Biometrics 61, 692–701
September 2005
DOI: 10.1111/j.1541-0420.2005.00357.x

Robust Tests for Treatment Effects Based on Censored Recurrent Event Data Observed over Multiple Periods

Richard J. Cook,[1,*] Wei Wei,[2] and Grace Y. Yi[1]

[1]Department of Statistics and Actuarial Science, University of Waterloo,
Waterloo, Ontario N2L 3G1, Canada
[2]Department of Biostatistics, School of Public Health, University of Michigan,
Ann Arbor, Michigan 48109-2029, U.S.A.
[*]*email:* rjcook@uwaterloo.ca

Summary. We derive semiparametric methods for estimating and testing treatment effects when censored recurrent event data are available over multiple periods. These methods are based on estimating functions motivated by a working "mixed-Poisson" assumption under which conditioning can eliminate subject-specific random effects. Robust pseudoscore test statistics are obtained via "sandwich" variance estimation. The relative efficiency of conditional versus marginal analyses is assessed analytically under a mixed time-homogeneous Poisson model. The robustness and empirical power of the semiparametric approach are assessed through simulation. Adaptations to handle recurrent events arising in crossover trials are described and these methods are applied to data from a two-period crossover trial of patients with bronchial asthma.

Key words: Conditional inference; Crossover design; Marginal models; Mixed-Poisson process; Rate function; Recurrent event.

1. Introduction

Medical studies are frequently designed based on clinically important events that may recur repeatedly over the course of follow-up. Examples include seizures in neurological studies (Thall and Vail, 1995), respiratory attacks in asthma trials (Sears et al., 1990), and exacerbations of symptoms in studies of cystic fibrosis (Fuchs et al., 1994). In such settings there are a variety of strategies available for the analyses of recurrent event data including intensity-based methods (Andersen et al., 1993), methods based on random effect models (Lawless, 1987), and marginal methods (Andersen and Gill, 1982; Wei, Lin, and Weissfield, 1989; Lawless and Nadeau, 1995). Marginal methods are perhaps the most widely adopted in clinical trials because they typically involve the fewest distributional assumptions and, therefore, provide a natural basis for making treatment comparisons. Even within the class of marginal methods, however, there are a variety of approaches one can adopt. Wei et al. (1989) describe methods for marginal analysis of multivariate failure time data which have been applied in settings with recurrent events (Hughes, 1997; Li and Lagakos, 1997). Marginal methods have also been developed based on rate functions (e.g., Andersen and Gill, 1982; Lawless and Nadeau, 1995), which can be generalized to incorporate stratification based on the cumulative number of events (Prentice, Williams, and Petersen, 1981; Pepe and Cai, 1993). Therneau and Grambsch (2000) provide an excellent discussion and illustration of the various marginal approaches. Cook and Lawless (2002) give a review of recent developments for the analysis of recurrent events.

Clinical trials often involve observation of subjects over two or more periods of interest. Following accrual, for example, patients may undergo a baseline period of observation during which events are recorded but no treatment is administered. Following this baseline period, a patient may be randomized to receive either an experimental or control treatment and then followed for the occurrence of events of interest (e.g., ACIP, 1992). In other settings baseline data representing the number of events occurring over a predefined period of interest are recorded retrospectively upon study entry (e.g., Fuchs et al., 1994). Crossover trials represent another setting where events are observed over multiple periods. Examples include a recent 1-year study of the prophylactic use of valcyclovir for outbreaks of symptoms in patients with Herpes simplex virus (Romanowski, Marina, and Roberts, 2003) and a study reported in Sears et al. (1990) on the efficacy of fenoterol versus placebo for the reduction of symptoms in patients with bronchial asthma. In long-term crossover trials such as these, complications can arise if patients do not switch from the treatment assigned in period 1 to the treatment assigned in period 2 at the scheduled time, or even if they simply withdraw from the study prematurely.

The purpose of this article is to develop robust methods suitable for use in clinical trials in which patients are observed under two or more treatment periods. We initially focus on settings in which a baseline period of observation is followed by a period in which randomized treatments are administered. The methods are developed to address problems arising from type I right censoring due to variable durations of observation

(Lawless, 2003). Extensions to deal with more complicated situations arising in crossover trials are then described.

The remainder of the article is organized as follows. In Section 2 we define notation, describe frameworks for inference, and examine asymptotic relative efficiency of different methods of analysis for mixed-Poisson processes. In Section 3 we review marginal methods based on rate functions and derive a new robust pseudoscore statistic based on a working assumption of a mixed-Poisson formulation. The frequency properties of the proposed test are studied via simulation in Section 4 and contrasted with those of some standard marginal analyses. Extensions for crossover trials are developed in Section 5 and applied to data from a recent asthma trial. Concluding remarks are made in Section 6.

2. Relative Efficiency

Here we investigate the relative efficiency of three methods of analysis for data from mixed-Poisson processes observed over a baseline period and a treatment period. We consider analyses based on relatively simple parametric models to provide some motivation for the developments that follow.

Consider a trial in which subjects undergo a common baseline period of observation denoted $(-\tau_R, 0]$, during which they receive standard care. Let R_i denote the number of events experienced by subject i over the baseline period, $i = 1, 2, \ldots, m$. Following this baseline period is a common treatment period denoted $(0, \tau]$, during which subjects receive either the experimental treatment or standard care as assigned by a balanced randomization scheme. Let $x_i = 1$ if subject i is randomized to receive the experimental treatment and $x_i = 0$ otherwise, and let N_i denote the number of events subject i experiences over $(0, \tau]$.

We assume $R_i \,|\, v_i \sim \text{Poisson}(v_i\rho)$ and that $N_i \,|\, v_i \sim \text{Poisson}(v_i\lambda\exp(\beta x_i))$, where R_i and N_i are independent given v_i, $i = 1, \ldots, m$. The term v_i is often thought of as a latent subject-specific effect which is introduced to characterize extra-Poisson variation. Typically v_i, $i = 1, \ldots, m$, are assumed to be independent random variables arising from a distribution $G(v_i; \phi)$ where $E(v_i) = 1$ and $\text{var}(v_i) = \phi$, $\phi > 0$, $i = 1, \ldots, m$. The gamma distribution is perhaps the most common choice because it is conjugate to the Poisson model. Here it leads to a negative trinomial joint probability mass function for (R_i, N_i) and a marginal negative binomial probability mass function for N_i (Lawless, 1987). Specifically,

$$P(r_i, n_i; \rho, \lambda, \beta, \phi) = \frac{\Gamma\left(\phi^{-1} + r_i + n_i\right)}{\Gamma(\phi^{-1}) r_i! n_i!} \times \frac{\rho^{r_i}(\lambda\exp(\beta x_i))^{n_i}}{(1 + \phi(\rho + \lambda\exp(\beta x_i)))^{\phi^{-1} + r_i + n_i}}, \quad (1)$$

where r_i, $n_i = 0, 1, 2, \ldots$, and

$$P(n_i; \lambda, \beta, \phi) = \frac{\Gamma(\phi^{-1} + n_i)}{\Gamma(\phi^{-1}) n_i!} \frac{(\phi\lambda\exp(\beta x_i))^{n_i}}{(1 + \phi\lambda\exp(\beta x_i))^{\phi^{-1} + n_i}}, \quad (2)$$

where $n_i = 0, 1, \ldots$. Cook and Wei (2003) explore efficiency gains realized by use of (1) instead of (2) for estimation of the regression coefficient.

A third valid probability mass function may be derived by recognizing that if we write

$$P(r_i, n_i \,|\, v_i; \rho, \lambda, \beta) = \frac{(v_i\rho)^{r_i}\exp(-v_i\rho)}{r_i!} \times \frac{(v_i\lambda\exp(\beta x_i))^{n_i}\exp(-v_i\lambda\exp(\beta x_i))}{n_i!},$$

$r_i + n_i$ is a minimal sufficient statistic for v_i and that one may therefore eliminate $v_1, \ldots, v_m$ by conditioning on $r_i + n_i$, $i = 1, \ldots, m$, respectively. Straightforward calculations give

$$P(r_i, n_i \,|\, r_i + n_i; \alpha, \beta) = \binom{r_i + n_i}{n_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)}\right)^{r_i} \times \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}\right)^{n_i}, \quad (3)$$

where $\alpha = \log(\lambda/\rho)$ and $n_i = 0, 1, \ldots, q_i$, with $q_i = r_i + n_i$, $i = 1, \ldots, m$. This conditioning argument suggests a logistic regression analysis based on m binomial samples where the ith sample has $r_i + n_i$ trials and n_i successes with the probability of success given by $\exp(\alpha + \beta x_i)/(1 + \exp(\alpha + \beta x_i))$ (Diggle et al., 2002; Jones and Kenward, 2003). Elimination of v_i means that the necessary distributional assumptions are somewhat weaker under this model.

Note that (1) is indexed by ρ, λ, β, and ϕ, (2) by λ, β, and ϕ, and (3) by α and β. We refer to the models based on (1), (2), and (3) as joint, marginal, and conditional, respectively, and we let $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ denote the consistent estimators obtained by maximizing the respective likelihood functions. The expected information matrices arising from these likelihoods lead to the following expressions for the limiting variances based on the joint, marginal, and conditional models, respectively:

$$\text{asvar}(\hat{\beta}_1) = \frac{2}{m} \cdot \left\{\frac{1}{\lambda} + \frac{1}{\lambda\exp(\beta)} + \frac{2\phi}{1 + \phi\rho}\right\}, \quad (4)$$

$$\text{asvar}(\hat{\beta}_2) = \frac{2}{m} \cdot \left\{\frac{1}{\lambda} + \frac{1}{\lambda\exp(\beta)} + 2\phi\right\}, \quad (5)$$

$$\text{asvar}(\hat{\beta}_3) = \frac{2}{m} \cdot \left\{\frac{1}{\lambda} + \frac{1}{\lambda\exp(\beta)} + \frac{2}{\rho}\right\}. \quad (6)$$

One can see that asymptotically the joint model is uniformly more efficient than the marginal and conditional models, and so this approach has considerable appeal on the basis of efficiency. Our ultimate goal, however, is to consider robust methods and it is challenging to consider robust analogues to (1) because joint distributions typically require a fuller model specification. It is, therefore, worthwhile to consider the relative efficiency of inferences based on the conditional and marginal models. To this end, we define

$$RE_{3:2} = \frac{\text{asvar}(\hat{\beta}_3)}{\text{asvar}(\hat{\beta}_2)} = \frac{\rho(1 + \exp(\beta)) + 2\lambda\exp(\beta)}{\rho(1 + \exp(\beta)) + 2\phi\rho\lambda\exp(\beta)} \quad (7)$$

as the relative efficiency function for the conditional versus the marginal approaches. Inspection of (7) reveals that whenever $\phi\rho > 1$ the conditional analysis leads to more efficient estimation of β than the marginal analysis. Further insight can be gained by considering the $100f\%$ relative efficiency contours defined as points in the parameter space where $RE_{3:2} = f \geq 0$.

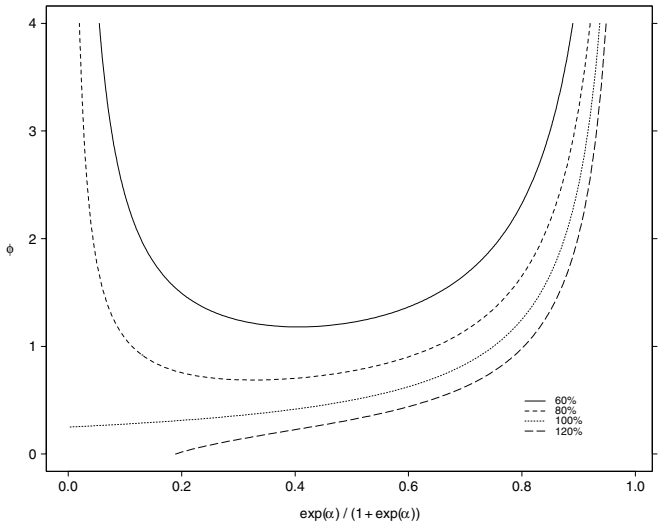

Figure 1. Contour plots of the relative efficiency of conditional versus marginal analyses under a mixed-Poisson model ($\rho + \lambda = 4$, $\exp(\beta) = 0.75$).

Suppose we set $\exp(\beta) = 0.75$ to represent a moderate treatment effect, and let $\rho + \lambda = 4$ to represent a moderate total number of events among control patients over the baseline and follow-up observation periods. Figure 1 displays a plot of the 60%, 80%, 100%, and 120% relative efficiency contours for β as a function of $\exp(\alpha)/(1 + \exp(\alpha)) = \lambda/(\lambda + \rho)$ and ϕ, for this scenario. The points for which the asymptotic variance under the conditional model is 20% lower than the asymptotic variance under the marginal model are denoted by the 80% relative efficiency contour. The asymptotic variance under the conditional model is 20% greater than the marginal model for points on the 120% contour. From Figure 1 it is clear that there is a large region in the parameter space in which the conditional analysis is more efficient than the marginal analysis and this region represents scenarios that one might reasonably expect to encounter in many biomedical settings. As the baseline mean becomes small (i.e., $\rho \to 0$) the marginal analysis leads to more efficient estimates than the conditional analysis, even when ϕ is large. Moreover, with very small ϕ, the marginal analysis is generally preferred. When the mean number of events in the follow-up period is comparable or smaller than the mean number of events in the baseline period (i.e., $\lambda/(\lambda + \rho) < 0.50$), however, even when the extent of extra-Poisson variation is relatively modest (i.e., $\phi < 1.0$) there can be as much as a 20% lower asymptotic variance under the conditional analysis. As one might expect, for any given α the gains from the conditional analysis become more substantial as ϕ increases.

3. Semiparametric Methods for Censored Data

3.1 *Robust Marginal Models*

Consider the setting in which patients are observed over a common baseline period $(-\tau_R, 0]$, are assigned to a treatment group, and are then followed for a treatment period of intended duration τ. The actual duration of observation during this treatment period may be less than this and vary from subject to subject due to serial patient accrual, administrative censoring, or early withdrawal, typical features of studies involving long-term follow-up. The aim in these settings is often to make robust inferences about treatment effects, and a common strategy is to employ robust marginal methods based on semiparametric analogues to (2) formed by multiplicative rate function models (Andersen and Gill, 1982; Lawless and Nadeau, 1995).

Let τ_i denote the time of last contact for subject i where $\tau_i \leq \tau$, and let $Y_i(t) = I(t \leq \tau_i)$ indicate that subject i is under observation at time t, $0 \leq t \leq \tau$. Let $N_i(t)$ be a right-continuous counting process that records the number of events experienced by subject i over the interval $(0, t]$ such that $dN_i(t) = 1$ if subject i experiences an event at time t and $dN_i(t) = 0$ otherwise. Let $\Lambda_i(t) = E(N_i(t) \,|\, x_i)$ and $d\Lambda_i(t) = E(dN_i(t) \,|\, x_i)$ denote the mean and rate functions for subject i, where $d\Lambda_i(t) = \Lambda_i(t) - \Lambda_i(t^-)$, $i = 1, \ldots, m$. Multiplicative marginal models for treatment effects based on rate functions often take the form

$$d\Lambda_i(t) = d\Lambda(t) \exp(\beta x_i), \quad t > 0,$$

where $d\Lambda(t)$ is a baseline rate function of an unspecified form, and β reflects the effect of treatment on the event rate. Andersen and Gill (1982) and Lawless and Nadeau (1995) proposed the use of estimating functions derived under the working assumption that the events are generated from a Poisson process. If $s_1, \ldots, s_k$ denote the k unique event times over all subjects in the sample we obtain estimating equations of the form

$$U_1(\beta, \Lambda(\cdot)) = \sum_{i=1}^{m} U_{1i}(\beta, \Lambda(\cdot)), \tag{8}$$

$$U_{2j}(\beta, \Lambda(\cdot)) = \sum_{i=1}^{m} U_{2ji}(\beta, \Lambda(\cdot)), \tag{9}$$

where $U_{1i}(\beta, \Lambda(\cdot)) = \int_0^\tau Y_i(s)\{dN_i(s) - d\Lambda(s)\exp(\beta x_i)\}x_i$ and $U_{2ji}(\beta, \Lambda(\cdot)) = Y_i(s_j)\{dN_i(s_j) - d\Lambda(s_j)\exp(\beta x_i)\}$, $j = 1, 2, \ldots, k$, $i = 1, \ldots, m$. These estimating functions may be shown to be unbiased under mild regularity conditions regarding the censoring distribution. As a result they provide consistent estimators for β and $d\Lambda(t)$, $0 < t < \tau$.

Setting (9) equal to zero and solving for $d\Lambda(s)$ gives the Breslow estimate

$$d\hat{\Lambda}^{\beta}(s) = \frac{\sum_{i=1}^{m} Y_i(s) dN_i(s)}{\sum_{i=1}^{m} Y_i(s)\exp(\beta x_i)}, \tag{10}$$

which may be inserted into (8) to obtain

$$U_1(\beta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s)\big\{dN_i(s) - d\hat{\Lambda}^{\beta}(s)\exp(\beta x_i)\big\}x_i. \tag{11}$$

This may in turn be rewritten as

$$U_1(\beta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s) W_i(s;\beta)\, dN_i(s),$$

where

$$W_i(s;\beta) = \left\{ x_i - \frac{\sum_{i=1}^{m} Y_i(s)\exp(\beta x_i)x_i}{\sum_{i=1}^{m} Y_i(s)\exp(\beta x_i)} \right\}.$$

A robust estimate of the variance of (11) may be derived by noting that

$$\operatorname{var}\left(\sqrt{m}^{-1}U_1(\beta)\right) = m^{-1}\sum_{i=1}^{m}\int_0^\tau\int_0^\tau Y_i(u)Y_i(v)W_i(u;\beta)W_i(v;\beta) \times \operatorname{cov}(dN_i(u), dN_i(v)),$$

and that this can be estimated consistently (Lawless and Nadeau, 1995) by

$$\widehat{\operatorname{var}}\left(\sqrt{m}^{-1}U_1(\beta)\right) = m^{-1}\sum_{i=1}^{m}\left\{ \int_0^\tau Y_i(u)W_i(u;\beta) \times \{dN_i(u) - d\hat{\Lambda}(u)\exp(\beta x_i)\}\right\}^2. \tag{12}$$

A marginal pseudoscore statistic for testing the hypothesis $H_0 : \beta = \beta_0$ is obtained by noting that as $m \to \infty$, $U_1^2(\beta_0)/\widehat{\operatorname{var}}(U_1(\beta_0))$ follows a chi-squared distribution with one degree of freedom under H_0, so that large values may be interpreted as providing evidence against the null hypothesis (Cook, Lawless, and Nadeau, 1996).

If interest lies in estimation, a consistent estimate for β, denoted $\hat{\beta}$, is obtained by solving $U_1(\beta) = 0$ where $U_1(\beta)$ is given by (11). A Taylor series expansion gives

$$\operatorname{var}(\sqrt{m}(\hat{\beta} - \beta)) \approx E\left(m^{-1}\partial U_1(\beta)/\partial\beta\right)^{-2}\operatorname{var}\left(\sqrt{m}^{-1}U_1(\beta)\right). \tag{13}$$

Interval estimates may be obtained by inserting a consistent estimate for β into the right-hand side of (12) leading to $(m)^{\frac{1}{2}}(\hat{\beta} - \beta) \sim N(0, \hat{C}^{-2}\hat{\Sigma})$, where

$$\hat{C} = m^{-1}\sum_{i=1}^{m}\int_0^\tau Y_i(u)W_i(u;\hat{\beta})\, d\hat{\Lambda}(u)x_i\exp(\hat{\beta}x_i)$$

is an empirical estimate of $E(m^{-1}\partial U_1(\beta)/\partial\beta)$ and

$$\widehat{\operatorname{var}}\left(\sqrt{m}^{-1}U_1(\beta)\right) = m^{-1}\sum_{i=1}^{m}\int_0^\tau Y_i(u)W_i(u;\hat{\beta}) \times \{dN_i(u) - d\hat{\Lambda}(u)\exp(\hat{\beta}x_i)\}^2$$

is an empirical estimate which we denote by $\hat{\Sigma}$.

3.2 *Robust Semiparametric Conditional Methods*

Let $t_{i1}, t_{i2}, \ldots, t_{in_i}$ denote the times of the $n_i(\tau_i) = n_i$ events experienced by subject i over $(0, \tau_i]$, $i = 1, \ldots, m$. Under the assumption that the data are generated by mixed-Poisson processes the joint distribution of R_i and $\{N_i(t), 0 < t < \tau_i\}$ given v_i, is

$$P(R_i, \{N_i(t), 0 < t < \tau_i\} \mid v_i; \rho, \lambda(\cdot), \beta) = \frac{(v_i\rho)^{r_i}\exp\{v_i\rho\}}{r_i!} \cdot \frac{(v_i\Lambda_i(\tau_i))^{n_i}\exp(-v_i\Lambda_i(\tau_i))}{n_i!}\prod_{j=1}^{n_i}\frac{\lambda(t_{ij})}{\Lambda(\tau_i)}$$

where $\Lambda_i(\tau_i) = \int_0^{\tau_i} d\Lambda(s)\exp(\beta x_i) = \Lambda(\tau_i)\exp(\beta x_i)$ and $N_i(\tau_i) = n_i$, $i = 1, 2, \ldots, m$. Conditioning on $r_i + n_i$ eliminates v_i and gives

$$P(R_i, N_i \mid r_i + n_i; \rho, \Lambda(\cdot), \beta) = \binom{r_i + n_i}{n_i}\left(\frac{1}{1 + \exp(\log\Lambda(\tau_i) - \log\rho + \beta x_i)}\right)^{r_i} \times \left(\frac{\exp(\log\Lambda(\tau_i) - \log\rho + \beta x_i)}{1 + \exp(\log\Lambda(\tau_i) - \log\rho + \beta x_i)}\right)^{n_i}, \tag{14}$$

which is analogous to (3). Note that under a time-homogeneous model (i.e., $d\Lambda(t) = \lambda\, dt$) (14) reduces to (3) but with linear predictor $\alpha + \beta x_i + \log\tau_i$ where $\log\tau_i$ is an offset for subject i, $i = 1, \ldots, m$. The validity of this adaptation depends on time homogeneity of the rate function $d\Lambda(t)$, however, and the developments that follow are directed at relaxing this assumption for greater generality.

Differentiating the log likelihood resulting from (14) with respect to β gives

$$U_1(\beta, \Lambda(\cdot), \rho) = \sum_{i=1}^{m} U_{1i}(\beta, \Lambda(\cdot), \rho) = \sum_{i=1}^{m}\left\{ n_i - (r_i + n_i) \times \left(\frac{\exp(\log\Lambda(\tau_i) - \log\rho + \beta x_i)}{1 + \exp(\log\Lambda(\tau_i) - \log\rho + \beta x_i)}\right)\right\}x_i \tag{15}$$

which may be viewed as a pseudoscore function for β in the same sense as (11). To conduct inference about β, however, one must again deal with the nuisance parameters ρ and $\Lambda(\cdot)$ in (15). Natural estimates of these parameters are obtained by the introduction of the following auxiliary estimating functions,

$$U_{20}(\rho) = \sum_{i=1}^{m} U_{20i}(\rho) \tag{16}$$

$$U_{2j}(\beta; \Lambda(\cdot)) = \sum_{i=1}^{m} U_{2ji}(\beta; \Lambda(\cdot)), \quad j = 1, 2, \ldots, k, \tag{17}$$

where $U_{20i}(\rho) = r_i - \rho$ and $U_{2ji}(\beta; \Lambda(\cdot)) = Y_i(s_j)\{dN_i(s_j) - d\Lambda(s_j)\exp(\beta x_i)\}$, $j = 1, \ldots, k$, $i = 1, \ldots, m$. If $\boldsymbol{\psi} = (\rho, \Lambda(\cdot))'$ and $\boldsymbol{\theta} = (\beta, \boldsymbol{\psi})'$, let $\boldsymbol{U}_{2i}(\boldsymbol{\theta}) = (U_{20i}(\boldsymbol{\theta}), U_{21i}(\boldsymbol{\theta}), \ldots, U_{2ki}(\boldsymbol{\theta}))'$ denote the vector of auxiliary estimating functions for subject i, $\boldsymbol{U}_2(\boldsymbol{\theta}) = (U_{20}(\boldsymbol{\theta}), U_{21}(\boldsymbol{\theta}), \ldots, U_{2k}(\boldsymbol{\theta}))'$, $\boldsymbol{U}_i(\boldsymbol{\theta}) = (U_{1i}(\boldsymbol{\theta}), \boldsymbol{U}'_{2i}(\boldsymbol{\theta}))'$, and $\boldsymbol{U}(\boldsymbol{\theta}) = (U_1(\boldsymbol{\theta}), \boldsymbol{U}'_2(\boldsymbol{\theta}))'$. Setting (16) and (17) equal to zero and solving leads to consistent estimates $\hat{\rho} = \bar{r}$ where $\bar{r} = \sum_{i=1}^{m} r_i/m$ and $d\hat{\Lambda}^{\beta}(s)$ given by (10). These may be inserted into (15) to give an estimating function $U_1(\beta, \hat{\boldsymbol{\psi}}^{\beta})$

where $\hat{\boldsymbol{\psi}}^{\beta} = (\hat{\rho}, d\hat{\Lambda}^{\beta}(\cdot))$, which may be used to obtain a consistent estimate of the regression coefficient.

The estimating functions derived in the previous section are motivated by a mixed-Poisson model, but β is interpretable more generally provided $E(dN_i(s) \,|\, x_i) = d\Lambda(s)\exp(\beta x_i)$. Robust variance estimates are therefore required for this result to be useful however. Let $A = E(-m^{-1}\partial \boldsymbol{U}/\partial \boldsymbol{\theta}')$ and $B = E(\boldsymbol{U}\boldsymbol{U}')$, which take the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} E\big(-m^{-1}\partial U_1/\partial\beta\big) & E\big(-m^{-1}\partial U_1/\partial\boldsymbol{\psi}'\big) \\ E\big(-m^{-1}\partial \boldsymbol{U}_2/\partial\beta\big) & E\big(-m^{-1}\partial \boldsymbol{U}_2/\partial\boldsymbol{\psi}'\big) \end{bmatrix}$$

and

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} E(U_1U_1) & E(U_1\boldsymbol{U}_2') \\ E(\boldsymbol{U}_2U_1) & E(\boldsymbol{U}_2\boldsymbol{U}_2') \end{bmatrix},$$

respectively. If $\hat{\boldsymbol{\theta}}$ denotes the solution to $\boldsymbol{U}(\boldsymbol{\theta}) = \mathbf{0}$, then one can show by Taylor series expansion (Breslow, 1990) that

$$\begin{aligned} \text{asvar}\big(\sqrt{m}^{-1}U_1(\hat{\boldsymbol{\theta}})\big) = {} & B_{11} - A_{12}A_{22}^{-1}B_{21} - B_{12}\big[A_{22}^{-1}\big]'A_{12}' \\ & + A_{12}A_{22}^{-1}B_{22}\big[A_{22}^{-1}\big]'A_{12}'. \end{aligned} \tag{18}$$

If we denote $\text{asvar}(\sqrt{m}^{-1}U_1(\hat{\boldsymbol{\theta}}))$ by Σ_U, then $\sqrt{m}^{-1}U_1(\hat{\boldsymbol{\theta}}) \sim N(0, \Sigma_U)$ and $m^{1/2}(\hat{\beta} - \beta) \sim N(0, \Sigma_\beta)$ asymptotically where $\Sigma_\beta = A_{11}^{-1}\Sigma_U\, A_{11}^{-1}$. Estimates for Σ_U and Σ_β can be obtained by using empirical estimates in place of the terms of A and B.

A robust pseudoscore test of $\mathrm{H}_0 : \beta = \beta_0$ versus $H_A : \beta \neq \beta_0$ can be carried out based on the above results. For example, if $\tilde{\boldsymbol{\psi}}^{\beta_0}$ denotes the solution to $\boldsymbol{U}_2(\beta_0, \boldsymbol{\psi}) = \mathbf{0}$, then $\text{asvar}(\sqrt{m}^{-1}U_1(\beta_0, \tilde{\boldsymbol{\psi}}^{\beta_0}))$ is given as above but with variance estimates computed at β_0 and $\tilde{\boldsymbol{\psi}}^{\beta_0}$. A pseudoscore statistic of $\mathrm{H}_0 : \beta = \beta_0$ versus $H_A : \beta \neq \beta_0$ is, therefore, given by

$$\bar{U}\big(\beta_0, \tilde{\boldsymbol{\psi}}^{\beta_0}\big) = \frac{U_1\big(\beta_0, \tilde{\boldsymbol{\psi}}^{\beta_0}\big)}{\sqrt{m\hat{\Sigma}_U\big(\beta_0, \tilde{\boldsymbol{\psi}}^{\beta_0}\big)}} \tag{19}$$

which asymptotically follows a standard normal distribution under the null hypothesis. Large absolute values provide evidence against the null hypothesis.

4. Simulation Studies

4.1 *Mixed-Poisson Processes*

Here we report on the design and results of simulation studies denoted at assessing the frequency properties of the proposed pseudoscore test. In this section we suppose baseline counts and events during the follow-up period are generated according to mixed-Poisson processes and described in Section 2. Specifically we randomly assign subjects to receive either the experimental or control treatment during the follow-up phase with equal probability. We simulate v_i for subject i as gamma distributed with mean 1 and variance ϕ. The baseline count is simulated from a Poisson distribution with mean $v_i\rho$ and the event times during the follow-up phase are simulated from a time-homogeneous Poisson process with rate $v_i\lambda\exp(\beta x_i)$. The maximum follow-up time is set at $\tau = 1$, but variable durations of follow-up are induced by an exponential censoring rate of $\log(10/9)$ implying that 10% of the sample will have less than 1 unit of follow-up. Sample sizes of $m = 50$, 100, 200, and 400 are considered.

We consider $\rho = \lambda = 1$ and 4 to represent scenarios with relatively infrequent and more frequent events and $\beta = 0$, $\log(0.7)$ and $\log(0.50)$ to correspond to no treatment effect, moderate, and strong treatment effects, respectively. We set $\phi = 0.5$, 1, 2, and 4 to represent mild to extreme forms of extra-Poisson variation. Note that the asymptotic relative efficiency findings of Section 2 suggest that in the absence of censoring, the conditional test of Section 3.2 would potentially lead to greater power when $\phi > 1$ because $\rho = 1$.

For each simulated trial, the data were analyzed by testing the null hypothesis $\mathrm{H}_0 : \beta = 0$ versus $H_A : \beta \neq 0$ at the 5% significance level. Robust marginal (*MARG*) analyses were carried out based on the methods described in Section 3.1 (Cook et al., 1996), as well as via the robust pseudoscore test based on the conditional semiparametric model (*COND*) of Section 3.2. For each parameter configuration 2000 datasets were simulated and the proportion of trials for which the null hypothesis was rejected was recorded. This represents the empirical type I error rate when $\exp(\beta) = 1.0$ and the empirical power when $\exp(\beta) < 1.0$.

The results reported in Table 1 for $\rho = \lambda = 1$ confirm that the marginal pseudoscore statistic has an empirical type I error rate consistent with the nominal rate of 0.05 and that the proposed test based on (18) performs well even for samples comprised of as few as 50 subjects. In terms of the empirical power, the results are broadly consistent with what one would expect based on the asymptotic results. When $\beta = \log(0.70)$, for example, under mild extra-Poisson variation ($\phi = 0.50$) the conditional test leads to lower power than the marginal tests, but when $\phi = 1$ the empirical powers of the tests are comparable, and when $\phi > 1$ there can be substantially higher power with the conditional test. The findings are broadly similar when $\rho = \lambda = 4$ so we do not report on them here.

We also consider the setting in which the baseline period of observation is used to screen subjects for inclusion in the study (Cook and Wei, 2002). In such settings patients are typically selected if $r_i \geq c_R$ where c_R is a specified selection threshold. Here we focus on settings $\rho = 1$ and $\phi = 1.0$ and consider $\lambda = 1$, 2, and 4 and $c_R = 1$ and 2. The frequency properties are examined under the same treatment effects as Table 1 to study both empirical type I error and power. The results, given in Table 2, indicate that the conditional test retains good control over the type I error in the presence of selection criteria, even when $m = 50$. By comparing the results in Table 2 with the results of Table 1 for $\phi = 1$, one can also see an appreciable gain in the empirical power from the introduction of the selection criteria. Moreover, when selection criteria are used, the conditional test often has greater power than the marginal test in settings where the reverse is true in Table 1 (see line 6 of Table 1 and lines 3 and 4 of Table 2).

4.2 *Mixed Renewal Processes*

The proposed test was derived under a working assumption of a mixed-Poisson model, but variance estimation is carried out to ensure robustness. Here we report on simulation studies

Table 1
Empirical rejection rates of robust marginal and conditional pseudoscore statistics under mixed-Poisson models ($\rho = \lambda = 1$; $\tau = 1$; 10% censoring)

		$m = 50$		$m = 100$		$m = 200$		$m = 400$	
$\exp(\beta)$	ϕ	MARG	COND	MARG	COND	MARG	COND	MARG	COND
1.0	0.5	0.055	0.058	0.048	0.057	0.041	0.050	0.046	0.055
1.0	1.0	0.050	0.043	0.051	0.056	0.054	0.051	0.052	0.048
1.0	2.0	0.068	0.057	0.057	0.053	0.049	0.054	0.051	0.046
1.0	4.0	0.049	0.054	0.052	0.044	0.050	0.050	0.047	0.047
0.7	0.5	0.162	0.128	0.263	0.215	0.463	0.355	0.763	0.650
0.7	1.0	0.144	0.118	0.195	0.218	0.371	0.350	0.652	0.626
0.7	2.0	0.108	0.119	0.146	0.203	0.262	0.344	0.489	0.624
0.7	4.0	0.077	0.112	0.094	0.197	0.183	0.348	0.314	0.610
0.5	0.5	0.407	0.308	0.483	0.627	0.782	0.902	0.974	0.995
0.5	1.0	0.337	0.287	0.319	0.591	0.570	0.883	0.847	0.994
0.5	2.0	0.242	0.284	0.194	0.569	0.360	0.868	0.622	0.992
0.5	4.0	0.152	0.239	0.107	0.495	0.206	0.806	0.372	0.977

MARG: marginal pseudoscore test; COND: conditional pseudoscore test.

designed to assess the robustness of the proposed test to departures from mixed-Poisson processes.

Data were simulated according to mixed renewal processes as follows. Subjects were randomly assigned to receive either the experimental or control treatment during the follow-up phase, as in Section 4.1. We let $x_i = 1$ if subject i is assigned to the treatment arm and $x_i = 0$ otherwise, $i = 1, 2, \ldots, m$. A subject-specific random effect was generated from a gamma distribution so that $E(v_i) = 1$ and $\text{var}(v_i) = \phi$. A sequence of interevent times $w_{11}, w_{12}, \ldots,$ were then simulated such that

$$f(w_{ij} \mid v_i, \boldsymbol{\gamma}) = \frac{w_{ij}^{\gamma_1 - 1} \exp(-w_{ij}/(v_i\gamma_2))}{\Gamma(\gamma_1)(v_i\gamma_2)^{\gamma_1}}, \quad w_{ij} > 0,$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$. For a subject with $v_i = 1$ then, $E(w_{ij} \mid v_i = 1) = \gamma_1\gamma_2$. We generate $w_{i1}, w_{i2}, \ldots,$ until $\sum_{j=1}^{k_i} w_{ij} > \tau_i$ at which point we record $R_i = k_i - 1$ as the baseline count observed over the interval $(-\tau_R, 0]$. For simplicity we ignore the backward recurrence time at $t = 0$ and simulate a sequence of follow-up interevent times $u_{i1}, u_{i2}, \ldots,$ such that

$$f(u_{ij} \mid v_i, x_i; \boldsymbol{\gamma}, \beta) = \frac{u_{ij}^{\gamma_1 - 1} \exp(-u_{ij}/(v_i\gamma_2 \exp(\beta x_i)))}{\Gamma(\gamma_1)(v_i\gamma_2 \exp(\beta x_i))^{\gamma_1}}, \quad u_{ij} > 0$$

until $\sum_{j=1}^{n_i} u_{ij} > \tau_i$ where τ_i is the follow-up time for subject i, $i = 1, 2, \ldots, m$. At this point we record $t_{ij} = \sum_{k=1}^{j} u_{ik}$, $j = 1, 2, \ldots, n_i$, as the event times during the follow-up phase.

Table 2
Empirical rejection rates of robust marginal and conditional pseudoscore statistics under mixed-Poisson model with selection based on $r_i \geq c_R$ ($\rho = 1$; $\phi = 1.0$; $\tau = 1$; 10% censoring)

			$m = 50$		$m = 100$		$m = 200$		$m = 400$	
λ	$\exp(\beta)$	c_R	MARG	COND	MARG	COND	MARG	COND	MARG	COND
1.0	1.0	1	0.059	0.059	0.047	0.056	0.047	0.059	0.054	0.062
1.0	1.0	2	0.063	0.060	0.066	0.055	0.048	0.045	0.050	0.048
1.0	0.7	1	0.187	0.209	0.324	0.358	0.577	0.622	0.851	0.885
1.0	0.7	2	0.256	0.278	0.433	0.469	0.709	0.767	0.939	0.961
1.0	0.5	1	0.496	0.535	0.788	0.823	0.977	0.986	1.000	1.000
1.0	0.5	2	0.635	0.665	0.907	0.925	0.997	0.997	1.000	1.000
2.0	1.0	1	0.069	0.041	0.060	0.043	0.058	0.052	0.048	0.053
2.0	1.0	2	0.059	0.045	0.057	0.054	0.057	0.043	0.057	0.051
2.0	0.7	1	0.236	0.264	0.390	0.493	0.665	0.809	0.921	0.978
2.0	0.7	2	0.306	0.354	0.516	0.626	0.797	0.903	0.976	0.997
2.0	0.5	1	0.596	0.701	0.871	0.944	0.993	1.000	1.000	1.000
2.0	0.5	2	0.724	0.811	0.946	0.986	0.999	1.000	1.000	1.000
4.0	1.0	1	0.060	0.043	0.056	0.047	0.051	0.049	0.050	0.041
4.0	1.0	2	0.067	0.052	0.053	0.060	0.053	0.055	0.061	0.043
4.0	0.7	1	0.238	0.399	0.422	0.711	0.697	0.942	0.947	0.999
4.0	0.7	2	0.293	0.486	0.508	0.781	0.805	0.976	0.979	1.000
4.0	0.5	1	0.651	0.884	0.915	0.997	0.996	1.000	1.000	1.000
4.0	0.5	2	0.746	0.938	0.959	0.999	1.000	1.000	1.000	1.000

MARG: marginal pseudoscore test; COND: conditional pseudoscore test.

Table 3
Empirical rejection rates of robust marginal and conditional pseudoscore statistics under mixed renewal model ($\gamma_1 = 2$; $\tau = 1$; 10% censoring)

			$m = 50$		$m = 100$		$m = 200$		$m = 400$	
$\exp(-\beta)$	γ_2	ϕ	MARG	COND	MARG	COND	MARG	COND	MARG	COND
1.0	0.50	0.1	0.065	0.042	0.052	0.055	0.056	0.053	0.054	0.052
1.0	0.50	0.2	0.059	0.050	0.049	0.059	0.037	0.047	0.044	0.054
1.0	0.125	0.1	0.058	0.053	0.051	0.064	0.046	0.051	0.053	0.042
1.0	0.125	0.2	0.066	0.056	0.044	0.053	0.050	0.044	0.048	0.048
0.9	0.50	0.1	0.085	0.058	0.102	0.073	0.153	0.101	0.240	0.174
0.9	0.50	0.2	0.084	0.069	0.089	0.081	0.127	0.114	0.215	0.184
0.9	0.125	0.1	0.135	0.109	0.174	0.171	0.306	0.284	0.546	0.547
0.9	0.125	0.2	0.102	0.125	0.117	0.190	0.211	0.337	0.361	0.584
0.7	0.50	0.1	0.335	0.209	0.505	0.397	0.810	0.642	0.984	0.914
0.7	0.50	0.2	0.278	0.221	0.431	0.416	0.752	0.702	0.958	0.930
0.7	0.125	0.1	0.684	0.647	0.916	0.926	0.999	0.997	1.000	1.000
0.7	0.125	0.2	0.489	0.683	0.745	0.932	0.951	0.998	0.999	1.000

MARG: marginal pseudoscore test; COND: conditional pseudoscore test.

As before we set $\tau_R = \tau = 1$ and have an exponential censoring process following accrual with rate log(10/9). We set $\gamma_1 = 2$ and consider $\gamma_2 = 1/2$ and $1/8$ to be roughly comparable to the scenarios of the mixed-Poisson simulations in terms of the expected number of events. We consider $\beta = 0$, 0.9^{-1}, and 0.7^{-1} to represent no treatment effect and progressively stronger treatment effects induced by increasing the mean interevent times. We consider $\phi = 0.1$ and 0.2 to represent mixed renewal processes with mild and moderate degrees of heterogeneity. Table 3 displays the empirical type I error rates for the robust marginal and conditional tests in the first four rows where $\exp(-\beta) = 1.0$. The findings suggest that even for a sample of 50 subjects, the empirical type I error rate is in very good agreement with the nominal level for mixed renewal processes.

The last eight rows of Table 3 report the empirical power under the two levels of treatment effect. Here the robust conditional analysis may be seen to have higher or lower empirical power than the marginal analysis, depending on the degree of heterogeneity. When there is mild heterogeneity ($\phi = 0.1$) the marginal methods lead to greater or comparable power, but when $\phi = 0.2$, however, there can be substantial gains in power with the conditional analysis. This is analogous to the findings for the mixed-Poisson setting where the greater the heterogeneity the greater the appeal of the conditional analysis.

5. Extensions to Crossover Trials

5.1 *Model Formulation and Inference*

Multi-period long-term crossover designs represent a natural setting in which to apply the conditional analyses proposed in Section 3.2. There are, however, some unique features which require consideration due to the particularities of crossover designs. These include the facts that patients receive treatments in an order determined by the assigned sequence group, the duration of two or more treatment periods may vary among subjects, and the fact that carry-over effects may arise. Here we consider generalizations of the procedure described in Section 3.2 to address these issues in the context of a two-period placebo control crossover trial.

Suppose a total of m subjects are randomly assigned to either sequence group 1, in which subjects receive the experimental treatment during period 1 and the placebo control during period 2, or sequence group 2, in which subjects receive the treatments in the reverse order. We let $g_i = k$ if subject i was assigned to group k, $k = 1, 2$, $i = 1, \ldots, m$. In long-term crossover trials treatment periods have a nominal duration, but there may be considerable variation in the actual

Table 4
Marginal and conditional robust analyses of data on the occurrence of coughing episodes from Sears et al. (1990)

		Marginal analysis			Conditional analysis		
Sample	Covariate	Est.	SE	p-value	Est.	SE	p-value
Overall	Treatment	0.294	0.115	0.011	0.284	0.108	0.008
	Carry-over	0.377	0.298	0.205	0.230	0.102	0.025
Male	Treatment	0.734	0.225	0.001	0.761	0.198	<0.001
	Carry-over	0.299	0.618	0.628	0.455	0.305	0.135
Female	Treatment	0.144	0.126	0.252	0.138	0.118	0.242
	Carry-over	0.344	0.337	0.307	0.189	0.094	0.045

durations of these periods due to complications in scheduling visits that may be required for the crossover to take place (i.e., if new medications are required), or due to early patient withdrawal. It is, therefore, important to give careful consideration to the actual treatment received at the time of events rather than the scheduled treatment.

Suppose the two treatment periods each has a nominal duration of τ units, but let τ_{i1} and τ_{i2} represent the actual duration of periods 1 and 2 for subject i, respectively, and let $\tau_i = \tau_{i1} + \tau_{i2}$, $i = 1, \ldots, m$. For subjects in sequence group 1, we may specify an interval $(\tau_{i1}, \tau_{i1} + c]$ over which there may be a residual carry-over effect of the experimental treatment received during period 1. For subjects in sequence group 2 there is no need to consider such an interval because for these patients a placebo treatment was received during period 1. Let $x_{i1}(t) = 1$ if subject i received the experimental treatment at time t and $x_{i1}(t) = 0$ otherwise, and let $x_{i2}(t) = 1$ if $g_i = 1$ and $\tau_{i1} < t \leq \tau_{i1} + c$, and $x_{i2}(t) = 0$. Therefore, $x_{i1}(t)$ indicates whether the current treatment is the experimental treatment and $x_{i2}(t)$ indicates whether the potential for a carry-over effect of the experimental treatment is present at time t. Given a subject-specific random effect v_i, we define the event rate at time s as $E(dN_i(s) \,|\, v_i;\, x_i(s)) = v_i d\Lambda_i(s)$ where $d\Lambda_i(s) = d\Lambda(s)\, \exp(\beta_1 x_{i1}(s) + \beta_2\, x_{i2}(s))$ and $d\Lambda(s)$ is a baseline rate function. We define $N_i(0,\, \tau_{i1}) = N_{i1}$, $N_i(\tau_{i1},\, \tau_{i1} + c) = N_{i21}$, $N_i(\tau_{i1} + c,\, \tau_i) = N_{i22}$, $N_i(\tau_{i1},\, \tau_i) = N_{i2}$ $(= N_{i21} + N_{i22})$, and $N_i(0,\, \tau_i) = N_i$, and use lower case letters to represent the corresponding realized values. Finally, we let $\Lambda(s,t) = \int_s^t d\Lambda(u)$, $\Lambda_i(s,t) = \int_s^t d\Lambda_i(u)$ and write $\Lambda(t)$ and $\Lambda_i(t)$ for $\Lambda(0,\, t)$ and $\Lambda_i(0,\, t)$, respectively.

The analogue of (3) for patients in sequence group 1 is,

$$P(n_{i1}, n_{i21}, n_{i22} \,|\, n_i) = \binom{n_i}{n_{i1} n_{i21} n_{i22}} \times \frac{\left(\Lambda(\tau_{i1}) e^{\beta_1}\right)^{n_{i1}} \left(\Lambda(\tau_{i1}, \tau_{i1} + c) e^{\beta_2}\right)^{n_{i21}} (\Lambda(\tau_{i1} + c, \tau_{i1}))^{n_{i22}}}{\Lambda_i(\tau_i)^{n_i}},$$

while for individuals in sequence group 2 we obtain

$$P(n_{i1}, n_{i2} \,|\, n_i) = \binom{n_i}{n_{i1} n_{i2}} \frac{(\Lambda(\tau_{i1}))^{n_i} \left(\Lambda(\tau_{i1}, \tau_i) e^{\beta_1}\right)^{n_{i2}}}{\Lambda_i(\tau_i)^{n_i}}.$$

The estimating functions for β_1 and β_2 are obtained as score functions from the resulting likelihood function and are given by

$$U_{11} = \sum_{i=1}^{m} U_{11i} \tag{20}$$

$$U_{12} = \sum_{i=1}^{m} U_{12i}, \tag{21}$$

where

$$U_{11i} = n_{ig_i} - n_i [\Lambda(\tau_{i1})]^{I(g_i=1)} [\Lambda(\tau_{i1}, \tau_i)]^{I(g_i=2)} \exp(\beta_1)/\Lambda_i(\tau_i)$$

is the contribution from the ith subject to the estimating function for β_1 and

$$U_{12i} = [n_{i21} - n_i \Lambda(\tau_{i1}, \tau_{i1} + c) \exp(\beta_2)/\Lambda_i(\tau_i)] I(g_i = 1)$$

is the corresponding contribution to the estimating function for β_2. A supplementary estimating function, given by

$$U_{2j} = \sum_{i=1}^{m} U_{2ij}, j = 1, 2, \ldots, k,$$

where

$$U_{2ij} = Y_i(s_j)[dN_i(s_j) - d\Lambda(s_j) exp(\beta_1 x_{i1}(s_j) + \beta_2 x_{i2}(s_j))],$$

gives

$$d\hat{\Lambda}^{\beta}(s_j) = \frac{\sum_{i=1}^{m} Y_i(s_j)\, dN_i(s_j)}{\sum_{i=1}^{m} Y_i(s_j) \exp(\beta_1 x_{i1}(s_j) + \beta_2 x_{i2}(s_j))}, \tag{22}$$

which may be substituted into (20) and (21) to obtain $\hat{\boldsymbol{\beta}}$, an estimate of $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Asymptotic variances may be obtained as in Section 3.2.

5.2 *Application to an Asthma Study*

Sears et al. (1990) report on a two-period crossover trial of asthma patients designed to compare the effects of regular use of an active inhaled bronchodilator (fenoterol hydrobromide) versus placebo for the treatment of symptoms of asthma. Patients were randomized to sequence groups. In one sequence group patients administered fenoterol four times daily for 24 weeks followed by another 24-week period during which they administered a matching placebo in a similar fashion; the other sequence group administered the placebo during the first 24 weeks followed by fenoterol. Patients recorded morning and evening peak expiratory flow rates, sputum production, chest tightness, use of rescue bronchodilators, and episodes of daytime and nighttime coughing and wheezing. Sixty-four patients were deemed eligible for inclusion in the primary analysis reported in Sears et al. (1990); reasons for subject exclusion are enumerated there. Among these patients the unique crossover times (number of subjects) were 164 (1), 165 (1), 166 (3), 167 (7), and 168 (52) days, and the unique total follow-up times were 335 (59), 336 (3), 338 (1), and 343 (1) days. Here we report on the analysis of data on a secondary outcome of nighttime coughing. In previous analyses (Ng and Cook, 1999) we found a difference in the effect of therapy for males ($m = 29$) and females ($m = 35$), so we report here on results overall and separately for males and females. Both the robust marginal analyses of Lawless and Nadeau (1995) and the robust conditional analyses based on Section 5.1 were performed.

Table 4 displays the estimates, standard errors, and Wald-based p-values arising from analyses based on the marginal method of Lawless and Nadeau (1995) and the analysis based on the robust conditional methods of Section 5.1. A 2-month duration was selected for possible carry-over effects; alternative durations were considered but the conclusions regarding the main treatment effect did not change so we simply report on these here. The standard errors based on the conditional analyses are smaller than from the robust marginal analysis reflecting an empirical gain in efficiency for this trial. The point estimate and 95% confidence intervals for relative risks reflecting the overall direct treatment effects from the two

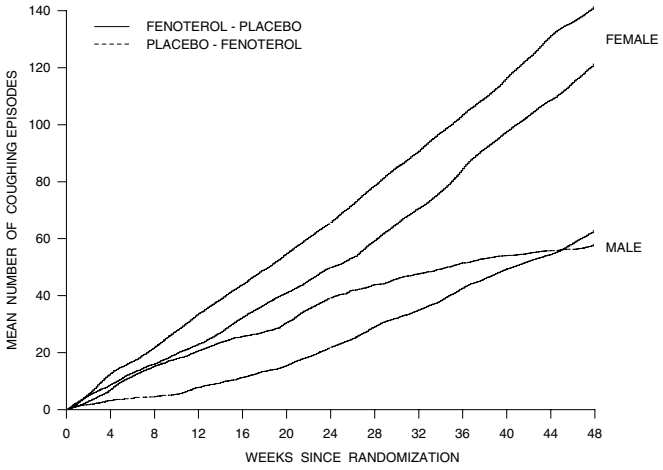

Figure 2. Nelson–Aalen estimates of the cumulative mean number of coughing episodes by sex and sequence group.

models are 1.34 (95% CI (1.07, 1.68)) and 1.32 (95% CI (1.07, 1.64)) for the marginal and conditional analyses, respectively. For males alone, the respective relative risks are 2.08 (95% CI (1.34, 3.23)) and 2.14 (95% CI (1.45, 3.16)), and for females they were 1.15 (95% CI (0.90, 1.48)) and 1.15 (95% CI (0.91, 1.45)). Figure 2 displays plots of the Nelson–Aalen estimates of the cumulative mean functions by sex over the course of follow-up. Note that the Nelson–Aalen plots of the mean functions for females in the two sequence groups are consistent with the presence of substantial carry-over effects from fenoterol received during period 1 because the mean functions do not converge in the second treatment period.

6. Discussion

We have developed semiparametric methods for efficient estimation of treatment effects based on recurrent events useful in long-term trials of patients observed over two or more treatments periods. The estimating functions we propose are derived under a "working" mixed-Poisson model but are valid provided the mean specification of the model is correct. The conditioning under the working model is motivated by the elimination of subject effects and a desire to make assessments of treatments based on within-subject comparisons as in the classical Gaussian framework. Robust variance estimates ensure the proposed methods work well in a broad range of contexts.

The study of the asymptotic relative efficiency revealed that under a mixed-Poisson model, the conditional approach is more efficient than the marginal approach when the baseline mean is greater than the inverse of the random effect variance and the simulation studies bear this out. When considering analysis strategies at the design stage in such settings, one can reduce the number of subjects required (e.g., Cook, 1995) by using the appropriate method of analysis. Sample size formulae have not been provided for the conditional model, but they are relatively straightforward to derive.

In many contexts, events are not observed in continuous time, but rather cumulative event counts are available from periodic assessments. Examples include radiographic studies in rheumatology where x-rays are required to count the number of newly damaged joints between clinic visits (Gladman et al., 1998) and studies of patients with cancer metastatic to bone where the outcome may be the development of new bone lesions only detectable upon bone scan (Hortobagyi et al., 1998). When counts are only observed after randomization, random effect models for data of this sort have been considered by a number of authors including Lawless and Zhan (1998), Staniswalis, Thall, and Salch (1997), and Dean and Balshaw (1997). The methods we propose could be adapted to deal with interval-censored recurrent event data. In this case, the estimate of the baseline mean function could be based on simple parametric models, or possibly piecewise constant rate functions possibly with the use of smoothing splines. Such approaches merit consideration in this context.

Acknowledgements

The authors thank Michael Kenward, an associate editor, and two referees for helpful comments, Ker-Ai Lee for programming assistance, and Drs Malcolm Sears and Peter Herbison for providing the data from the asthma trial. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes for Health Research (CIHR). R. J. Cook is a CIHR Investigator.

References

ACIP Investigators. (1992). Asymptomatic cardiac ischaemia pilot study (ACIP). *American Journal of Cardiology* **70,** 744–747.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics* **10,** 1100–1120.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Breslow, N. E. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* **85,** 565–571.

Cook, R. J. (1995). The design and analysis of randomized trials with recurrent events. *Statistics in Medicine* **14,** 2081–2098.

Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods for Medical Research* **11,** 141–166.

Cook, R. J. and Wei, W. (2002). Selection effects in randomized trials with count data. *Statistics in Medicine* **21,** 515–531.

Cook, R. J. and Wei, W. (2003). Conditional analysis of mixed Poisson processes with baseline counts. *Biostatistics* **4,** 479–494.

Cook, R. J., Lawless, J. F., and Nadeau, J. C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52,** 732–739.

Dean, C. B. and Balshaw, R. (1997). Efficiency lost by analysing counts rather than event times in Poisson and

overdispersed Poisson regression models. *Journal of the American Statistical Association* **92,** 1387–1398.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of Longitudinal Data*. London: Oxford University Press.

Fuchs, H. J., Borowitz, D. S., Christiansen, D. H., Morris, E. M., Nash, M. L., Ramsey, B. W., Rosenstein, B. J., Smith, A. L., and Wohl, M. E. (1994). Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *New England Journal of Medicine* **331,** 637–642.

Gladman, D. D., Farewell, V. T., Kopciuk, K. A., and Cook, R. J. (1998). HLA markers and progression in psoriatic arthritis. *Journal of Rheumatology* **25,** 730–733.

Hortobagyi, G. N., Theriault, R. L., Lipton, A., Porter, L., Blayney, D., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., Knight, R. D., Heffernan, M., Mellars, K., and Reitsma, D. J. (1998). Long-term prevention of skeletal complications of metastatic breast cancer with pamidronate. *Journal of Clinical Oncology* **16,** 2038–2044.

Hughes, M. D. (1997). Power considerations for clinical trials using multivariate time-to-event data. *Statistics in Medicine* **16,** 865–882.

Jones, B. and Kenward, M. G. (2003). *Design and Analysis of Cross-Over Trials*, 2nd edition. New York: Chapman and Hall/CRC.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics* **15,** 209–225.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition. New York: Wiley.

Lawless, J. F. and Nadeau, J. C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168.

Lawless, J. F. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics* **4,** 549–565.

Li, Q. H. and Lagakos, S. W. (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16,** 911–924.

Ng, E. T. M. and Cook, R. J. (1999). Robust inference for bivariate point processes. *Canadian Journal of Statistics* **27,** 509–524.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* **88,** 811–820.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.

Romanowski, B., Marina, R. B., and Roberts, J. N. (2003). Patients' preference of valacyclovir once-daily suppressive therapy versus twice-daily episodic therapy for recurrent genital herpes: A randomized study. *Sexually Transmitted Diseases* **30,** 226–231.

Sears, M. R., Taylor, D. R., Print, C. G., Lake, D. C., Li, Q., Flannery, M., Yates, D. M., Lucas, M. K., and Herbison, G. P. (1990). Regular inhaled beta-agonist treatment in bronchial asthma. *Lancet* **336,** 1391–1396.

Staniswalis, J. G., Thall, P. F., and Salch, J. (1997). Semiparametric regression analysis for recurrent event interval counts. *Biometrics* **53,** 1334–1353.

Thall, P. F. and Vail, S. C. (1995). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46,** 657–671.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Received January 2004. *Revised December* 2004.
Accepted January 2005.