Mixed Poisson Likelihood Regression Models for Longitudinal Interval Count Data

Peter F. Thall

Biometrics, Vol. 44, No. 1 (Mar., 1988), 197-209.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198803%2944%3A1%3C197%3AMPLRMF%3E2.0.CO%3B2-J

Biometrics is currently published by International Biometric Society.

BIOMETRICS 44, 197–209
March 1988

Mixed Poisson Likelihood Regression Models for Longitudinal Interval Count Data

Peter F. Thall

The Biostatistics Center, Department of Statistics/Computer and Information Systems, George Washington University, 6110 Executive Boulevard, Suite 750, Rockville, Maryland 20852, U.S.A.

SUMMARY

In many longitudinal studies it is desired to estimate and test the rate over time of a particular recurrent event. Often only the event counts corresponding to the elapsed time intervals between each subject's successive observation times, and baseline covariate data, are available. The intervals may vary substantially in length and number between subjects, so that the corresponding vectors of counts are not directly comparable. A family of Poisson likelihood regression models incorporating a mixed random multiplicative component in the rate function of each subject is proposed for this longitudinal data structure. A related empirical Bayes estimate of random-effect parameters is also described. These methods are illustrated by an analysis of dyspepsia data from the National Cooperative Gallstone Study.

1. Introduction

In many longitudinal studies the exact occurrence times of a particular recurrent event of interest are not available. Rather, only the counts corresponding to the intervals between each subject's successive observation times are known, along with possibly some baseline covariate data. It is often the case that both the total number of counts and their specific observation times for each subject vary substantially across the sample. A basic difficulty in analyzing such data is that the counts are not directly comparable between subjects, since they do not correspond to the same time intervals.

This type of data frequently arises in clinical trials, where the recurrence rate of an important nonfatal event may serve as an index of disease morbidity. Examples are episodes of dyspepsia in patients with gallstones, hypoglycemia in diabetics, angina pectoris in coronary patients, and seizures in epileptics. Patients are usually required to report the number of episodes occurring between clinic visits. In practice, patients are early, late, or miss scheduled visits, and each patient's data are subject to right censoring. A central issue is whether the time-dependent event rate differs between treatment groups. In any case, an estimator of the covariate-adjusted rate function is desired.

The data structure for a single subject may be described formally as follows. Let $0 = \tau_0 < \tau_1 < \cdots < \tau_J$ denote the successive observation times and N_j the event count reported at τ_j, corresponding to the preceding interval $A_j = (\tau_{j-1}, \tau_j]$, $j = 1, \ldots, J$. Thus, A_j has length $d_j = \tau_j - \tau_{j-1}$ and midpoint $m_j = (\tau_{j-1} + \tau_j)/2$. Denote $\mathbf{N} = (N_1, \ldots, N_J)^T$ and $\mathbf{A} = (A_1, \ldots, A_J)^T$ for convenience, with $\mathbf{X}$ a p-vector of baseline covariates. Subjects will be represented by the subscript i, so that N_{ij} is the jth count of the ith subject, etc.

Key words: Clinical trials; Generalized linear models; Longitudinal count data; Maximum likelihood; Nonlinear regression.

Thall and Lachin (1988) discuss nonparametric methods for analysis of this data structure in the absence of covariates. Korn and Whittemore (1979) analyze longitudinal binary response data with covariates by estimating logistic regression coefficients for each subject and then averaging these across the sample. This two-stage method is generalized by Stiratelli, Laird, and Ware (1984), who develop an empirical Bayes approach. Gilmour, Anderson, and Rae (1985) derive a quasi-likelihood for binomial data with covariates by averaging over multivariate normally distributed random effects. Hinde (1982) and Breslow (1984) each take a similar approach to univariate Poisson count data with covariates $\mathbf{X}$, including an unobserved normally distributed random effect in $\mathbf{X}$ to account for so called "extra-Poisson" variation not explained by the observed covariates. Mixed Poisson likelihoods and Poisson regression models have been employed by a number of other authors in various contexts. Weber (1971) uses iteratively reweighted least squares (IRLS) to fit a mixed homogeneous Poisson process to traffic accident count data. More general formulations of Poisson regression models showing the equivalence of IRLS and the Fisher scoring method of maximum likelihood have been given by Frome, Kutner and Beauchamp (1973) and Frome (1983).

The likelihood employed here may be regarded as a mixed generalized linear model (GLM) having multivariate response. Since their introduction by Nelder and Wedderburn (1972), the class of GLMs has been successively enlarged by a number of authors. These include Wedderburn (1974), who introduced quasi-likelihoods, Thompson and Baker (1981), Jorgensen (1983), and Green (1984), who provides a very general proof of the equivalence of Fisher scoring and IRLS in maximum likelihood regression models. More recently, Liang and Zeger (1986) and Zeger and Liang (1986) have adapted the quasi-likelihood approach to longitudinal data. In addition to dealing with the statistical problems at hand, it is my intention here to further illustrate via this application the breadth and utility of the ideas developed by the authors cited above. In particular, maximum likelihood regression methods can be applied to longitudinal data having inherent dependencies within each response vector.

For the present problem, the criteria shall be that the model (i) account for variability between patients, (ii) provide an explicit representation of the dependence among the counts of each patient, (iii) allow the event rate to vary as a function of time, (iv) be sufficiently tractable and flexible to allow broad application, especially in the clinical trial setting described, and (v) incorporate covariate data. The likelihood given in Section 2 adapts the Poisson likelihood regression approach by incorporating a multiplicative variance component γ into each patient's event rate process and mixing according to the distribution of γ. Parameter estimation is discussed in Section 3, including solution of the likelihood equations, estimation of the covariate-adjusted rate function, analysis of deviance, and a related empirical Bayes approach to estimation of random (subject)-effect parameters. An illustration applying these methods to data from the National Cooperative Gallstone Study is presented in Section 4. Computational aspects are discussed briefly in Section 5.

2. The Likelihood

Regard each subject's sequence of events as a doubly stochastic Poisson process (cf. Cox, 1955). Specifically, the rate process of the ith subject shall be assumed to be of the form

$$\gamma_i \lambda(t; \boldsymbol{\beta}, \mathbf{X}_i), \quad i = 1, \ldots, n, \tag{2.1}$$

where $\gamma_1, \ldots, \gamma_n$ are iid unobserved nonnegative random subject effects with λ an explicit function of time t, the baseline covariate vector $\mathbf{X}_i$, and a parameter vector $\boldsymbol{\beta}$. In the standard GLM formulation for univariate Poisson response data (cf. McCullagh and Nelder, 1983, Chaps 2 and 6) $\mathbf{X}_i$ and $\boldsymbol{\beta}$ are of the same dimension and λ_i depends on them only

through their inner product $\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}$, usually via the log-linear link $\lambda_i = \exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})$. In general, however, $\boldsymbol{\beta}$ may contain elements that characterize the functional shape of λ aside from $\mathbf{X}_i$, and λ may take on any nonnegative differentiable form appropriate to the given situation, provided the model is identifiable.

To adapt this to interval count data, define the cumulative hazard

$$\Lambda_i(A) = \int_A \lambda(t; \boldsymbol{\beta}, \mathbf{X}_i)\, dt \tag{2.2}$$

for any time interval A. Denote $\Lambda_{ij} = \Lambda_i(A_{ij})$, corresponding to the interval preceding the jth visit. Although λ may vary with time, $\mathbf{X}_i$ should contain only baseline covariates, since updated covariate information is a time-dependent response. The inclusion of such data in regression models of the sort considered here risks adjusting one response variable by another and so obfuscating the effect of interest (cf. Kalbfleisch and Prentice, 1980, Chap. 5.3). We thus define the p-vector $\mathbf{Z}_{ij}^{\mathrm{T}} = (\mathbf{f}_{ij}^{\mathrm{T}}, \mathbf{X}_i^{\mathrm{T}})$, where $\mathbf{f}_{ij}$ is a $(p_1 \times 1)$-vector of functions of time, $\mathbf{X}_i$ is $(p_2 \times 1)$ with $p = p_1 + p_2$, and the $J_i \times p$ covariate matrix is

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}_{i1}^{\mathrm{T}} \\ \vdots \\ \mathbf{Z}_{iJ_i}^{\mathrm{T}} \end{bmatrix}, \quad i = 1, \ldots, n.$$

In all that follows we use the approximation $\Lambda_{ij} = d_{ij}\lambda(m_{ij}; \mathbf{X}_i, \boldsymbol{\beta}) = d_{ij}\exp(\mathbf{f}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{(1)} + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^{(2)})$. For simplicity, index $\boldsymbol{\beta}^{(1)} = (\beta_1, \ldots, \beta_{p_1})^{\mathrm{T}}$ and $\boldsymbol{\beta}^{(2)} = (\beta_{p_1+1}, \ldots, \beta_p)^{\mathrm{T}}$. The first term of $\log(\lambda)$ is its underlying time-varying component. For example, the form $t^{\beta_2}\exp(\beta_1 t)$ is represented by $\mathbf{f}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{(1)} = \beta_1 m_{ij} + \beta_2 \log(m_{ij})$, using m_{ij} in place of t. For a polynomial function, this term would be $\beta_1 m_{ij} + \cdots + \beta_{p_1} m_{ij}^{p_1}$, with possibly $\log(m_{ij})$ in place of m_{ij}. A plot of the mean empirical rate function should provide a basis for initially determining a reasonable form for the underlying rate as a function of time, and subsequently for assessing the fitted model. The ith subject's empirical rate function at time t is defined by

$$\hat{\lambda}_{\mathrm{E}}(t; i) = \sum_{j=1}^{J_i} (N_{ij}/d_{ij}) I(t \in A_{ij}),$$

and undefined for $t \notin \bigcup_j A_{ij}$. The sample mean empirical rate, likewise defined only for t in the time period of the study, is given by

$$\hat{\lambda}_{\mathrm{E}}(t) = \frac{\sum_{i=1}^{n} \hat{\lambda}_{\mathrm{E}}(t; i)}{\sum_{i=1}^{n} \sum_{j=1}^{J_i} I(t \in A_{ij})}. \tag{2.3}$$

For each t, the ith summand in the numerator is taken to be 0 if $\hat{\lambda}_{\mathrm{E}}(t; i)$ is undefined. The denominator is simply the number of subjects in the study at time t, hereafter denoted by n_t.

Let $g(x; \boldsymbol{\phi})$ be the common density function of the γ_i's, with $\boldsymbol{\phi}$ its q-dimensional parameter vector. The model is thus parameterized by $\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\phi}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})$, where $\boldsymbol{\phi}$ may be regarded as a vector of mixing, variance component, or nuisance parameters and $\boldsymbol{\beta}$ a vector of covariate and rate parameters. Denoting $P(k; x) = e^{-x}x^k/k!$, $k = 0, 1, 2, \ldots$, our likelihood for n subjects is the product over i of the mixed Poissons

$$\mathscr{L}_i = \int_0^{\infty} \prod_{j=1}^{J_i} P(N_{ij}; x\Lambda_{ij}) g(x; \boldsymbol{\phi})\, dx. \tag{2.4}$$

This expression is similar to (3.1) of Stiratelli et al. (1984), with the essential differences that their objective is to estimate random effects and components of variance, and they

integrate out fixed effects under a normality assumption, subsequently carrying out a two-stage procedure employing the EM algorithm. Our likelihood is also analogous to (3.5) of Harville and Mee (1984), who treat a mixed model for categorical data, also assuming normality.

Denote $N_{i\cdot} = \Sigma_j N_{ij}$, $\Lambda_{i\cdot} = \Sigma_j \Lambda_{ij}$, $J_\cdot = \Sigma_i J_i$. Under a gamma mixing distribution $g(x; \alpha, \nu) = x^{\alpha-1}\exp(-x/\nu)/[\nu^\alpha \Gamma(\alpha)]$, $x > 0$, the likelihood takes the form

$$\mathscr{L}_{G,i}(\boldsymbol{\theta}) = \frac{\Gamma(\alpha + N_{i\cdot})}{\Gamma(\alpha)} \nu^{-\alpha}(\Lambda_{i\cdot} + \nu^{-1})^{-(N_{i\cdot}+\alpha)} \prod_{j=1}^{J_i} (\Lambda_{ij})^{N_{ij}}/N_{ij}!, \tag{2.5}$$

with $\boldsymbol{\phi}^{\mathrm{T}} = (\alpha, \nu)$ and $q = 2$. The gamma distribution is employed primarily for its flexibility. The intention here is not to construct a model that is "correct," but rather to obtain a likelihood function that is both realistic and tractable, and that will provide reasonable estimators of the covariate-adjusted rate.

It is important to note that this is a conditional likelihood in $\mathbf{N}_i$ given the covariates $\mathbf{X}_i$ and clinic visit process $\{\mathbf{A}_i, J_i\}$. If we assume the latter terms to follow some distribution h parameterized by $\boldsymbol{\theta}^*$, the total likelihood of the ith patient is the product

$$\mathscr{L}_i(\mathbf{N}_i \mid \mathbf{A}_i, J_i, \mathbf{X}_i)h(\mathbf{A}_i, J_i, \mathbf{X}_i).$$

The log-likelihoods in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ are thus maximized separately, subject to the key assumption that $\mathscr{L}_i$ does not involve $\boldsymbol{\theta}^*$ and h does not involve $\boldsymbol{\theta}$. While treating covariates as fixed is a standard approach in regression, the assumption that the distribution of the covariates and reporting time process is parameterized separately from the distribution of the responses for longitudinal data is often made implicitly without mention. In this article our concern shall be fitting the product of the conditional likelihoods (2.5).

To verify that the likelihood is identifiable in $\boldsymbol{\theta}$ for our data structure, suppress i and take $J = 1$ initially, with $\Lambda(\boldsymbol{\beta}) = \Lambda$ treated as a single parameter. In this case the distribution is

$$f(n; \Lambda, \alpha, \nu) = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)} \frac{\Lambda^n}{n!} \nu^{-\alpha}(\Lambda + \nu^{-1})^{-(n+\alpha)}, \quad n = 0, 1, 2, \ldots, \tag{2.6}$$

and it is easily verified that f is identifiable only in α and $\nu\Lambda$, i.e., ν and Λ are not separately identifiable. Generalizing to J intervals, it is likewise the case that f is identifiable in α and $\nu\Lambda_1, \ldots, \nu\Lambda_J$. However, since each $\Lambda_j = \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j)$, $\nu_1\exp(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{Z}_j) = \nu_2\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{Z}_j)$, for each $\mathbf{Z}_j$ implies that $\nu_1 = \nu_2$ and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, provided that the rows of the matrix

$$\begin{bmatrix} 1 & \cdots & 1 \\ \mathbf{Z}_1 & \cdots & \mathbf{Z}_J \end{bmatrix}$$

are linearly independent. Note that an additive constant β_0 in the exponent would violate this condition. One may regard such a term as being absorbed into ν, a conceptualization that is in accordance with the initial formulation (2.1) of the rate and the fact that $\mathrm{E}(\gamma) = \alpha\nu$.

The mixture formulation accounts for both extra-Poisson variation (overdispersion) and dependence among the elements of $\mathbf{N}_i$. Specifically, N_{ij} has mean $\mu_{ij} = \mathrm{E}(\gamma)\Lambda_{ij}$ and variance $\sigma_{ij}^2\,\mathrm{E}(\gamma)\Lambda_{ij} + \mathrm{var}(\gamma)\Lambda_{ij}^2$, and $\mathrm{cov}(N_{ij}, N_{ij'}) = \mathrm{var}(\gamma)\Lambda_{ij}\Lambda_{ij'}$ for $j \neq j'$, $i = 1, \ldots, n$, with $\mathbf{N}_1, \ldots, \mathbf{N}_n$ mutually independent. The multiplicative term γ_i becomes additive in the exponent as $\log(\gamma_i) = \varepsilon_i$. This formulation is comparable to those of Hinde (1982), who assumes $\{\varepsilon_i\}$ to be iid $\mathrm{N}(0, \sigma^2)$ and estimates σ^2 via the EM algorithm, and Breslow (1984), who proceeds with an assumption of approximate normality and iterates alternately between σ and $\boldsymbol{\beta}$ to obtain estimates. However, both authors deal with the simpler case in which the counts are univariate, conditional upon the covariates.

The commonly used approach to iid Poisson data in which each count's entire mean is mixed independently cannot be applied here. Due to the dependence, inequality of

observation intervals, and variation of λ over time, it is clearly inappropriate to mix all $E(N_{ij})$ with the same distribution. The simple device of mixing the multiplicative effect γ_i for each $E(\mathbf{N}_i)$ avoids these difficulties and provides a model that meets the criteria (i)–(v) given earlier.

3. Estimation

3.1 *Fitting the Likelihood*

Denote $L_i = \log(\mathscr{L}_i)$, $L = \Sigma_i L_i$, $\mathbf{N}^{\mathrm{T}} = (\mathbf{N}_1^{\mathrm{T}}, \ldots, \mathbf{N}_n^{\mathrm{T}})$, $\mathbf{\Lambda}^{\mathrm{T}} = (\mathbf{\Lambda}_1^{\mathrm{T}}, \ldots, \mathbf{\Lambda}_n^{\mathrm{T}})$ and $\mathbf{W} = \partial L/\partial \mathbf{\Lambda} = (\mathbf{W}_1^{\mathrm{T}}, \ldots, \mathbf{W}_n^{\mathrm{T}})^{\mathrm{T}}$, where each $\mathbf{W}_i = \partial L/\partial \mathbf{\Lambda}_i$. We shall also require the $J_\cdot \times p$ matrix

$$\mathbf{D} = \frac{\partial \mathbf{\Lambda}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_n \end{bmatrix}, \tag{3.1}$$

where $\mathbf{D}_i = \partial \mathbf{\Lambda}_i/\partial \boldsymbol{\beta} = \operatorname{diag}(\Lambda_{ij})\mathbf{Z}_i$. The log-likelihood is given by

$$\begin{aligned} L_i = I[N_{i\cdot} > 0] \sum_{r=0}^{N_{i\cdot}-1} \log(\alpha + r) - \alpha \log \nu \\ - (N_{i\cdot} + \alpha)\log(\Lambda_{i\cdot} + \nu^{-1}) + \sum_{j=1}^{J_i} N_{ij} \log(\Lambda_{ij}), \quad i = 1, \ldots, n, \end{aligned} \tag{3.2}$$

aside from terms not involving $\boldsymbol{\theta}$. The score vector $\mathbf{U}(\boldsymbol{\theta}) = \partial L/\partial \boldsymbol{\theta} = ((\partial L/\partial \boldsymbol{\phi})^{\mathrm{T}}, (\partial L/\partial \boldsymbol{\beta})^{\mathrm{T}})^{\mathrm{T}} = (\mathbf{U}(\boldsymbol{\phi})^{\mathrm{T}}, \mathbf{U}(\boldsymbol{\beta})^{\mathrm{T}})^{\mathrm{T}}$, say, with

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{D}^{\mathrm{T}}\mathbf{W} = \sum_{i=1}^{n} \mathbf{D}_i^{\mathrm{T}}\mathbf{W}_i. \tag{3.3}$$

Since $W_{ij} = N_{ij}/\Lambda_{ij} - (N_{i\cdot} + \alpha)/(\Lambda_{i\cdot} + \nu^{-1})$, the ith summand of (3.3) is

$$\mathbf{Z}_i^{\mathrm{T}}\left(\mathbf{N}_i - \frac{N_{i\cdot} + \alpha}{\Lambda_{i\cdot} + \nu^{-1}} \mathbf{\Lambda}_i\right). \tag{3.4}$$

Computation of $\mathbf{U}(\boldsymbol{\phi})$ is straightforward.

Solution of the likelihood equations $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ may be achieved either by a Gauss–Newton approach, with some step size modification, or a partial Fisher scoring version of this method. The second derivative (gradient) matrix may be partitioned in the form

$$\mathbf{\Delta} = -\frac{\partial^2 L}{\partial \boldsymbol{\theta} \boldsymbol{\theta}^{\mathrm{T}}} = \begin{bmatrix} \Delta_{\phi\phi} & \Delta_{\phi\beta} \\ \Delta_{\beta\phi} & \Delta_{\beta\beta} \end{bmatrix}, \tag{3.5}$$

where $\Delta_{\phi\phi} = -\partial^2 L/\partial \boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}$, $\Delta_{\phi\beta} = -\partial^2 L/\partial \boldsymbol{\phi}\boldsymbol{\beta}^{\mathrm{T}} = \Delta_{\beta\phi}^{\mathrm{T}}$, and $\Delta_{\beta\beta} = -\partial^2 L/\partial \boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}$. The iterative step $\boldsymbol{\theta}^{(r)} \longrightarrow \boldsymbol{\theta}^{(r+1)}$ is given in general by

$$\boldsymbol{\theta}^{(r+1)} = s_r \Delta_r^{-1} \mathbf{U}(\boldsymbol{\theta}^{(r)}) + \boldsymbol{\theta}^{(r)}, \tag{3.6}$$

with Δ_r computed at $\boldsymbol{\theta}^{(r)}$ and the step size s_r chosen to avoid overstepping the solution (maximizing) point, especially at early stages of the process. A preliminary evaluation of L for a selected array of values of $\boldsymbol{\theta}$ is also advisable, since we found the algorithm quite sensitive to choice of a starting value for our data set. In this regard, a modified

Gauss–Newton approach using only first-order derivatives of L might be useful at the initial stages, although we did not adopt this method for our analysis.

As an alternative to the matrix Δ of exact derivatives, $\Delta_{\beta\beta}$ may be replaced by its expected value. Denote $E(-\partial^2 L_i/\partial \Lambda_i \Lambda_i^T) = E(\mathbf{W}_i \mathbf{W}_i^T) = \mathbf{B}_i$ and $E(\Delta_{\beta\beta}) = \mathbf{D}^T\mathbf{B}\mathbf{D} = \Sigma_i \mathbf{D}_i^T \mathbf{B}_i \mathbf{D}_i$, where $\mathbf{B}$ is the $J_. \times J_.$ block-diagonal matrix $\text{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_n)$. It follows easily that

$$\mathbf{B}_i = \alpha\nu\{\text{diag}(\Lambda_{ij}^{-1}) - \frac{\nu}{1 - \nu\Lambda_{i.}} \mathbf{1}\mathbf{1}^T\},$$

hence

$$\mathbf{D}_i \mathbf{B}_i \mathbf{D}_i = \mathbf{Z}_i^T[\alpha\nu \text{ diag}(\Lambda_{ij}) - \frac{\nu}{1 + \nu\Lambda_{i.}} \Lambda_i \Lambda_i^T)\mathbf{Z}_i. \tag{3.7}$$

For the example given in Section 4 to follow, ϕ is reparameterized as $\phi^T = (a, b) = (\log \alpha, \log \nu)$ to allow unconstrained optimization. The empirical rate plots suggest that $t^{\beta_2}\exp(\beta_1 t)$ should be sufficiently flexible to serve as the time-varying component of the fitted rate function. The estimated rate at time t for a subject with covariate vector $\mathbf{X}$ is thus given by

$$\hat{\lambda}(t; \mathbf{X}) = \exp\{\hat{a} + \hat{b} + \hat{\beta}_1 t + \hat{\beta}_2 \log(t) + \mathbf{X}^T\hat{\boldsymbol{\beta}}^{(2)}\}. \tag{3.8}$$

3.2 *Analysis of Deviance*

Denote $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iJ_i})^T$. Since each $\mu_{ij} = E(\gamma)\Lambda_{ij}$, it is a function of the nuisance parameters $\boldsymbol{\phi}$ as well as $\boldsymbol{\beta}$. In the usual generalized linear model formulation the deviance would be given by

$$-2 \sum_{i=1}^{n} \{L_i(\boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}); \mathbf{N}_i) - L_i(\hat{\boldsymbol{\mu}}_i; \mathbf{N}_i)\},$$

with $\hat{\mu}_{ij} = N_{ij}$ and the full model likelihood a function of the data alone. Due to the presence of $\boldsymbol{\phi}$ in the model, there is no satisfactory baseline $L(\hat{\boldsymbol{\mu}}; \mathbf{N})$ relative to which a residual deviance may be defined. We thus consider differences in $-2L(\hat{\boldsymbol{\theta}})$ for comparison of nested models.

It seems reasonable here to deal with a maximal model that includes $\boldsymbol{\phi}$, rate function parameters $\boldsymbol{\beta}^{(1)}$, and the parameter vector $\boldsymbol{\beta}^{(2)}$ corresponding to the largest set of baseline covariates initially considered. Our minimal model thus includes only $\boldsymbol{\phi}$ and $\boldsymbol{\beta}^{(1)}$; specifically, it is the four-parameter likelihood in which $\Lambda_{ij} = d_{ij}\exp(\beta_1 m_{ij} + \beta_2 \log m_{ij})$, with $\boldsymbol{\beta}^{(1)} = (\beta_1, \beta_2)^T$. Each model may be considered to have three components: (1) the random effects $\{\gamma_i\}$ corresponding to the parameters $\boldsymbol{\phi}$; (2) the underlying time-dependent rate function $\lambda_0(t) = t^{\beta_2}\exp(\beta_1 t) = \exp(\beta_1 t + \beta_2 \log t)$; and (3) a subset of prognostic covariates $\mathbf{X}_i$ chosen from those of the maximal model and parameterized by $\boldsymbol{\beta}^{(2)}$.

3.3 *Empirical Bayes Estimation*

Although $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ are estimated together, one may regard our methodology as part of an empirical Bayes approach in which $\gamma_1, \ldots, \gamma_n$ are random parameters, g their prior, and $\boldsymbol{\phi}$ the hyperparameters of g. From this point of view, our likelihood $\mathcal{L}_G$ is the marginal of $\mathbf{N}$, and a posterior for $\boldsymbol{\gamma}$ is given by $\prod_{i,j} P(N_{ij}; \gamma_i \Lambda_{ij})/\mathcal{L}_G(\mathbf{N})$. Denoting the maximum likelihood estimators by $\hat{\alpha}, \hat{\nu}, \hat{\boldsymbol{\beta}}$, an estimator of each γ_i is given by the mean of its posterior, which is gamma with parameters $\hat{\alpha} + N_{i.}$ and $[\Lambda_{i.}(\hat{\boldsymbol{\beta}}) + \hat{\nu}^{-1}]^{-1}$. Thus, the estimator of γ_i is

$$\hat{\gamma}_i = \frac{\hat{\alpha} + N_{i.}}{\hat{\nu}^{-1} + \Lambda_{i.}(\hat{\boldsymbol{\beta}})}.$$

When the distribution of subject effects is of interest, $\hat{\gamma}_1, \ldots, \hat{\gamma}_n$ may serve as an estimated sample for computation of descriptive statistics, an empirical cdf, etc.

4. Example

In the National Cooperative Gallstone Study (NCGS) (cf. Schoenfield et al., 1981), a major concern was the effect of the drug chenodiol on biliary symptoms commonly associated with gallstones. For illustration, we present an analysis of the incidence of dyspepsia over the first 2 years of the study for the 111 patients who had floating stones and were assigned to either the high-dose ($n_1 = 63$) or placebo ($n_2 = 48$) groups. All patients were scheduled for routine clinic center visits at 1, 2, 3, 6, 9, 12, 16, 20, and 24 months of follow-up, and at 28 months if gallstone dissolution was observed. Thus, N_{ij} is the number of episodes of dyspepsia reported by patient i at his or her jth visit. Baseline covariates included in the analysis are AGE in years, SEX (1 = female, 0 = male), alcohol drinking status ALCAT (1 = current drinker, 0 = not current drinker), and a treatment group indicator TRT (1 = high dose, 0 = placebo). The observed number of visits per patient varied from 1 to 12, with a median, mean, and standard deviation of 9, 7.9, and 2.18, respectively.

The data were initially fit using the likelihood specified by

$$\Lambda_{ij} = d_{ij}\exp[\beta_1 m_{ij} + \beta_2 \log(m_{ij}) + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^{(2)}].$$

The maximal model (I) included all four baseline covariates, with successive submodels obtained by deleting the least significant covariate at each stage of a simple stepdown procedure. Parameter estimates and their standard errors for each model are given in Table 1. To assess goodness of fit, the generalized Pearson statistic (cf. Liang and Zeger, 1986, §3.3)

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{J_i}(N_{ij} - \hat{\mu}_{ij})^2/\hat{\sigma}_{ij}^2}{J_\cdot - p} \tag{4.1}$$

for longitudinal data was employed. Since γ has mean $\alpha\nu$ and variance $\alpha\nu^2$, $\mu_{ij} = \alpha\nu\Lambda_{ij}$ and $\sigma_{ij}^2 = \alpha\nu\Lambda_{ij}(1 + \nu\Lambda_{ij})$ for the mixed Poisson model (2.4).

Table 1
Parameter estimates and standard errors for models obtained via stepdown procedure

Parameter (Covariate)	Model I (maximal)	II	III	IV	V (Minimal)	III* (Poisson)
a	−1.7269 (.1613)	−1.7352 (.1612)	−1.7606 (.1604)	−1.8029 (.1592)	−1.8725 (.1576)	
b	4.4403 (1.9021)	3.8831 (1.7111)	4.4756 (1.7709)	1.4904 (.4043)	1.0674 (.3080)	
β_0						1.1470 (.1730)
β_1 (t)	.0265 (.0021)	.0265 (.0021)	.0264 (.0021)	.0265 (.0020)	.0266 (.0020)	.0213 (.0019)
β_2 (log(t))	−.6575 (.0562)	−.6572 (.0562)	−.6565 (.0562)	−.6591 (.0562)	−.6632 (.0560)	−.6596 (.0537)
β_3 (TRT)	−.9148 (.5249)	−.8636 (.5036)	−1.0949 (.4764)	−1.1696 (.4782)	—	−.8009 (.0598)
β_4 (AGE)	−.0658 (.0336)	−.0594 (.0320)	−.0599 (.0337)	—	—	−.0302 (.0028)
β_5 (ALCAT)	.7497 (.5128)	.7030 (.4902)	—	—	—	—
β_6 (SEX)	−.4105 (.4934)	—	—	—	—	—
$\tilde{\sigma}^2$	3.36	3.21	3.30	3.26	2.61	28.36

As a baseline for comparison, a "naive" Poisson regression model (III*) ignoring within-subject dependencies was also fit to the data. The same covariate set as that of model III was used, with additive parameter β_0 in the linear component. This is the usual GLM for Poisson counts, with $\mu_{ij} = \sigma_{ij}^2 = d_{ij}\exp[\beta_0 + \beta_1 m_{ij} + \beta_2 \log(m_{ij}) + \mathbf{X}_i^T \boldsymbol{\beta}^{(2)}]$.

The four parameters a, b, β_1, and β_2 were highly significant in each of models I–V, based on the approximately normal statistics $\hat{\theta}/s(\hat{\theta}_j)$. Likelihood-ratio comparisons for these models are summarized in the analysis of deviance given in Table 2. Both the likelihood-ratio tests and stepdown procedure indicate that TRT and AGE are significant, while the covariates SEX and ALCAT may be dropped from the model. It is notable that $\hat{b}$ sharply decreased when each of the significant variables AGE and TRT was removed. This might have been anticipated given the large negative correlations between $\hat{b}$ and the coefficients of these variables, i.e., $\hat{\rho}_{\text{III}}(\hat{b}, \hat{\beta}_4) = -.9743$ and $\hat{\rho}_{\text{II}}(\hat{b}, \hat{\beta}_3) = -.6593$. In contrast, $\hat{a}$ remained relatively constant. Equivalently, the mean of γ dropped while its coefficient of variation decreased only slightly.

Table 2
Analysis of deviance

Model	$-2L(\hat{\theta})$	Difference (P-value)	Components
V (minimal)	3,642.35		λ_0, γ
IV	3,636.57	5.78 (.016)	TRT$\mid\lambda_0$, γ
III	3,633.20	3.37 (.066)	AGE$\mid\lambda_0$, γ, TRT
II	3,631.20	2.00 (.173)	ALCAT$\mid\lambda_0$, γ, TRT, AGE
I (maximal)	3,630.50	.70 (.403)	SEX$\mid\lambda_0$, γ, TRT, AGE, ALCAT

A comparison of the Poisson GLM III* to the mixed Poisson model III shows a striking reduction in $-2L(\hat{\boldsymbol{\theta}})$, from 6,920.12 to 3,633.20. Moreover, the Poisson model is clearly overdispersed ($\tilde{\sigma}^2 = 28.36$), with a reduction of one order of magnitude in $\tilde{\sigma}^2$ under the mixed Poisson model. There is a marked increase in the standard error of the estimated coefficients of AGE and TRT due to fitting the mixed Poisson model, compared to the naive Poisson GLM. This is apparently due to the fact that, in accounting for between-subject variability, inclusion of random effects in the model reduces the explanatory power of the observed covariates, as is the case in the general linear mixed model. In contrast, the standard errors of the estimated coefficients of t and $\log(t)$ are virtually the same under the Poisson and mixed Poisson models. This may be due to the fact that clinic visit time, which enters the model as the covariates m_{ij} and $\log(m_{ij})$, varies primarily within subjects.

Although the form of $\lambda_0(t)$ is quite flexible, the above formulation allows that the rates in the two treatment groups differ only by the multiplicative term $\exp(\beta_3)$, aside from the other covariates. The more general case in which $\lambda_0(t)$ has the different shapes $t^{\beta_{21}} e^{\beta_{11} t}$ and $t^{\beta_{20}} e^{\beta_{10} t}$ in the high-dose and placebo groups, respectively, is considered next, especially since the mean empirical rate plots appear to have different shapes in the two treatment groups (Figures 1 and 2). The mixed Poisson model VI thus generalizes model III by replacing $\beta_1 m_{ij} + \beta_2 \log(m_{ij})$ with the sum of treatment × time interaction terms $\beta_{11}\text{TRT}\, m_{ij} + \beta_{21}\text{TRT}\log(m_{ij}) + \beta_{10}(1 - \text{TRT})m_{ij} + \beta_{20}(1 - \text{TRT})\log(m_{ij})$. Although this interaction model does not fit into the hierarchy of models summarized in Tables 1 and 2, it does contain III as a submodel via the hypothesis H: $\beta_{11} = \beta_{10}$, $\beta_{21} = \beta_{20}$. For the fitted interaction model, $-2L_{\text{VI}}(\hat{\boldsymbol{\theta}}) = 3{,}524.30$ so that the likelihood-ratio statistic for comparing models III and VI is 108.9 on 2 df, indicating a substantial improvement in fit due to including this generalization of $\lambda_0(t)$.

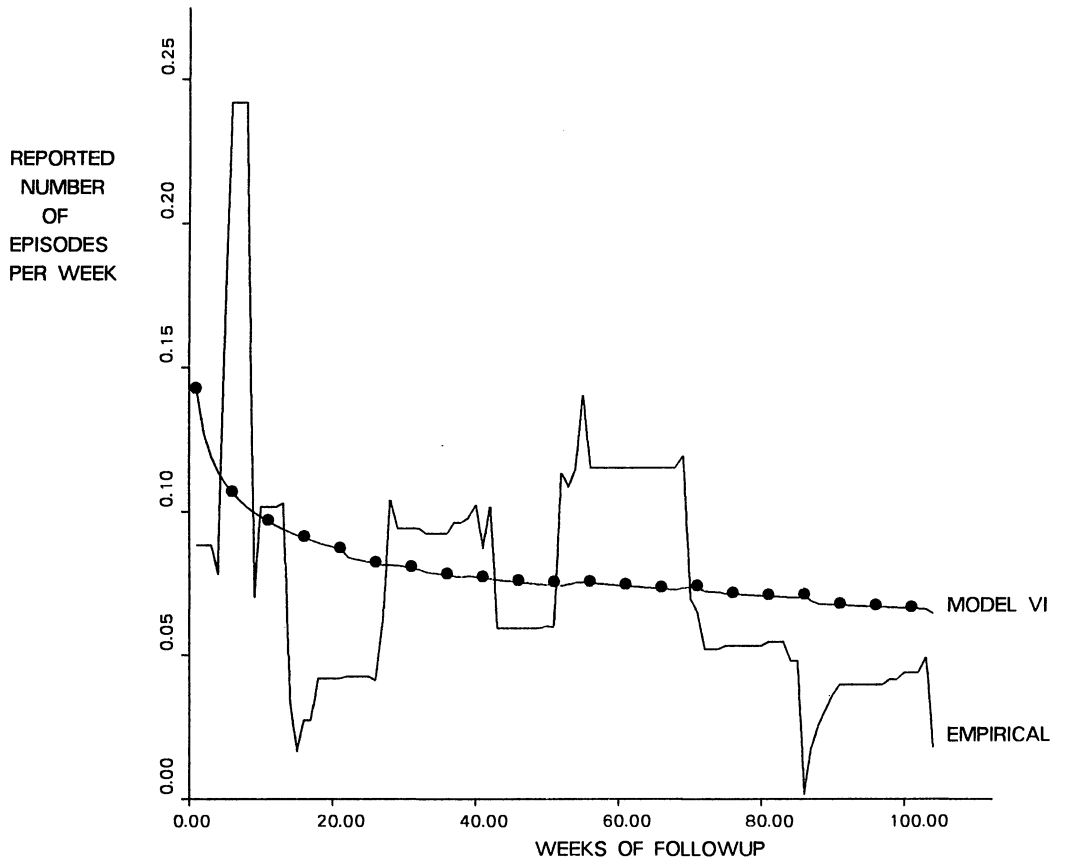

Figure 1. Mean empirical and fitted rates of dyspepsia for high-dose chenodiol treatment group.

Inspection of the fitted parameter values of model VI, summarized in the first column of Table 3, indicates that the coefficient β_{11} of t in the high-dose group is nonsignificant, based on the statistic $\hat{\beta}_{11}/s(\hat{\beta}_{11}) = -.0131$ ($P = .9895$). However, elimination of this parameter yields the reduced model VII, with 1-df likelihood-ratio statistic $-2[L_{\text{VII}}(\hat{\boldsymbol{\theta}}) - L_{\text{VI}}(\hat{\boldsymbol{\theta}})] = 2.58$ ($P = .1082$). In addition to the problem that these tests give rather conflicting results, dropping β_{11} from the model produces a nontrivial change in the parameter estimates $\hat{\beta}_{21}$ and $\hat{\beta}_3$. Comparisons of the mean covariate-adjusted rate to the mean empirical rate in the high-dose group showed a clearly better fit under model VI than under VII, especially for t in the range 1 to 26 weeks.

The coefficient β_3 of the high-dose group indicator TRT in models VI and VII is not a treatment "effect" in the usual sense, since the underlying rate functions have different shapes in the two treatment groups. Although the term β_3TRT in the linear component of these models multiplies the rate function of the high-dose group by $\exp(\beta_3)$, it does so in the presence of the time × treatment interaction terms. The two groups are properly compared under model VI in terms of their underlying rates, regarded as functions of time. These are

$$t^{\beta_{21}}\exp(\beta_{11}t + \beta_3) \quad \text{(High dose)}$$

and

$$t^{\beta_{20}}\exp(\beta_{10}t) \quad \text{(Placebo)}$$

aside from the common term $a + b + \beta_4$AGE.

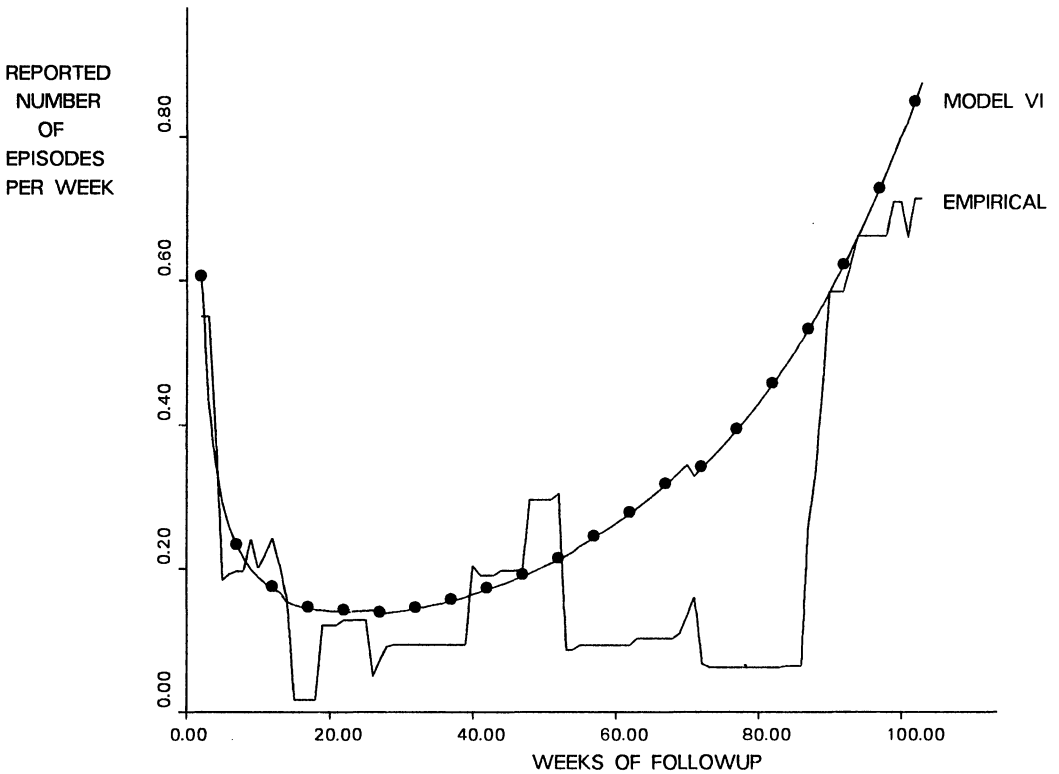

Figure 2. Mean empirical and fitted rates of dyspepsia for placebo treatment group.

An alternate approach to dealing with treatment interaction is simply to fit the model separately in the two groups. This yields the ten-parameter model VIII, also summarized in Table 3. In this model all parameters a, b, β_1, β_2, and β_4 are treatment group-specific, with no β_3TRT term in the linear component. This produces a reduction in $-2L(\hat{\theta})$ from model VI of .34 on 2 df, with only the parameter estimates $\hat{a}_1$ and $\hat{b}_1$ changing appreciably, apparently to adjust for the deleted β_3 in the high-dose group. Specifically, the sum $\hat{a} + \hat{b} + \hat{\beta}_3 = 1.4535$ in $\log(\hat{\lambda})$ for the high-dose group under model VI is rather close to the comparable term $\hat{a}_1 + \hat{b}_1 = 1.3671$ under model VIII. Recall that the identifiability condition given in Section 2 precludes an additive constant β_0 in the linear component, with $b = \log(\nu)$ fulfilling this role. Thus, the terms $(a + b + \beta_3, a + b)$ under model VI and $(a_1 + b_1, a_0 + b_0)$ under model VIII each are analogous to the intercepts in a two-group analysis of covariance model with heterogeneous intercepts and slopes. Models VI and VIII yield roughly the same fit to these data, with model VI requiring fewer parameters.

Based on these results, model VI was chosen as the final model, the instability of inferences regarding $\hat{\beta}_{11}$ notwithstanding. The lower triangle of the estimated correlation matrix of $\hat{\theta} = (\hat{a}, \hat{b}, \hat{\beta}_{11}, \hat{\beta}_{21}, \hat{\beta}_{10}, \hat{\beta}_{20}, \hat{\beta}_3, \hat{\beta}_4)^T$, with variances along the diagonal, is

$$\begin{bmatrix}
.0257 & & & & & & & \\
-.0933 & 3.0827 & & & & & & \\
-.0049 & -.0106 & 1.3 \times 10^{-5} & & & & & \\
.0037 & .0036 & -.8836 & .0097 & & & & \\
-.0084 & .0343 & .0002 & -.0001 & 6.9 \times 10^{-6} & & & \\
-.0036 & -.0378 & -.0003 & .0001 & -.8762 & .0051 & & \\
-.0037 & .0145 & .2685 & -.3704 & -.1431 & .2157 & .2867 & \\
.0027 & -.9730 & .0114 & -.0041 & .0099 & -.0270 & -.1663 & .0011
\end{bmatrix}.$$

Table 3
Parameter estimates and standard errors for treatment interaction models

Parameter (Covariate)	Model VI	VII	VIII	VI* (Poisson)
High-dose				
a_1	−1.7636 (.1603)	−1.7626 (.1603)	−1.6696 (.2267)	
b_1	5.0223 (1.7558)	5.0327 (1.7525)	3.0367 (1.9571)	β_0 1.4387 (.1816)
β_{11} (t)	-4.7×10^{-5} (.0036)	—	-5.6×10^{-5} (.0036)	-9.9×10^{-5} (.0035)
β_{21} (log(t))	−.1618 (.0987)	−.0052 (.0017)	−.1616 (.0987)	−.1304 (.0998)
β_3 (TRT)	−1.8052 (.5355)	−2.1363 (.4974)	—	−1.9137 (.2540)
β_{41} (AGE)	−.0666 (.0333)	−.0668 (.0333)	−.0650 (.0355)	−.0293 (.0028)
$\tilde{\sigma}_1^2$	3.19	3.27	3.38	17.13
Placebo				
a_0	−1.7636 (.1603)	−1.7626 (.1603)	−1.8537 (.2282)	
b_0	5.0223 (1.7558)	5.0327 (1.7525)	5.4516 (4.0458)	β_0 1.4387 (.1816)
β_{10} (t)	.0414 (.0026)	.0414 (.0026)	.0414 (.0026)	.0323 (.0024)
β_{20} (log(t))	−.9252 (.0712)	−.9251 (.0712)	−.9247 (.0713)	−.9249 (.0657)
β_{40} (AGE)	−.0666 (.0333)	−.0668 (.0333)	−.0732 (.0782)	−.0293 (.0028)
$\tilde{\sigma}_0^2$	3.30	3.30	3.13	39.92
p	8	7	10	7
$-2L(\hat{\theta})$	3,524.30	3,526.88	3,523.96	6,854.86

It is notable that the coefficients of t and log(t) were highly negatively correlated in both treatment groups (−.8836 in the high-dose and −.8762 in the placebo group). The highest correlation (−.9730) was between $\hat{b}$ and $\hat{\beta}_4$, as was the case in the noninteraction model III.

The mean covariate-adjusted rate

$$\hat{\lambda}_{VI}(t) = \sum_{i=1}^{n} \hat{\lambda}(t; \mathbf{X}_i)/n_t$$

was computed for each treatment group to assess goodness of fit via graphical comparison to the mean empirical rates. Analogously to the computation of $\hat{\lambda}_E(t; i)$, each $\hat{\lambda}(t; \mathbf{X}_i)$ was computed only for t within the time period where patient i actually had data. These plots appear in Figures 1 and 2. The fits in each group appear to be quite good, especially given the variability in reporting times and use of the interval midpoint m_{ij} in place of t for computation of Λ_{ij}. A more detailed version of this graphical comparison would be to overlay plots of $\hat{\lambda}_E(t; i)$ and $\hat{\lambda}(t; \mathbf{X}_i)$ separately for each subject. Given the impracticality of presenting $n = 111$ such graphs here, the plots of their means within each treatment group give reasonable pictures of the average fit across the sample.

Use of the estimated posterior mean $\hat{\gamma}_i$ in place of $\exp(\hat{a} + \hat{b})$ for estimation of the ith patient's rate gave rather poor fits for many patients and on average. This is perhaps due

to the fact that the distribution of $\hat{\gamma}_i$, $i = 1, \ldots, 111$, is heavily skewed to the right, varying from .079 to 367.27 with the following quantiles:

Percent	5	10	25	50	75	90	95
Quantile	.170	.227	.624	2.10	16.28	97.62	145.03

We note the distinction between the distribution of these estimated posterior means and the distribution of the random effects themselves.

In addition to the large negative correlation between $\hat{b}$ and $\hat{\beta}_4$, it is interesting that the term $\exp(\hat{\beta}_4 \text{AGE}) = (.9356)^{\text{AGE}}$ shows a substantial decrease over the range of ages (26 to 75 years) of patients considered. Comparing the extremes, the ratio of reported incidence of dyspepsia of a 26-year-old patient to that of a 75-year-old patient is 26.10 for the fitted model VI.

The present modelling approach exhibits several advantages over simpler methods, at least for the data set considered here. A large increase in the fitted log-likelihood was achieved by passing from the Poisson GLM, which ignores within-subject response dependence, to the mixed Poisson model. A considerable reduction in overdispersion was also achieved, in addition to elimination of an apparent heteroscedasticity exhibited by the estimated dispersion parameters $\tilde{\sigma}_1^2 = 17.13$ and $\tilde{\sigma}_0^2 = 39.92$ for the interaction Poisson GLM VI*. We note that the overdispersion was not entirely eliminated, however, since the final model had estimated dispersion parameters $\tilde{\sigma}_1^2 = 3.19$ and $\tilde{\sigma}_0^2 = 3.30$ (with pooled value 3.21), rather than values near 1. Although the empirical rate plots provide a reasonable informal comparison of the treatment groups, they ignore all other covariate information. In this regard our model enjoys the advantages of any parametric regression model, i.e., formal tests of treatment or other covariate effects and prediction of the rate function for given $\mathbf{X}$ are available. Based on our numerical results, the necessary model assumptions and computational effort appear to be justified.

5. Computation

Each fitted likelihood was required to satisfy three convergence criteria: (i) that each component of $\boldsymbol{\theta}^{(r+1)} - \boldsymbol{\theta}^{(r)}$ be smaller than $\pm 10^{-12}$; (ii) that each component of $\mathbf{U}(\boldsymbol{\theta}^{(r+1)})$ be smaller than $\pm 10^{-12}$; and (iii) that $|L(\boldsymbol{\theta}^{(r+1)}) - L(\boldsymbol{\theta}^{(r)})| < 10^{-8}$. Using the golden section line search method to obtain a step size, criteria (i) and (iii) were easily met, subsequent to a preliminary search for a starting point as described earlier. The Gauss–Newton approach without use of optimum step size selection worked poorly or not at all in early stages of the iterative scheme. The step size subroutine was dropped to save time near convergence, where the criterion (ii) became the one most difficult to satisfy. This was apparently due to the existence of a plateau in a one- or two-dimensional submanifold of the space of $\mathbf{U}(\boldsymbol{\theta})$ for some of the models. When this occurred, convergence was obtained quite easily by fixing all parameters but those corresponding to the entries of $\mathbf{U}(\boldsymbol{\theta})$ not satisfying (ii) and maximizing in the remaining parameters. When selective use of both the line search for optimum step size and maximization in a subspace were employed, convergence was obtained in fewer than 10 iterations. Use of the modified Fisher scoring method did not appear to affect either the speed or accuracy of the algorithm.

All computations were carried out in SAS on an IBM 4361.

Acknowledgements

This work was partially supported by National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases research grant #R01-AM-35952. The author wishes to thank John M. Lachin for suggesting this problem and providing the exemplary data set,

and M. Hamdy for assisting with numerical computations. I am also grateful to three referees and an associate editor for many constructive comments and suggestions.

RÉSUMÉ

Dans beaucoup d'études temporelles on désire estimer le taux d'arrivée d'un événement récurrent particulier. On ne dispose souvent que des comptages correspondant au temps écoulé entre deux observations de chaque sujet, ainsi que de covariable accompagnatrices. Les intervalles peuvent varier beaucoup en durée et en nombre suivant les sujets, si bien que les vecteurs de comptage correspondants ne sont pas directement comparables. On propose, pour cette structure de données temporelles, une famille de modèle de régression, à l'aide d'une vraisemblance poissonnienne, qui comprend un terme multiplicatif dans la fonction de taux pour chaque sujet. On décrit aussi l'estimateur bayesien empirique associé des paramètres aléatoires. Ces méthodes sont illustrées par l'analyse de données concernant la dyspepsie, provenant de la National Cooperative Gallstone Study.

REFERENCES

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33,** 38–44.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Series B* **17,** 129–164.

Frome, E. L. (1983). The analysis of rates using Poisson regression models. *Biometrics* **39,** 665–674.

Frome, E. L., Kutner, M. H., and Beauchamp, J. J. (1973). Regression analysis of Poisson distributed data. *Journal of the American Statistical Association* **68,** 935–940.

Gilmour, A. R., Anderson, R. D., and Rae, A. L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika* **72,** 593–599.

Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society, Series B* **46,** 149–192.

Harville, D. A. and Mee, R. W. (1984). A mixed-model procedure for analyzing ordered categorical data. *Biometrics* **40,** 393–408.

Hinde, J. (1982). Compound Poisson regression models. In *GLIM 82: Proceedings of the International Conference in Generalized Linear Models*, R. Gilchrist, (ed.), 109–121. Berlin: Springer-Verlag.

Jorgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika* **70,** 19–28.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Korn, E. L. and Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics* **35,** 795–802.

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

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* New York: Chapman and Hall.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135,** 370–384.

Schoenfield, L. J., Lachin, J. M., The Steering Committee, and the NCGS Group (1981). Chenodiol (chenodeoxycholic acid) for dissolution of gallstones: The National Coooperative Gallstone Study. *Annals of Internal Medicine* **95,** 257–282.

Stiratelli, R., Laird, N., and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40,** 961–971.

Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random interval count data. *Journal of the American Statistical Association* **83,** in press.

Thompson, R. and Baker, R. J. (1981). Composite link functions in generalized linear models. *Applied Statistics* **30,** 125–131.

Weber, D. C. (1971). Accident rate potential: An application of multiple regression analysis of a Poisson process. *Journal of the American Statistical Association* **66,** 285–288.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61,** 439–447.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.

Received May 1986; *revised February and July* 1987.