. On the other hand, when time-varying, X_{ist} can be derived as a function of X_{is} and X_{it}, for example, a difference, lag, sum, product, ratio, etc. The correlation between two measurements Y_{is} and Y_{it} is then calculated as $\rho_{i,st} = \mathrm{Cov}(Y_{is}, Y_{it}) / \sqrt{\mathrm{Var}(Y_{is})\mathrm{Var}(Y_{it})}$.

3 Datasets

Two real-life longitudinal datasets are analyzed herein, namely, the Epilepsy data and the Jimma Infant Growth (or Survival) Study. We describe these two datasets in turn.

3.1 Epilepsy Data

This dataset is presented and analyzed by Leppik et al (1985), and Thall and Vail (1990), among others. The data were obtained from a placebo-controlled clinical trial of 59 patients with epilepsy. These patients, suffering from simple or complex partial seizures, were enrolled in a randomized clinical trial that aimed at studying the effect of the anti-epileptic drug known as progabide on the number of epileptic seizures over time. In the study, 31 epileptic patients were randomized to the group that received progabide while 28 patients received a placebo, as an adjuvant to the standard anti-epileptic chemotherapy. Progabide is an anti-epileptic drug whose primary mechanism of action is to enhance gamma-aminobutyric acid (GABA) content. GABA is the primary inhibitory neurotransmitter in the brain. Prior to receiving treatment, baseline data on the number of epileptic seizures during the preceding 8-week interval were recorded. Counts of epileptic seizures during 2-week intervals before each of four successive post-randomization clinic visits were recorded. The dataset also contains information on the patient identification, treatment (0=Placebo, 1=Progabide), age, baseline 8 week seizure count, and the seizure count during the first, second, third, and fourth 2-week time interval. Figure 1 shows the evolution of the number of seizures for each epileptic patient over the study period, while Figure 2 shows the distribution of the seizure counts over all week intervals and both treatment groups. The evolution of the average and median number of epileptic seizures between the consecutive two-weeks period by treatment are shown in Figures 3 and 4, respectively. There are differences in the seizure counts within patients but also between patients over time. Specifically, one patient seems to have an extreme number of seizure counts at all time points relative to the other profiles while another patient registered a rather distant number (76) of seizures at the third visit. We also observe from Figure 2 that the distribution of the seizure counts is quite skewed and that the majority of the counts were between 0 and about 20, although there was a count of up to 102 seizures in the first two-week interval (see Figure 1). From Figures 3 and 4, it can be seen that the progabide group has lower (mean or median) seizure counts except at the second two-weeks interval. The seizure counts seem to reduce, on average, over the study period for both treatment arms. It is common for longitudinal studies to have cases that at some point in the study drop out or miss some of the visits. For this dataset, however, all patients were observed at all the visits.

3.2 Jimma Infant Growth Study

This dataset, also referred to as the Jimma Infant Survival Differential Longitudinal Growth Study, has been analyzed by Lesaffre et al (1999) in the linear mixed models

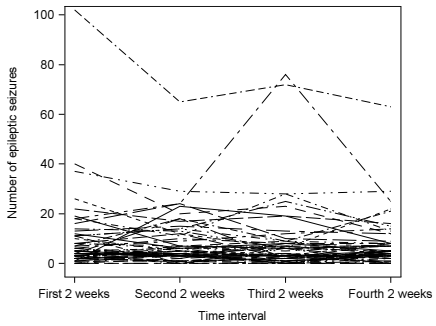

Fig. 1 Epilepsy Data: Subject-specific profiles of the number of epileptic seizures over study weeks.

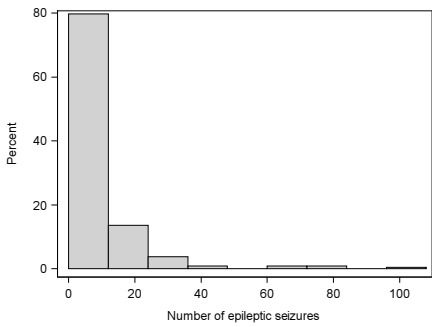

Fig. 2 Epilepsy Data: Distribution of the number of epileptic seizures.

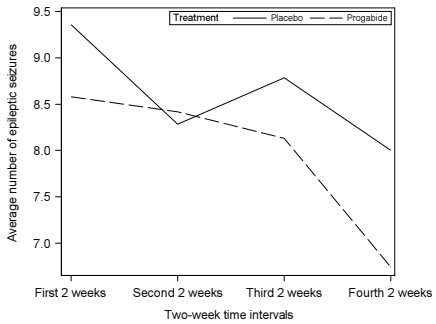

Fig. 3 Epilepsy Data: Average evolution of the number of epileptic seizures over study weeks by treatment.

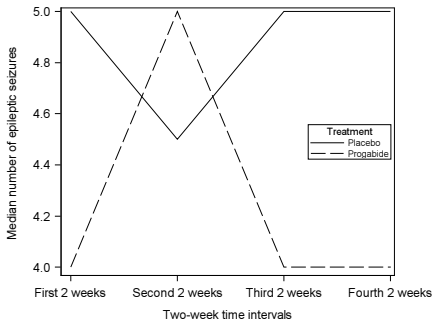

Fig. 4 Epilepsy Data: Median evolution of the number of epileptic seizures over study weeks by treatment.

context, while Kassahun et al (2012) have used it in the binary data framework in which they sought to identify risk factors for children being overweight, based on a dichotomization of the Body Mass Index (BMI). It is an Ethiopian study, set up to establish risk factors affecting infant survival and to investigate socio-economic, maternal, and infant-rearing factors that contribute most to children's early survival. Children born in Jimma, Keffa and Illubabor, located in Southwestern Ethiopia were examined for their first year growth characteristics. At baseline (birth), there were a total of 7969 infants enrolled in the study, both singleton and twin live births inclusive. However, only singleton live births (7872 infants) at baseline are considered. The children were followed-up every two months, until the age of one year, thus $age = 0, 2, 4, 6, 8, 10$ and 12 months. Herein, we are interested in modeling the total number of days of diarrheal illness as a function of gender (1=Male, 0=Female), whether mother continued breastfeeding (1=Yes, 0=No) for the 12 months, whether mother sought medical help (1=Yes, 0=No), and place of residence (1=rural, 2=urban, 3=semi-urban). Figure 5 shows the evolution of the number of days of illness over the 12-month period for 399 (5%) randomly sampled infants while Figure 6 depicts the average number of days of illness over the 12 months by gender. From Figure 5, we observe a tendency of the number of days of diarrheal illness to increase as the

infant grows older. There are also a lot of variability observable within an infant, and likewise between infants as they evolve. Figure 6 further shows an increasing trend in the average number of days of diarrheal illness as the infants get older, with the females always having lower average counts than the males. Table 1 shows the num-

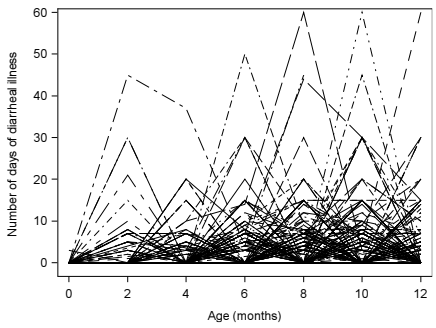

Fig. 5 Jimma Data: Infant-specific profiles of the number of days of diarrheal illness over age.

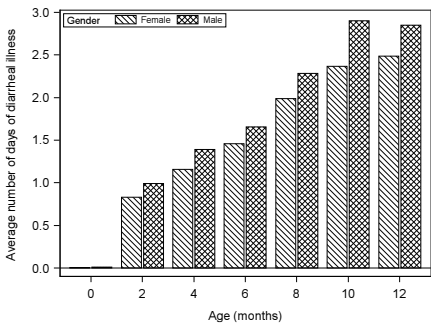

Fig. 6 Jimma Data: Average number of days of diarrheal illness by gender over age.

ber of infants whose responses were recorded over the 12-month period, by gender. As is typical of longitudinal studies, there is a reduction over time in the number of infants. As mentioned in Section 3.1, longitudinal studies very often have missing

Table 1 *Jimma data: Number of infants with observations by gender and age.*

	Age (months)						
Gender	0	2	4	6	8	10	12
Female	3865	3706	3570	3488	3401	3351	2920
Male	4007	3798	3656	3536	3455	3388	2972
Total	7872	7504	7226	7024	6856	6739	5892

data. The Jimma study is no exception. Figure 7 shows some (20 out of 59) of the missingness patterns present in the dataset. In general, both intermittent missingness and dropout as well as the first infant visit not having been at age=$0, 2, 4$ or 6 months are present. Because it is not our intention to deal with missing data in this paper, we have assumed that the missingness mechanism is not related to the number of days of diarrheal illness observed and have excluded 525 infants with intermittent missingness.

4 Data Analysis

To analyze the epilepsy dataset presented in Section 3.1, the following covariates are considered: baseline (the 8-week pre-randomization seizure count), age (years), treatment and time (visit), denoted as $B^\star$ $A^\star$, $T^\star$ $t^\star$, respectively. As mentioned in Section 3.1, one patient is observed in Figure 1 to have a seemingly outlying profile. However, Thall and Vail (1990) find no clinical basis to tag the patient as an extreme case. The following analyses of the Epilepsy data therefore use data for all the 59 patients in the study.

Considering no time-varying covariates for the covariance $\mathrm{E}(W_{ist}) = \eta_{ist}$, we fitted the model

$$\begin{aligned} \log(\eta_{is}) &= \beta_{0s} + \beta_1 t_{is}^\star + \beta_2 T_i^\star + \beta_3 (t_{is}^\star \times T_i^\star) + \beta_4 B_i^\star + \beta_5 A_i^\star, \\ \log(\eta_{it}) &= \beta_{0t} + \beta_1 t_{it}^\star + \beta_2 T_i^\star + \beta_3 (t_{is}^\star \times T_i^\star) + \beta_4 B_i^\star + \beta_5 A_i^\star, \\ \log(\eta_{ist}) &= \beta_{0st} + \beta_2 T_i^\star + \beta_5 A_i^\star. \end{aligned} \tag{17}$$

In general, the covariates used for modeling $\log(\eta_{ist})$ may be the same or different from those used for η_{is} and η_{it}. Table 2 shows the parameter estimates and standard errors corresponding to Model 17. The model with time-varying covariate *visit* is

$$\begin{aligned} \log(\eta_{is}) &= \beta_{0s} + \beta_1 t_{is}^\star + \beta_2 T_i^\star + \beta_3 (t_{is}^\star \times T_i^\star) + \beta_4 B_i^\star + \beta_5 A_i^\star, \\ \log(\eta_{it}) &= \beta_{0t} + \beta_1 t_{it}^\star + \beta_2 T_i^\star + \beta_3 (t_{is}^\star \times T_i^\star) + \beta_4 B_i^\star + \beta_5 A_i^\star, \\ \log(\eta_{ist}) &= \beta_{0st} + \beta_1 \Phi(t_{is}^\star, t_{it}^\star) + \beta_2 T_i^\star + \beta_5 A_i^\star, \end{aligned} \tag{18}$$

where $\varphi(t_{is}^\star, t_{it}^\star)$ denotes a function applied to the time-varying covariate *time*, in this case, a difference between *time* at point s and t $(t_{is}^\star - t_{it}^\star)$. Other possibilities for $\varphi(\cdot)$ may include, for example, the lag, ratio, sum, product, etc. From Table 2, the interaction between the visits and treatments is not significant ($p = 0.7977$) despite the fact that the mean profiles in Figure 3 suggest otherwise. The discrepancy between the observation in Section 3.1 and this finding is related to that one patient whose profile seems more extreme relative to the others. Mean profiles based on data without this potentially outlying patient (not shown) also suggested no interaction between the treatments and visits. Considering age as a time-varying covariate when modeling the covariance (Model 18) changed the results slightly but the conclusions remained similar to when only time stationary covariates are used to model the covariance. Table 5 shows the intervals of the minimum and maximum correlations, for the placebo and progabide groups, obtained from fitting (17) and (18). The correlations range over a wide interval with the progabide group having even wider ranges. By modeling the covariance between two measurements using visit as a time-varying covariate and a difference as the time-varying function seems to have a minor impact on the parameter estimates but also on correlations.

To analyze the Jimma study, denote the covariates age, sex, breastfeeding, help and place as $a^\star$, $s^\star$, $b^\star$, $h^\star$ and $p^\star$, respectively. The model fitted considering *age* as

the time-varying covariate is

$$\begin{aligned}
\log[\mathrm{E}(W_{is})] &= \beta_{0s} + \beta_1 a_{is}^\star + \beta_2 s_i^\star + \beta_3 b_i^\star + \beta_4 h_i^\star + \beta_5 p_{i,\text{rural}}^\star + \beta_6 p_{i,\text{semi-urban}}^\star, \\
\log[\mathrm{E}(W_{it})] &= \beta_{0t} + \beta_1 a_{it}^\star + \beta_2 s_i^\star + \beta_3 b_i^\star + \beta_4 h_i^\star + \beta_5 p_{i,\text{rural}}^\star + \beta_6 p_{i,\text{semi-urban}}^\star, \\
\log[\mathrm{E}(W_{ist})] &= \beta_{0st} + \beta_1 \varphi(a_{is}^\star, a_{it}^\star) + \beta_2 s_i^\star + \beta_3 b_i^\star + \beta_4 h_i^\star,
\end{aligned} \tag{19}$$

where $\varphi(a_{is}^\star, a_{it}^\star) = a_{is}^\star - a_{it}^\star$. Table 4 confirms the observations made in Section 3.2, namely, that that the average number of days of diarrheal illness increases as the infants grow older with the female infants having lower counts of days compared to the males. We also find that not breastfeeding is positively related to the number of days of diarrheal illness ($p = 0.0150$) while not seeking medical help is also highly statistically signicant in increasing the number of days of having diarrhea in the infants. Table 6 shows the minimum and maximum estimates of the correlation by gender obtained from fitting (19). The ranges of the correlation are a bit narrower than those from the epilepsy but there are minor differences in the correlation estimates between males and females. Unlike the Epilepsy data case where the minimum and maximum estimates of the correlation seem not to change much over time, the Jimma dataset reflects a decreasing trend in the correlations as the infants get older, in the sense that measurements close together are more correlated than those further apart.

5 Concluding Remarks

In this paper, we have worked on estimating equations that can be used for modeling longitudinal data with the goal of making inference on (sub)populations. These estimating equations model the dependence of the mean response on covariates of interest, without specifying the joint distribution of the vector of responses from a subject. Should scientific interest lie only in the estimation of the so-called population averaged parameters, the approach of Liang and Zeger (1986) is quite sufficient and one need not worry about more involved methods. Because in practice, the method of Liang and Zeger (1986) is limited should interest lie also in the association structure, alternatives have been proposed. For example, Prentice (1988) proposed simultaneous estimation of the marginal mean and association structure permitting inference also on the parameters characterizing the association, in the context of binary data. As has been shown in this paper, the binary case is special as the model for the association is fully determined by the mean and covariance. For count data, however, this issue is a bit more involved and proposed solutions such as in Prentice and Zhao (1991) come to the rescue. They estimate the parameters of the marginal mean and association simultaneously without making the orthogonality assumption made by Zhao and Prentice (1990). This, however, is computationally unbecoming since it involves third- and higher-order moments.

We have presented estimating equations at pair level of the vector of responses for each subject in the context of correlated count data. The proposal incorporates the bivariate Poisson distribution which allows the modeling of the covariance between two measurements. It is formulated such that the variance-covariance matrix of the

outcome variable is not a nuisance but one on which inference can be made while the standard errors are estimated using a sandwich estimator. The method allows for time-stationary as well as time-varying covariates and gives the user the flexibility to determine which function to use for the time-varying covariates. Possibilities may be a lag, ratio, difference, sum, product, etc. A SAS macro has been written to implement this method and is available at `http://ibiostat.be/software/longitudinal` or `http://ibiostat.be/software/count`. Using a 64-bit Windows 8.1 operating system computer with 8GB RAM and 2.80GHz processor, Model 17 converged, based on a dataset of 236 observations, after 14 iterations with a real time of 0.45 seconds. Similarly, (18) took 0.38 seconds (real time) and converged after 14 iterations. Finally, Model 19 was fitted on a dataset of about $46,000$ observations and converged after 13 iterations and 59.39 seconds.

Note that while the standard errors are estimated using the sandwich estimator, the SAS macro reports both the model-based and sandwich (also sometimes known as robust or empirical) standard errors. In the results shown, there is a considerable discrepancy between these two types of standard errors for some parameters. This routinely implies that there are discrepancies between the true and assumed higher-order moments. Fortunately, in such cases, the sandwich estimator provides valid inferences nevertheless.

Table 2 *Epilepsy data: Parameter estimates and standard errors when a time stationary covariate is considered (Model 17).*

		Model-based			Sandwich or empirical		
Effect	Est.	s.e.	χ^2	p	s.e.	χ^2	p
Intercept (β_{0s})	−1.1904	0.2157	30.45	<.0001	1.2655	0.88	0.3468
Intercept (β_{0t})	−1.2119	0.2474	24.00	<.0001	1.2563	0.93	0.3347
Intercept (β_{0st})	1.3148	0.0929	200.17	<.0001	0.4836	7.39	0.0066
visit (β_1)	−0.1724	0.0396	18.95	<.0001	0.0500	11.87	0.0006
trt(Placebo) (β_2)	0.2782	0.0452	37.89	<.0001	0.2490	1.25	0.2638
trt*visit (β_3)	0.0404	0.0390	1.07	0.3004	0.1576	0.07	0.7977
Baseline (β_4)	0.0379	0.0015	629.29	<.0001	0.0088	18.46	<.0001
Age (β_5)	0.0082	0.0027	8.93	0.0028	0.0129	0.40	0.5257

Table 3 *Epilepsy data: Parameter estimates and standard errors when a time-varying covariate is considered (Model 18).*

		Model-based			Sandwich or empirical		
Effect	Est.	s.e.	χ^2	p	s.e.	χ^2	p
Intercept (β_{0s})	−1.3907	0.2221	39.20	<.0001	1.3798	1.02	0.3135
Intercept (β_{0t})	−1.6024	0.2518	40.50	<.0001	1.4621	1.20	0.2731
Intercept (β_{0st})	1.2180	0.0992	150.64	<.0001	0.5099	5.71	0.0169
visit (β_1)	−0.0457	0.0246	3.45	0.0633	0.0468	0.95	0.3289
trt(Placebo) (β_2)	0.3037	0.0448	46.04	<.0001	0.2456	1.53	0.2163
trt*visit (β_3)	0.0066	0.0375	0.03	0.8603	0.1550	0.00	0.9660
Baseline (β_4)	0.0378	0.0015	619.46	<.0001	0.0091	17.27	<.0001
Age (β_5)	0.0085	0.0027	9.72	0.0018	0.0129	0.44	0.5094

Table 4 *Jimma data: Parameter estimates and standard errors when a time-varying covariate is considered (Model 19).*

		Model-based			Sandwich or empirical		
Effect	Est.	s.e.	χ^2	p	s.e.	χ^2	p
Intercept (β_{0s})	−1.3373	0.0090	21968.71	<.0001	0.2232	35.89	<.0001
Intercept (β_{0t})	−0.9529	0.0098	10068.03	<.0001	0.1791	28.31	<.0001
Intercept (β_{0st})	−0.3984	0.0091	3006.32	<.0001	0.1775	5.04	0.0248
age (β_1)	0.1260	0.0007	28725.74	<.0001	0.0083	228.11	<.0001
sex(Female) (β_2)	−0.1732	0.0034	2556.79	<.0001	0.0307	31.93	<.0001
bf(No) (β_3)	0.2379	0.0099	575.10	<.0001	0.0978	5.92	0.0150
help(No) (β_4)	2.0924	0.0038	308352.00	<.0001	0.0380	3034.68	<.0001
place(Rural) (β_5)	−0.0448	0.0064	48.82	<.0001	0.1060	0.18	0.6729
place(Semi-urban)(β_6)	−0.2088	0.0077	744.55	<.0001	0.1095	3.63	0.0566

Table 5 *Epilepsy data: Minimum and maximum correlations from fitting Model 17 (top panel) and Model 18 (bottom panel) for the two treatments.*

	Placebo				Progabide			
Visit	1	2	3	4	1	2	3	4
1	[1.00,1.00]				[1.00,1.00]			
2	[0.18,0.92]	[1.00,1.00]			[0.05,0.92]	[1.00,1.00]		
3	[0.19,0.93]	[0.20,0.93]	[1.00,1.00]		[0.05,0.93]	[0.06,0.94]	[1.00,1.00]	
4	[0.20,0.93]	[0.22,0.94]	[0.23,0.94]	[1.00,1.00]	[0.06,0.93]	[0.06,0.94]	[0.07,0.95]	[1.00,1.00]
1	[1.00,1.00]				[1.00,1.00]			
2	[0.20,0.93]	[1.00,1.00]			[0.05,0.93]	[1.00,1.00]		
3	[0.21,0.93]	[0.21,0.93]	[1.00,1.00]		[0.06,0.93]	[0.06,0.93]	[1.00,1.00]	
4	[0.22,0.94]	[0.22,0.94]	[0.21,0.93]	[1.00,1.00]	[0.06,0.94]	[0.06,0.94]	[0.06,0.93]	[1.00,1.00]

Table 6 *Jimma data: Minimum and maximum correlations from fitting Model 19 by gender.*

Gender	Age	N	0	2	4	6	8	10	12
Female	0	3453	[1.00,1.00]						
	2	3453	[0.27,0.81]	[1.00,1.00]					
	4	3336	[0.18,0.76]	[0.21,0.77]	[1.00,1.00]				
	6	3255	[0.14,0.69]	[0.16,0.71]	[0.18,0.72]	[1.00,1.00]			
	8	3174	[0.11,0.62]	[0.12,0.64]	[0.15,0.66]	[0.17,0.69]	[1.00,1.00]		
	10	3102	[0.08,0.53]	[0.09,0.56]	[0.10,0.58]	[0.12,0.60]	[0.13,0.62]	[1.00,1.00]	
	12	2667	[0.06,0.40]	[0.06,0.47]	[0.07,0.50]	[0.08,0.52]	[0.09,0.54]	[0.10,0.56]	[1.00,1.00]
Male	0	3549	[1.00,1.00]						
	2	3549	[0.27,0.81]	[1.00,1.00]					
	4	3411	[0.19,0.76]	[0.21,0.77]	[1.00,1.00]				
	6	3303	[0.14,0.69]	[0.16,0.71]	[0.20,0.72]	[1.00,1.00]			
	8	3219	[0.10,0.62]	[0.12,0.64]	[0.13,0.66]	[0.15,0.67]	[1.00,1.00]		
	10	3135	[0.08,0.53]	[0.09,0.56]	[0.10,0.60]	[0.11,0.60]	[0.13,0.62]	[1.00,1.00]	
	12	2721	[0.06,0.40]	[0.07,0.47]	[0.08,0.50]	[0.09,0.52]	[0.10,0.54]	[0.11,0.56]	[1.00,1.00]

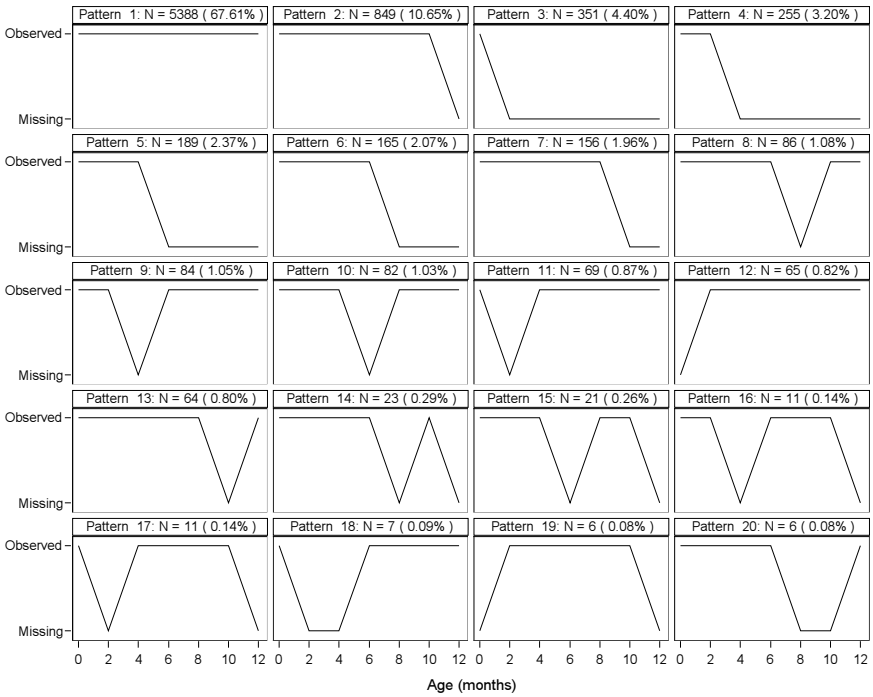

Fig. 7 20 of the 59 missing patterns in the Jimma dataset (N is the number of infants with the pattern under consideration).

References

Agresti A (2002) Categorical Data Analysis, 2nd edn. John Wiley & Sons, New York

Breslow N (1984) Extra-poisson variation in log-linear models. Applied Statistics 33:38–44

Breslow N, Clayton D (1993) Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88:9–25

Crowder M (1995) On the use of a working correlation matrix in using generalised linear models for repeated measures. Biometrika 82:407–410

Diggle P, Heagerty P, Liang K, Zeger S (2002) Analysis of Longitudinal Data, 2nd edn. Oxford Science Publications, Clarendon Press, Oxford

Fahrmeir L, Tutz G (1994) Multivariate Statistical Modelling Based on Generalized Linear Models. Springer-Verlag, Heidelberg

Fahrmeir L, Tutz G (2001) Multivariate Statistical Modeling Based on Generalized Linear Models, 2nd edn. Springer Series in Statistics, Springer-Verlag, New York

Fan J, Huang T, Li R (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. Journal of the American Statistical Association 102(478):632–641, URL `http://dx.doi.org/10.1198/016214507000000095`, pMID: 19707537

Fitzmaurice G, Laird N, Ware J (2004) Applied Longitudinal Analysis. John Wiley & Sons, New York

Hardin J, Hilbe J (2003) Generalized Estimating Equations. Chapman and Hall/CRC, Boca Raton

Iddi S, Molenberghs G (2013) A marginalized model for zero-inflated, overdispersed and correlated count data. Electronic Journal of Applied Statistical Analysis 6:149–165

Karlis D (2003) An em algorithm for multivariate poisson distribution and related models. Journal of Applied Statistics 30:63–77

Kassahun W, Neyens T, Molenberghs G, Faes C, Verbeke G (2012) Modeling overdispersed longitudinal binary data using a combined beta and normal random-effects model. Archives of Public Health 70, issue: 1, Chapter: 7, Article: 7

Kim H, Shults J (2010) %qls SAS macro: A SAS macro for analysis of correlated data using quasi-least squares. Journal of Statistical Software 35:1–22

Kocherlakota S, Kocherlakota K (1992) Bivariate Discrete Distributions. CRC Press, Boca Raton

Kocherlakota S, Kocherlakota K (2001) Regression in the bivariate poisson distribution. Communications in Statistics, Theory & Methods 30:815–825

Lakshminarayana J, Pandit S, Rao K (1999) On a bivariate poisson distribution. Communications in Statistics, Theory & Methods 28:267–276

Lambert D (1992) Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34:1–13

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

Lee Y, Nelder J (2004) Conditional and marginal models: Another view. Statistical Science 19:219–228

Leng C, Zhang W, Pan J (2010) Semiparametric mean-covariance regression analysis for longitudinal data. Journal of the American Statistical Association 105(489):181–193, URL `http://dx.doi.org/10.1198/jasa.2009.tm08485`

Leppik I, Dreifuss F, Bowman-Cloyd T, et al (1985) A double-blind crossover evaluation of progabide in partial seizures. Neurology 35:285

Lesaffre E, Asefa M, Verbeke G (1999) Assessing the goodness-of-fit of the Laird and Ware model-an example: the Jimma infant survival differential longitudinal study. Statistics in Medicine 18:835–854

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

Liang K, Zeger S, Qaqish B (1992) Multivariate regression analyses for categorical data. Journal of the Royal Statistical Society, Series B 54:3–40

Lindsey J, Lambert P (1998) On the appropriateness of marginal models for repeated measurements in clinical trials. Statistics in Medicine 17:447–469

Lipsitz S, Laird N, Harrington D (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. Biometrika 78:153–160

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

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

Molenberghs G, Verbeke G, Demétrio C (2007) An extended random effects approach to modeling repeated overdispersed count data. Lifetime Data Analysis 13:513–531

Molenberghs G, Verbeke G, Demétrio C, Vieira A (2010) A family of generalized linear models for repeated measures with normal and conjugate random effects. Statistical Science 25:325–347
Nelder J, Wedderburn R (1972) Generalized linear models. Journal of the Royal Statistical Society B 135:370–384
Prentice R (1988) Correlated binary regression with covariates specific to each binary observation. Biometrics 44:1033–1048
Prentice R, Zhao L (1991) Estimating equations for parameters in means and covariances of multivariate discrete and continuous response. Biometrics 47:825–839
Ridout M, Hinde J, Demetrio C (2001) A score test for a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. Biometrics 57:219–233
Solis-Trapala I, Farewell V (2005) Regression analysis of overdispersed correlated count data with subject specific covariates. Statistics in Medicine 24:2557–2575
Sun W, Shults J, Leonard M (2009) A note on the use of unbiased estimating equations to estimate correlation in analysis of longitudinal trials. Biometrical Journal 51:5–18
Thall P, Vail S (1990) Some covariance models for longitudinal count data with overdispersion. Biometrics 46:657–671
Wang Y, Carey V (2004) Unbiased estimating equations from working correlation models for irregularly timed repeated measures. Journal of the American Statistical Association 99:845–852
Winkelmann R (2008) Econometric Analysis of Count Data. Springer, Berlin
Wolfinger R, O'Connell M (1993) Generalized linear mixed models: A pseudo-likelihood approach. Journal of Statistical Computing and Simulation 48:233–243
Ye H, Pan J (2006) Modelling of covariance structures in generalized estimating equations for longitudinal data. Biometrika 93:927–941
Zhao L, Prentice R (1990) Correlated binary regression using a quadratic exponential model. Biometrika 77:642–648
Ziegler A (2011) Generalized Estimating Equations. Springer, New York