Analysis of Repeated Events and Panel Count Data

E. Juarez-Colunga[1,2], and C. B. Dean[3]

[1]Department of Biostatistics and Informatics, University of Colorado Denver, Aurora, CO 80045, USA.

[3]Department of Statistical and Actuarial Sciences, Western University, Western Science Centre, London, ON N6A 5B7, Canada.

[2]Corresponding address: Elizabeth.Juarez-Colunga@ucdenver.edu

Abstract

Recurrent event data arise when there is an event that may repeat over time. This type of data is common in many areas of the sciences, social sciences, medicine and elsewhere. In this article we review counting process models, in particular, Poisson and mixed Poisson models for recurrent event data, as well as models for aggregated data, where only the number of events occurring in intervals of time, called *panel data*, are available. Estimation is conducted through likelihood and estimating equations. We illustrate the methods through an example of recurrence of superficial bladder cancer tumors in a clinical trial. We also briefly discuss optimal recurrent event designs.

Key Words : clinical trial; counting process; design; life-history data; Poisson regression.

1   Introduction

In the last few decades there has been great interest in monitoring longitudinal processes. One type of longitudinal process arises when there is an event that may repeat over time; data arising from such a process are called *recurrent event data*, and are common in many fields. For example, in business, interest may be in monitoring insurance claims, in engineering, in monitoring lines of faulty code, and in medical research, in studying the occurrence of adenomas in the colon, or the occurrence of superficial bladder cancer tumors.

An example considered in greater detail later typifies the sort of scenario through which recurrent data arise in clinical trials. This study, conducted by the Veterans Administrative Co-operative Urological Research Group, investigated the

effects of placebo pills, pyridoxine pills, and periodic instillation of thiotepa into the bladder on the frequency of recurrence of bladder cancer (Byar et al., 1977). The data appear in Andrews and Herzberg (2000). All 116 patients had bladder cancer when they entered the study; the tumors were removed and the patients were randomly assigned to one of the three treatments. Two covariates were considered which may reflect cancer severity at baseline: the number of tumors and the size, in centimetres, of the largest tumor. Figure 1 displays the recurrences for the three treatment groups.

One of the main classifications of recurrent event analyses is based on whether the focus is on analyzing the timing between events, or the counts of events. Both are based on counting processes. While the first scenario focuses on modeling the hazard between events, the second develops models for the intensity, or the mean number of events. When the system is sharply different each time a new event occurs, either because of the effect of the event itself, or the effect of the interventions employed after events occur, other models than the ones considered here are utilized. In these cases, time to first event, then time between first and second event, etc., may be considered in separate models. Cook and Lawless (2007) provides a comprehensive source on models and methods for recurrent event analyses.

Often, in order to monitor longitudinal processes, individuals are examined at specific followup times, and only the number of events that occurred between followup visits are recorded; these give rise to what is termed *panel data*. In this article we focus on inference for both recurrent event data and panel data based on counting processes. We discuss estimation of recurrent event models through likelihood and estimating equations. Estimating equations provide a robust estimation

approach. In Section 2 we introduce the counting process likelihood, as well as Poisson and mixed Poisson models, for both recurrent and panel count data. In Section 3 we discuss estimation of recurrent event models, while in Section 4 we illustrate these methods. We close with a discussion of other considerations in the analysis of recurrent event data in Section 5. In this section we outline a key result regarding optimal design of panel data studies, building on the robust approach to estimation developed in this paper using estimating equations.

2   Models Based on Counting Processes

2.1   Counting Process Likelihood

Models for recurrent events are typically formulated through their intensity functions, which are the probability distribution of the events occurring in a small interval of time $[t, t + \delta t)$, given the history of events up to time t (Cook and Lawless, 2007). Let $\{N(t), t \geq 0\}$ be the right-continuous counting process that records the number of events for an individual over the interval $[0, t]$; we define $N(0) = 0$ for simplicity. If $H(t) = \{N(s) : 0 \leq s < t\}$ represents the history if the process up to time t, then the intensity process function is defined as

$$\lambda(t|H(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{\Delta N(t) = 1|H(t)\}}{\Delta t}. \tag{1}$$

It is assumed that the probability of two or more events occurring over the interval $[t, t + \Delta t)$ is $o(\Delta t)$, so $\Pr(\Delta N(t) = 1|H(t)) = \lambda(t|H(t))\Delta t + o(\Delta t)$, and $\Pr(\Delta N(t) = 0|H(t)) = 1 - \lambda(t|H(t))\Delta t + o(\Delta t)$. The likelihood, obtained via product integration (Kalbfleisch and Prentice, 2002), based on n event times $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq \tau$ occurring during the time $[0, \tau]$ for fixed τ is

3

$$L = \left[\prod_{j=1}^{n} \lambda(t_j|H(t)) \exp\left(-\int_{t_{j-1}}^{t_j} \lambda(u|H(u))du \right) \right] \exp\left(-\int_{t_n}^{\tau} \lambda(u)du \right) \quad (2)$$

$$= \left[\prod_{j=1}^{n} \lambda(t_j|H(t)) \right] \exp\left(-\int_{0}^{\tau} \lambda(u|H(u))du \right). \quad (3)$$

Maximum likelihood methods can then be employed for inference by invoking usual asymptotic theory.

The definition of the intensity function in (1) can be extended to include covariate information up to and including time t through an extended definition of the history of the process, specifically $H(t) = \{N(s) : 0 \leq s < t; \boldsymbol{x}(s) : 0 \leq s \leq t\}$, where $\boldsymbol{x}(s)$ is the covariate vector at time s.

2.2 Poisson and Mixed Poisson Models

The specification of the model through the intensity function in (1) is very general, but in practice assumptions are often made regarding the history of the process based on scientific input and these may aid estimation. Poisson processes are often utilized because of their memoryless property; they simplify the history to only that at time t, i.e. the instantaneous probability of the occurrence of a new event in a small window of time depends on the history only through t. In this case, the intensity has the following form:

$$\lambda(t|H(t)) = h(t), \quad (4)$$

which implies that the number of events $N(s,t)$ in a time interval $(s,t]$ follows a Poisson distribution, i.e.

$$Pr\{N(t) - N(s) = n\} = \frac{\left\{ \int_{s}^{t} \lambda(u)du \right\}^{n} \exp\left\{ -\int_{s}^{t} \lambda(u)du \right\}}{n!}, \quad (5)$$

where the mean $E(N(s, t))$ is $\mu(s, t) = \int_s^t \lambda(u)du$. The expected cumulative number of events occurring from time 0 to time t is known as the cumulative intensity function or cumulative mean function, and has the form

$$\mu(t) = \int_0^t \lambda(u)du. \tag{6}$$

Another important property of the Poisson process is that the number of events occurring in disjoint intervals are independent random variables.

Since under a Poisson model the intensity function does not depend on the history of the process, we have $\lambda(t|H(t))dt = h(t)dt = E[dN(t)]$ (see Equation 4), interpreted as both a conditional and a marginal probability, or an expectation. The function $h(t)$ is sometimes referred to as the rate function.

Covariate information is in practice commonly included in a multiplicative form as

$$\lambda(t|H(t)) = \rho(t)g(\boldsymbol{x}(t); \boldsymbol{\beta}), \tag{7}$$

where $\rho(t)$ is known as the baseline intensity function, and it may modeled non-parametrically or parametrically. Here we will focus on parametric forms through a parameter $\boldsymbol{\alpha}$ of dimension d_α, i.e. $\rho(t) = \rho(t; \boldsymbol{\alpha})$; forms such as the exponential $(\exp(\alpha t))$ and Weibull $(\alpha t^{\alpha-1})$ are often used. The function $g(\boldsymbol{x}(t); \boldsymbol{\beta})$ is often specified as $g(\boldsymbol{x}(t); \boldsymbol{\beta}) = \exp(\boldsymbol{x}'(t)\boldsymbol{\beta})$ to ensure $\lambda(t|H(t))$ is positive, and in this case, the coefficients $\boldsymbol{\beta}$ may be interpreted as log-relative risks. For simplicity in presentation, we will consider time-independent covariates, i.e. $\lambda(t; \boldsymbol{x}) = \rho(t; \boldsymbol{\alpha})\exp(\boldsymbol{x}'\boldsymbol{\beta})$.

Consider M individuals, and assume each individual is observed up to time τ_i referred to as the *termination time*, $i = 1, \ldots, M$. Let the observation process for the $i-$th individual $\{Y_i(t), t \geq 0\}$ be 1 if individual i is under study at t

and 0 otherwise, and assume that $\{Y_i(t), t \geq 0\}$ is independent of the counting process $N_i(t)$. Thus the observed counting process may be defined as $\bar{N}_i(t) = \int_0^t Y_i(u)dN_i(u)$, $t \in [0, \tau_i]$ for individual i, $i = 1, \ldots, M$. The Poisson counting process has intensity $\lambda(t; \boldsymbol{x}_i) = \rho(t; \boldsymbol{\alpha}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, which implies that $\mathrm{E}(N_i(t)) = \mu_i(t) = R_i(t; \boldsymbol{\alpha}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, $i = 1, \ldots, M$, where $R_i(t; \boldsymbol{\alpha}) = \int_0^t Y_i(t)\rho(u; \boldsymbol{\alpha})du$. Let the total aggregated count of events for individual i be n_i, i.e. $N(\tau_i) = n_i$. For n_i events observed at times $\{t_{i1}, t_{i2}, \ldots, t_{in_i}\}$, the likelihood for individual i has the following form (Lawless, 1987b)

$$L_i = \prod_{j=1}^{n_i} \lambda(t_{ij}; \boldsymbol{x}_i) \exp\left\{-\int_0^{\tau_i} \lambda(u; \boldsymbol{x}_i)\right\} \tag{8}$$

$$= \left\{\prod_{j=1}^{n_i} \frac{\rho(t_{ij}; \boldsymbol{\alpha})}{R_i(\tau_i; \boldsymbol{\alpha})}\right\} \times \{R_i(\tau_i; \boldsymbol{\alpha}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}^{n_i} \exp\left\{-R_i(\tau_i; \boldsymbol{\alpha}) \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right\}. \tag{9}$$

The first term in the likelihood (9) corresponds to the conditional distribution of the event times given the number of events, while the second is the likelihood kernel for the distribution of total number of events n_i. The Poisson model has the constraint that $\mathrm{Var}(N_i(\tau_i)) = \mathrm{E}(N_i(\tau_i)) = \mu_i(\tau_i)$, which is too limiting in many cases. Often, the variability observed exceeds what can be explained through covariates available. In these cases, is common to use a mixed Poisson model, in which the rate function for subject i is $\nu_i\lambda(t; \boldsymbol{x}_i)$ where ν_i's are independent positive random variables that follow a distribution $G(\nu; \phi)$ such that $\mathrm{Var}(\nu_i) = \phi$. The function $\lambda(t; \boldsymbol{x})$ is now interpreted as a population average rate function among subjects with covariate vector $\boldsymbol{x}$, since $\mathrm{E}(dN_i(t)|\boldsymbol{x}) = \lambda(t; \boldsymbol{x})dt$. As well, the random effect ν_i represents the effect of covariates which are unaccounted for in the regression model. Note that ν_i may also be a cluster effect, taking the same value for all individuals within the same cluster. This can be used to account

for unknown hospital effects, for example, where patients are clustered within hospitals. When ν_i follows a Gamma distribution, the marginal distribution of n_i is negative binomial. In this case, the variance has the form $\mathrm{Var}(N_i(\tau_i) = \mu_i(\tau_i) + \mu_i^2(\tau_i)\phi$ accommodating extra-Poisson variation in the second term. In addition, if $s_1 < t_2$, $\mathrm{Cov}(N_i(t_1, s_1), N_i(t_2, s_2)) = \phi\mu_i(t_1, s_1)\mu_i(t_2, s_2)$. The parameter ϕ reflects both the degree of overdispersion and the degree of association between disjoint interval counts.

2.3   Likelihood for Panel Count Data

Panel data arises when counts of events are recorded at specified followup times. Such data are quite common as continuous followup is expensive; when it is not ethically required, or seen to be too invasive given the clinical context, panel studies are employed. For example, studies investigating recurrence of superficial bladder cancer tumors commonly collect information in a panel followup framework; as an illustration, Vancouver Coastal Health in Canada is currently conducting a clinical trial on bladder cancer with followup every 3 months for the first two years, every 6 months for the next two years, and yearly thereafter (Vancouver Coastal Health, 2012).

Let the $e_i + 1$ individual-specific *panel followup times* be denoted by $T_{i,0} = 0 < T_{i,1} < T_{i,2} < \ldots < T_{i,e_i} = \tau_i$. Panel counts for individual i are denoted as $n_{ip} = \bar{N}_i(T_{i,p}) - \bar{N}_i(T_{i,p-1})$, $p = 1, 2 \ldots, e_i$, and the total aggregated count of events for individual i is $n_{i+} = \sum_{p=1}^{e_i} n_{ip}$.

The counting process $N_i(t)$ is again modeled here as a mixed Poisson process with intensity function $\lambda_i(t) = \nu_i \rho(t; \boldsymbol{\alpha}) \exp(\boldsymbol{x}_i' \boldsymbol{\beta})$, given ν_i, an individual-specific random effect accounting for overdispersion. We set $\mathrm{E}(\nu_i) = 1$ without loss of

generality, and let $\text{var}(\nu_i) = \phi$. A particular scenario that may be of interest, as in the bladder cancer study, is where it is required to test treatment efficacy in reducing the number of recurrences. Let $\boldsymbol{x}_i$ be a $k \times 1$ treatment indicator vector for the i-th individual, such that $x_{i1} = 1$ represents an intercept term, and $x_{ij} = 1$ if individual i received treatment j, or 0 otherwise, $j = 2, \ldots, k$. Thus the corresponding β's are parametrized such that the treatment effects are measured relative to treatment 1; β_1 reflects the overall mean, and $\boldsymbol{\alpha}$ describes the shape of the baseline intensity function $\rho(t, \boldsymbol{\alpha})$.

Writing the cumulative baseline intensity function for the entire followup time for individual i as $R_i = \int_0^{\tau_i} Y_i(t)\rho(t; \boldsymbol{\alpha})dt$, then $\mu_{i+} = \mu_i(\tau_i) = R_i \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$. Similarly, defining the cumulative baseline intensity function in panel period p as $R_{ip} = \int_{T_{i,p-1}}^{T_{i,p}} Y_i(t)\rho(t; \boldsymbol{\alpha})dt$, we have $\mu_{ip} = \mu_i(T_{i,p-1}, T_{i,p}) = \text{E}(n_{ip}) = R_{ip} \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$.

The likelihood based on panel count data may be written as the product of two terms 1) a conditional distribution of the number of events in each panel n_{ip} given the total number of events n_{i+}, and 2) the distribution of the total number of events n_{i+}. Specifically, the likelihood has the following form:

$$
L_p(\boldsymbol{\theta}) = \prod_{i=1}^{M} \left[\binom{n_{i+}}{n_{i1}, \ldots, n_{ie_i}} \prod_{p=1}^{e_i} \left(\frac{R_{ip}}{R_i} \right)^{n_{ip}} \right] \times \prod_{i=1}^{M} \int_0^{\infty} \frac{(\nu_i \mu_{i+})^{n_{i+}} e^{-\nu_i \mu_{i+}}}{n_{i+}!} G(\nu_i)d\nu_i.
$$

(10)

Note that if there is a single panel, $L_p(\boldsymbol{\theta})$ (10) will reduce to the simple mixed Poisson kernel.

3 Use of Estimating Equations for Inference in Recurrent Event Models

Inference for Poisson and mixed Poisson models may be based on likelihood methods, and alternatively, on more robust approaches only assuming the form of the mean. These alternative approaches estimate the variance by using either a robust model-based variance or a robust empirical variance. In this section we provide details on how estimation can be carried out using estimating equations; we build the estimating equations for regression parameters from the negative binomial distribution for the counts of events.

Let the intensity function of the counting process of the events, given the subject-specific random effect ν_i, be

$$\lambda_i(t) = \nu_i \rho(t; \boldsymbol{\alpha}) \exp\left\{ \boldsymbol{x}_i' \boldsymbol{\beta} \right\}. \tag{11}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}', \phi)'$, and let ω_{ipl} be the time of the l-th event, from the start of the study, for the i-th individual in panel period p, $i = 1, \ldots, M$, $p = 1, \ldots, e_i$, $l = 1, \ldots, n_{ip}$. The likelihood kernel based on either the full data (subscripted by $d = f$) or the panel data (subscripted by $d = p$) factorizes as:

$$L_d(\boldsymbol{\theta}) = L_{\alpha,d}(\boldsymbol{\alpha}) \times \prod_{i=1}^{M} \int_0^\infty \frac{(\nu_i \mu_{i+})^{n_{i+}} e^{-\nu_i \mu_{i+}}}{n_{i+}!} G(\nu_i) d\nu_i, \quad d \in \{f, p\} \tag{12}$$

where

$$L_{\alpha,f}(\boldsymbol{\alpha}) = \prod_{i=1}^{M} \prod_{p=1}^{e_i} \prod_{l=1}^{n_{ip}} \frac{\rho(\omega_{ipl}; \boldsymbol{\alpha})}{R_i}, \tag{13}$$

and

$$L_{\alpha,p}(\boldsymbol{\alpha}) = \prod_{i=1}^{M} \left[\binom{n_{i+}}{n_{i1}, \ldots, n_{ie_i}} \prod_{p=1}^{e_i} \left(\frac{R_{ip}}{R_i}\right)^{n_{ip}} \right]. \tag{14}$$

Lawless (1987a) provides the first and second derivatives of the negative binomial likelihood with respect the parameters, which provides the basis for likelihood estimation along with (12).

Let $\boldsymbol{g}_d = (\boldsymbol{g}'_\beta, \boldsymbol{g}'_{\alpha,d}, g_\phi) = \boldsymbol{0}'$ denote the full set of estimating equations for the panel $(d = p)$ or the full $(d = f)$ data. Since the likelihood is a function of $\boldsymbol{\beta}$ only through the second term in (12), the estimating equations for this parameter are developed here as the usual quasi-likelihood equations $(\partial \boldsymbol{\mu}/\partial \boldsymbol{\beta})' U_o^{-1}(\mathbf{n} - \boldsymbol{\mu}) = \boldsymbol{0}$, where $U_o = \mathrm{diag}\{\mu_{i+}(1 + \phi\mu_{i+}), i = 1, \ldots, M\}$, $\mathbf{n} = (n_{1+}, \ldots, n_{M+})'$ is a vector of counts, and $\boldsymbol{\mu} = (\mu_{1+}, \ldots, \mu_{M+})'$ is the vector of their expected values. Defining $U = \mathrm{diag}\{\mu_{i+}, i = 1, \ldots, M\}$, this becomes

$$\boldsymbol{g}_\beta = X'UU_o^{-1}(\mathbf{n} - \boldsymbol{\mu}) = \boldsymbol{0}. \tag{15}$$

We obtain an estimating equation for $\boldsymbol{\alpha}$ by combining $\partial \log L_{\alpha,d}/\partial \boldsymbol{\alpha}$, $d = f, p$, with quasi-likelihood estimation as both first and second terms in (12) depend on $\boldsymbol{\alpha}$, yielding

$$\boldsymbol{g}_{\alpha,d} = \frac{\partial \log L_{\alpha,d}(\boldsymbol{\alpha})}{\partial \alpha} + W'UU_o^{-1}(\mathbf{n} - \boldsymbol{\mu}) = \boldsymbol{0}, \tag{16}$$

where W is a matrix with entries

$$w_{ia} = \frac{\partial \log R_i}{\partial \alpha_a}, \qquad i = 1, \ldots, M, \text{ and } a = 1, \ldots, d_\alpha. \tag{17}$$

Several choices may be considered for the estimating equation of the overdispersion parameter ϕ. In our examples, we use the pseudo-likelihood estimator, which has been popular since its introduction by Davidian and Carroll (1987). It has performed well in simulation studies for simple overdispersed count analyses and has documented optimality properties (Nelder and Lee, 1992). The pseudo-likelihood

estimating equation for ϕ is

$$g_\phi = \sum_{i=1}^{M} \frac{(n_{i+} - \mu_{i+})^2 - (1 - h_i)\mu_{i+}(1 + \phi\mu_{i+})}{(1 + \phi\mu_{i+})^2} = 0, \qquad (18)$$

where $h_i = \text{diag}(U^{1/2}V'(V'UV)^{-1}V'U^{1/2})$, $V = (\ X\ \ W\ )$; h_i is the diagonal of the hat matrix and represents a correction to reduce small sample bias in this simple second moment equation.

Let the estimator of $\boldsymbol{\theta}$ from either full $(d = f)$ or panel $(d = p)$ data be denoted by $\hat{\boldsymbol{\theta}}_d$. Under standard conditions for the application of asymptotic results to estimating equations, $\sqrt{M}(\hat{\boldsymbol{\theta}}_d - \boldsymbol{\theta})$ is asymptotically normal with asymptotic covariance

$$E\left(-\lim_{M\to\infty}\frac{\partial \boldsymbol{g}_d}{\partial \boldsymbol{\theta}}\right)^{-1} E\{\lim_{M\to\infty}\boldsymbol{g}_d\boldsymbol{g}_d'\}\left\{E\left(-\lim_{M\to\infty}\frac{\partial \boldsymbol{g}_d}{\partial \boldsymbol{\theta}}\right)^{-1}\right\}'. \qquad (19)$$

Finite sample variance estimates are obtained by substituting $\hat{\boldsymbol{\theta}}_d$ for $\boldsymbol{\theta}$ and omitting the expressions $\lim_{M\to\infty}$. In this case there are two options for approximating the expectation of the terms in (19). The first is a model-based approach, which in this case requires specification of 3rd and 4th moments of the counts. The second is an empirical approach, which substitutes $E\{\sum_{i=1}^{M} g_{id}g_{id}'\}$ by $\{\sum_{i=1}^{M} g_{id}g_{id}'\}$; where g_{id} denotes the contribution to the score equation from individual i.

4   Analysis of Bladder Cancer Data

In this section we estimate the treatment effects for the bladder cancer study (Andrews and Herzberg, 2000) discussed in Section 1 under a design with continuous followup as well as a panel design, for illustrative purposes, with 2 equally spaced scheduled followup visits over 64 months; for the panel design, we record informa-

tion on event recurrences at the scheduled followup times 32 and 64 months, and at termination times.

Figure 1 displays the recurrences of events. Note that the rate of occurrence of events seems to be slightly lower in the thiotepa group. Figure 1 shows that the three treatment groups are very similar in terms of the distributions of the termination times, the number of tumors, and the size of the largest tumor at baseline.

Table 1 reports parameter estimates for a model with Weibull baseline, and their standard errors: likelihood based, robust model-based and robust empirical, under a 2-panel design, as well as from an analysis of the full data with continuous followup. Since the estimated coefficient corresponding to the size of the largest tumor at baseline is not significant, this variable has been excluded in the subsequent discussion. The estimates under the negative binomial and quasi-likelihood are quite close, except for the overdispersion parameter, which is slightly larger based on the negative binomial analysis. The standard errors are also similar, with likelihood values larger than the robust model-based estimates. The robust empirical estimates are data-driven (not model-based) and are quite close to those from the negative binomial analysis. The difference is again more pronounced for estimates of the overdispersion parameter. The thiotepa treatment may have a significant protective effect on recurrences, while the pyridoxine treatment effect is non-significant. There is substantial overdispersion in the data, and the estimate of the Weibull shape parameter α is quite close to unity.

Importantly, note that the estimates of the treatment effects (relative to placebo) from the panel design are quite close to those from the full data analysis, indicating cost efficiency would be gained from well-designed panel designs with relatively

little loss of information regarding treatment effects. There are larger differences observed in panel versus full data analyses for estimates of other parameters. For example, the estimate of the standard error of $\hat{\alpha}$ is considerably higher in the 2-panel design than that of the corresponding estimate from the full data analysis.

5 Other Approaches and Topics

Dependent termination time of the recurrent event process is one of the topics that has drawn more attention in the last few years. This topic requires appropriate modeling when the end of the study, for example death, is related to the occurrence of the events. A popular approach to model this dependency is through shared individual frailties, where the recurrent event process and the survival process are linked (Liu et al., 2004).

This approach of shared individual frailties linking recurrent event and survival outcomes has also been used to model dependency when considering multivariate recurrent event processes (He et al., 2008; Zhao et al., 2012). This may be useful when modeling recurrent event processes that may evolve differently over time, but are likely to be intrinsically associated.

In summary, in this article we have presented briefly some counting based models, and have discussed approaches where we have both full data available, through the timing of events, and where only panel counts are available; we also discussed inference based on likelihood and estimating equations in both situations. Juarez-Colunga, Balshaw, and Dean (Juarez-Colunga et al.) built from this inference framework an investigation of the loss of efficiency of a treatment estimator in panel designs as compared to designs using continuous followup. A main result

of this investigation states that little is lost when the distributions of both the termination times and the covariates adjusted for in the analysis are close across treatment groups. A simple example of similarity in such distributions would be if the termination times over the covariates are identical across the treatment groups; for example, if the covariate is sex, and there are two treatments (A and B), this would mean that the same number of females and males are allocated to the treatment groups, and that the termination times for each gender under treatment A are the same as for those under treatment B. Juarez-Colunga, Balshaw, and Dean (Juarez-Colunga et al.) provides details on conditions for optimal panel designs, broadly as well as several illustrative examples.

Acknowledgments

References

Andrews, D. F. and Herzberg, A. M. (2000). *Data: A collection of problems from many fields for the student and research worker*. Springer-Verlag, New York.

Byar, D., Blackard, C., and the Veterans Administration Co-operative Urological Research Group (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* **10,** 556–561.

Cook, R. and Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer, New York.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* **82,** 1079–1091.

He, X., Tong, X., and Sun, J. (2008). Regression analysis of multivariate panel count data. *Biostatistics (Oxford, England)* **9,** 234–248.

Juarez-Colunga, E., Balshaw, R., and Dean, C. B. Efficient panel designs for longitudinal recurrent event studies recording panel counts. *Biostatistics* To appear.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The statistical analysis of failure time data*. John Wiley & Sons, 2nd edition.

Lawless, J. F. (1987a). Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* **15,** 209–225.

Lawless, J. F. (1987b). Regression methods for poisson process data. *Journal of the American Statistical Association* **82,** 808–815.

Liu, L., Wolfe, R., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal events. *Biometrics* **60,** 747–756.

Nelder, J. A. and Lee, Y. (1992). Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **54,** 273–284.

Vancouver Coastal Health (2012). A study for treatment of superficial bladder cancer using ogx-427. Clinical Trials NIH Registry.

Zhao, X. X., Liu, L. L., Liu, Y. Y., and Xu, W. W. (2012). Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. *Biometrical Journal* **54,** 585–599.

Figure 1: Bladder cancer data: the lines in the plot display the observation times ending at τ_i, termination times, for individuals, while the '+' indicates event times. For the display of the covariates (size and tumors) and individual-specific rates, the length of the lines are proportional to the size of the largest tumor, the number of tumors, and the rate of events, respectively; additionally, lighter colors are used to highlight large values of these variables.
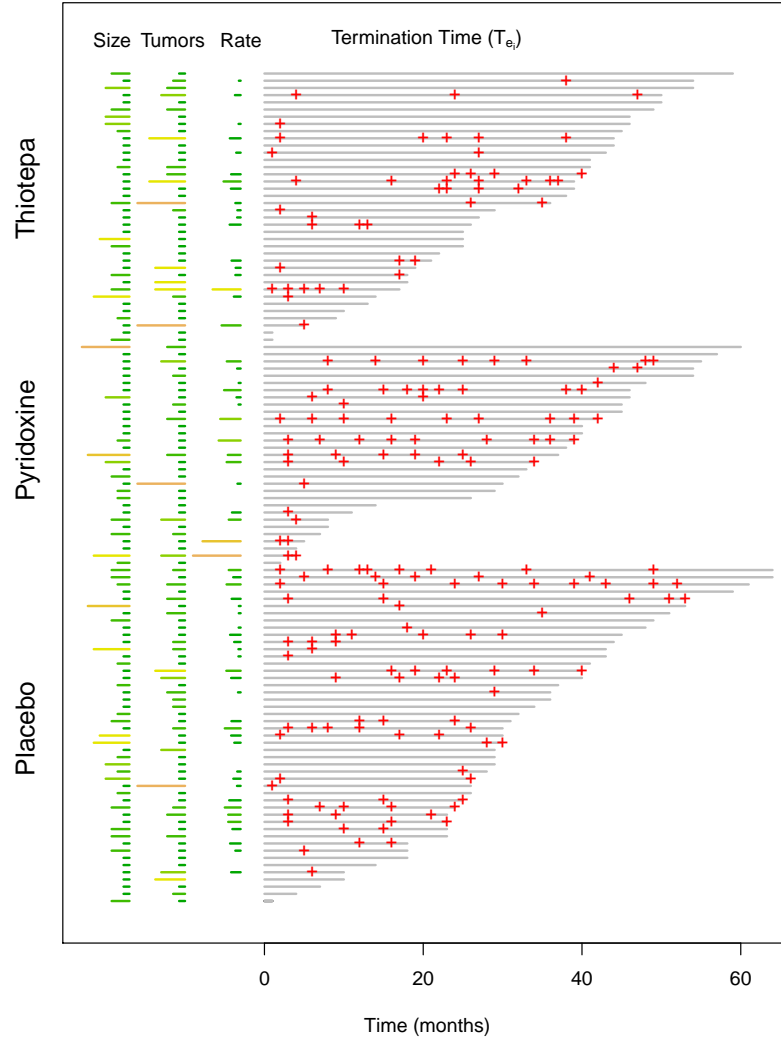
Table 1: Parameter estimates and their standard errors, resulting from the likelihood (NB) and quasi-likelihood (QL) analyses, using both a robust model-based (MB) and a robust empirical (EMP) approach, fit to the bladder cancer data. The regression parameters $\beta_1, \beta_2, \beta_3$ correspond to the three treatment groups, parametrized with respect to the placebo, and β_4 to the covariate, number of tumors at baseline.

	Full Data					2-Panel Data				
	Estimates		Standard Errors			Estimates		Standard Errors		
	NB	QL	NB	MB	EMP	NB	QL	NB	MB	EMP
β_1	-3.449	-3.428	0.348	0.275	0.333	-2.976	-2.947	0.503	0.477	0.486
β_2	0.133	0.118	0.316	0.302	0.287	0.118	0.108	0.313	0.301	0.286
β_3	-0.541	-0.540	0.317	0.264	0.291	-0.553	-0.551	0.315	0.263	0.290
β_4	0.245	0.240	0.072	0.057	0.065	0.240	0.236	0.072	0.056	0.065
α	1.019	1.015	0.069	0.055	0.068	0.886	0.880	0.124	0.117	0.121
ϕ	1.144	0.848	0.285	0.193	0.275	1.122	0.846	0.282	0.193	0.275