Statistical Modelling 2003; **3**: 99–108

Likelihood-based analysis of longitudinal count data using a generalized Poisson model

Peter J Toscas[1] and Malcolm J Faddy[2]
[1]CSIRO Mathematical and Information Sciences, Cleveland, Australia
[2]School of Mathematics and Statistics, University of Birmingham, Birmingham, UK

Abstract: Models based on a generalization of the simple Poisson process are discussed and illustrated with an analysis of some longitudinal count data on frequencies of epileptic fits. The models enable a broad class of discrete distributions to be constructed, which cover a variety of dispersion properties that can be characterized in an intuitive and appealing way by a simple parameterization. This class includes the Poisson and negative binomial distributions as well as other distributions with greater dispersion than Poisson, and also distributions underdispersed relative to the Poisson distribution. Comparing a number of analyses of the data shows that some covariates have a more significant effect using this modelling than from using mixed Poisson models. It is argued that this could be due to the mixed Poisson models used in the other analyses not providing an appropriate description of the residual variation, with the greater flexibility of the generalized Poisson modelling generally enabling more critical assessment of covariate effects than more standard mixed Poisson modelling.

Key words: longitudinal data; negative binomial distribution; overdispersion; Poisson distribution; transition models; underdispersion

Data and Programs available from: http://stat.uibk.ac.at/SMIJ
Received: December 2000; **revised**: September 2001, August 2002, October 2002; **accepted**: October 2002

1 Introduction

There has been much attention in recent years on analysis of count data that show overdispersion relative to the Poisson distribution. Liang and Zeger (1986) and Zeger and Liang (1986) presented a generalized estimating equation approach, which can be applied to both count and noncount data. This approach is robust in that any distribution for the data does not have to be completely specified, only the first- and second-order moments being used, but these may not necessarily be good descriptors of the underlying distribution, particularly if the distribution is strongly skewed.

Generalized linear mixed models can also be used, where additional normal random effects are included in the linear predictor to account for the overdispersion. Serial correlations for longitudinal data can be incorporated by specifying appropriate covariances for the random effects, as with the standard linear mixed model. Breslow and Clayton (1993) used a penalized quasi-likelihood approach to estimation using numerical approximations for the mixed distributions.

Address for correspondence: CSIRO Mathematical and Information Sciences, PO Box 120, Cleveland, QLD 4163, Australia. E-mail: peter.toscas@csiro.au

10.1191/1471082X03st050oa

Hougaard and colleagues (1997) took an approach to model count data by describing the counts in terms of Poisson processes with randomly distributed rate parameters. They proposed a three-parameter distribution for these rates that includes the Gamma and the inverse Gaussian as special cases, and they accounted for between-subject variation by this distribution of rates. These models are also mixed models, but with a variety of different mixing distributions possible. However, Hougaard *et al.* (1997) did not include any covariate dependence in their paper.

In this paper a methodology suggested by Faddy (1997a) (see also Bosch and Ryan, 1998) is extended to model between-subject variation in a longitudinal analysis of count data in the presence of a number of covariates. The serial dependence is described by *transition models* (Diggle *et al.*, 1994, Chapter 10) where the conditional probabilities of temporal transitions are modelled rather than the marginal probability distributions. The construction of the probability distributions is done by taking a stochastic process in continuous time and determining the transient probabilities at an arbitrary time. The resulting probability distributions form a very general class including distributions underdispersed relative to the Poisson, the Poisson distribution itself, and distributions overdispersed relative to the Poisson including the negative binomial distribution. The use of such a broad class of distributions can contribute to more robust inferential procedures as the modelling allows for a more flexible assessment of residual dispersion.

2 The models

Faddy (1997a) constructed discrete probability distributions by considering a generalization of the Poisson process, or the Markov pure birth process $\{Y(t); t \geq 0\}$ with $Y(0) = 0$, defined by:

$$\mathrm{P}\{Y(t+\delta t) = n+1 \,|\, Y(t) = n\} = \lambda_n \delta t + o(\delta t)$$

and (2.1)

$$\mathrm{P}\{Y(t+\delta t) = n \,|\, Y(t) = n\} = 1 - \lambda_n \delta t + o(\delta t)$$

where λ_n are rate parameters. From this definition it can be seen that the probability of an increase occurring in the time interval $(t, t+\delta t)$ is dependent on the current state of the process at time t. When λ_n is constant (i.e., no such dependence), the simple Poisson process results with Poisson distributed $Y(t)$. Ball (1995) showed that when this rate is increasing with n the distribution of $Y(t)$ is overdispersed relative to the Poisson distribution and when it is decreasing as n increases the distribution of $Y(t)$ is underdispersed relative to the Poisson distribution. In particular, if λ_n is *linearly* increasing then the negative binomial distribution results. Other useful forms for λ_n will be suggested, leading to a general family of discrete distributions being developed.

Faddy (1997a) points out that the probabilities $p_n(t) = \mathrm{P}\{Y(t) = n\}$, for $n = 0, 1, 2, \ldots$, can be calculated using a matrix exponential formulation of the solution of (2.1) (Cox and Miller, 1965, p. 182):

$$\begin{pmatrix} p_0(t) & p_1(t) & p_2(t) & \cdots & p_n(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \exp(\mathbf{Q}t) \qquad (2.2)$$

where the matrix $\mathbf{Q}$ has the form:

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\lambda_2 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -\lambda_n \end{pmatrix}$$

and n here is arbitrary but finite. Notice that because the process defined by (2.1) only allows increases in $Y(t)$, equation (2.2) gives the exact probabilities $p_0(t), p_1(t), \ldots, p_n(t)$, for *any* n. The matrix exponential in (2.2) can be calculated numerically using routines that are generally available (e.g., MATLAB®, 1996 or GAUSS™, 1996). The time t in (2.2) may be taken without loss of generality to be 1.

Faddy (1997a) also showed that for *any* discrete distribution with support on $\{0, 1, 2, \ldots\}$ there is an equivalent sequence $\lambda_0, \lambda_1, \lambda_2, \ldots$ such that the distribution of $Y(1)$ from (2.2) is the same as this discrete distribution. This generality of the formulation gives the resulting models considerable appeal in data analysis. With this in mind, the following simple form for the rate parameters $\lambda_n (n \geq 0)$ was suggested by Faddy (1997a):

$$\lambda_n = a(b + n)^c \tag{2.3}$$

where $b > 0$, $a > 0$, and $c \leq 1$. When $c = 0$ in (2.3) the Poisson distribution results from (2.2), and for $c = 1$ the negative binomial distribution. For other values of c, distributions with a variety of dispersion properties can be constructed including underdispersion relative to the Poisson distribution ($c < 0$) as well as overdispersion ($c > 0$). The restriction $c \leq 1$ is necessary to produce *honest* probability distributions from (2.2) and (2.3), but nevertheless a very broad class of distributions is constructed using this process (see Faddy, 1997b, for another more complicated form of λ_n that admits convex increasing sequences).

Equation (2.2) gives the probabilities in terms of the λ_n parameterization (2.1), and moments would have to be calculated numerically as general expressions for them are not available. However, a diffusion approximation to the process (2.1) from Faddy (1997a) does give an approximation for the mean:

$$E\{Y(1)\} \approx b\left[\left\{1 + \frac{a(1-c)}{b^{1-c}}\right\}^{\{1/(1-c)\}} - 1\right]. \tag{2.4}$$

See Faddy (1997a) for an approximation to the variance of $Y(1)$.

The expression in (2.4) gives the exact mean for the Poisson ($c = 0$) and negative binomial ($c \to 1$) distributions but is an approximation for other c values. Faddy (1997a) used such a diffusion approximation to the general process (2.1) to demonstrate that a *concave* increasing sequence of λ_n results in a distribution of $Y(1)$ that has dispersion between that associated with the Poisson and negative binomial

distributions, while a *convex* increasing sequence gives even more variation than the negative binomial distribution.

By setting the expression on the right-hand side of (2.4) equal to the mean, μ say, and solving for a, the equation

$$a = \frac{(\mu + b)^{(1-c)} - b^{(1-c)}}{(1-c)} \tag{2.5}$$

is obtained. Replacing a in (2.3) with the expression in (2.5) and using (2.2) gives a distribution with approximate mean μ.

In the context of longitudinal data observed from a number of subjects over time, the rate for subject i at time j from (2.3) may be represented as

$$\lambda_{ij,n} = a_{ij}(b_j + n)^{c_j} \tag{2.6}$$

or using (2.5) and the mean μ_{ij} for subject i at time j, put

$$a_{ij} = \frac{(\mu_{ij} + b_j)^{(1-c_j)} - b_j^{(1-c_j)}}{(1-c_j)} \tag{2.7}$$

in (2.6) and then use (2.2). Note that distributions with different means are constructed here using (2.2) for each time and subject, with the b_j and c_j parameters accounting for the random between-subject variation at time j.

The mean μ_{ij} could be represented in terms of any covariates as:

$$\log(\mu_{ij}) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\gamma} \tag{2.8}$$

where $\mathbf{x}_i^{\mathrm{T}}$ is a row vector containing nontime-dependent covariates for subject i and $\mathbf{z}_{ij}^{\mathrm{T}}$ is a row vector containing any time-dependent covariates for subject i at time j; in this formulation the vector $\mathbf{z}_{ij}^{\mathrm{T}}$ will typically involve the subject's observations from previous times, so that the appropriate conditional distribution can be constructed. The inclusion of $\mathbf{z}_{ij}^{\mathrm{T}}$ here to incorporate longitudinal dependence results in models similar to the transition models of Diggle *et al.* (1994, Chapter 10) but with more general conditional distributions for the residual variation.

This formulation (2.8), (2.7), and (2.6) has used the approximate form for the mean given in (2.4), and can be regarded as a reparameterization. The effect of the approximation is only that the exact means will *not* be log-linear as specified by (2.8), and as such may be of little concern. Shown in Figure 1 are plots of the exact means, calculated numerically using the probabilities (2.2) for suitably large n, as functions of the linear predictor (right-hand side of (2.8)) over a representative range: it can be seen that these means are not far from being log-linear (the b and c parameter values that were used here corresponded to those estimated from the data analyses presented in the next section).

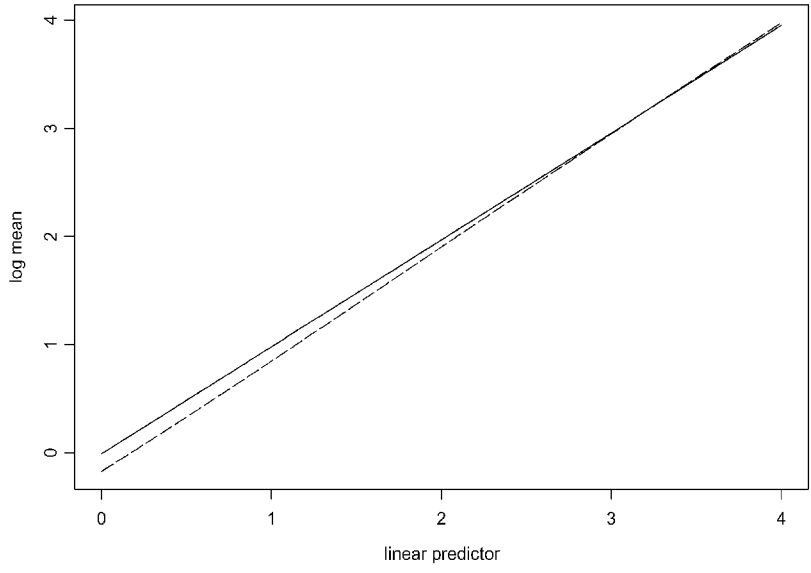

Figure 1 Exact means as functions of linear predictor: $b = 1.83$, $c = 0.74$: —; and $b = 0.046$, $c = 0.16$: – – –

3 Example

In this section the modelling is applied to data from Thall and Vail (1990) on a clinical trial of 59 epileptics, 31 of whom were randomly assigned to a treatment group and received the anti-epileptic drug progabide, with the remaining 28 given a placebo. Before treatment began the numbers of seizures suffered by each subject over an eight-week period were recorded for baseline purposes. After treatment began the numbers of seizures were recorded as responses over four successive two-week periods, giving longitudinal data, so appropriate time-dependent covariates will have to be considered. The main aim of the study was to assess the effectiveness of the treatment, while allowing for a number of confounding factors.

All the models fitted to these data had four common nontime-dependent covariates, in addition to the constant term: treatment, log(age), log{(baseline count)/4} and treatment × log{(baseline count)/4}. Time-dependent covariates were included in the analyses in two ways. First, *cf.* Model (3) of Chapter 10 of Diggle *et al.* (1994), where the lag k covariates for subject i at time j:

$$z_{ij(k)} = \log\left(\frac{y_{i,j-k} + 0.1}{\bar{y}_{j-k} + 0.1}\right), \quad \text{for } j = k + 1, \ldots, 4 \text{ and } k = 1, 2, 3$$

where $y_{i,j}$ is the seizure count for subject i in the jth two-week period and $\bar{y}_j$ is the jth two-week period mean over subjects, were used. Secondly, a covariate based on the average of a subject's previous counts:

$$z_{ij} = \log\left(\frac{\sum_{k=1}^{j-1}\frac{y_{i,k}}{j-1} + 0.1}{\sum_{k=1}^{j-1}\frac{\bar{y}_k}{j-1} + 0.1}\right),$$

for subject i at time j was used. Here, 0.1 has been arbitrarily added because some of the subjects had zero counts in some of the two-week periods; however, parameter estimates were not overly influenced by the choice of this value. Intuitively, this second approach to constructing the longitudinal dependence might be similar to that of a mixed model, as the averages of the subjects' previous counts would be informative about any random effect over subjects.

For any parameterization $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ specifying the covariate dependence (2.8), the necessary $\lambda_{ij,n}$ can be determined from (2.6) and (2.7) for each patient and time period, and the (conditional) probabilities of their responses calculated using (2.2). The product of these probabilities will then give the likelihood of the data, and iterative maximum likelihood estimation of all parameters including the b_j and c_j in (2.6) can be carried out (GAUSS^{TM}, 1996). Table 1 shows the results of maximum likelihood fitting of a number of models described by (2.8), (2.7), (2.6), and (2.2).

The most general model in Table 1 is M1: in this model all three possible time-dependent covariates are included. Model M2 differs from model M1 by not having the lag 3 time-dependent covariate: this effect is not significant ($2 \times$ log-likelihood ratio $= 0.90$ on 1 d.f.). However, the effect of the lag 2 term is more significant ($2 \times$ log-likelihood ratio $= 3.59$ on 1 d.f. with p-value ≈ 0.06, comparing M3 with M2), and that of the lag 1 term highly significant ($2 \times$ log-likelihood ratio $= 19.35$ on 1 d.f. with p-value < 0.001, comparing M4 with M3). Both the resulting serial correlations were positive. The second and fourth two-week periods resulted in estimates of the c-parameter between 0 and 1, suggesting that the corresponding data are (conditionally)

Table 1 Models fitted to the epilepsy data

Model identification	b-Estimates	c-Estimates	Lag terms included	Log-likelihood
M1	$\hat{b}_1 = 2.62$, $\hat{b}_2 = 2.00$ $\hat{b}_3 = 2.01$, $\hat{b}_4 = 0.064$	$\hat{c}_1 = 1$, $\hat{c}_2 = 0.77$ $\hat{c}_3 = 1$, $\hat{c}_4 = 0.16$	1, 2, and 3	−629.10
M2	$\hat{b}_1 = 2.70$, $\hat{b}_2 = 1.83$ $\hat{b}_3 = 2.04$, $\hat{b}_4 = 0.046$	$\hat{c}_1 = 1$, $\hat{c}_2 = 0.74$ $\hat{c}_3 = 1$, $\hat{c}_4 = 0.16$	1 and 2	−629.55
M3	$\hat{b}_1 = 2.81$, $\hat{b}_2 = 1.67$ $\hat{b}_3 = 2.02$, $\hat{b}_4 = 0.042$	$\hat{c}_1 = 1$, $\hat{c}_2 = 0.72$ $\hat{c}_3 = 1$, $\hat{c}_4 = 0.17$	1	−631.34
M4	$\hat{b}_1 = 3.33$, $\hat{b}_2 = 1.28$ $\hat{b}_3 = 1.62$, $\hat{b}_4 = 0.089$	$\hat{c}_1 = 1$, $\hat{c}_2 = 0.70$ $\hat{c}_3 = 1$, $\hat{c}_4 = 0.26$	None	−641.02
M5	$\hat{b}_1 = 2.48$, $\hat{b}_2 = 1.88$ $\hat{b}_3 = 2.05$, $\hat{b}_4 = 0.0013$	$\hat{c}_1 = 1$, $\hat{c}_2 = 0.76$ $\hat{c}_3 = 1$, $\hat{c}_4 = 0.096$	Average	−629.47

overdispersed relative to the Poisson distribution but not as dispersed as a negative binomial distribution.

For the case where the time-dependent covariate was based on the average of a subject's previous counts the model in Table 1 is M5. The log-likelihood for this model is comparable to the earlier model M2, and the estimates of c_2 and c_4 were also between 0 and 1.

Some reduction in c as the trial progressed might have been expected as more information became available on the subjects for inclusion in the conditional distributions. Fitting models with declining c_i such as $c_i = 1 - \theta(i-1)^{\phi}$ for $i = 1, 2, 3$, and 4 gave maximized log-likelihoods ≥ -630.43 (M2) and ≥ -630.36 (M5) for $\phi \geq 5$, close to those shown in Table 1. However, estimating both θ and ϕ here gave $\hat{\theta} \to 0$ and $\hat{\phi} \to \infty$ with $\hat{c}_1 = \hat{c}_2 = \hat{c}_3 = 1$ and $\hat{c}_4 < 1$.

The generalized Pearson statistic values $\sum(\text{observed count} - \text{mean})^2/\text{variance}$ for models M2 and M5 are listed in Table 2; these are both large relative to their degrees of freedom (note that here *exact* means and variances were calculated using the probabilities from (2.2)).

An examination of probability estimates from these models reveals that subjects 227, 207 and, in particular, 225 have anomalously high observations for the third, first, and second two-week periods, respectively. The probability of greater than or equal to the observed value for the second two-week period of subject 225 is less than 10^{-4}, while the corresponding probabilities for the anomalous observations of the other two subjects are closer to 0.01. These three observations contribute 76.16 (M2) and 101.97 (M5) to the generalized Pearson statistic values in Table 2. The poor fit of these models to the epilepsy data can be attributed to these three observations, especially the second two-week period observation for subject 225, which contributes 46.82 (M2) and 68.81 (M5) to these generalized Pearson statistics. Table 2 in Thall and Vail (1990) also reveals that these three observations are large relative to what would have been expected from the baseline covariable. In addition, Breslow and Clayton (1993) highlighted subjects 227 and 225 as having high 'overall count levels relative to expectation based on covariables.'

Table 2 Generalized Pearson statistics for models fitted to the epilepsy data

Model	Generalized Pearson statistic	Degrees of freedom
M2	276.40	227
M5	288.60	228

Table 3 Generalized Pearson statistics for models fitted to the epilepsy data with observations for subject 225 removed

Model	Generalized Pearson statistic	Degrees of freedom
M2(−225)	248.46	223
M5(−225)	244.85	224

Table 4 Nontime-dependent covariate coefficient estimates and their approximate standard errors, for models with subject 225 removed

	Model	
Parameters	M2(−225)	M5(−225)
Constant	−0.52 (0.71)	−0.09 (0.70)
Treatment	−0.89 (0.24)	−0.75 (0.24)
log(baseline/4)	0.71 (0.08)	0.63 (0.08)
log(age)	0.39 (0.20)	0.29 (0.19)
Treatment × log(baseline/4)	0.33 (0.10)	0.27 (0.10)

With all four observations of subject 225 removed from the analysis, the generalized Pearson statistic values for models M2(−225) and M5(−225) are as listed in Table 3. The results in Table 3 suggest that with subject 225 removed, both models fit the data reasonably well. For both models M2(−225) and M5(−225) the estimates of c_2 and c_4 were still between 0 and 1.

Table 4 lists the parameter estimates for the nontime-dependent covariates considered by Thall and Vail (1990) and their standard errors (determined from the usual asymptotic maximum likelihood theory using the inverse of the hessian matrix) for models M2(−225) and M5(−225). The results reveal that the coefficients for the covariates treatment, log(baseline/4) and treatment × log(baseline/4) are all significantly different from zero (p-values <0.01). However, the log(age) covariate does not have such a significant effect, with p-value approximately 0.06 from model M2(−225) and 0.12 from model M5(−225). If log(age) is dropped from these models then the other factors, including the interaction, remain significant with p-values <0.03. In comparison, the best-fitting model of Thall and Vail (1990), model 22 in their Table 4, and the random effects models of Breslow and Clayton (1993), suggest that both age and the interaction between baseline and treatment are non-significant covariates. The significance of the interaction between baseline and treatment in the models described here is an interesting inference, as it suggests that the treatment effect is less for those patients with high baseline numbers of seizures.

Model 22 of Thall and Vail (1990) and Model II of Breslow and Clayton (1993) both reveal a significant drop in the *mean* number of epileptic fits suffered by the subjects in the fourth two-week period. The incorporation of such a covariate for the fourth two-week period into models M2 and M5 (Table 1) and M2(−225) and M5(−225) had only a small effect ($2 \times$ log-likelihood ratio ≤ 0.42 on 1 d.f.), so that the only significant effect due to this fourth period using the modelling described here is on the *dispersion.* This is an intriguing aspect of the analysis: it is conceivable that these differences are due to some sort of 'trade-off' between mean and dispersion properties of the data in the model fitting process, and the more flexible mean/dispersion properties of the models described here have admitted these possibilities. The models of Breslow and Clayton (1993) do have a more rigid structure involving normal random effects, and they do not fit the data as well as any of the models in Table 1: log-likelihood – number of estimated parameters being −664.87 for Model IV of Breslow and Clayton (1993) compared with −642.55 for model M2.

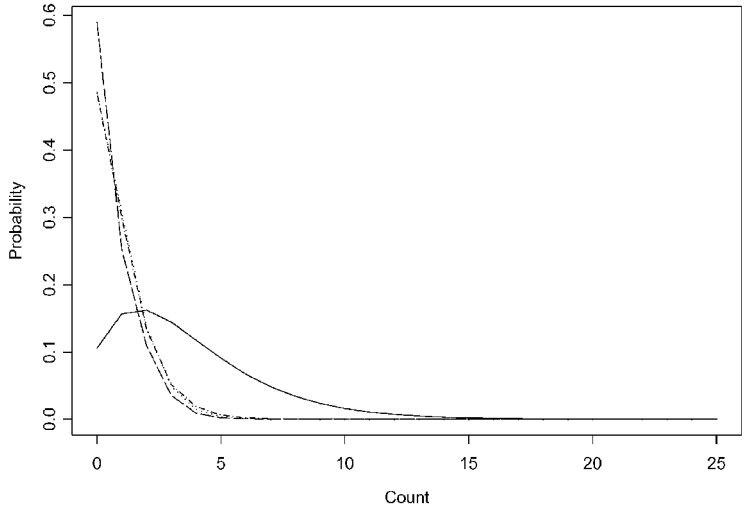

Figure 2 The estimated two-week period probabilities for subject 232 from model M5(−225): first: —, second: · · ·, third: –·–, and fourth: – – –

A feature of the probability distributions estimated from these models is their change over time. Breslow and Clayton (1993) identify subject 135 as one who improved over time and subject 232 as one with low or zero counts. The estimated (conditional) probabilities from model M5(−225) of seizure counts of 10 or more for subject 135 dropped from 0.1 in the first two-week period to 0.02 in the fourth. Figure 2 reveals that for subject 232, from the second two-week period the conditional probability of this subject suffering more than one or two seizures in a two-week period had dramatically fallen.

4 Concluding remarks

The modelling described in this paper was based on constructing discrete distributions in terms of a Markov process (2.1). The theoretical evidence for the shape of the sequence of rate parameters in (2.1) influencing the dispersion of the resulting distribution from (2.2) is not complete (Ball, 1995; Faddy, 1997a) but, when combined with empirical comparison of the data and fitted distributions described in Faddy (1997a,b), the evidence is certainly compelling.

One feature of the data analysis presented in section 3 is that there was evidence from the data that (conditional) distributions less dispersed than the negative binomial were required. This is an important observation, as the use of normally distributed random effects in the log-linear predictor (2.8) may result in distributions with even more dispersion than the negative binomial (numerical investigations of log-normally mixed Poisson distributions do suggest that the equivalent sequences of rates λ_n ($n \geq 0$) from (2.1) are convex increasing in n). So routine use of such random effects in the log-linear predictor may seriously overstate the residual dispersion, and it is interesting to note

that in analyses using mixed models (Thall and Vail, 1990; Breslow and Clayton, 1993) the estimated coefficients of the covariates in (2.8) were generally larger than those reported here, and their relative precision considerably less with fewer significant effects apparent. These differences could be due to the latter analyses being based on a formulation that uses an inappropriate description of the residual variation, whereas the approach adopted here has allowed more flexible modelling of the mean and variance. Later models of Breslow and Clayton (1993) do admittedly include additional terms that allow the residual variation to change with time period, but these do not permit the range of dispersion properties offered by the models described here.

Alternative estimation procedures such as generalized estimating equations do enable analyses based on models with a wide range of dispersion properties to be carried out, but Crowder (1995) has pointed out that this can result in inconsistent parameter estimates if the 'working' correlation matrix is inappropriately defined.

Finally, the generality of the models described in this paper cannot be understated with a great range of distributional properties covered by the formulation (2.3). The equivalent of differently distributed random effects could be incorporated by allowing the parameters b and c to vary according to a particular design configuration, in much the same way that the analysis presented in section 3 allowed different distributions for the different time periods in the study. However, the precise correspondence between distributions constructed from (2.1), (2.2), and (2.3) with $c > 0$ and mixed Poisson distributions is not immediately obvious and would benefit from further study.

Acknowledgements

The preparation of this paper has benefited from discussions with Richard Morton, David Smith, and You-Gan Wang. Thanks are also due to the referees and editor for many helpful suggestions.

References

Ball F (1995) A note on variation in birth processes. *Mathematical Scientist*, **20**, 50–55.

Bosch RJ, Ryan LM (1998) Generalised Poisson models arising from Markov processes. *Statistics and Probability Letters*, **39**, 205–12.

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

Cox DR, Miller HD (1965) *The Theory of Stochastic Processes.* London: Methuen.

Crowder MJ (1995) On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, **82**, 407–10.

Diggle PJ, Liang K-Y, Zeger SL (1994) *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

Faddy MJ (1997a) Extended Poisson process modelling and analysis of count data. *Biometrical Journal*, **39**, 431–40.

Faddy MJ (1997b) On extending the negative binomial distribution, and the number of weekly winners of the UK national lottery. *Mathematical Scientist*, **22**, 77–82.

GAUSS™ Command Reference (1996) Maple Valley, Washington: Aptech Systems, Inc.

Hougaard P, Lee M-LT, Whitmore GA (1997) Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics*, **53**, 1225–38.

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

MATLAB® (1996) Using MATLAB. Natick, Massachusetts: The Math Works Inc.

Thall PF, Vail SC (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–71.

Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–30.