BIOMETRICS 46, 1017–1034
December 1990

Analysis of Multi-Type Recurrent Events in Longitudinal Studies; Application to a Skin Cancer Prevention Trial

Hasan Abu-Libdeh

Mathematics Department, Birzeit University,
P.O. Box 14, Birzeit, West Bank, Via Israel

Bruce W. Turnbull*

School of Operations Research and Industrial Engineering,
Cornell University, Ithaca, New York 14853, U.S.A.

and

Larry C. Clark

Arizona Cancer Center, University of Arizona,
Tucson, Arizona 85724, U.S.A.

SUMMARY

We consider the statistical modeling and analysis of replicated multi-type point process data with covariates. Such data arise when heterogeneous subjects experience repeated events or failures which may be of several distinct types. The underlying processes are modeled as nonhomogeneous mixed Poisson processes with random (subject) and fixed (covariate) effects. The method of maximum likelihood is used to obtain estimates and standard errors of the failure rate parameters and regression coefficients. Score tests and likelihood ratio statistics are used for covariate selection. A graphical test of goodness of fit of the selected model is based on generalized residuals. Measures for determining the influence of an individual observation on the estimated regression coefficients and on the score test statistic are developed. An application is described to a large ongoing randomized controlled clinical trial for the efficacy of nutritional supplements of selenium for the prevention of two types of skin cancer.

1. Introduction

In this paper we are concerned with the statistical analysis of an experiment, planned or observational, in which the subjects may experience a series of recurrent events of differing types. Thus the data for each subject consist of multiple failure times, each failure being labeled as one of a set of possible types. The failure times can be represented as the realization of a stochastic point process. For failures considered to be all of a single type there have been applications to many fields: in medical trials, for example, to studies of epileptic seizures, asthma, bladder cancer (Freedman, Sylvester, and Byar, 1989), gallstones (Thall, 1988); in animal carcinogenicity experiments to incidence of mammary tumors (Gail, Santner, and Brown, 1980); in engineering, to reliability of repairable systems (Ascher and Feingold, 1984; Gaver and O'Muircheartaigh, 1987; Follman and Goldberg, 1988); in

* To whom correspondence should be sent.

Key words: Empirical Bayes; Generalized residuals; Influence analysis; Multiple failure times; Multivariate; Point processes; Poisson regression; Random effects; Regression diagnostics; Reliability.

sociology, to studies of mobility, unemployment, fertility, etc. (Allison, 1984). In another application, Prentice, Williams, and Peterson (1981) discuss a study of infections following bone marrow transplantation. At the end of their paper, they introduce the multi-type problem in which the events, namely infections, are classified as being of bacterial, fungal, or viral origin.

Our own application of interest is to the analysis of a triple-blind placebo-controlled randomized clinical trial (Clark, Turnbull, and Combs, unpublished manuscript) currently in progress to study the efficacy of a nutritional supplement of selenium in the prevention of skin cancer in high-risk patients. Starting in September 1983, patients with previous history of either multiple basal cell epithelioma (BCE) or a squamous cell carcinoma (SCC) or otherwise at high risk were recruited into the trial in one of seven clinics and assigned at random to either treatment or placebo. The clinics are mainly situated in the southeastern United States where levels of selenium in the soil are low and the risk of skin cancer is high. The patients are followed closely for a minimum of 4 years by means of frequent and regular clinic visits. The patients are also instructed on self-examination for appearance of skin lesions and are told to report to the clinic if they see any suspicious symptoms without waiting for the next scheduled appointment. For each patient the date, type (BCE or SCC), and site of any new diagnosed lesion are recorded. A number of endpoints are monitored for purposes of safety and efficacy; however, here we will be concerned only with occurrences of BCEs and SCCs. A number of baseline covariates are recorded for each patient at the time of randomization. These include amongst others: clinic, gender (M/F), age, childhood farm exposure (Y/N), smoker (Y/N), amount of sun damage on skin, tanning ability, and number of previous skin tumors diagnosed in the previous 5 years.

In Section 2, we present a generalization of a parametric model described by Lawless (1987), who has surveyed earlier literature on the analysis of replicated point process data with covariates and some related Poisson regression models. In many situations the model we consider can be viewed as an adequate but parsimonious method for describing the data, and serve as a basis for assessing the significance of treatment and other covariate effects, prediction, and subgroup identification analysis. The model can be viewed alternatively as a multi-type Poisson process regression model with mixed random and fixed effects, as a multivariate negative binomial regression model, or as an empirical Bayes model. We also discuss the application of several possible generalizations of the model. In Section 3, we discuss how ideas of regression diagnostics, residual and influence analysis, can be applied in this model. In the final section we use the procedure to analyze the skin tumor trial data set described above.

2. Mixed Point Process Models

We suppose that we have I independent subjects or individuals, each with an associated row vector $\mathbf{x}_i$ $(1 \leqslant i \leqslant I)$ of baseline covariate values. Subject i is observed over time period $[0, T_i]$, where time is measured from a defined starting point for that subject. In our skin cancer example, this origin is the randomization date. Individuals experience repeated "events" or "failures," each of which can be any of J different types. In the skin cancer example, these could represent the $J = 2$ tumor types (BCEs or SCCs) or they could represent J different sites on the body. Suppose that $K_{ij} \geqslant 0$ events of type j $(1 \leqslant j \leqslant J)$ are observed to occur on individual i $(1 \leqslant i \leqslant I)$ at times $0 = t_{ij0} < t_{ij1} < \cdots < t_{ijK_{ij}}$. For simplicity we suppose that there are no ties among events of the same type in the same individual. Also gap times $\{z_{ijk}: 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J, 1 \leqslant k \leqslant K_{ij}\}$ are defined by $z_{ijk} = t_{ijk} - t_{ijk-1}$ for $1 \leqslant k \leqslant K_{ij}$. The final gap time $z_{ijK_{ij}+1} = T_i - t_{ijK_{ij}} \geqslant 0$ is considered censored and used as such in the analysis.

We now describe a class of models for the point process data which is a generalization of that described by Lawless (1987, §3). The mixture or empirical Bayes model is

as follows:

1. Associated with subject i ($1 \leq i \leq I$) are $J+1$ nonnegative parameters θ_i, ξ_{ij} ($1 \leq j \leq J$) where $\sum_{j=1}^{J} \xi_{ij} = 1$ for all i. Given θ_i and $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iJ})$, the event processes of type j ($1 \leq j \leq J$) in individual i ($1 \leq i \leq I$) are independent nonhomogeneous Poisson processes with respective intensity functions $\lambda_{ij}(t; \mathbf{x}) = \theta_i \xi_{ij} \lambda_0(t) \exp(\mathbf{x}_i \boldsymbol{\beta})$. Here $\lambda_0(t) > 0$ is a baseline intensity function, $\boldsymbol{\beta}$ is a column p-vector of regression coefficients, and time t is measured from the ith subject's origin or starting point. If $\lambda_0(t) = \delta t^{\delta - 1}$, then we have a "Weibull" or "power-law" intensity function which has been used by Crow (1974), Lee and Lee (1978), Ascher and Feingold (1984), Lawless (1987), and others. If $\delta = 1$, then we have the usual homogeneous Poisson process.
2. The parameters $(\theta_i, \xi_{i1}, \xi_{i2}, \ldots, \xi_{iJ})$ corresponding to individual i are considered random effects. The values $\theta_1, \theta_2, \ldots, \theta_I$ are considered an independent identically distributed (i.i.d.) sample from the gamma distribution with scale parameter γ and shape parameter ν (Johnson and Kotz, 1970, p. 166, eqn (2)). The vectors $(\boldsymbol{\xi}_i: 1 \leq i \leq I)$ are considered an i.i.d. sample from a Dirichlet distribution (Johnson and Kotz, 1972, p. 233, eqn (31)) with parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_J)$, and independent of the $\{\theta_i\}$.
3. All individuals are assumed statistically independent.

The gamma mixing distribution for the $\{\theta_i\}$ is a convenient and flexible choice; it has been used by Gaver and O'Muircheartaigh (1987), Lawless (1987), and Thall (1988), amongst others. Of special interest is the degenerate case when $\nu \to \infty$ so that the θ_i are equal and constant.

As the conjugate prior, the Dirichlet distribution is the natural choice for $(\boldsymbol{\xi}_i: 1 \leq i \leq I)$. A case of special interest is the degenerate case where all $\alpha_j \to \infty$ ($1 \leq j \leq J$) such that $\boldsymbol{\xi}_i$ can be considered as a constant vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_J)$. Then the number of parameters to be estimated is reduced by 1, and it is implied that whereas individuals' combined total baseline event rates $\{\theta_i\}$ may vary substantially, the distribution among event types given an overall level is homogeneous over subjects.

It should be noted that with a redefinition of the time parameter t, much of what follows can be easily adapted to a model in which the underlying nonhomogeneous Poisson process in assumption 2 above is replaced by a renewal process where the gap or interevent times are independent with hazard functions $\lambda_{ij}(t; \mathbf{x})$. The power-law choice for $\lambda_0(t)$ would be now somewhat analogous to the Weibull–Markov models used by Gail et al. (1980). The two time scales for t are discussed by Prentice et al. (1981).

Other more complex parametric models could be considered. For example, we might consider a model where the random variables $\{\theta_i \xi_{ij}\}$ are all mutually independent or where regression coefficients $\{\beta\}$ are event-type-dependent. Typically such models might simply lead to separate univariate analyses for each event type. In this paper our aim is to construct a family of models that is not too elaborate and can give a parsimonious description of the data and yet a family that is rich and flexible enough to account for the heterogeneity observed in the data.

2.1 *Estimation of Unknown Parameters*

Define the baseline cumulative intensity function by $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$. Then, given θ_i and $\boldsymbol{\xi}_i$, the contribution of the ith subject to the conditional likelihood is

$$\prod_{j=1}^{J} \left\{ \prod_{k=1}^{K_{ij}} \theta_i \xi_{ij} \lambda_0(t_{ijk}) e^{\mathbf{x}_i \boldsymbol{\beta}} \right\} \exp\{-\theta_i \xi_{ij} \Lambda_0(T_i) e^{\mathbf{x}_i \boldsymbol{\beta}}\}$$

$$= \left\{ \prod_{j=1}^{J} \xi_{ij}^{K_{ij}} \right\} \{\theta_i e^{\mathbf{x}_i \boldsymbol{\beta}}\}^{K_{i\cdot}} \exp\{-\theta_i \Lambda_0(T_i) e^{\mathbf{x}_i \boldsymbol{\beta}}\} \prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} \lambda_0(t_{ijk}),$$

where we have used the fact that $\sum_{j=1}^{J} \xi_{ij} = 1$ for all i, and $K_{i\cdot} = \sum_{j=1}^{J} K_{ij}$ is the total number of events in individual i. If $K_{ij} = 0$ for any i, j, the empty product is defined to be unity.

Notice that this conditional likelihood factors into two parts: $L_1(\boldsymbol{\xi}_i)L_2(\theta_i, \lambda_0, \boldsymbol{\beta})$. Since the mixing distributions for $\boldsymbol{\xi}_i$ and θ_i are taken to be independent, this implies that the problem of estimating the parameters $\boldsymbol{\alpha}$ of the Dirichlet mixing distribution can be treated separately from the estimation of the other parameters, ν, γ, $\boldsymbol{\beta}$, and $\lambda_0(\cdot)$. We take a parametric approach and use the power-law intensity function mentioned earlier, so that $\lambda_0(t) = \delta t^{\delta-1}$. The contribution of the ith subject to the marginal likelihood function can be obtained by integrating out θ_i and $\boldsymbol{\xi}_i$ in the above expression. This contribution is given by

$$\begin{aligned}
&\int_0^1 \cdots \int_0^1 \int_0^\infty \Gamma\left(\sum_{j=1}^{J} \alpha_j\right) \prod_{j=1}^{J} \left[\frac{\xi_{ij}^{K_{ij}+\alpha_j-1}}{\Gamma(\alpha_j)}\, d\xi_{ij}\right] \int_0^\infty \left[\prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} \lambda_0(t_{ijk})\right] \\
&\qquad \times \exp(K_{i\cdot}\mathbf{x}_i\boldsymbol{\beta})\, \frac{\theta_i^{K_{i\cdot}+\nu-1} \exp[-\theta_i(T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1/\gamma)]}{\Gamma(\nu)\gamma^\nu}\, d\theta_i \\
&= \left[\frac{\Gamma(\sum_{j=1}^{J} \alpha_j)}{\Gamma(K_{i\cdot} + \sum_{j=1}^{J} \alpha_j)} \prod_{j=1}^{J} \frac{\Gamma(K_{ij} + \alpha_j)}{\Gamma(\alpha_j)}\right] \\
&\qquad \times \left[\frac{\Gamma(K_{i\cdot} + \nu)}{\Gamma(\nu)} \prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} t_{ijk}^{\delta-1} \frac{[\delta\gamma \exp(\mathbf{x}_i\boldsymbol{\beta})]^{K_{i\cdot}}}{[\gamma T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^{K_{i\cdot}+\nu}}\right] \\
&= L_{1i}(\boldsymbol{\alpha})L_{2i}(\nu, \gamma, \delta, \boldsymbol{\beta}), \quad \text{say.}
\end{aligned} \tag{1}$$

The total log likelihood is, therefore,

$$\mathscr{L}(\boldsymbol{\alpha}, \nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \{\mathscr{L}_{1i}(\boldsymbol{\alpha}) + \mathscr{L}_{2i}(\nu, \gamma, \delta, \boldsymbol{\beta})\}, \tag{2}$$

where $\mathscr{L}_{1i}$ and $\mathscr{L}_{2i}$ are given by

$$\mathscr{L}_{1i}(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \left[\sum_{k=1}^{K_{ij}} \log(\alpha_j + k - 1)\right] - \sum_{s=1}^{K_{i\cdot}} \log\left(\sum_{j=1}^{J} \alpha_j + s - 1\right), \tag{3}$$

$$\begin{aligned}
\mathscr{L}_{2i}(\nu, \gamma, \delta, \boldsymbol{\beta}) &= K_{i\cdot}[\log(\gamma\delta) + \mathbf{x}_i\boldsymbol{\beta}] + (\delta - 1) \sum_{j=1}^{J} \sum_{k=1}^{K_{ij}} \log(t_{ijk}) \\
&\quad + \sum_{s=1}^{K_{i\cdot}} \log(\nu + s - 1) - (K_{i\cdot} + \nu)\log[\gamma T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1].
\end{aligned} \tag{4}$$

Empty sums are taken to be zero.

For inference on $\boldsymbol{\alpha}$, expressions for the score vector, $\mathbf{U}(\boldsymbol{\alpha})$, and for the sample information matrix, $\mathbf{I}(\boldsymbol{\alpha})$, are given in the Appendix. For inference on ν, γ, δ, and $\boldsymbol{\beta}$, the score vector and sample information matrix are also given in the Appendix; however, for use in Section 3 we provide here $\mathbf{U}_{\beta_j}(\nu, \gamma, \delta, \boldsymbol{\beta})$ and $\mathbf{I}_{\beta_j\beta_k}(\nu, \gamma, \delta, \boldsymbol{\beta})$:

$$\begin{aligned}
\mathbf{U}_{\beta_j}(\nu, \gamma, \delta, \boldsymbol{\beta}) &= \sum_{i=1}^{I} \left\{K_{i\cdot}x_{ij} - \frac{\gamma(K_{i\cdot} + \nu)x_{ij}T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta})}{\gamma T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1}\right\} \\
&= \sum_{i=1}^{I} \left\{\frac{x_{ij}[K_{i\cdot} - \nu\gamma T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta})]}{\gamma T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1}\right\},
\end{aligned} \tag{5}$$

$$\mathbf{I}_{\beta_j\beta_k}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{\frac{\gamma(K_{i\cdot} + \nu)T_i^\delta x_{ij}x_{ik}\exp(\mathbf{x}_i\boldsymbol{\beta})}{[\gamma T_i^\delta \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2}\right\}. \tag{6}$$

Maximum likelihood estimates (MLEs) of the parameters of interest, $\boldsymbol{\alpha}$, ν, γ, δ, and $\boldsymbol{\beta}$, are obtained by setting the components of the score vector equal to zero and solving for the unknown values. An iterative procedure such as the Newton–Raphson method can be employed to solve the likelihood equations. For numerical stability, it is sometimes helpful to reparametrize the gamma prior in terms of ν and $\mu = \nu\gamma$ instead of ν and γ because of the asymptotic independence of $\hat{\nu}$ and $\hat{\mu}$.

2.2 *Estimation of θ_i and ξ_{ij}*

The parameters $\{\theta_i\}$ and $\{\boldsymbol{\xi}_i\}$ associated with each individual can be estimated by their respective a posteriori means. Since the priors for θ and $\boldsymbol{\xi}$ are taken as independent, and the likelihood factors, the posterior distributions for θ_i and $\boldsymbol{\xi}_i$ are also independent. Now, given $\mathbf{K}_i = (K_{i1}, \ldots, K_{iJ})$, it is straightforward to show that θ_i has a gamma distribution with shape parameter $K_{i\cdot} + \nu$ and mean $(K_{i\cdot} + \nu)/(T_i^{\delta}\exp(\mathbf{x}_i\boldsymbol{\beta}) + \gamma^{-1})$ and that $\boldsymbol{\xi}_i$ has a Dirichlet distribution with parameters $(\mathbf{K}_i + \boldsymbol{\alpha})$. Therefore, we estimate θ_i and ξ_{ij} by

$$\hat{\theta}_i = \frac{K_{i\cdot} + \hat{\nu}}{T_i^{\hat{\delta}}\exp(\mathbf{x}_i\hat{\boldsymbol{\beta}}) + \hat{\gamma}^{-1}}, \quad \hat{\xi}_{ij} = \frac{K_{ij} + \hat{\alpha}_j}{K_{i\cdot} + \sum_{s=1}^{J}\hat{\alpha}_s}. \tag{7}$$

These estimators will be useful later when we consider generalized residuals.

3. Regression Diagnostics

The problem of assessing the adequacy of fit of a heterogeneous point process regression model has received scant attention so far in literature. An exception is Lawless (1987, §5), who discussed generalized residuals. He suggested replacing the "random effects" by their posterior estimators and then using the unit exponential distribution as a reference frame for judging the adequacy of the model at hand. A simple extension to Lawless' work is described below. Given θ_i and ξ_{ij}, the type j specific cumulative intensity function for subject i is given by $\Lambda_{ij}(t\,;\,\mathbf{x}_i) = \theta_i\xi_{ij}T_i^{\delta}\exp(\mathbf{x}_i\boldsymbol{\beta})$, for $t \in [0, T_i]$. Therefore, under the Poisson process assumption, the quantities

$$e_{ijk} = \theta_i\xi_{ij}\exp(\mathbf{x}_i\boldsymbol{\beta})[t_{ijk}^{\delta} - t_{ijk-1}^{\delta}], \quad k = 1, \ldots, K_{ij},$$

are independent unit exponential variates. Also, when $k = K_{ij} + 1$, this quantity can be regarded as a censored observation from the unit exponential distribution. Replacing the unknown parameters by their estimates, we define the generalized residuals as

$$\hat{e}_{ijk} = \hat{\theta}_i\hat{\xi}_{ij}\exp(\mathbf{x}_i\hat{\boldsymbol{\beta}})[t_{ijk}^{\hat{\delta}} - t_{ijk-1}^{\hat{\delta}}], \quad k = 1, \ldots, K_{ij} + 1, \tag{8}$$

where, from (7), $\hat{\theta}_i$ and $\hat{\xi}_{ij}$ are the estimated posterior means of θ_i and ξ_{ij}, respectively. The quantities $\{\hat{e}_{ij}\}$ can be treated approximately as a censored random sample from the unit exponential distribution. Therefore, a cumulative hazard plot (Nelson, 1972) of the $\{\hat{e}_{ij}\}$ should produce an approximate straight line through the origin with slope 1 under the assumptions of this model.

Turning to the problem of the influence of observations, several methods have been proposed in the literature to investigate the consequence of modest perturbations on the analysis of a statistical model (see Cook, 1986). A common approach is the case deletion approach where the impact of deleting the ith observation on some quantity of interest is evaluated.

Consider a very general setting and let $\mathbf{w} = (w_1, \ldots, w_N)'$ be a vector of weights associated with the random sample of size N and assume that a weighted analysis is carried out. Let $\hat{\boldsymbol{\eta}}(\mathbf{w})$ be the MLE of the vector of unknown parameters, $\boldsymbol{\eta}$ say. In this discussion we confine our attention to the special weights $\mathbf{w} = (1, \ldots, w_i, \ldots, 1)'$; therefore, $\hat{\boldsymbol{\eta}}(\mathbf{w}) = \hat{\boldsymbol{\eta}}(w_i)$ can be thought of as a function of w_i. We restrict the choice of w_i so that $\hat{\boldsymbol{\eta}}(1) = \hat{\boldsymbol{\eta}}$

and $\hat{\boldsymbol{\eta}}(0) = \hat{\boldsymbol{\eta}}_{(i)}$, where $\hat{\boldsymbol{\eta}}_{(i)}$ is the MLE of $\boldsymbol{\eta}$ when observation i is deleted. The quantity $\boldsymbol{\Delta}_i = \hat{\boldsymbol{\eta}} - \hat{\boldsymbol{\eta}}_{(i)}$ measures the change exerted on $\hat{\boldsymbol{\eta}}$ upon deleting observation i. A substantial change in $\hat{\boldsymbol{\eta}}$ upon deleting observation i alerts us to possible errors in recording data, lack of sufficient data to fit a meaningful model, a need to modify the fitted model, or that this observation is unusual and deserves special consideration (Storer and Crowley, 1985). $\boldsymbol{\Delta}_i$ can, of course, be obtained by physically removing observation i and recomputing $\hat{\boldsymbol{\eta}}$. This approach, however, is very expensive in most cases and not practical especially if an iterative method is needed to obtain $\hat{\boldsymbol{\eta}}$. Closed-form expressions or good approximations for $\boldsymbol{\Delta}_i$ are available in many regression problems but, when $\hat{\boldsymbol{\eta}}$ must be obtained iteratively, analytic expressions are no longer possible. However, one-step approximations can be obtained for regression problems where the inverse of the information matrix can be expressed in terms of the inverse of the information matrix from the full model without the ith observation and quantities related to observation i alone (Storer and Crowley, 1985).

For the Cox proportional hazards model, Cain and Lange (1984) used the empirical influence curve to analyze the change in a regression coefficient estimate when observation i is deleted. They expanded $\hat{\boldsymbol{\eta}}(w_i)$ around $w_i = 1$ and obtained an approximation to $\boldsymbol{\Delta}_i$ for a Cox proportional hazards regression model. In this subsection we derive an approximation to $\boldsymbol{\Delta}_i$ for our models using this approach. As mentioned before, we consider $\hat{\boldsymbol{\eta}}$ with score vector $\mathbf{U}(\hat{\boldsymbol{\eta}}) = 0$ as functions of w_i. Following Cain and Lange (1984), we expand $\hat{\boldsymbol{\eta}}(w_i)$ around $w_i = 1$:

$$\hat{\boldsymbol{\eta}}(w_i) \approx \hat{\boldsymbol{\eta}}(1) + (w_i - 1)\left.\frac{\partial \hat{\boldsymbol{\eta}}(w_i)}{\partial w_i}\right|_{w_i=1}.$$

Evaluating at $w_i = 0$, we obtain $\boldsymbol{\Delta}_i$ in terms of the first derivative of $\hat{\boldsymbol{\eta}}(w_i)$ with respect to w_i evaluated at $w_i = 1$. Using the implicit relationship between the score vector and w_i, writing $\mathbf{U}[\hat{\boldsymbol{\eta}}(w_i)]$ as $\mathbf{U}[\hat{\boldsymbol{\eta}}(w_i), w_i]$, and taking the derivative of $\mathbf{U}[\hat{\boldsymbol{\eta}}(w_i), w_i]$ with respect to w_i, we obtain

$$\frac{\partial \mathbf{U}}{\partial w_i}\{\hat{\boldsymbol{\eta}}(w_i), w_i\} = 0 = \frac{\partial \mathbf{U}}{\partial \hat{\boldsymbol{\eta}}(w_i)} \cdot \frac{\partial \hat{\boldsymbol{\eta}}(w_i)}{\partial w_i} + \frac{\partial \mathbf{U}}{\partial w_i}.$$

Provided $[\partial \mathbf{U}/\partial \hat{\boldsymbol{\eta}}]^{-1}$ exists, we have

$$\boldsymbol{\Delta}_i \approx \frac{\partial \hat{\boldsymbol{\eta}}(w_i)}{\partial w_i} = \left[-\frac{\partial \mathbf{U}}{\partial \hat{\boldsymbol{\eta}}(w_i)}\right]^{-1} \cdot \frac{\partial \mathbf{U}}{\partial w_i} = \mathbf{I}^{-1}(\hat{\boldsymbol{\eta}}) \cdot \frac{\partial \mathbf{U}(\hat{\boldsymbol{\eta}})}{\partial w_i},$$

where the above expressions are evaluated at $w_i = 1$ and where $\mathbf{I}(\cdot)$ is the sample information matrix.

This measure is useful in its present form only when $\boldsymbol{\eta}$ is one-dimensional. In this case we propose plotting the points (i, Δ_i) and declaring an observation to be influential if it is "separated" from the rest of the points. When the dimension of $\boldsymbol{\eta}$ is greater than 1, then a scalar measure of influence can be obtained by normalizing $\boldsymbol{\Delta}_i$. One possible normalization is to use the inverse of the estimated covariance matrix to get the scalar quantity $\nabla_i = \boldsymbol{\Delta}_i' \mathbf{I}(\hat{\boldsymbol{\eta}})\boldsymbol{\Delta}_i$. This choice of normalizing matrix is motivated by the discussion presented in Section 3.5 of Cook and Weisberg (1982, pp. 113–135). To apply the above derivations to our model we need to specify the weights such that $w_i = 1$ gives the full model and $w_i = 0$ gives the full model without the ith observation. A simple weighting scheme that satisfies the above requirements allocates the weights to the log likelihood contribution, $\mathscr{L}_i(\boldsymbol{\eta})$, from the ith observation so that the total log likelihood is $\sum w_i \mathscr{L}_i(\boldsymbol{\eta})$.

We now return to our model for replicated multi-type point processes. We will concentrate on the influence of the ith subject on the regression coefficients $\boldsymbol{\beta}$ $(=\boldsymbol{\eta})$. As noted

before, the likelihood function factors and we just use the component $\mathscr{L}_2(\nu, \gamma, \delta, \boldsymbol{\beta})$ of the log likelihood where $\mathscr{L}_2(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \mathscr{L}_{2i}(\nu, \gamma, \delta, \boldsymbol{\beta})$. Suppose first that δ, ν, and γ are completely specified and define the following scalars, vectors, and matrices:

$$a_i = a_i(T_i) = (\gamma T_i^{\delta} \exp(\mathbf{x}_i \boldsymbol{\beta}) + 1)^{-1}, \quad \mathbf{A} = \operatorname{diag}(a_1, \ldots, a_I),$$

$$s_i = K_{i\cdot} - \nu\gamma T_i^{\delta} \exp(\mathbf{x}_i \boldsymbol{\beta}), \quad \mathbf{S} = (s_1, \ldots, s_I)',$$

$$v_i = \gamma(K_{i\cdot} + \nu) T_i^{\delta} \exp(\mathbf{x}_i \boldsymbol{\beta})[\gamma T_i^{\delta} \exp(\mathbf{x}_i \boldsymbol{\beta}) + 1]^{-1}, \quad \mathbf{V} = \operatorname{diag}(v_1, \ldots, v_I),$$

$$\mathbf{x}_i = (x_{i1}, \ldots, x_{ip}), \quad \mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_I')', \quad p = \dim(\boldsymbol{\beta}),$$

$$\mathbf{z}_i = (z_{i1}, \ldots, z_{ip}), \quad \mathbf{Z} = (\mathbf{z}_1', \ldots, \mathbf{z}_I')' = \mathbf{A}^{1/2}\mathbf{V}^{1/2}\mathbf{X},$$

$$\mathbf{Y} = (y_1, \ldots, y_I)' = \mathbf{V}^{-1/2}\mathbf{A}^{1/2}\mathbf{S}.$$

Then the sample information matrix, $\mathbf{I}(\boldsymbol{\beta})$, can be written as $\mathbf{I}(\boldsymbol{\beta}) = \mathbf{X}'\mathbf{A}\mathbf{V}\mathbf{X} = \mathbf{Z}'\mathbf{Z}$ and

$$\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial w_i} = \frac{\partial \mathscr{L}_{2i}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} x_{i1} a_i s_i \\ \vdots \\ x_{ip} a_i s_i \end{bmatrix} = \mathbf{x}_i' a_i s_i = \mathbf{z}_i' y_i.$$

Using (5) and (6) and substituting the MLE values for the unknown $\boldsymbol{\beta}$ yields

$$\boldsymbol{\Delta}_i = [\mathbf{I}(\hat{\boldsymbol{\beta}})]^{-1} \left[\frac{\partial \mathbf{U}(\hat{\boldsymbol{\beta}})}{\partial w_i} \right] = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_i' y_i$$

and $\nabla_i = y_i \mathbf{z}_i (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_i' y_i = y_i^2 h_{ii}$, where h_{ii} is the ith diagonal element of the projection matrix $\mathbf{H} = (h_{ij}) = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. Notice that aside from a scaling factor, this quantity is precisely the familiar Cook distance (Cook, 1977) for detecting influential observations in linear regression. This suggests that the measure ∇_i may be used in the same way as we use the Cook distance in linear regression. When nuisance parameters ν, γ, δ, and $\boldsymbol{\beta}$ are unknown, we can simply replace them with their MLEs from the full model, although, more precisely, the appropriate elements of the inverse of the partitioned full information matrix should be employed [see, e.g., Cox and Hinkley (1974, p. 323); for rigorous details, see Abu-Libdeh (unpublished Ph.D. dissertation, Cornell University, 1988). Hence, to investigate the influence of each observation on the MLEs we proceed as follows:

1. Construct the vectors and matrices **S**, **A**, **V**, **Z**, and **Y**, using parameter values estimated from the full data.
2. Regress **Y** and **X** and evaluate the prediction matrix **H**.
3. Evaluate the values $\nabla_i = y_i^2 h_{ii}$ $(1 \leq i \leq I)$.
4. Plot the pairs $\{(i, \nabla_i): i = 1, \ldots, I\}$ and declare observation i to be influential if it is "separated" from the remaining points. Alternatively, when plotting, the ∇_i $(1 \leq i \leq I)$ can be reordered by values of a specific covariate or by values of the subject's crude failure rate, $K_{i\cdot}/T_i$.

If interest lies in checking the significance of covariates included in the model, rather than the point estimation of regression coefficients, we may also wish to examine the effect of deleting the ith subject upon the score test statistic. Lustbader and Moolgavkar (1985) have used this approach in matched case–control and survival studies. Based on the log likelihood $\mathscr{L}_2$, a scalar measure of influence of the ith subject on the score test statistic is given by the difference Sc_i in the score statistic for the full model and that obtained when the ith subject is omitted. The differences $\{Sc_i: 1 \leq i \leq I\}$ can be plotted in the same way as the $\{\nabla_i\}$.

4. Application to Skin Cancer Prevention Trial

Here we apply our methods to the skin cancer prevention trial described in Section 1. Included in the analysis are 770 patients, all of whom have had at least one follow-up visit. The patients were recruited from seven clinics and constitute those available at the second of seven scheduled interim analyses. Here, for the purpose of illustration, we treat the data as coming from a fixed single-sample design although, according to the protocol, all analyses will be adjusted for the sequential nature of the study using the methods of Jennison and Turnbull (1984). The patients are randomized into two groups. In one, the patients receive a daily nutritional supplement of 200 mcg of selenium as a selenium-enriched brewer's yeast; the other group receives a placebo. At this interim analysis, even though the coding of the treatment indicator covariate is blinded, the principal investigator requested that we deliberately change some treatment assignment indicator covariate values in using this data set to illustrate our methods. This was done to prevent premature conclusions and speculation on the treatment effect. A number of authors have prescribed this; see, for example, Geller and Pocock (1987, p. 219). In an interesting study, Green, Fleming, and O'Fallon (1987) find that mere publication of interim results can precipitate a decline in patient accrual and physician participation. We believe that the data set will serve our purpose well in illustrating the steps of statistical modeling and analysis we have proposed. However, because of the data modifications, it cannot be used to assess the significance or effect of the selenium treatment.

The events of interest are the successive incidence times of the two types of tumors, where we define an "event" or a "failure" of the first type to be the detection of one or more basal cell epithelioma (BCE) at a clinic visit. Similarly, a "failure" of the second type is the medical detection of one or more squamous cell carcinoma (SCC) at a clinic visit. At a later stage with more complete data, we intend to enlarge the number of types to distinguish the sites on the body where the tumors occur. Table 1 summarizes the distribution of patients according to the number of failures of each type. The first column of the table shows the exact number of failures of each type that the patient had.

The covariates entering the analysis are coded as follows: "Group" = 1 or 2; "Age" = age in years at randomization date; "Gender" = 1 for males, 0 for females; "Smoking" = 1 if the patient was a smoker as of 1 month before randomization date, 0 otherwise; "Farming" = 1 if the patient lived or worked on a farm until the age of 18, 0 otherwise;

Table 1
Frequency distribution of the number of patients according to the number of failures from each type

	Group 1 (386 patients)		Group 2 (384 patients)	
Number of failures	Type 1 (BCE)	Type 2 (SCC)	Type 1 (BCE)	Type 2 (SCC)
0	292	347	279	328
1	62	27	52	44
2	17	9	24	7
3	8	1	12	5
4	5	2	9	
5	2		2	
6			3	
7			2	
8			0	
9			0	
10			1	
Total failures:	150	56	224	73

"Sun-burn" = answer to the question "when you go out in the sun during early summer for 30–45 minutes, do you sun-burn? 1 = always, 2 = sometimes, 3 = never"; Ptum = number of BCE or SCC tumors that were recorded in the 5 years preceding randomization; Clinic-site = a vector of six dummy variables indicating which of the seven clinics treated the patient. As mentioned above, some group covariate values were deliberately switched; the remaining covariates and the response variables in the data set represent the actual information from the trial participants.

The analysis of this data set was carried out on an IBM-AT personal computer using the Newton–Raphson procedure in a portable FORTRAN program. The starting values for the parameters of the random effect distributions were calculated by a crude method of moments. The starting values for the regression coefficients were chosen to be zero. All results were checked by comparison with a program written independently in GAUSS.

4.1 *Estimation of the Parameters of the Mixing Distributions*

From (2) and (3), we use the first part of the log likelihood function to estimate the parameters α_1 and α_2 of the Dirichlet mixing distribution, which now reduces to a beta distribution since $J = 2$ here. The maximum likelihood estimates of these parameters and their standard deviations (in parentheses) are 1.6771 (.2685) for α_1 and .6230 (.1266) for α_2. The estimated proportion of BCEs is then $\hat{\alpha}_1/(\hat{\alpha}_1 + \hat{\alpha}_2) = .7291$, as opposed to the crude proportion of $(150 + 224)/(150 + 56 + 224 + 73) = .7435$. The latter would be the estimate of ξ_1 if this fraction was assumed constant across all patients. However, a histogram of the $\hat{\xi}_{i1}$ was not too revealing because so many $K_{i\cdot}$ were equal to zero.

4.2 *Covariate Selection*

The analysis can be viewed as a test of treatment effect with adjustment for the other covariates; thus all models included the "Group" covariate. For the other variables, the following stepwise procedure was employed:

1. Fit models with two covariates only—namely, the group indicator and each of the remaining covariates.
2. Calculate the significance level or *P*-value for each of the covariates added in step 1 and retain in the model the covariate that satisfies at least one of the following conditions:
   - **i.** Its *P*-value is the smallest among all calculated *P*-values and is also smaller than .25.
   - **ii.** The inclusion of that covariate also results in a practical change in the *Z*-value for the group effect.
   - **iii.** The covariate is considered biologically important.
3. Now continue to add in covariates one at a time using the above selection criteria until no omitted covariate satisfies any of the conditions 2(i)–2(iii).

Using the above criteria for covariate selection, a nonhomogeneous Poisson process mixture regression model was selected in four steps. It included the covariates: Group, Previous number of tumors (Ptum), Clinic-site, and Gender. Estimates of the parameters and associated standard errors (in parentheses) for this model are displayed in the "Model 5" column of Table 2. The maximum likelihood estimates of the parameters of the model selected at each step are also displayed in Table 2 (Models 1–4). The standard errors were obtained from the square roots of the diagonal elements of the inverse of the sample information matrix. Notice that the inclusion of the covariate Ptum (Model 3) has produced a significant reduction in the *Z*-value for the group effect (from 2.1472 to 1.4762).

Table 2
Covariate selection results for Poisson mixture regression models

Par.	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
ν	.902	.923	1.272	1.483	1.526	1.479	1.640
	(.146)	(.151)	(.243)	(.301)	(.313)	(.298)	(.379)
$\mu = \nu\gamma$	.00076	.00067	.00060	.00073	.00052	.00090	.00091
	(.00021)	(.00019)	(.00017)	(.00023)	(.00018)	(.00020)	(.00020)
δ	1.089	1.087	1.055	1.094	1.097	1	1
	(.043)	(.043)	(.042)	(.045)	(.045)	—	—
Group		.270	.179	.150	.166	.175	.151
		(.126)	(.122)	(.118)	(.118)	(.118)	(.117)
Ptum			.066	.075	.072	.073	.081
			(.011)	(.011)	(.011)	(.011)	(.013)
Clinic 2				−.495	−.414	−.349	−.400
				(.206)	(.207)	(.205)	(.203)
Clinic 3				−.229	−.135	−.106	−.105
				(.194)	(.197)	(.196)	(.194)
Clinic 4				−.822	−.715	−.618	−.639
				(.240)	(.244)	(.239)	(.237)
Clinic 5				−.743	−.770	−.798	−.804
				(.315)	(.315)	(.314)	(.312)
Clinic 6				−.715	−.749	−.710	−.872
				(.367)	(.366)	(.365)	(.395)
Clinic 7				−.948	−.904	−.799	−.823
				(.246)	(.246)	(.241)	(.238)
Gender					.361	.354	.339
					(.139)	(.139)	(.138)
X^2		4.604	48.704	72.807	79.666	79.243	64.711
(P-value)		(.032)	(≪.01)	(≪.01)	(≪.01)	(≪.01)	(≪.01)

[a] Model 7 based on data with 5 outlier subjects deleted; see Section 4.4.

Moreover, this covariate has drastically increased the significance of the regression model. These results indicate that this covariate is a very strong predictor, as might be expected. Examining the values of this covariate, we see that there are 76 patients who have 10 or more prerandomization tumors and one patient with a value of 68! The exclusion of these patients from the analysis resulted in a drastic drop in the Z-value for this covariate (from approximately 6 to approximately 2). The inclusion of the Clinic-site covariate in the regression model (see Model 4 in Table 2) reduced further the significance of the group effect. The signs of the maximum likelihood estimates for this variable indicate that in comparison with baseline clinic (#1), the relative risk of these clinics is smaller. Moreover, the Z-values of some of these clinics indicate the possible existence of heterogeneity among patients of these clinics and/or clinic populations. The Z-value for the Gender covariate in the "Model 5" column of Table 2 shows that males are at significantly higher risk of skin cancer compared with females. This finding might be explained by the fact that men are more likely to be occupationally exposed to the sun than women.

Also from Table 2, a likelihood ratio test of time homogeneity of the Poisson failure process (H_0: $\delta = 1$) is barely significant at the 5% level. Given the fact that the sample size (770) is large and that $\hat{\delta}$ is very close to 1 ($1.0 < \hat{\delta} < 1.1$) in every model considered, it appears that the time-homogeneity assumption on the individual Poisson failure processes is an acceptable one. To examine the effect of dropping δ from our model, we performed a complete analysis under the assumption that $\delta = 1$. As might be expected, the stepwise selection of covariates produced the same sequence of models as before; the parameter estimates of the selected model are displayed in the "Model 6" column in Table 2. These

estimates are also very close to that of the model with $\delta \neq 1$. Hence the simpler time-homogeneous Model 6 was chosen as a final model and we proceed with a sensitivity analysis. (Model 7 will be discussed in Section 4.4.)

The correlation matrix for the final homogeneous Poisson mixture regression model is displayed in Table 3; the diagonal elements of the matrix are the estimated standard deviations of the parameters of the selected regression model (Model 6 of Table 2).

A histogram of the 770 $\{\hat{\theta}_i\}$ values, estimated using (7) and Model 6 parameter estimates, is shown in Figure 1. Its shape is consistent with the gamma mixing distribution assumption.

Table 3

Correlation matrix for the selected homogeneous Poisson process mixture model

(.2981)	.084	−.029	−.170	.028	−.008	−.003	.008	.006	.013	−.045
	(.000199)	−.324	−.132	−.634	−.714	−.578	−.357	−.306	−.513	−.587
		(.1177)	−.078	.045	.017	.042	.020	.011	.021	.060
			(.0114)	−.225	.019	−.151	−.027	−.008	−.132	−.053
				(.2046)	.666	.586	.404	.344	.564	.149
					(.1961)	.579	.413	.356	.557	.181
						(.2393)	.343	.292	.480	.174
							(.3143)	.227	.343	−.033
								(.3648)	.294	−.032
									(.2408)	.072
										(.1391)

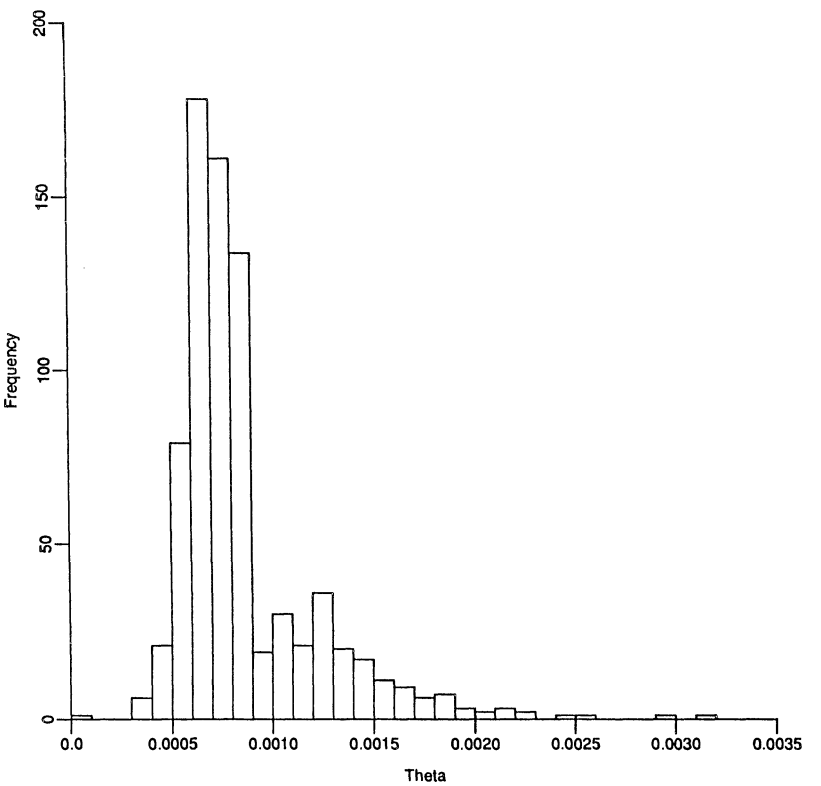


Figure 1. Frequency histogram of the posterior estimated failure rates $\{\hat{\theta}_i\}$ for 770 subjects.

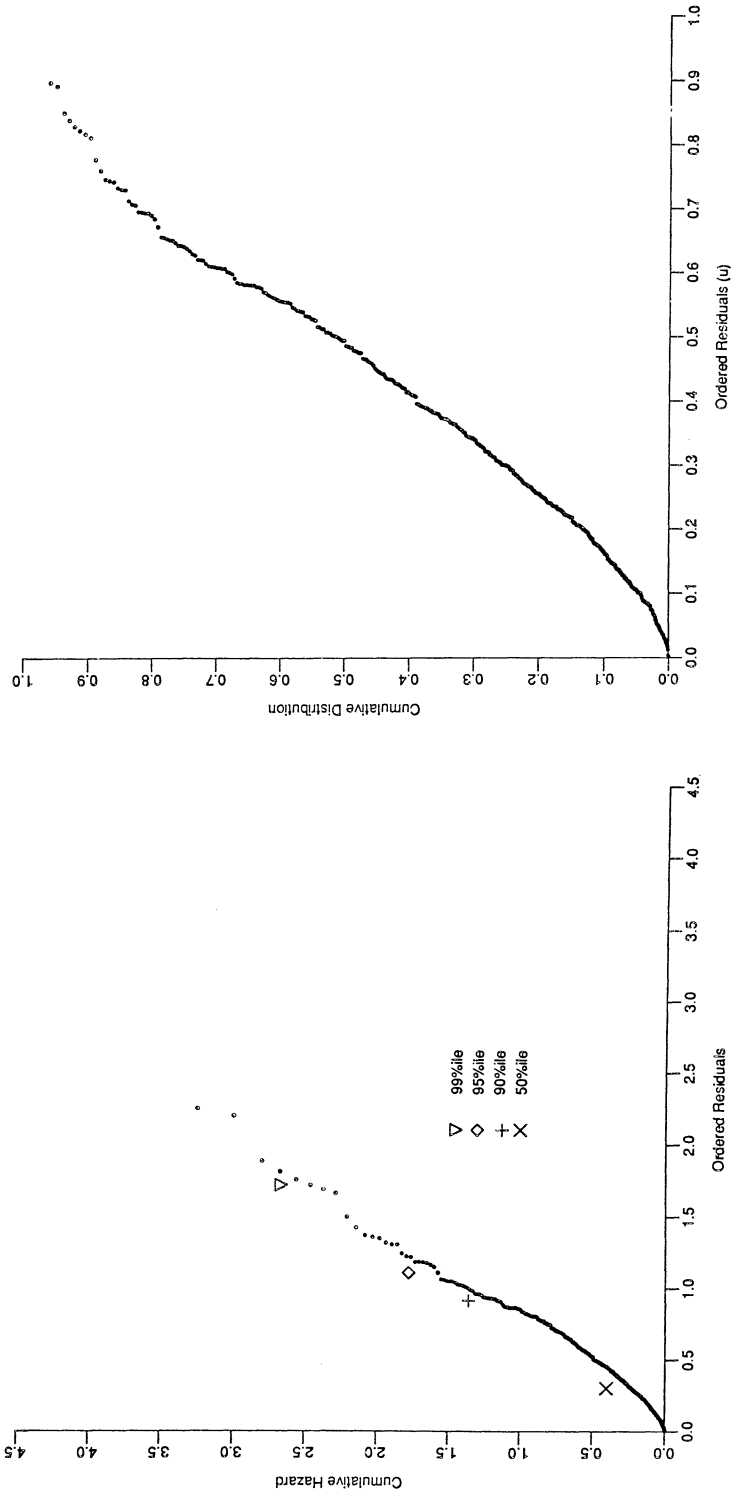
The MLE for the mean of the gamma mixing distribution for $\theta = 9.0 \times 10^{-4}$/day. The average of the a posteriori estimates $\hat{\theta}_i$ is 8.6×10^{-4}/day. In a homogeneous model, when baseline failure rates are considered equal for all patients, the common value for θ is estimated by 7.4×10^{-4}/day.

4.3 *Generalized Residuals*

Following Lawless (1982, p. 282; 1987, §5), we evaluated all 2,043 generalized residuals $\{\hat{e}_{ijk}\}$ for the 770 patients using equation (8) and the covariates of the selected model (see Model 6 of Table 2). Of the 2,043 residuals, 503 are uncensored corresponding to the 503 tumor occurrences and the remainder are treated as censored. The empirical cumulative hazard function of the residuals is plotted in Figure 2a. The plot exhibits some curvature in the upper tail, although roughly 90% of the residuals form an acceptable approximation to a straight line with slope 1 through the origin. The residuals in the upper tail come mainly from some patients with large numbers of previous tumors (Ptum), suggesting that the model may not be adequate for such patients. However Figure 2a may be deceiving. Because of the increasing increments between values of successive expected exponential order statistics (the "exponential scores"), the figure appears to give more weight to the upper tail where there are few observations. This "exaggerated visual effect in the upper tail" has been noted by Elandt-Johnson and Smith (1989, p. 708). Instead, when concerned with "overall" fit, they recommend plotting so-called u-residuals, which should be approximately uniformly distributed under the assumed model. The Kaplan–Meier estimate (Miller, 1981, p. 46) of the distribution function of the 2,043 transformed residuals $\hat{u}_{ijk} = 1 - \exp(-\hat{e}_{ijk})$ is shown in Figure 2b. This is approximately equivalent (Miller, 1981, p. 66) to the cumulative hazard plot, Figure 2a, when scales of both axes have been subjected to this transformation. When graphed this way (Figure 2b), the plot appears much closer to the desired straight line with unit slope through the origin, the distribution function of the uniform distribution. This result is consistent with the fact that when the selected model (Model 6 in Table 2) is compared to a larger "perturbed" model (Model 5) in which it is embedded, there is only a marginal practical improvement in fit. An alternative way to transform the scales of the cumulative hazard plot, Figure 2a, to avoid the exaggerated visual effects of the upper tail is to use the empirical transformation by which the percentiles of the observed ordered residuals are equally spaced. Here, the resulting plot is similar to that in Figure 2b.

4.4 *Influence Analysis*

In this subsection using the selected Model 6 of Table 2, we investigate the influence of individual observations on both the estimated regression coefficients and the score test for testing the null hypothesis that these regression coefficients are zero. Recall that we defined an observation to be influential on the estimated regression coefficients if the point (i, ∇_i) is clearly separated from the remaining points in a plot of $\{(i, \nabla_i): i = 1, \ldots, I\}$. Such an observation is usually an outlier in the response or factor spaces. Hence, it is reasonable to expect ∇_i to be large for those patients with a noticeably large number of failures or a covariate value that is considerably different from the remaining values of that covariate. The plot of the points (i, ∇_i) and the points (i, Sc_i) are displayed in Figures 3 and 4, respectively. In both of these figures, the patients are arranged in rank order of their crude failure rates, where the crude failure rate for the ith patient is defined as $K_{i\cdot}/T_i$, ties being decided in favor of the patient with larger value of T_i. In Figure 3, the four most influential observations in order were patients 335, 686, 746, and 754, whereas in Figure 4, the most influential four were patients 615, 686, 746, and 754. The influence measures for the five patients mentioned are summarized in Table 4. Upon inspection of the observed data on



(**a**) Exponential residuals (**b**) Transformed uniform (u-) residuals

Figure 2. Generalized residual plots.

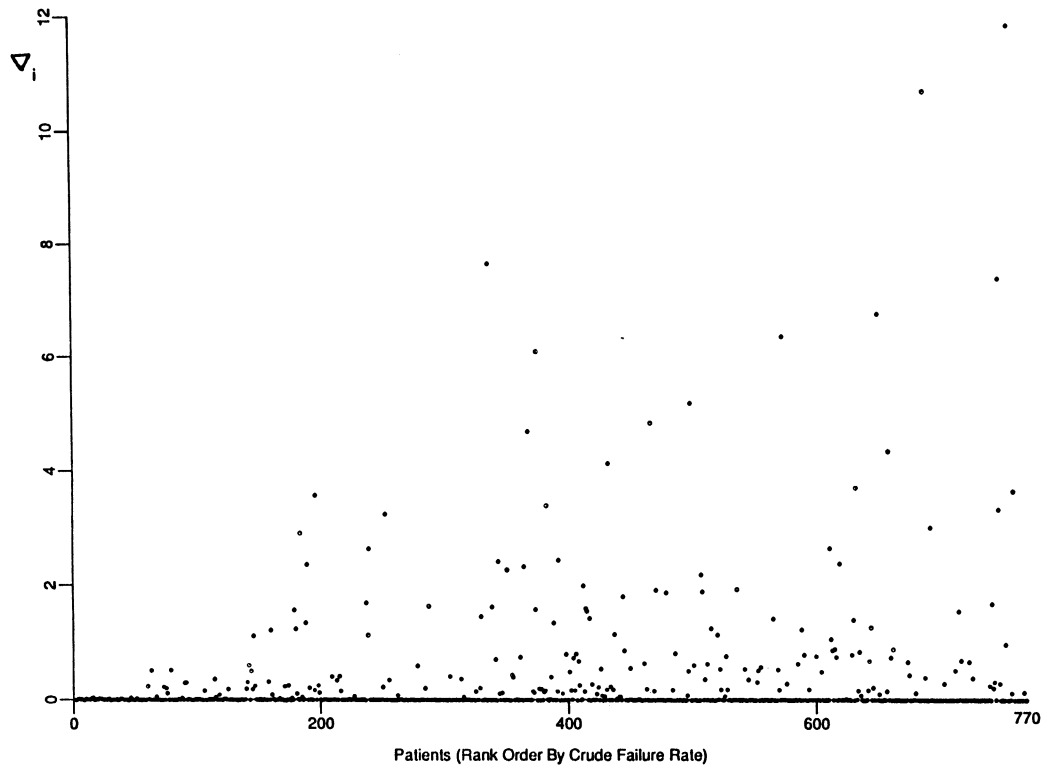


Figure 3. Scatter plot of the scalar measure of influence, ∇_i, on the estimated regression parameters. The horizontal axis represents the patients in rank order of crude failure rate.

these patients, we observe the following:

1. Except for patient 335, each of these patients has a large number of failures and a large value for Ptum. In particular, patient 615 has a value of Ptum = 68. However, patient 335 has only 3 failures and a value of Ptum = 0.
2. The total follow-up time on each of these patients is less than 2.9 years except for patient 335 who has a follow-up time of 4 years.
3. All these patients belong to group 2 and have high sun-burn values.
4. All these patients are male nonsmokers except for patient 746 who smokes.
5. Patients 615, 686, and 754 have worked on farms, where they were probably exposed to arsenical pesticides, which are known skin carcinogens.

Next, we compare the X^2 value of the selected model with that obtained after deleting each of the observations flagged by Δ_i one at a time and then when all the four observations are deleted simultaneously. Upon deleting observation 686, X^2 dropped by 5.81, for observation 746 the drop is by 7.4, for observation 754 the drop is 6.37, whereas for observation 335 the X^2 value has increased by .84. However, when all four observations were deleted, the drop in X^2 value was by 19.33. Turning to the observations flagged by Sc_i, the drop in the score test statistic was by 8.69, 10.20, 8.19, and −20.52, respectively, when each of the observations 686, 746, 754, and 615 was individually deleted. Notice that the deletion of observation 615 has increased the significance of the selected model. This is in agreement with the sign of Sc_i for this observation. The deletion of all of the four observations simultaneously has an effect of dropping the value of the score test statistic by

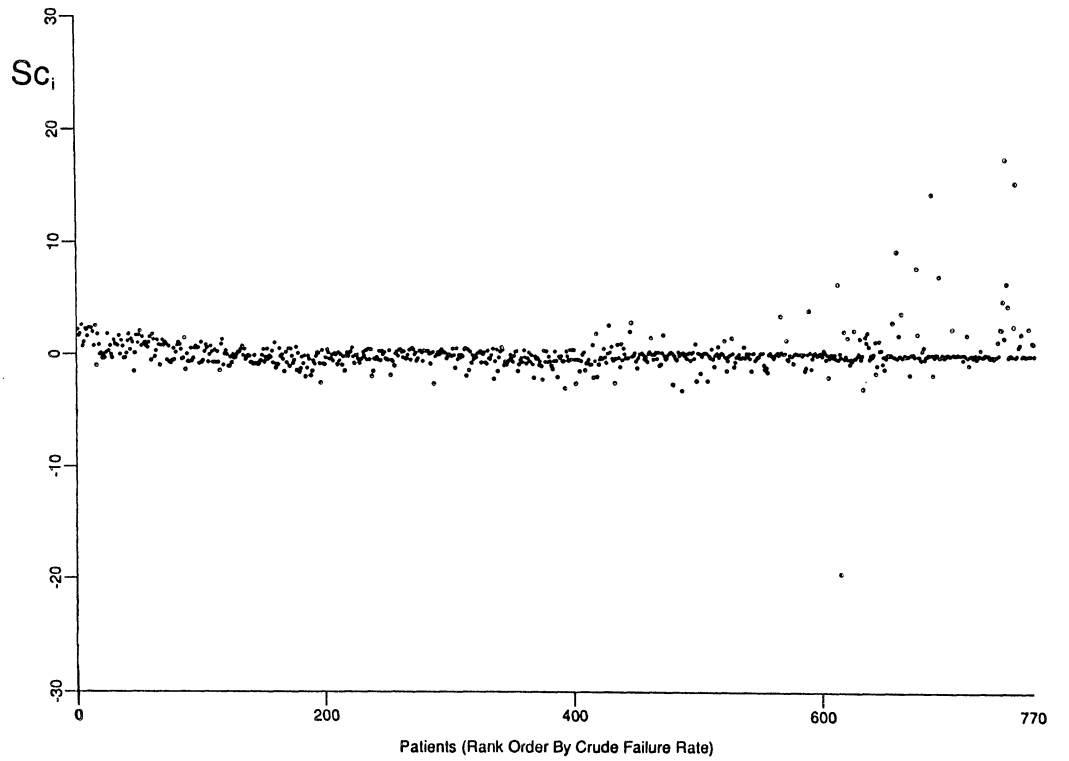


Figure 4. Scatter plot of the scalar measure of influence, Sc_i, on the score test. The horizontal axis represents the patients in rank order of crude failure rate.

Table 4
Values of the influence measures for the most influential observations

	Patient number (in rank order of crude failure rate)				
	335	615	686	746	754
Failure rate	.0030	.0064	.0102	.0144	.0177
∇_i	7.664	.898	10.712	7.409	11.873
Sc_i	−2.133	−19.489	14.432	17.593	15.487

12.22. However, when observation 615 was retained in the data, this drop was by 26.78. Parameter estimates and test statistics that result from deleting all five observations are displayed in the "Model 7" column of Table 2.

5. Final Remarks

The Poisson process assumption for this model can be relaxed to allow for wider applicability. An important generalization is to allow for type- (or site-) specific covariate effects. It might also be desirable to allow for time-dependent covariates. As mentioned in Section 2, a renewal process model can be considered by changing the time scale in the functional form of the intensity function (Prentice et al., 1981). A further generalization of the model can be made by allowing multiplicities or tied event times and modeling the failure process by a generalized Poisson process. This generalization, however, requires some specifications on the distributional aspects of these multiplicities. A topic related to

the treatment of multiplicities concerns the analysis of point processes when the data come in the form of interval counts. This would involve extensions of the work of Thall (1988) or of Abu-Libdeh (unpublished Ph.D. dissertation, Cornell University, 1988). It would also be desirable to develop extensions to multi-type point processes of semiparametric procedures based on partial likelihoods as suggested by Prentice et al. (1981, p. 379). The influence measures developed in this article appear to be sensitive to outlying observations. More work is needed to study their accuracy and distributional properties.

Acknowledgements

The authors are grateful to Professor D. R. Cox for some very helpful discussions. This research was supported in part by grants from the U.S. National Institutes of Health and from the American Cancer Society.

Résumé

Nous étudions la modélisation et l'analyse statistiques de processus ponctuels multiformes avec covariables. De telles données sont obtenues avec des sujets hétérogènes chez lesquels on peut observer plusieurs événements de différente nature. Le processus sous-jacent est modélisé par des processus de Poisson mixtes et non homogènes, à effets fixes (les covariables) et aléatoires (les individus). On obtient, par maximum de vraisemblance, des estimations des taux d'apparition des événements, des coefficients de régression et de laurs écarts-types. La sélection des covariables est opérée à l'aide de score-tests et de rapports de vraisemblance. A l'aide de résidus généralisés, on construit un test graphique d'ajustement de modèle proposé. On développe des évaluations de l'influence d'une observation particulière tant sur les coefficients de la régression que sur la statistique du score test. On décrit une application à un important essai clinique actuellement en cours, sur l'efficacité d'une alimentation enrichie en Selenium pour la prévention de deux types de cancer de la peau.

References

Allison, P. D. (1984). *Event History Analysis*: *Regression for Longitudinal Data*. Beverly Hills, California: Sage Publications.

Ascher, H. and Feingold, H. (1984). *Repairable Systems Reliability*: *Modeling, Inference, Misconceptions and Their Causes*. New York: Marcel Dekker.

Cain, K. C. and Lange, N. T. (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics* **40,** 493–499.

Cook, D. R. (1977). Detection of influential observations in linear regression. *Technometrics* **19,** 15–18.

Cook, D. R. (1986). Assessment of local influence (with Discussion). *Journal of the Royal Statistical Society, Series B* **42,** 133–169.

Cook, D. R. and Weinberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

Crow, L. H. (1974). Reliability analysis of complex, repairable systems. In *Reliability and Biometry*, F. Proschan and R. J. Serfling (eds), 379–410. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Elandt-Johnson, R. C. and Smith, F. B. (1989). Graphical generalized residuals in fitting distributions: Applications to epidemiological follow-up data. *Statistics in Medicine* **8,** 703–723.

Follman, D. A. and Goldberg, M. S. (1988). Distinguishing heterogeneity from decreasing hazard rates. *Technometrics* **30,** 389–396.

Freedman, L., Sylvester, R., and Byar, D. P. (1989). Using permutation tests and bootstrap confidence limits to analyze repeated events data from clinical trials. *Controlled Clinical Trials* **10,** 129–141.

Gail, M. H., Santner, T. J., and Brown, C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36,** 255–266.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics* **29,** 1–13.

Geller, N. L. and Pocock, S. J. (1987). Interim analysis in randomized clinical trials: Ramifications and guidelines for practitioners. *Biometrics* **43,** 213–223.

Green, S. J., Fleming, T. R., and O'Fallon, J. R. (1987). Policies for study monitoring and interim reporting of results. *Journal of Clinical Oncology* **5,** 1477–1484.
Jennison, C. and Turnbull, B. W. (1984). Repeated confidence intervals for group sequential clinical trials. *Controlled Clinical Trials* **5,** 33–45.
Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics*: *Continuous Univariate Distributions I.* Boston: Houghton Mifflin.
Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions I.* New York: Wiley.
Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* New York: Wiley.
Lawless, J. F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association* **82,** 808–815.
Lee, L. and Lee, K. (1978). Some results on inference for the Weibull process. *Technometrics* **20,** 41–45.
Lustbader, E. D. and Moolgavkar, S. H. (1985). A diagnostic statistic for the score test. *Journal of the American Statistical Association* **80,** 375–379.
Miller, R. G. (1981). *Survival Analysis.* New York: Wiley.
Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14,** 945–966.
Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.
Storer, B. E. and Crowley, J. (1985). A diagnostic for Cox regression and general conditional likelihoods. *Journal of the American Statistical Association* **80,** 139–147.
Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44,** 197–209.

Received September 1988; *revised September* 1989; *accepted January* 1990.

Appendix

Components of Score Vector and Information Matrix for Likelihood Based on Random Effects Poisson Process Model of Section 2

For inference on $\boldsymbol{\alpha}$, the score vector, $\mathbf{U}(\boldsymbol{\alpha})$, and sample information matrix $\mathbf{I}(\boldsymbol{\alpha})$, are given by

$$U_{\alpha_j}(\boldsymbol{\alpha}) = \sum_{i=1}^{I} \left\{ \sum_{k=1}^{K_{ij}} (\alpha_j + k - 1)^{-1} - \sum_{s=1}^{K_{i\cdot}} \left(\sum_{j=1}^{J} \alpha_j + s - 1 \right)^{-1} \right\},$$

$$\mathbf{I}_{\alpha_j\alpha_k}(\boldsymbol{\alpha}) = \sum_{i=1}^{I} \left\{ \sum_{k=1}^{K_{ij}} -\delta_{jk}(\alpha_j + k - 1)^{-2} + \sum_{s=1}^{K_{i\cdot}} \left(\sum_{j=1}^{J} \alpha_j + s - 1 \right)^{-2} \right\},$$

where δ_{jk} is the Kronecker delta.

For inference on ν, γ, δ, and $\boldsymbol{\beta}$, the score vector and sample information matrix are given by

$$U_{\nu}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \sum_{s=1}^{K_{i\cdot}} (\nu + s - 1)^{-1} - \log[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1] \right\},$$

$$U_{\gamma}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{K_{i\cdot}}{\gamma} - \frac{(K_{i\cdot} + \nu) T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta})}{\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1} \right\},$$

$$U_{\delta}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{K_{i\cdot}}{\gamma} - \frac{\gamma(K_{i\cdot} + \nu) T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) \log(T_i)}{\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1} + \sum_{j=1}^{J} \sum_{k=1}^{K_{ij}} \log(t_{ijk}) \right\},$$

$$U_{\beta_j}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ K_{i\cdot} x_{ij} - \frac{\gamma(K_{i\cdot} + \nu) x_{ij} T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta})}{\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1} \right\},$$

$$\mathbf{I}_{\nu\nu}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \sum_{s=1}^{K_{i\cdot}} (\nu + s - 1)^{-2} \right\},$$

$$\mathbf{I}_{\nu\gamma}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta})}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} \right\},$$

$$\mathbf{I}_{\nu\delta}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) \log(T_i)}{\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1} \right\},$$

$$\mathbf{I}_{\nu\beta_j}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{\gamma x_{ij} T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta})}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} \right\},$$

$$\mathbf{I}_{\gamma\gamma}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ (K_{i\cdot} + \nu) \left[\frac{T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta})}{\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1} \right]^2 - \frac{K_{i\cdot}}{\gamma^2} \right\},$$

$$\mathbf{I}_{\gamma\delta}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{(K_{i\cdot} + \nu) T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) \log(T_i)}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} \right\},$$

$$\mathbf{I}_{\gamma\beta_j}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{(K_{i\cdot} + \nu) x_{ij} T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta})}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} \right\},$$

$$\mathbf{I}_{\delta\delta}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{\gamma(K_{i\cdot} + \nu) T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) [\log(T_i)]^2}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} + \frac{K_{i\cdot}}{\delta^2} \right\},$$

$$\mathbf{I}_{\delta\beta_j}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{\gamma(K_{i\cdot} + \nu) T_i^{\delta} x_{ij} \exp(\mathbf{x}_i\boldsymbol{\beta}) \log(T_i)}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} \right\},$$

$$\mathbf{I}_{\beta_j\beta_k}(\nu, \gamma, \delta, \boldsymbol{\beta}) = \sum_{i=1}^{I} \left\{ \frac{\gamma(K_{i\cdot} + \nu) T_i^{\delta} x_{ij} x_{ik} \exp(\mathbf{x}_i\boldsymbol{\beta})}{[\gamma T_i^{\delta} \exp(\mathbf{x}_i\boldsymbol{\beta}) + 1]^2} \right\}.$$