Point Process Regression Models for Multiple Events with Random Effects and Measurement Error

Wenxin Jiang
Department of Statistics, Northwestern University, Evanston, IL 60208,
Bruce W. Turnbull
School of Operations Research and Statistics Center, 227 Rhodes Hall, Cornell University, Ithaca, NY 14853-3801
and
Larry C. Clark
Arizona Cancer Center, University of Arizona, Tucson AZ 85724

March 5, 1997

Abstract

Statistical methodology is presented for the analysis of multiple events with random effects and measurement error. We model multiple events in a general space using a random measure, and define point process regression models with residual random effects. Our approach to parameter estimation and significance testing is to start with a simple naive model of Poisson process regression, and then to adjust for random effects and any possible covariate measurement error. We illustrate the techniques with data from a randomized clinical trial for the prevention of recurrent skin tumors.

1 Introduction

Consider the statistical analysis of a follow-up (or reliability) study in which individual subjects (or units) may experience a series of recurrent events over time. Subjects are heterogeneous and covariates are available from each subject. The event times for each subject can be viewed as a realization of a stochastic point process. To analyze such data, a variety of parametric point process regression models have been proposed — a good discussion appears in Lawless (1987). A more flexible approach is to use a semiparametric approach, in which the functional form of the baseline intensity function is unspecified. Four such approaches which may all be considered generalizations of the Cox (1972) model for survival data. These are: the counting process formulation of Andersen and Gill (1982); the marginal approach of Wei, Lin and Weissfeld (1989); the conditional approach of Prentice, Williams and Peterson (1981); and the cumulative mean function approach of Lawless and Nadeau (1995).

Statistical software to implement parametric and semi-parametric methods are available in standard computer packages, e.g. STATA 5.0 (StataCorp 1997) and Splus 3.3 (Statistical Science Inc. 1995).

There are occasions when we encounter a more general setting than multiple events occurring over time for which we may need to consider point processes in a higher dimensional space. For example, the events occurring over time may be of different types or severity — for an example, see Abu-Libdeh et al. (1990). In this case, the corresponding point process is defined on a product space of the time and the type (or severity) of the events. In this paper we introduce a flexible model for multiple events based on point processes in this general setting. We concentrate on the issue of model misspecification, specifically presence of omitted covariates and random effects and presence of covariate measurement error.

For the special case of survival analysis, where subjects can experience at most a single event during their followup time, a number of researchers have examined the problem of covariate measurement error in a partial likelihood analysis using Cox's (1972) model. Prentice (1982) and Pepe et. al. (1989) considered the induced hazard function conditional on the observed covariate instead of its true value. Nakamura (1992) and Buzas (1996) used procedures based on constructing unbiased or approximately unbiased estimating equations from the partial likelihood score equations. The methods of Raboud (1991) and Raboud et al. (1993) were based on examining the root of the asymptotic score equation of the naive partial likelihood. There is also an extensive literature on misspecification of the Cox model due to the presence of frailties or omitted covariates. A recent review has been given by Keiding, Andersen and Klein (1997).

The effects of random subject heterogeneity and covariate measurement error on fully parametric analyses of recurrent events, based on time-homogeneous Poisson process with gamma mixture, have been discussed by Turnbull, Jiang and Clark (1997), of which the present paper may be regarded as a semi-parametric analog.

Sandwich-type robust asymptotic variances have been discussed by several authors, including Wei, Lin and Weissfeld (1989), Lin and Wei (1989), and Lawless and Nadeau (1995). In this paper, we will show that the robust variances can account for extra variation from random effects and from covariate measurement error, and can consequently lead to diminished $\mathcal{Z}$-values for testing the significance of the regression coefficients.

First, in Section 2, we introduce the random measure approach of point processes in a discrete version. In Section 3 we discuss the naive likelihood-based analysis based on Poisson processes. In the subsequent sections we examine the effects of random heterogeneity and covariate measurement error on parameter estimation and significance tests and discuss how to adjust for these effects. In Section 8, we illustrate our results with data on recurrent skin cancers from the Nutritional Prevention of Cancer (NPC) clinical trial (Clark et al. 1996). We close in Section 9 with a discussion of the assumptions underlying the models.

2 Modeling Multiple Events With Residual Random Effects

Let $(\tilde{E}, \mathcal{E})$ be a measurable space where $\mathcal{E}$ is a σ-field on the carrier space $\tilde{E}$, which could be the time axis, a two-dimensional plane, or a product space of event type and event time, for example. We are interested in multiple events on $\tilde{E}$. The outcome variable for a single subject is a random measure $\tilde{Y}(\cdot)$ on $(\tilde{E}, \mathcal{E})$. If $\tilde{Y}(\cdot)$ is non-negative integer valued, it is called a point process (Karr 1991, Chap. 1). In that case, $\tilde{Y}(B)$ counts the random number of points in a set B in $\mathcal{E}$. The expectation $\Phi(\cdot) = E\tilde{Y}(\cdot)$ forms a mean measure. When $\tilde{E}$ is the one dimensional time axis, the mean measure is specified by a cumulative intensity function $\Lambda(t) = \Phi((0, t])$.

It is convenient to formulate the problem in a discrete version. Suppose the carrier space has a finite partition $\tilde{E} = \bigcup_{k \in \mathcal{K}} \tilde{E}_k$. Choose the σ-field $\mathcal{E} = \sigma(\{\tilde{E}_k\})$ generated by all subsets of $\{\tilde{E}_k\}_{k \in \mathcal{K}}$. On $\mathcal{E}$, the random measure $\tilde{Y}(\cdot)$ is determined by the collection of random variables $\tilde{Y}_k \equiv \tilde{Y}(\tilde{E}_k)$, $k \in \mathcal{K}$; the mean measure $\Phi(\cdot)$ is determined by $\Phi_k \equiv \Phi(\tilde{E}_k)$, $k \in \mathcal{K}$, and $\Phi_k = E\tilde{Y}_k$. Therefore Φ_k is the expected number of events on "patch" $\tilde{E}_k$. Because of censoring, not all patches may be under observation. We define 'at risk' indicators: $H_k = 1$ if "patch" $\tilde{E}_k$ is under observation, 0 otherwise, for $k \in \mathcal{K}$. Then $Y_k = H_k \tilde{Y}_k$ denotes the *observed* random measure outcome on "patch" $\tilde{E}_k$. Note that $H_k Y_k = Y_k$, since $H_k^2 = H_k$.

For multiple events, we postulate a relationship between the outcome process $\tilde{Y}(\cdot)$ and a collection of p covariates, using a specified regression model. However, it may well be that the outcome process also depends on factors *not* included in the model. Suppose the random measure outcome $\{\tilde{Y}_k, k \in \mathcal{K}\}$ is related to a set of predictors, (Z, O, H). Here $H = \{H_k; k \in \mathcal{K}\}$ are the at risk indicators defined above, $Z = \{Z_k; k \in \mathcal{K}\}$ where Z_k is the value of the *included* covariate $p-$vector taken on by the subject on patch $\tilde{E}_k$, and $O = \{O_k; k \in \mathcal{K}\}$ where O_k is the value of a vector of *omitted* covariates taken on by the subject on patch $\tilde{E}_k$. Often the vectors $\{Z_k\}$ (and perhaps $\{O_k\}$) will be constant of k; for example, if $\tilde{E}$ is the time axis and the covariates are non time varying. The mean measure is now defined as the conditional mean $\Phi_k = E\{\tilde{Y}_k | Z, O, H\}$, $k \in \mathcal{K}$, conditional on all covariates. A regression model specifies a link function, which is the expectation conditional on the covariates *included* in the model. For example, an exponential link function is defined as

$$E\{\tilde{Y}_k | Z, H\} = \tilde{\Lambda}_k exp(Z_k' \beta) \quad \text{or} \quad E\{Y_k | Z, H\} = H_k \tilde{\Lambda}_k exp(Z_k' \beta), \tag{1}$$

for some scalar $\tilde{\Lambda}_k$ and regression coefficient vector β of dimension p. Note that in the second equation in (1), $Y_k = H_k \tilde{Y}_k$, the *observed* outcome on $\tilde{E}_k$ and that (1) provides a special example of the *multiplicative* mean measure model.

Equation (1) implies that the mean process (conditional on Z_k's) for the outcome $\{\tilde{Y}_k\}$ is independent of the observing process $\{H_k\}$. More details of the implication will be discussed in Section 9.

The *residual random effect* is defined as the ratio $\psi_k = \Phi_k / E\{\tilde{Y}_k | Z, H\}$ (assuming the positiveness of the denominator). The rationale of this definition comes from the fact that the mean measure Φ_k's are not fully determined by the covariate Z_k's included in the model, but also depend on the omitted covariates, O. Therefore $\Phi_k = E\{\tilde{Y}_k | Z, O, H\}$ differs from the link function $E\{\tilde{Y}_k | Z, H\}$, introducing a residual random effect $\psi_k = \Phi_k / E\{\tilde{Y}_k | Z, H\}$. Note that

$E(\psi_k|Z,H) = 1$. In general we can write $\Phi_k = \psi_k E\{\tilde{Y}_k|Z,H\}$. (E.g., $\Phi_k = \psi_k \tilde{\Lambda}_k exp(Z'_k\beta)$ for the exponential link.)

Now consider the situation where there are n subjects, which are n i.i.d. replicates of the system described above, labeled by $i = 1, 2, ..., n$. With the obvious extension of notation, Y_{ik}, H_{ik}, Z_{ik}, O_{ik} denote the corresponding quantities for patch k of system i, $i = 1, 2, \ldots, n$, and $k \in \mathcal{K}$. Similarly Φ_{ik} is the mean measure or the 'intensity' of events for patch k of system i. We will concentrate on an inhomogeneous Poisson process regression model with residual random effects. This is an example of the multiplicative mean measure model with exponential link. Specifically this model assumes:

> Conditional Poisson model: Conditional on $\{Z_{ik}, O_{ik}, H_{ik}; i = 1, 2, \ldots, n,\ k \in \mathcal{K}\}$, the observed random variables Y_{ik} are independent Poisson distributed with means $H_{ik}\psi_{ik}\tilde{\Lambda}_k exp(Z'_{ik}\beta),\ \{i = 1, 2, \ldots, n,\ k \in \mathcal{K}\}$.

When a system of patches is composed of line segments on a time axis, we are essentially describing a point process in discrete time. In the counting process formulation, the mean measure Φ_{ik} becomes the discrete version of intensity function $d\tilde{\Lambda}_i(t)$, noting that k labels the time intervals. Similarly Y_{ik} becomes $dN_i(t)$, Z_{ik} becomes $Z_i(t)$, and H_{ik} becomes $\Delta_i(t)$ which indicates whether subject i is at risk at time t. The results we obtain in the following sections can be transformed into this continuous time formulation. Of course, the patch indices $\{k\}$ can also label the lattice squares on a plane, so that $\tilde{\Lambda}_{ik}$ is basically the 2-dimensional intensity function $\tilde{\Lambda}_i(x, y)$. We can also include multi-type processes in time, for example, by using an index set $\mathcal{K} = \{1, 2\} \times \{1, 2, 3, ..., 1000\}$ where each $k \in \mathcal{K}$ is an integer-valued vector with its first component indicating the type of events (1 or 2, say), and the second component indicating the location on the time axis (intervals 1,2,...,1000, say). However, since $\mathcal{K}$ is finite, from now on we will take $\mathcal{K} = \{1, 2, \ldots, K\}$ for some K, without loss of generality. Usually discretization will not cause any problem in real applications. For example, in the NPC clinical trial described in Section 8, event times were recorded only as falling on a given day; and thus "days" formed a natural discrete time scale.

3 Naive Log-likelihood and Naive MLE

In the conditional Poisson model of the previous section, the residual random effects $\{\psi_{ik}; 1 \leq i \leq n, 1 \leq k \leq K\}$ are unobserved. The likelihood function based on the observed information $W_{ik} = (Y_{ik}, H_{ik}, Z_{ik}), 1 \leq i \leq n, 1 \leq k \leq K$, involves an integration over the unknown ψ_{ik}'s, which can be very difficult. Instead, we begin by considering a naive analysis in which the presence of the random effects is neglected (all ψ_{ik}'s are taken as 1). This leads to a misspecified or "naive" log likelihood function $R = R(s)$ which, up to a constant of argument $s' \equiv (\Lambda_1, ..., \Lambda_K, b')$, is given by

$$R = \sum_{i=1}^{n} \log \prod_{k=1}^{K} \{(\Lambda_k e^{Z'_{ik}b})^{H_{ik}Y_{ik}} exp(-H_{ik}\Lambda_k e^{Z'_{ik}b})\} = \sum_{i,k} H_{ik}\{Y_{ik}Z'_{ik}b + Y_{ik} log\ \Lambda_k - \Lambda_k e^{Z'_{ik}b}\}.$$

the true parameter $\theta' \equiv (\tilde{\Lambda}_1, ..., \tilde{\Lambda}_K, \beta')$ with $s' \equiv (\Lambda_1, ..., \Lambda_K, b')$ to emphasize the fact that we are using a misspecified model in which the parameters may not have the same

interpretation. We note that R is a sum of n i.i.d. copies of the function

$$\rho = \rho(s) = \sum_k H_k\{Y_k Z_k' b + Y_k \log \Lambda_k - \Lambda_k e^{Z_k' b}\} \tag{2}$$

where we have suppressed the index i for the i.i.d. systems.

For fixed b, R is maximized at $\Lambda_k = (\sum_i Y_{ik})(\sum_i H_{ik} e^{Z_{ik}' b})^{-1}$. Substituting this value of Λ_k into R, we find that the function R on the "ridge" is just the log partial likelihood function $\mathcal{L}$ of the argument b, up to some constant, where

$$\mathcal{L}(b) = \log \prod_{i,k} \left(\frac{e^{Z_{ik}' b}}{\sum_j H_{jk} e^{Z_{jk}' b}}\right)^{Y_{ik}}.$$

Similar arguments in counting process theory can be seen in Andersen et al. (1993, Sec.VII.2.1, Page 482). Hence R is maximized by the "naive" maximum likelihood estimates (MLEs) $b = \hat{b}$ and $\Lambda_k = \hat{\Lambda}_k$, $k = 1, ..., K$, where

$$\hat{b} = \arg\max \mathcal{L}(b) \quad \text{and} \quad \hat{\Lambda}_k = (\sum_i Y_{ik})(\sum_i H_{ik} e^{Z_{ik}' \hat{b}})^{-1}, \tag{3}$$

where the $\hat{\Lambda}_k$'s form a discrete version of the Nelson-Aalen estimates (*cf.* Andersen et al. 1993, Sec.VII.2).

4 Asymptotic Properties of the Naive MLE

The asymptotic properties of such naive likelihood estimates have been discussed by various authors including Huber (1967), White (1994), Jiang (1996), Turnbull et al. (1997). The principal results can be summarized as follows:

Proposition: Suppose W_i $i = 1, ..., n$ are *i.i.d.* copies of W, where W has a probability distribution $P_W^{(\theta)}$. Let $R_n(s) = \sum_{i=1}^n \rho(W_i; s)$, which could be a naive log-likelihood function with argument s. Under regularity conditions (Jiang 1996, Turnbull et al. 1997),

A. $\hat{s}_n \equiv \arg\max_s R_n(s)$ is strongly consistent to $s^0(\theta) = \arg\max_s E_\theta \rho(W; s)$, and $\sqrt{n}(\hat{s}_n - s^0(\theta)) \xrightarrow{\mathcal{D}} N(0, I^{-1} V I^{-1})$ where $I = -\mathrm{E}\nabla^2 \rho(W; s)\Big|_{s^0(\theta)}$ and $V = \mathrm{E}\nabla\rho\nabla\rho'\Big|_{s^0(\theta)}$;

B. An inverse function $(s^0)^{-1}$ exists, $\hat{\theta}_n = (s^0)^{-1}(\hat{s}_n)$ is strongly consistent to the original parameter θ, and $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, D' I^{-1} V I^{-1} D)$ with $D = \nabla (s^0)^{-1}\Big|_{s^0(\theta)}$;

C. $s = s^0(\theta)$ satisfies the estimating equation

$$\nabla \mathrm{E}_\theta \rho(W; s) = 0. \tag{4}$$

In the following discussion, when convenient, we may omit the subscript n for $\hat{s}_n$, $\hat{\theta}_n$ and R_n, and denote the asymptotic limit of the naive MLE $s^0(\theta)$ as $s(\theta)$ or simply s, which

we sometimes refer to as the "naive" parameter. The relation $s^0(\theta)$ expresses the naive parameter in terms of the original parameter, which is termed as the "bridge relation" (Jiang 1996).

Note that the naive MLE $\hat{s}$ is in general inconsistent for the original parameter θ, but instead consistently estimates the naive parameter $s(\theta)$. When $s(\theta)$ is invertible, we can form a consistent estimator $\hat{\theta}$ for θ by inverting the relation $s(\theta)$, as suggested by result (B) of the Proposition. The general strategy therefore is to attempt to find the bridge relation $s(\theta)$ using result (A) or (C) of the proposition, and then assuming the relation is invertible, use result (B) to obtained an "adjusted" estimator of θ.

The asymptotic normality of $\hat{s}$ and $\hat{\theta}$ from the above proposition show how standard errors and test statistics can be constructed. The "naive" asymptotic variance (matrix) of $\hat{s}$ that ignores the model misspecification is $(nI)^{-1}$, the inverted naive information matrix. The correct asymptotic variance of $\hat{s}$ is $n^{-1}I^{-1}VI^{-1}$, the so-called "sandwich formula" — Huber (1967), Carroll et al. (1995, page 263). The asymptotic variance of the adjusted estimate $\hat{\theta}$ is thus given by $n^{-1}D'I^{-1}VI^{-1}D$, which we term as the "double-sandwich formula". If the expectations in I and V are not available then quantities based on sample averages can be used in the usual way, *e.g.* for I use $\hat{I} = -n^{-1}\sum_i \nabla^2\rho(W_i;s)\,|_{\hat{s}}$, for V use $\hat{V} = n^{-1}\sum_i(\nabla\rho(W_i;s))(\nabla\rho(W_i;s))'\,|_{\hat{s}}$. To test the significance of a particular regressor variable, the jth say, using the Wald method, the naive test statistic that ignores model misspecification would be the $\mathcal{Z}$-value: $\mathcal{Z}_N = \hat{s}_j/\sqrt{Avar_N(\hat{s}_j)}$ where $Avar_N(\hat{s}_j)$ is the jth diagonal element of $(nI)^{-1}$. The correct $\mathcal{Z}$-statistic is: $\mathcal{Z}_{\text{Adj.}} = \hat{\theta}_j/\sqrt{Avar(\hat{\theta}_j)}$. where $Avar(\hat{\theta}_j)$ is the jth diagonal element of $n^{-1}D'I^{-1}VI^{-1}D$.

In the present formalism, W is the collection of variables $\{H_k, Y_k, Z_k; k \in \mathcal{K}\}$, $\theta = (\{\tilde{\Lambda}_k\}_{k\in\mathcal{K}}, \beta')'$, and $s = (\{\Lambda_k\}_{k\in\mathcal{K}}, b')'$. The first problem is to find the consistent limit $s^0(\theta)$ of the naive MLE by using equation (4), which involves taking derivatives of ρ for all components of s. Taking the expectation by first conditioning on the H_k's and Z_k's, the equations become

$$\begin{aligned} \partial_b E\rho &= \sum_k E\ H_k Z_k(Y_k - \Lambda_k e^{Z_k'b}) = \sum_k E\ H_k Z_k(\tilde{\Lambda}_k e^{Z_k'\beta} - \Lambda_k e^{Z_k'b}) = 0 \\ \text{and } \partial_{\Lambda_k} E\rho &= \Lambda_k^{-1} E\ H_k(Y_k - \Lambda_k e^{Z_k'b}) = \Lambda_k^{-1} E\ H_k(\tilde{\Lambda}_k e^{Z_k'\beta} - \Lambda_k e^{Z_k'b}) = 0. \end{aligned} \tag{5}$$

An obvious solution is the trivial one $b = \beta$ and $\Lambda_k = \tilde{\Lambda}_k$. This solution can actually be proven unique by first solving the second equation for Λ_k as a function of b and substituting into the first one. We can then prove that the first equation, which now only has b as the unknown, is the gradient of a concave function in b under some non-degeneracy conditions for the random variables H_k and Z_k, $1 \le k \le K$.

Hence the bridge relation is trivially $s^0(\theta) = \theta$. The naive MLE's of Section 3, obtained by neglecting random effects, give consistent estimates for $\tilde{\Lambda}_k$ and β. Therefore the consistent estimates for the original parameters are just $\hat{\theta} = \hat{s}$, *i.e.* $(\hat{\tilde{\Lambda}}_k, \hat{\beta}) = (\hat{\Lambda}_k, \hat{b})$. Also we notice that the naive MLE's are just the Nelson-Aalen-like estimates for baseline mean measure $\tilde{\Lambda}_k; 1 \le k \le K$, and the partial likelihood estimates for regression parameter β. These hold in general settings where the mean measure defined by Equation (1) of Section 2 is correctly specified; and do not depend on specific probability models such as the conditional Poisson model which we have been considering. Similar ideas have been applied to point

processes in time by Lawless and Nadeau (1995), from the viewpoint of estimating equations and conditional mean function specifications. Incidentally, in Section 7 where we consider covariate measurement error, we will find that the bridge relation is no longer a trivial one.

We now turn to the problem of estimating the asymptotic variances of the consistent estimates $\hat{b}$ and $\hat{\Lambda}_k$. In the next two sections, we will see that the usual asymptotic variance estimates of $\hat{b}$ based on the inverted partial likelihood information matrix is inadequate. In an adjusted analysis using the robust asymptotic variance, the $\mathcal{Z}$-value for assessing the significance of a component of the regression coefficient β can diminish, despite the consistency of the naive estimate $\hat{b}$, since the robust variance can pick up the extra variation from the residual random effects and become larger than the naive variance estimator.

5 Asymptotic Variances

In the previous section we have found that the naive MLE's are consistent for the original parameters. By part (B) of the Proposition, this implies that the double sandwich formula for their asymptotic variance reduces to the usual sandwich formula. First we obtain the naive sample information matrix $n\hat{I} = -\nabla^2 R|_{\hat{s}}$ from the second order derivatives of R. Note that the parameter (row)-vector s' can be partitioned into a K-dimensional sub-vector $\Lambda' = (\Lambda_1, ..., \Lambda_K)$ and a p-dimensional sub-vector b. Inverting the $(K+p) \times (K+p)$ dimensional naive sample information matrix $n\hat{I}$, we obtain the partitioned matrix:

$$(n\hat{I})^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{6}$$

where sub-matrix C_{11} is $K \times K$, $C_{12} = C_{21}'$ is $K \times p$, and C_{22} is $p \times p$. In (6),

$$C_{22} \equiv \hat{\mathcal{I}}^{-1} \text{ where } \hat{\mathcal{I}} = \sum_k (\sum_i H_{ik} Y_{ik})(S_k^{(2)}/S_k^{(0)} - E_k E_k'). \tag{7}$$

where we have used the fairly standard notation:

$$S_k^{(0)} = \sum_i H_{ik} e^{Z_{ik}'\hat{b}}, \;\; S_k^{(1)} = \sum_i H_{ik} Z_{ik} e^{Z_{ik}'\hat{b}}, \;\; S_k^{(2)} = \sum_i H_{ik} Z_{ik} Z_{ik}' e^{Z_{ik}'\hat{b}}, \;\; E_k = S_k^{(1)}/S_k^{(0)}.$$

Of course, $\hat{\mathcal{I}} = -\partial_b \partial_b' \mathcal{L}$ is exactly the same as the sample information matrix obtained by taking second order derivative of the partial log likelihood $\mathcal{L}$ with respect to b. Thus

$$C_{22} \equiv \hat{\mathcal{I}}^{-1} = \widehat{Avar}_N(\hat{b}) \tag{8}$$

is the asymptotic variance matrix used in the naive partial likelihood analysis that ignores residual random effects, which is not robust.

The other sub-matrices in (6), are given by

$$\begin{aligned} (C_{11})_{lk} &= \hat{\Lambda}_k (\sum_i H_{ik} e^{Z_{ik}'\hat{b}})^{-1} \delta_{lk} + \hat{\Lambda}_l E_l' \hat{\mathcal{I}}^{-1} E_k \hat{\Lambda}_k, \;\; l, k = 1, ..., K, \\ (C_{12})_k &= -\hat{\Lambda}_k E_k' \hat{\mathcal{I}}^{-1}, \; (C_{21})_k = (C_{12})_k', \;\; k = 1, \; ... \; , \; K. \end{aligned}$$

and C_{11} is the naive asymptotic variance-covariance estimator for vector $\hat{\Lambda}' \equiv (\hat{\Lambda}_1, ..., \hat{\Lambda}_K)$.

In order to use the sandwich formula, we also need to calculate the sample variance for the naive score function $n\hat{V} = \sum_i (\nabla \rho(W_i; s))(\nabla \rho(W_i; s))'|_{\hat{s}}$. We obtain

$$n\hat{V} = \sum_i \begin{bmatrix} J_{i,11} & J_{i,12} \\ J_{i,21} & J_{i,22} \end{bmatrix} \tag{9}$$

Here the sub-matrices, for $i = 1, \ldots, n$, are given by:

$$\begin{aligned} (J_{i,11})_{l,k} &= H_{il} H_{ik} \hat{\Lambda}_l^{-1} \hat{\Lambda}_k^{-1} (Y_{il} - \hat{\Lambda}_l e^{Z'_{il}\hat{b}})(Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}}), \text{ for } l,\ k\ =\ 1,\ ...\ ,\ K, \\ (J_{i,12})_l &= \sum_k H_{il} \hat{\Lambda}_l^{-1} H_{ik} Z'_{ik} (Y_{il} - \hat{\Lambda}_l e^{Z'_{il}\hat{b}})(Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}}), \text{ for } l = 1, \ldots,\ K, \\ J_{i,22} &= \{\sum_k H_{ik} Z_{ik} (Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}})\}^{\otimes^2}, \end{aligned}$$

and $J_{i,21} = J'_{i,12}$, where the operation $\otimes^2$ of a column vector v denotes $v^{\otimes^2} = vv'$.

By the sandwich formula, we are now able to estimate the robust asymptotic variance by $\widehat{Avar}(\hat{\theta}) = (n\hat{I})^{-1}(n\hat{V})(n\hat{I})^{-1}$. where $\hat{\theta}' = (\hat{\Lambda}', \hat{\beta}') = (\hat{\Lambda}', \hat{b}')$. From (6) and (9), we obtain:

$$\widehat{Avar} \begin{bmatrix} \hat{\Lambda} \\ \hat{b} \end{bmatrix} = \sum_i \{\sum_k H_{ik} (Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}}) \begin{bmatrix} A_{ik} \\ B_{ik} \end{bmatrix}\}^{\otimes^2}$$

where the K-dimensional sub-vectors $\{A_{ik}\}$ are given by

$$(A_{ik})_l = \delta_{lk} (\sum_j H_{jl} e^{Z'_{jl}\hat{b}})^{-1} - \hat{\Lambda}_l E'_l \hat{\mathcal{I}}^{-1} (Z_{ik} - E_k), \quad l = 1, \ldots, K$$

and the p-dimensional sub-vector $(B_{ik}) = \hat{\mathcal{I}}^{-1}(Z_{ik} - E_k)$. Here δ_{lk} is the Kronecker delta.

These equations enable us to calculate the variance estimators for $\hat{\tilde{\Lambda}}_k = \hat{\Lambda}_k$ and $\hat{\beta} = \hat{b}$. We can show that the robust asymptotic variance estimator for $\hat{\beta}$ (or $\hat{b}$) can be put in the sandwich form

$$\widehat{Avar}(\hat{b}) = \hat{\mathcal{I}}^{-1} \hat{\mathcal{V}} \hat{\mathcal{I}}^{-1} \text{ where } \hat{\mathcal{V}} = \sum_i \{\sum_k H_{ik} (Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}})(Z_{ik} - E_k)\}^{\otimes^2}. \tag{10}$$

Here $\mathcal{V}$ is the sample variance estimator for the score function corresponding to the partial log likelihood $\mathcal{L}$.

Consider now the case of processes in time, when $\mathcal{K}$ indexes points on a discrete time axis, and we can define a cumulative intensity function $\tilde{\Lambda}_0(t) = \sum_{k \le t} \tilde{\Lambda}_k$. We can then estimate $\tilde{\Lambda}_0(t)$ by $\hat{\Lambda}_0(t) = \sum_{k \le t} \hat{\tilde{\Lambda}}_k = \sum_{k \le t} \hat{\Lambda}_k$. The asymptotic variance for $\hat{\Lambda}_0(t)$ can then be expressed in terms of those of $\hat{\Lambda}_k$'s. The asymptotic covariance matrix estimator of the vector $(\hat{\Lambda}_0(t), \hat{b}')'$ and $(\hat{\Lambda}_0(s), \hat{b}')'$ is given by:

$$\widehat{Acov}\{\begin{bmatrix} \hat{\Lambda}_0(t) \\ \hat{b} \end{bmatrix}, \begin{bmatrix} \hat{\Lambda}_0(s) \\ \hat{b} \end{bmatrix}\} \equiv \begin{bmatrix} \widehat{Acov}(\hat{\Lambda}_0(t), \hat{\Lambda}_0(s)) & \widehat{Acov}(\hat{\Lambda}_0(t), \hat{b}) \\ \widehat{Acov}(\hat{b}, \hat{\Lambda}_0(s)) & \widehat{Avar}(\hat{b}) \end{bmatrix} = \sum_i G_i(t) G_i(s)' \tag{11}$$

$$\text{where } G_i(t) \equiv \sum_k H_{ik}(Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}}) \begin{bmatrix} \sum_{l \le t}(A_{ik})_l \\ B_{ik} \end{bmatrix}.$$

In particular,

$$\widehat{Avar}(\hat{\Lambda}_0(t)) = \sum_i \{\sum_{l \le t} \sum_k H_{ik}(Y_{ik} - \hat{\Lambda}_k e^{Z'_{ik}\hat{b}})(\delta_{lk}(\sum_j H_{jl} e^{Z'_{jl}\hat{b}})^{-1} - \hat{\Lambda}_l E'_l \hat{\mathcal{I}}^{-1}(Z_{ik} - E_k))\}^{\otimes^2}. \quad (12)$$

On the other hand, the naive asymptotic variance can be obtained by

$$\begin{aligned} \widehat{Avar}_N(\hat{\Lambda}_0(t)) &= \sum_{l \le t} \sum_{k \le t} Acov_N(\hat{\Lambda}_l, \hat{\Lambda}_k) = \sum_{l \le t} \sum_{k \le t} (C_{11})_{lk} \\ &= \sum_{k \le t} \hat{\Lambda}_k (\sum_i H_{ik} e^{Z'_{ik}\hat{b}})^{-1} + \sum_{l \le t} \sum_{k \le t} \hat{\Lambda}_l E'_l \hat{\mathcal{I}}^{-1} E_k \hat{\Lambda}_k. \end{aligned} \quad (13)$$

In the continuous limit (13) is consistent with the formula on Page 505 of Andersen et al. (1993). When there are residual random effects, however, we should use the robust variance estimator in (12).

Now that we have obtained the asymptotic variance estimators, we can compute confidence intervals and test statistics for $\hat{\beta} = \hat{b}$ and $\hat{\tilde{\Lambda}}_0(t) = \hat{\Lambda}_0(t)$, based on the asymptotic normality of the MLE's. As would be expected, the robust variances turn out to be larger than the naive ones because the former account for the extra uncertainty induced by the random effects. This in turn leads to wider confidence intervals and larger P-values (see *e.g.* Jiang 1996, Page 128-30). In the next section we examine an example.

6 Example of Variance Inflation

We will compare the naive and robust asymptotic variances of $\hat{b}$ constructed from $p \times p$ matrices $\mathcal{I}$ and $\mathcal{V}$, which are, respectively, the consistent limits of $\hat{\mathcal{I}}$ and $\hat{\mathcal{V}}$ as given by (7) and (10). Let b and Λ_k, $1 \le k \le K$ denote those values that maximize $E\rho$ where ρ is given by (2). is just at $b = \beta$ and $\Lambda_k = \tilde{\Lambda}_k$, We obtain

$$\mathcal{I} = \sum_k (EY_k)\{\frac{EZ_k Z'_k H_k e^{Z'_k b}}{EH_k e^{Z'_k b}} - (\frac{EZ_k H_k e^{Z'_k b}}{EH_k e^{Z'_k b}})^{\otimes^2}\} \quad (14)$$

$$\text{and } \mathcal{V} = E\{\sum_k (Y_k - H_k \Lambda_k e^{Z'_k b})(Z_k - \frac{EZ_k H_k e^{Z'_k b}}{EH_k e^{Z'_k b}})\}^{\otimes^2}. \quad (15)$$

The robust asymptotic variance of $\hat{\beta}$ $(= \hat{b})$ can be put in the sandwich form $Avar(\hat{\beta}) = \frac{1}{n}\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1}$.

The same result can be derived by a rigorous calculation of $Avar(\hat{\beta})$ (instead of its estimate), which can be obtained by repeating the matrix algebras in Section 5, working on the limiting function $E\rho$ instead of $n^{-1}R$. Here we would calculate I, I^{-1}, V, and $I^{-1}VI^{-1}$, instead of $\hat{I}$, $\hat{I}^{-1}$, $\hat{V}$, and $\hat{I}^{-1}\hat{V}\hat{I}^{-1}$ as we did in Section 5. $Avar(\hat{\beta})$ would be obtained by the sub-matrix of $n^{-1}I^{-1}VI^{-1}$ restricted to the β dimensions, which would become $\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1}$, instead of $\hat{\mathcal{I}}^{-1}\hat{\mathcal{V}}\hat{\mathcal{I}}^{-1}$.

An alternative expression for $\mathcal{I}$ is

$$\mathcal{I} = E\{\sum_k H_k \Lambda_k e^{Z_k' b}(Z_k - \frac{EZ_k H_k e^{Z_k' b}}{EH_k e^{Z_k' b}})^{\otimes^2}\}. \tag{16}$$

Here we have used the fact that $EY_k = EH_k\Phi_k = EH_k\psi_k\Lambda_k e^{Z_k' b} = \Lambda_k EH_k e^{Z_k' b}$, since $E(\psi_k|H_k'\text{s},\ Z_k'\text{s}) = 1$. Expanding the direct product $(\ )^{\otimes^2}$ in the second expression shows that it is equivalent to (14).

Let us now consider the case when the predictor Z_k is constant, $Z_k = Z$, and Z is statistically independent of H_k. This may be assumed, for example, when Z is a treatment assignment indicator in a randomized clinical trial. In this case, (16) leads to

$$\mathcal{I} = E\{\sum_k E(Y_k|H,Z)(Z - \frac{EZe^{Z'b}}{Ee^{Z'b}})^{\otimes^2}\},$$

using $E(Y_k|H,Z) = E(H_k\psi_k\Lambda_k e^{Z_k' b}|H,Z) = H_k\Lambda_k e^{Z_k' b}$. can express $\mathcal{I}$ as

$$\mathcal{I} = E\{E(Y^+|Z,H)(Z - \frac{EZe^{Z'b}}{Ee^{Z'b}})^{\otimes^2}\} \tag{17}$$

where $Y^+ \equiv \sum_k Y_k$ is the total number of events observed for a system. Similarly, from (15) we obtain

$$\mathcal{V} = E\{[Y^+ - E(Y^+|H,Z)]^2(Z - \frac{EZe^{Z'b}}{Ee^{Z'b}})^{\otimes^2}\} = E\{Var(Y^+|Z,H)(Z - \frac{EZe^{Z'b}}{Ee^{Z'b}})^{\otimes^2}\}. \tag{18}$$

Note that, if there were no random effect, Y^+ would have a Poisson distribution conditional on Z and H, since it is a sum of conditionally independent Poisson random variables. Therefore the conditional variance $Var(Y^+|Z,H) = E(Y^+|Z,H)$ the conditional mean. Hence $\mathcal{I} = \mathcal{V}$, and the sandwich formula is reduced to $\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1} = (\mathcal{I})^{-1}$.

However when there exist residual random effects, the extra variation from the ψ_k's makes $Var(Y^+|Z,H) > E(Y^+|Z,H)$. Suppose $\psi_k = \psi$, which is constant for a system (independent of k), then $Y^+|H,Z,\psi \sim Poisson(\psi\Lambda^+ e^{Z'b})$ where $\Lambda^+ \equiv \sum_k H_k\Lambda_k$. By first conditioning on (H,Z,ψ), we obtain $E(Y^+|Z,H) = \Lambda^+ e^{Z'b}$ and $Var(Y^+|Z,H) = \Lambda^+ e^{Z'b} + \kappa(\Lambda^+ e^{Z'b})^2$ where $\kappa = Var(\psi|H,Z)$. Therefore

$$Var(Y^+|Z,H) - E(Y^+|Z,H) = \kappa(\Lambda^+ e^{Z'b})^2 > 0.$$

Hence

$$(\mathcal{V} - \mathcal{I}) = E\{\kappa(\Lambda^+ e^{Z'b})^2(Z - \frac{EZe^{Z'b}}{Ee^{Z'b}})^{\otimes^2}\}$$

is positive definite, and so is $\frac{1}{n}\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1} - (n\mathcal{I})^{-1} = \frac{1}{n}\mathcal{I}^{-1}(\mathcal{V} - \mathcal{I})(\mathcal{I}^{-1})'$. Therefore, for any linear combination of the components of $\hat{b}$, its robust variance is greater than the naive variance based on $\mathcal{I}^{-1}$. For example, for a component b_j, we have

$$Avar(\hat{b}_j) - Avar_N(\hat{b}_j) = [n^{-1}\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1} - (n\mathcal{I})^{-1}]_{jj} > 0$$

where the $[\ ]_{jj}$ denotes the jth diagonal element of a matrix. Obviously the increment is proportional to κ, the extra variation introduced by random mixing.

Let us use these results to derive the asymptotic variance of treatment effect estimator $\hat{\beta}$ in a randomized clinical trial, where Z is a scalar taking value 0 or 1 with equal probability 0.5. We follow patients (systems) over time and use the time of randomization to treatment assignment as the origin of the time axis. Then the information contained in $\{H_k\}$ is the same as the follow-up time T for a patient. In this case, $\Lambda^+ = \Lambda_0(T)$, which is the baseline cumulative intensity (*i.e.* for a patient with $Z = 0$) at end of the follow-up period. We treat T as a random variable. Now $\Lambda^+ = \Lambda_0(T)$ leads to $E(Y^+|Z,H) = \Lambda_0(T)e^{Z'b}$ and $Var(Y^+|Z,H) = \Lambda_0(T)e^{Z'b} + \kappa(\Lambda_0(T)e^{Z'b})^2$. Substituting into (17) and (18), we are left with expectations taken over T and Z only. Taking expectation with respect to Z first, noting that Z is independent of T and takes value 0 or 1 with equal probability, we find

$$\mathcal{I} = \frac{1}{2}E\Lambda_0(T)(\frac{e^\beta}{1+e^\beta}) \text{ and } \mathcal{V} = \frac{1}{2}E\Lambda_0(T)(\frac{e^\beta}{1+e^\beta}) + \kappa E\{(\Lambda_0(T))^2\}(\frac{e^\beta}{1+e^\beta})^2.$$

The naive asymptotic variance of $\hat{\beta}$ can then be obtained as

$$Avar_N(\hat{\beta}) = (n\mathcal{I})^{-1} = \frac{2}{n}((E\Lambda_0(T)e^\beta)^{-1} + (E\Lambda_0(T))^{-1}) \tag{19}$$

present). The robust asymptotic variance, on the other hand, is obtained by $\frac{1}{n}\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1}$ as

$$Avar(\hat{\beta}) = \frac{2}{n}\{((E\Lambda_0(T)e^\beta)^{-1} + (E\Lambda_0(T))^{-1}) + 2\kappa E\Lambda_0^2(T)(E\Lambda_0(T))^{-2}\}. \tag{20}$$

The inflation of $Avar(\hat{\beta})$ due to extra variance κ leads to an increased sample size in study planning, comparing to the usual Poisson process regression. Consider the problem of testing $H_0 : \beta = 0$ vs $H_1 : \beta = \pm\delta$, where δ is a medically significant difference. In order for the Wald test to achieve specified type I error rate α_I and type II error rate α_{II}, the total sample size (including both treatment groups) should be

$$n \approx 2(z_{\alpha_I/2} + z_{\alpha_{II}})^2\delta^{-2}\{((E\Lambda_0(T)e^\beta)^{-1} + (E\Lambda_0(T))^{-1}) + 2\kappa E\Lambda_0^2(T)(E\Lambda_0(T))^{-2}\} \tag{21}$$

based on the robust asymptotic variance in (20) instead of

$$n_{Naive} \approx 2(z_{\alpha_I/2} + z_{\alpha_{II}})^2\delta^{-2}\{(E\Lambda_0(T)e^\beta)^{-1} + (E\Lambda_0(T))^{-1}\}$$

based on (19), which happens to be the sample size formula obtained from using the pure Poisson regression model ($\kappa = 0$).

Now $E\Lambda_0(T)e^\beta$ and $E\Lambda_0(T)$ are the expected number of events for a patient in treatment group ($Z = 1$) and the group ($Z = 0$) respectively. If we can make the assumption that the point process is approximately homogeneous in time, these become r_1ET and r_0ET, respectively, where $r_1 = r_0e^\beta$, for some $r_0 > 0$. estimate, assuming that T's are approximately equal for all (21) becomes

$$n \approx 2(z_{\alpha_I/2} + z_{\alpha_{II}})^2\delta^{-2}\{((r_1ET)^{-1} + (r_0ET)^{-1}) + 2\kappa(1 + Var(T)(ET)^{-2})\}.$$

The parameters r_0, r_1, κ are unknown but may be estimated if results from a pilot study are available. The parameters r_0, r_1, κ in a pilot sample can be estimated by the method of, say, negative binomial regression (Abu-Libdeh et al. 1990). Alternatively, we can estimate them by the method of moments as $r_0 \approx \sum_{i:Z_i=0} Y_i^+ / \sum_{i:Z_i=0} T_i$ and analogously for r_1. A method of moments estimate for κ can be obtained as

$$\kappa \approx \frac{\sum_i ((Y_i^+)^2 - Y_i^+))}{\sum_{i\in(1)}(r_1 T_i)^2 + \sum_{i\in(0)}(r_0 T_i)^2} - 1.$$

A similar approach to the planning the study duration for recurrent events data is taken by Cook (1995), based on a model of constant recruiting rate and exponential censoring.

7 Covariates Measured with Error

Suppose the observed covariate Z is a surrogate for a true covariate, X, which is not observed. Suppose Z, X and the random effect ψ are all constant across different patches of any one system, and the distribution of $(Z, X, \psi, \{H_k\}_{k\in\mathcal{K}})$ does not involve the parameter of interest β. Now a "model" consists of two parts – (a) a response model which relates the outcome variables $\{Y_k; k \in \mathcal{K}\}$ to X, ψ and the H_k's, and (b) a covariate error structure model which relates the observed covariate Z to the true one X.

To describe the part of the model that relates to the covariate error structure, we first need to partition the true covariate vector as $X' = (\tilde{X}', A')$. Here the components represent those covariates $(\tilde{X})$ measured with error, and those covariates (A) measured without error, which are of dimensions r, q, respectively, say. We correspondingly partition the observed surrogate variable vector as $Z' = (\tilde{Z}', A')$. A covariate error structure model is one which specifies the joint distribution of X and Z. In our application in Section 8, we shall use a simple normal additive error model (NADD), where $\tilde{Z} = \tilde{X} + U$, $\tilde{X} \sim N(0, \Sigma_{\tilde{X}})$, $U \sim N(0, \Sigma_U)$ and are independent. Both $\tilde{X}$ and U are assumed to be independent of A which is measured without error. This assumption might be valid if, for example, A is treatment assignment in a randomized trial. It is worth noting that this implies the conditional distribution:

$$\tilde{X} \mid Z \quad \sim \quad N(\Omega'\tilde{Z},\ \Sigma) \quad \text{where } \Omega = \Sigma_{\tilde{Z}}^{-1}\Sigma_{\tilde{X}}, \quad \Sigma = \Sigma_U \Sigma_{\tilde{Z}}^{-1}\Sigma_{\tilde{X}}, \quad \Sigma_{\tilde{Z}} = \Sigma_{\tilde{X}} + \Sigma_U.$$

We will call the matrix Ω the "attenuation" matrix and it plays an important role. With little added effort, it is possible to apply the techniques we describe to more general models for the joint distribution of X and Z — for example to the conditional normal (CN) model in which the conditional distribution of $\tilde{X}$ given Z is normal with mean a linear function of Z and constant variance: *i.e.* $\tilde{X} \mid Z \quad \sim \quad N(C_0' + C'Z,\ \Sigma)$ where $C'Z = \Omega'\tilde{Z} + C_A' A$ for general vector C_0 and general matrices C, Σ, C_A, and Ω. Clearly the NADD model is a special case, in which, in particular, $C_0 = 0$.

For the response model we take the conditional Poisson specification:

$$Y_k | H, X, \psi \sim \text{Poisson}(\psi H_k \tilde{\Lambda}_k e^{X'\beta}). \tag{22}$$

We can make inference on β and $\tilde{\Lambda}_k$ by using the recipe introduced in Section 4. That is: we start by performing a naive analysis ignoring the presence of random effects and

measurement error, and then adjust the resulting estimates and variances to account for the misspecification. The naive log-likelihood function is constructed by taking ψ and X in model as 1 and Z, respectively, in the log-likelihood function derived from (22). The resulting R and ρ have exactly the same form as in Section 3. However the asymptotic naive score equation (4) now leads to

$$\partial_b E\rho = \sum_k E\ H_k Z_k (Y_k - \Lambda_k e^{Z'b}) = \sum_k E\ H_k Z_k (\tilde{\Lambda}_k E(e^{X'\beta}|Z) - \Lambda_k e^{Z_k' b}) = 0 \qquad (23)$$

and

$$\partial_{\Lambda_k} E\rho = \Lambda_k^{-1} E\ H_k (Y_k - \Lambda_k e^{Z'b}) = \Lambda_k^{-1} E\ H_k (\tilde{\Lambda}_k E(e^{X'\beta}|Z) - \Lambda_k e^{Z_k' b}) = 0, \quad \text{for} \quad k \in \mathcal{K}. \qquad (24)$$

Here we have assumed that the follow-up process $\{H_k\}$ is noninformative for the covariate error structure, in the sense that the conditional distribution of X given Z and H depends on Z only. We partition the regression parameter by $\beta' = (\tilde{\beta}', \gamma')$, and the asymptotic limit of the naive MLE $b' = (\tilde{b}', g')$, corresponding to the partition $Z' = (\tilde{Z}', A')$. Suppose that the measurement error model is specified as for (CN) above. We can solve the asymptotic naive score equations, (23) and (24) by noting that

$$E(e^{X'\beta}|Z) = E(e^{\tilde{X}'\tilde{\beta} + A' C_A}|Z) = \exp\{(C_0\tilde{\beta} + \frac{1}{2}\tilde{\beta}'\Sigma\tilde{\beta}) + \tilde{Z}'(\Omega\tilde{\beta}) + A'(\gamma + C_A\tilde{\beta})\},$$

which comes from the conditional normality of $\tilde{X}$ given Z. A solution can be easily read off as

$$g = \gamma + C_A\tilde{\beta}, \quad \tilde{b} = \Omega\tilde{\beta}, \quad \Lambda_k = \tilde{\Lambda}_k \exp\{C_0\tilde{\beta} + \frac{1}{2}\tilde{\beta}'\Sigma\tilde{\beta}\}. \qquad (25)$$

The uniqueness of this solution can be demonstrated in a similar way as earlier for the solution of equation (5). Notice that when measurement error is present, the naive MLE's are no longer consistent for the original parameters. When $\tilde{Z}$ is one-dimensional and we use the NADD model, we have $\Omega = \Sigma_{\tilde{X}}/(\Sigma_{\tilde{X}} + \Sigma_U) < 1$ and the second equation in (25) exhibits the feature of 'attenuation'.

The inverted relations of (25) are

$$\gamma = g - C_A\Omega^{-1}\tilde{b}, \quad \tilde{\beta} = \Omega^{-1}\tilde{b}, \quad \tilde{\Lambda}_k = \Lambda_k \exp\{-C_0\Omega^{-1}\tilde{b} - \frac{1}{2}\tilde{b}'\Pi\tilde{b}\}, \qquad (26)$$

where $\Pi \equiv (\Omega^{-1})'\Sigma\Omega^{-1}$. For processes in time, when $\mathcal{K}$ indexes points on a discrete time axis, the third equations in (25) and (26) lead to the following relation for the cumulative baseline hazard $\tilde{\Lambda}_0(t) = \sum_{k \le t} \tilde{\Lambda}_k$:

$$\Lambda_0(t) = \tilde{\Lambda}_0(t) \exp\{C_0\tilde{\beta} + \frac{1}{2}\tilde{\beta}'\Sigma\tilde{\beta}\} \text{ and } \tilde{\Lambda}_0(t) = \Lambda_0(t) \exp\{-C_0\Omega^{-1}\tilde{b} - \frac{1}{2}\tilde{b}'\Pi\tilde{b}\}. \qquad (27)$$

The equations (26) and (27) tell us how to form consistent estimators for the parameters of interest from the naive MLE's, namely the usual partial likelihood estimates for the regression coefficients and the Nelson-Aalen estimate for the baseline cumulative intensity.

Naive calculation of the asymptotic variance of the naive MLE $\hat{b}$ based on $(n\mathcal{I})^{-1}$ can be proven smaller than the adjusted one based on $\frac{1}{n}\mathcal{I}^{-1}\mathcal{V}\mathcal{I}^{-1}$. Now, an argument similar to that in Section 6 leads to

$$(\mathcal{V}-\mathcal{I}) = E\{(\kappa(\Lambda^+)^2 E(e^{2X'\beta}|Z) + (\Lambda^+)^2 Var(e^{X'\beta}|Z))(Z - \frac{EZe^{Z'b}}{Ee^{Z'b}})^{\otimes^2}\}$$

which is positive definite. Hence, as before, for any linear combination of the components of $\hat{b}$, the robust variance is greater than the naive variance based on $\mathcal{I}^{-1}$. Now the increment contains two terms: one term proportional to κ due to random mixing, as well as another proportional to $Var(e^{X'\beta}|Z)$ contributed by the measurement error in Z. In particular, when $b=(\tilde{b},g)$ consists of two scalars, we have

$$Avar(\hat{b}) > Avar_N(\hat{b}) \quad \text{and} \quad Avar(\hat{g}) > Avar_N(\hat{g}). \tag{28}$$

Since the naive MLE's are not consistent for $(\tilde{\Lambda}_k,\beta)$, we must use the double-sandwich formula to estimate the asymptotic variances of the adjusted estimates $\hat{\tilde{\Lambda}}_k$ and $\hat{\beta}$ in general. However when A is the treatment variable in a randomized clinical trial, in which $\tilde{X}$ is assumed independent of A conditional on $\tilde{Z}$, we have $C_A=0$ and $\gamma=g$. Therefore the asymptotic variances of particular interest are $Avar(\hat{\gamma})=Avar(\hat{g})$ and $Avar(\hat{\tilde{\beta}}) = \Omega^{-1}Avar(\hat{\tilde{b}})(\Omega^{-1})'$. (These can be obtained either by directly using the Δ-method or formally calculating the D-matrix in Section 4). The asymptotic variance of $\hat{\tilde{\Lambda}}_0(t)$ can also be obtained from (27) as

$$\begin{aligned} Avar(\hat{\tilde{\Lambda}}_0(t)) &= e^{-\tilde{b}'\Pi\tilde{b}}\{Avar(\hat{\Lambda}_0(t)) + \Lambda_0^2(t)\tilde{b}'\Pi'(Avar(\hat{\tilde{b}}))\Pi\tilde{b} \\ &\quad -\Lambda_0(t)Acov(\hat{\Lambda}_0(t),\hat{\tilde{b}})\Pi\tilde{b} - \Lambda_0(t)\tilde{b}'\Pi'Acov(\hat{\tilde{b}},\hat{\Lambda}_0(t))\}. \end{aligned} \tag{29}$$

In the above expressions, the asymptotic variances and covariances for the naive estimates can be obtained from (11), and $\tilde{b}$ and $\Lambda_0(t)$ can be estimated by the naive MLE's, $\hat{\tilde{b}}$ and $\hat{\Lambda}_0(t)$, respectively.

Consider again the case where γ and $\tilde{\beta}$ are scalars. For inference on γ, asymptotically, the magnitude of the adjusted $\mathcal{Z}$ value

$$|\mathcal{Z}_{Adj.}| = \frac{|\hat{\gamma}|}{\sqrt{Avar(\hat{\gamma})}} = \frac{|\hat{g}|}{\sqrt{Avar(\hat{g})}} < \frac{|\hat{g}|}{\sqrt{Avar_N(\hat{g})}} = |\mathcal{Z}_N|$$

even though the adjusted estimate is the same as the naive one, due to the inflation of the adjusted asymptotic variance. For inference on $\tilde{\beta}$, the regression coefficient for the predictor measured with error, asymptotically

$$|\mathcal{Z}_{Adj.}| = \frac{|\hat{\tilde{\beta}}|}{\sqrt{Avar(\hat{\tilde{\beta}})}} = \frac{|\Omega^{-1}\hat{\tilde{b}}|}{\sqrt{Avar(\Omega^{-1}\hat{\tilde{b}})}} = \frac{|\hat{\tilde{b}}|}{\sqrt{Avar(\hat{\tilde{b}})}} < \frac{|\hat{\tilde{b}}|}{\sqrt{Avar_N(\hat{\tilde{b}})}} = |\mathcal{Z}_N|$$

by (28), even though the adjusted estimate $\hat{\tilde{\beta}}$ is greater in magnitude then the naive one, since the adjusted asymptotic variance more than compensates for this.

In summary, we find that the usual partial likelihood procedure will give a consistent estimate for the treatment effect γ. However its naive asymptotic variance from $\hat{\mathcal{I}}^{-1}$ tends to be an underestimate. When the robust asymptotic variance is used instead, a smaller $\mathcal{Z}$-value for the Wald statistic will usually result. Estimates of regression coefficients for variables measured with error are attenuated. All these features are exhibited in the application in the next section.

8 Illustration: Skin cancer recurrences in a clinical trial

Clark et al. (1996) have reported on the results of the "Nutritional Prevention of Cancer" (NPC) trial which studied the long-term safety and efficacy of a daily 200μg nutritional supplement of selenium (Se) for the prevention of cancer. This was a double-blind, placebo-controlled randomized clinical trial with 1312 patients accrued since it started in 1983. Each patient was randomly assigned to either the selenium (coded as A=1) or placebo (A=0) group. A number of endpoints were considered, but here we shall concentrate on one of the primary endpoints — namely squamous cell carcinoma (SCC) of the skin. The results for this endpoint are of particular interest because Clark et al. (1996) found a negative (but not statistically significant, P = 0.15) effect of selenium supplementation. This was opposite to previous expectations, and contrasted sharply with findings of highly significant positive benefits of the selenium supplementation in preventing a number of other types of cancers. However, for the SCC endpoint, the original analysis presented in Clark et al. (1996) considered only the time to *first* occurrence of an SCC in each subject. Also, neither random effects nor effects of covariate measurement error were taken into account. In this section, we reanalyze the data to illustrate the methods we have discussed.

For each patient, the time (measured from date of randomization) of each new occurrence of an SCC was observed. At randomization, a number of baseline covariates were also recorded. Of course, the most important of these was the treatment assignment (Se or placebo), but others included such variables as age, clinic, gender, smoking status, previous history of skin cancer, and blood biochemical levels, in particular plasma Se status. While some of these variables are recorded accurately, others, such as plasma Se status, are subject to measurement error. Even with these predictors, it is likely that there are still remaining factors influencing the study outcomes of a patient. We ascribe them to the residual random effect, with respect to a collection of predictors taken into the model. As our first analysis, we will take the $\{0,1\}$-valued treatment assignment indicator A as the only predictor: $Z = A$. We will see how the robust inference based on naive MLE's can be performed for this point process regression model on Z with residual random effects.

8.1 Random Effects

The data set for the NPC trial contained records on 1312 patients, with the longest follow-up period being 4618 days. We take the set of patches as days: $\mathcal{K} = \{1, 2, 3, \ldots, 4618\}$. The data set contains the elapsed time in days of each SCC occurrence, counted from the date

of randomization. For example, a patient with ID number 1001557 had SCC occurrences on Day 178, Day 286, Day 549, and Day 1018. This patient was followed until Day 2268 ("censoring day"), and was in Se supplemented group (with $Z = A = 1$, constant in time). In the formalism of Section 2, the data for this patient can be summarized as, labeling him as the ith patient, $Y_{i1} = 0$,, $Y_{i2} = 0$, ..., $Y_{i,178} = 1$, $Y_{i,179} = 0$, ... , $Y_{i,286} = 1$, $Y_{i,287} = 0$, ... , $Y_{i,549} = 1$, $Y_{i,550} = 0$, ... , $Y_{i,1018} = 1$, $Y_{i,1019} = 0$, ... , $Y_{i,4618} = 0$; $Z_{i,1} = Z_{i,2} = ... = Z_{i,4618} = 1$; $H_{i,1} = H_{i,2} = ... = H_{i,2268} = 1$, $H_{i,2269} = ... = H_{i,4618} = 0$. We can form the naive log-likelihood function R in Section 3 and calculate the naive MLE's $\hat{b}$ (here it only has one component $\hat{g}$) and $\hat{\Lambda}_0(t) = \sum_{k=1}^{t} \hat{\Lambda}_k$ for $t = 1, 2, ..., 4618$, by using the partial likelihood estimator and Nelson-Aalen estimator as in Section 3. The results in Section 4 imply these naive MLE's are consistent for the original parameters β (here containing only the one component γ, the regression coefficient for the treatment assignment indicator) and $\tilde{\Lambda}_0(t)$. The naive asymptotic variance estimates of the naive MLE's using $(n\hat{I})^{-1}$ are obtained from (8) and (13). The robust variances using the sandwich formulae come from (10) and (12). The results are summarized as Table 1 and Figure 1.

Table 1: Naive and adjusted analyses of NPC trial SCC data with treatment as the only predictor: Point estimates and $\mathcal{Z}$-values

	Estimate	Standard Error	$\mathcal{Z}$-value
(Naive)	$\hat{g} = 0.1180$	0.0586	2.0137
(Adjusted)	$\hat{\gamma} = 0.1180$	0.1237	0.9539

Note: Naive = ignoring random effects; Adjusted = adjusted for random effects.

From Table 1, we note that the naive analysis using multiple recurrences ascribes a significant adverse effect of Se supplementation. (This may be compared with the analysis reported in Clark et al. (1996, Table 2). They reported a larger log relative risk, namely $\log(1.14) = 0.131$, but this is less significant (logrank test P = 0.15), it being less sensitive because the data used was only the time to first tumor. Table 1 also shows that when the extra variability due to patient heterogeneity is accounted for, the point estimate of the treatment effect (log relative risk) is unchanged but its significance is greatly diminished.

[Figure 1 about here.]

Figure 1 illustrates the estimates of the baseline cumulative intensity, which has now the interpretation of the expected number of SCC's increasing over time (in days), for a typical placebo group patient. Notice that $\hat{\tilde{\Lambda}}_0(t) = \hat{\Lambda}_0(t)$, but the naive analysis yields pointwise 95% confidence bands that are much narrower than those based on the robust variance formula (12).

Figure 1 also suggests that a parametric model with a constant intensity rate may be appropriate, where it is assumed that $\tilde{\Lambda}_0(t) = t \exp(\beta_0)$, as in Turnbull et al. (1997). Under this model, the naive MLE's $(\hat{b}_0, \hat{g})$ obtained from neglecting random effects are again consistent, and are same as the adjusted estimates $(\hat{\beta}_0, \hat{\gamma})$. The reduction of $\mathcal{Z}$-values in the adjusted analysis also takes place. The results are summarized in Table 2.

Table 2: Naive and adjusted analyses of NPC trial SCC data with treatment as the only predictor: Point estimates and $\mathcal{Z}$-values (parametric constant intensity model)

	Estimate	Standard Error	$\mathcal{Z}$-value
(Naive)	$\hat{g} = 0.1239$	0.0586	2.1143
	$\hat{b}_0 = 1.1320$	0.0429	26.3780
(Adjusted)	$\hat{\gamma} = 0.1239$	0.1237	1.0018
	$\hat{\beta}_0 = 1.1320$	0.0919	12.3226

Note: Naive = ignoring random effects; Adjusted = adjusted for random effects.

8.2 A Covariate Measured with Error

As well as treatment indicator A, we now also include in our model as a covariate the log baseline serum plasma selenium level Z_1. This variable was missing for 26 patients and so we only include the remaining 1286 patients in the following analyses. The variable Z_1 is measured with error, with its hypothetical true value X_1 unknown. We use the simple normal additive model (NADD), whereby $Z_1 = X_1 + U$, as introduced in Section 7. For the parameters of this model we use values $\Sigma_X = 0.106^2$, $\Sigma_U = 0.151^2$. These were obtained from an internal validation data set based on replicate plasma Se measurements in placebo patients. The same data set also validated assumptions of the NADD model. The details are given in Turnbull et al. (1997), but their estimates differ slightly from ours since they used an earlier and somewhat smaller interim data set. From these values, we obtain the attenuation factor of $\Omega^{-1} = 3.01$. In the formalism of Section 7, $\tilde{Z} = Z_1$, $\tilde{X} = X_1$, $\tilde{b} = b_1$ and $\tilde{\beta} = \beta_1$ and are all scalars. We may take Z_1 and X_1 to have zero means by subtracting suitable constants.

According to procedure of Section 7, we start with an naive analysis based on Poisson process with no random effects, and using surrogate Z_1 in place of the true covariate X_1. We obtain the naive MLE's $\hat{g}$, $\hat{b}_1$ and $\hat{\Lambda}_0(t)$ from the partial likelihood estimates and the Nelson-Aalen estimates from Section 3, and calculate their naive asymptotic variances from (8) and (13) based on $(nI)^{-1}$. Because A the treatment assignment variable is randomized and thus independent of X_1, we see from Section 7 the naive MLE $\hat{g}$ for the treatment effect γ is consistent and equal to the adjusted estimator $\hat{\gamma}$. However, the naive MLEs of $\hat{b}_1$ of β_1 and $\hat{\Lambda}_0(t)$ of $\tilde{\Lambda}_0(t)$ are both inconsistent and must be adjusted using (26) and (27). (In fact, $\hat{\Lambda}_0(t)$ overestimates $\tilde{\Lambda}_0(t)$, as can be seen from (27) taking $C_0 = 0$ as implied by the NADD model.) The naive and adjusted estimates of the regression coefficients (log relative risks) along with standard errors based on naive and robust variances are given in Tables 3 and 4. As before, the point estimate of log relative risk for treatment $\hat{\gamma}$ is unchanged, but its significance is greatly diminished when the robust standard error is used. The log relative risk for log baseline plasma Se level increases in magnitude when the adjusted estimate is used, but so does the standard error. However, even recognizing the increased uncertainty in a model with random effects and measurement error, it remains a highly significant variable, suggesting that high plasma Se levels are associated with lowered risk of SCC. The results

in Tables 3 and 4 exhibit the typical features summarized at the end of Section 7.

Table 3: Naive analysis of NPC trial SCC data with treatment A and log baseline Selenium level X_1 as predictors, ignoring random effects and measurement error in X_1 : Maximum likelihood estimates and $\mathcal{Z}$-values

		Estimate	$\sqrt{Avar}_N$	Z_N
(Treatment)	g	0.1169	0.0592	1.9747
(Baseline Se)	b_1	-0.6897	0.1456	-4.7370

Table 4: Analysis of NPC trial SCC data with treatment A and log baseline Selenium level X_1 as predictors, adjusted for measurement error and random effects: Point estimates and $\mathcal{Z}$-values

		Estimate	$\sqrt{Avar}$	$\mathrm{Z}_{Adj.}$
(Treatment)	γ	0.1169	0.1253	0.9330
(Baseline Se)	β_1	-2.0760	0.9632	-2.1553

[Figure 2 about here.]

Figure 2 shows the naive and adjusted estimates of the cumulative intensity function. The naive MLE $\hat{\Lambda}_0(t)$ is larger than the adjusted consistent estimator $\hat{\tilde{\Lambda}}_0(t)$ as stated in Section 7 although the difference is extremely small and hard to distinguish. This is because the ratio of the two estimators given by $e^{\frac{1}{2}(\hat{b}_1)'(\Omega^{-1})'\Sigma\Omega^{-1}\hat{b}_1} = 1.0164$ is very close to one. In general, when this happens, the influence of covariate measurement error on the baseline cumulative hazard estimator is negligible. In Figure 2 we have also plotted the 95% pointwise confidence bands for $\hat{\Lambda}_0(t)$ based on the naive asymptotic variance obtained from (13) based on $(nI)^{-1}$, as well as bands based on the robust asymptotic variance (29) for the consistent estimator $\hat{\tilde{\Lambda}}_0(t)$. The plots indicate that although the naive estimates and adjusted estimates are very close, the robust confidence band is much wider than that obtained from the naive analysis.

A parametric model of constant intensity rate $\tilde{\Lambda}_0(t) = t\exp(\beta_0)$, as may be suggested by Figure 2, yields much the same results. The results, corresponding to those in Tables 3 and 4 for the semiparametric model, are given in Tables 5 and 6. The results in these two tables would be the same as Tables 1 and 2 of Turnbull et al. (1997) but differ somewhat because their results were based on an earlier interim data set.

It should be noted that Tables 4 and 6 ignore the uncertainty in the attenuation factor Ω and treat it as known. In fact, it was estimated from the internal validation data as described

Table 5: Naive analysis of NPC trial SCC data with treatment A and log baseline Selenium level X_1 as predictors, ignoring random effects and measurement error in X_1 : Maximum likelihood estimates and $\mathcal{Z}$-values (constant intensity model)

		Estimate	$\sqrt{Avar}_N$	Z_N
(Treatment)	g	0.1218	0.0592	2.0581
(Baseline Se)	b_1	-0.7245	0.1454	-4.9825
	b_0	1.1226	0.0435	25.8344

Table 6: Analysis of NPC trial SCC data with treatment A and log baseline Selenium level X_1 as predictors, adjusted for measurement error and random effects: Point estimates and $\mathcal{Z}$-values (constant intensity model)

		Estimate	$\sqrt{Avar}$	$\mathrm{Z}_{Adj.}$
(Treatment)	γ	0.1218	0.1253	0.9717
(Baseline Se)	β_1	-2.1807	0.9633	-2.2638
	β_0	1.1047	0.0924	11.9521

above. Using that data, application of the delta method leads to a standard error for $\hat{\Omega}^{-1}$ of 0.46. The sensitivity of the qualitative conclusions can be examined by repeating the above analyses with Ω^{-1} set equal to $\hat{\Omega}^{-1} \pm s.e.$, for example. However the $\hat{\gamma}$, the estimate of log relative risk for treatment and its robust standard error will be unchanged. Also, because the robust standard error adjusts by the same proportion as the parameter estimate, the $\mathcal{Z}$-values in Tables 4 and 6 for β_1 will also be unchanged.

9 Discussion

The consistency of the maximum partial likelihood estimators of regression coefficients and Nelson-Aalen-type estimators of the baseline intensity function ultimately comes from the mean measure specification (1). There are two issues that arise when considering this model.

The first issue concerns the requirement $E(\tilde{Y}_k|Z,H) = E(\tilde{Y}_k|Z)$ implied by (1). For recurrent events data, such as for the example of the NPC trial of the previous section, this would be satisfied if the length of followup or observation time for each subject is independent of the event process of that subject. However certain situations are excluded, such as for example when: (i) a subject is more likely to leave the study earlier if he has a higher frequency of events (here SCCs); or (ii) a subject is withdrawn from the study as soon as he has experienced a fixed number, r say, of events. In such a case, subjects with higher frailties (higher event rates) would be less likely to be still at risk at later time periods, resulting in an underestimate of the intensity function there. Lawless and Nadeau (1995, p.164) propose several tests of the independent observation time assumption. For our NPC trial data, we performed their two Wald-type tests which are based on including an extra covariate of either (a) the length of observation time or follow-up time, τ say; or (b) an indicator on whether τ is longer than the median (here 2795 days). The second test (b) yielded a $\mathcal{Z}$-value of -0.65 for the new covariate, which is not significant. This indicates no evidence against the independence assumption. On the other hand, the first test (a) did indicate a marginally significant negative relation between the event frequency and the follow-up length. However as Lawless and Nadeau (1995) point out, this test is highly sensitive to influential observations. We calculated the correlation between the event rate (number of events divided by τ) and the follow-up length (τ). This was done separately for the placebo and Se group patients. The correlations are -0.0194 and -0.0931 respectively, whereas the 3% trimmed correlations were $+0.0039$ and $+0.0032$. Thus the data seemed generally consistent with the independence assumption. Also there was nothing in the protocol of the design or conduct of the followup that would lead to a suspicion that there should be nonindependence of the event processes and the observation times.

A second issue concerns the interpretation of the regression coefficients in the model specification (1), in which, being averaged over the frailties or omitted covariates, β represents a "population" effect of the covariates Z. An alternate formulation would be to postulate a model in which

$$E(\tilde{Y}_k|H,Z,O) = \tilde{\Gamma}_k exp(Z'\zeta + O'\omega), \tag{30}$$

where $\tilde{\Gamma}_k$ is a baseline mean measure, and ζ and ω are regression coefficients. Z and O are assumed to be constant over k for convenience. By conditioning on O, the parameter ζ

represents the effect of Z matching the value of the omitted covariate, or the frailty. Thus ζ represents an "individual" effect of the covariates Z. Chastang, Byar and Piantadosi (1988, p.1254) give a good discussion of these two models and argue that (30) is more useful. Of course, in general O or the frailties are not observed and so an analysis based on (30) is not possible. However, under the condition $E(e^{O'\omega}|Z,H) = E(e^{O'\omega})$, the "population" parameter β is actually the same as the "individual" one (ζ). This can be seen by taking the expectation of (30) conditional on Z and H and comparing with (1). Such a condition might be considered reasonable for randomized clinical trials with recurrent events data, where Z represents a randomized treatment assignment indicator. In this case, in addition to the independent observation time assumption as in the previous paragraph, it may be assumed that conditional on the follow-up process H, Z and O are independent. Note that for *survival* data, such an assumption might be unreasonable, because those subjects surviving (all H_k's= 1) on an inferior treatment will be associated with superior frailties leading to a diminution or underestimate of the apparent treatment effect — Chastang *et al.* (1988), Keiding *et al.* (1997).

Our method is based on large sample asymptotic theory. Although the data set we used in our illustration was quite large, it would be of interest to investigate finite sample properties of the procedure. Our method uses a simplified naive model as the starting point, and corrects any asymptotic biases on the resulting naive likelihood estimates afterwards. It may not be fully efficient when compared with the method which uses a properly specified yet more complicated likelihood. Yet it has the advantage of being able to make use of simple naive partial likelihood estimates and Nelson-Aalen estimates, often available in standard computer software packages, in a wide variety of multiple events problems, with effects of random heterogeneity and measurement error explicitly assessed in the procedure. This method of starting from a simple naive likelihood function has also been used to consider other nonlinear regression models, such as negative binomial regression (Turnbull et al. 1997), exponential regression and logistic regression (Jiang 1996), for the effects of model misspecification which could arise from the presence of latent variables, random effects, measurement error, omitted covariates, and generally incomplete data.

ACKNOWLEDGMENT

This research was supported in part by a grant from the U.S. National Institutes of Health.

REFERENCES

Abu-Libdeh, H., Turnbull, B. W., and Clark, L. C. (1990), "Analysis of Multi-type Recurrent Events in Longitudinal Studies; Application to a Skin Cancer Prevention Trial," *Biometrics*, 46, 1017-1034.

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

Andersen, P. K. and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1100-1120.

Buzas, J. (1996), "Unbiased Scores in Cox Regression with Covariate Measurement Error," Preprint, University of Vermont.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, Chapman and Hall, New York.

Chastang, C., Byar, D. and Piantadosi, S. (1988), "A Quantitative Study of the Bias in Estimating the Treatment Effect Caused by Omitting a Balanced Covariate in Survival Models," *Statistics in Medicine*, 7, 1243-1255.

Clark, L.C., Combs, G. F., Turnbull, B.W., Slate, E.H., Chalker, D.K., Chow, J., Davis, L.S., Glover, R.A., Graham, G.F., Gross, E.G., Krongrad, A., Lesher, J.L., Park, H.K., Sanders, B.B., Smith, C.L., Taylor, J.R. and the Nutritional Prevention of Cancer Study Group. (1996). "Effects of Selenium Supplementation for Cancer Prevention in Patients with Carcinoma of the Skin: A Randomized Clinical Trial." *Journal of the American Medical Association*, 276 (24), 1957-1963. (Editorial: p1984-5.)

Cook, R. (1995), "The Design and Analysis of Randomized Trials with Recurrent Events," *Statistics in Medicine*, 14, 2081-2098.

Cook, R. J., Lawless, J. F. and Nadeau, C. (1996) "Robust Tests for Treatment Comparisons Based on Recurrent Event Responses," *Biometrics*, 52, 557-571.

Cox, D. R. (1972), "Regression Models and Life-tables," *Journal of the Royal Statistical Society B*, 34, 187-207.

Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," *Proceedings of the 5th Berkeley Symposium on Probability and Statistics 1*, University of California Press, 221-233.

Jiang, W. (1996), *Aspects of Misspecification in Statistical Models: Applications to Latent Variables, Measurement Error, Random Effects, Omitted Covariates and Incomplete Data*, Ph. D. Thesis, Cornell University.

Karr, A. F. (1991) *Point Processes and Their Statistical Inference, 2nd Ed.*, Marcel Dekker, New York.

Keiding, N., Andersen, P. K. and Klein, J. P. (1997), "The Role of Frailty Models and Accelerated Failure Time Models in Describing Heterogeneity Due to Omitted Covariates," *Statistics in Medicine*, 16, 215-224.

Lawless, J. F. (1987), "Regression Methods for Poisson Process Data," *The Journal of the American Statistical Association*, 82, 808-815.

Lawless, J. F. and Nadeau, C. (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158-168.

Lin, D. Y. and Wei, L. J. (1989), "The Robust Inference for the Cox Proportional Hazard Model," *The Journal of the American Statistical Association*, 84, 1074-1078.

Nakamura, T. (1992), "Proportional Hazards Model with Covariates Subject to Measurement Error," *Biometrics*, 48, 829-838.

Oakes, D. (1992), " Frailty Models for Multiple Event Times," *Survival Analysis: State of the Arts*, Ed. J. P. Klein and P. K. Goel, Netherlands: Kluwer Academic Publishers, pp. 371-9.

Pepe, M. S., Self, S. G. and Prentice, R. L. (1989), "Further Results on Covariate Measurement Errors in Cohort Studies with Time to Response Data," *Statistics in Medicine*, 8, 1167-1178.

Prentice, R. L. (1982), "Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model," *Biometrika*, 69, 331-342.

Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981), "On the Regression Analysis of Multivariate Failure Time Data," *Biometrika*, 68, 373-379.

Raboud, J. M. (1991), *The Effects of Errors in Measurement in Survival Analysis*, Ph. D. Thesis, University of Toronto.

Raboud, J., Reid, N., Coates, R. A. and Farewell, V. T. (1993), "Estimating Risks of Progressing to Aids when Covariates are Measured with Error," *Journal of the Royal Statistical Society A*, 156, 393-406.

StataCorp (1997). *Stata Statistical Software, Release 5.0* Stata Corporation, College Station, Texas.

Statistical Sciences Inc. (1995). *S-PLUS Statistical Software* Seattle, Washington.

Turnbull, B. W., Jiang, W. and Clark, L. C. (1997), "Regression Models for Recurrent Event Data: Parametric Random Effects Models with Measurement Error," *Statistics in Medicine*, (in press).

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *The Journal of the American Statistical Association*, 84, 1065-1073.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge, England.

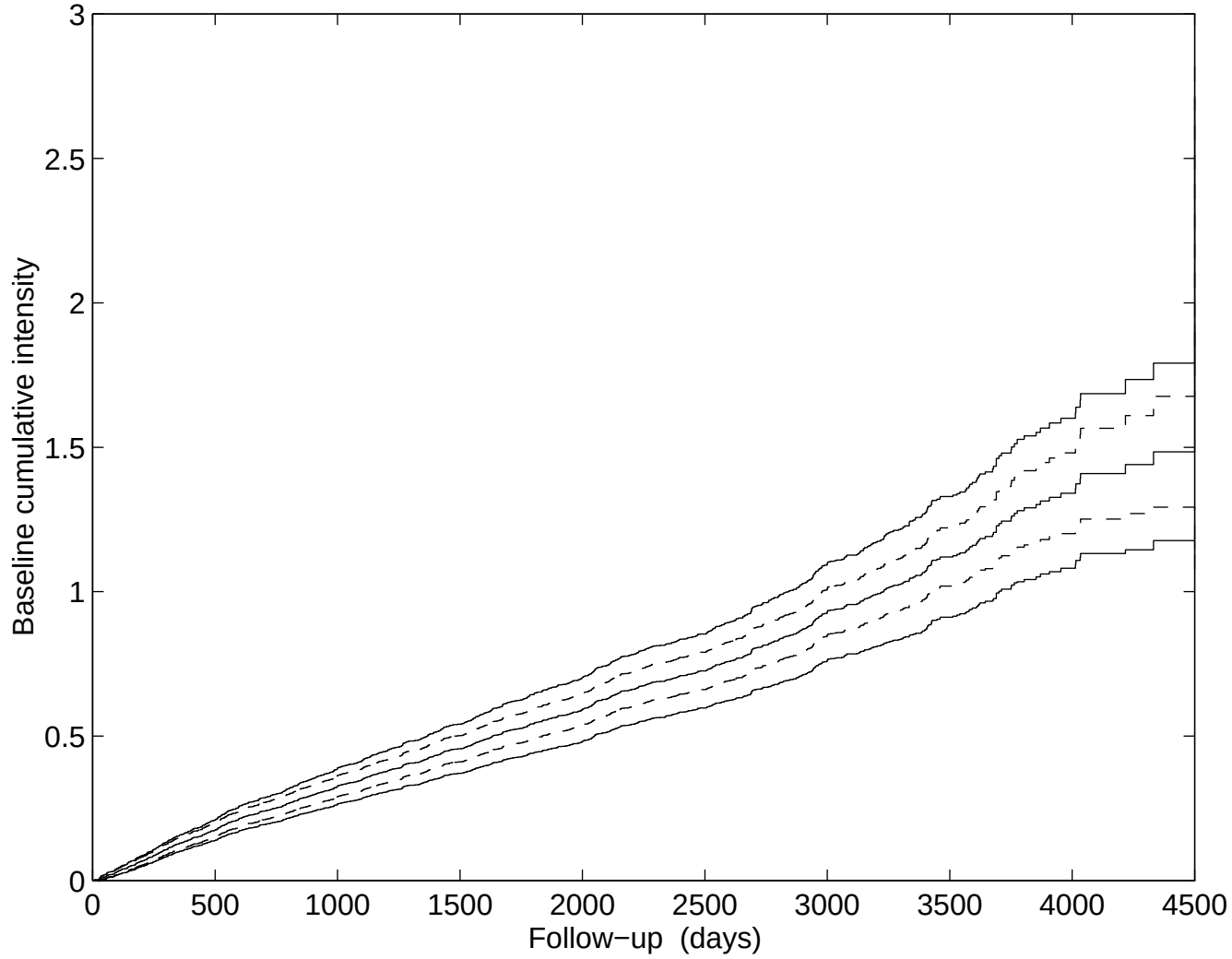

Figure 1: Estimate of baseline cumulative intensity function for SCC recurrence data for NPC trial. The outer dashed lines are the pointwise 95% confidence limits based on the naive analysis whereas the outer solid lines represent the pointwise 95% confidence limits based on a robust analysis adjusted for random effects.

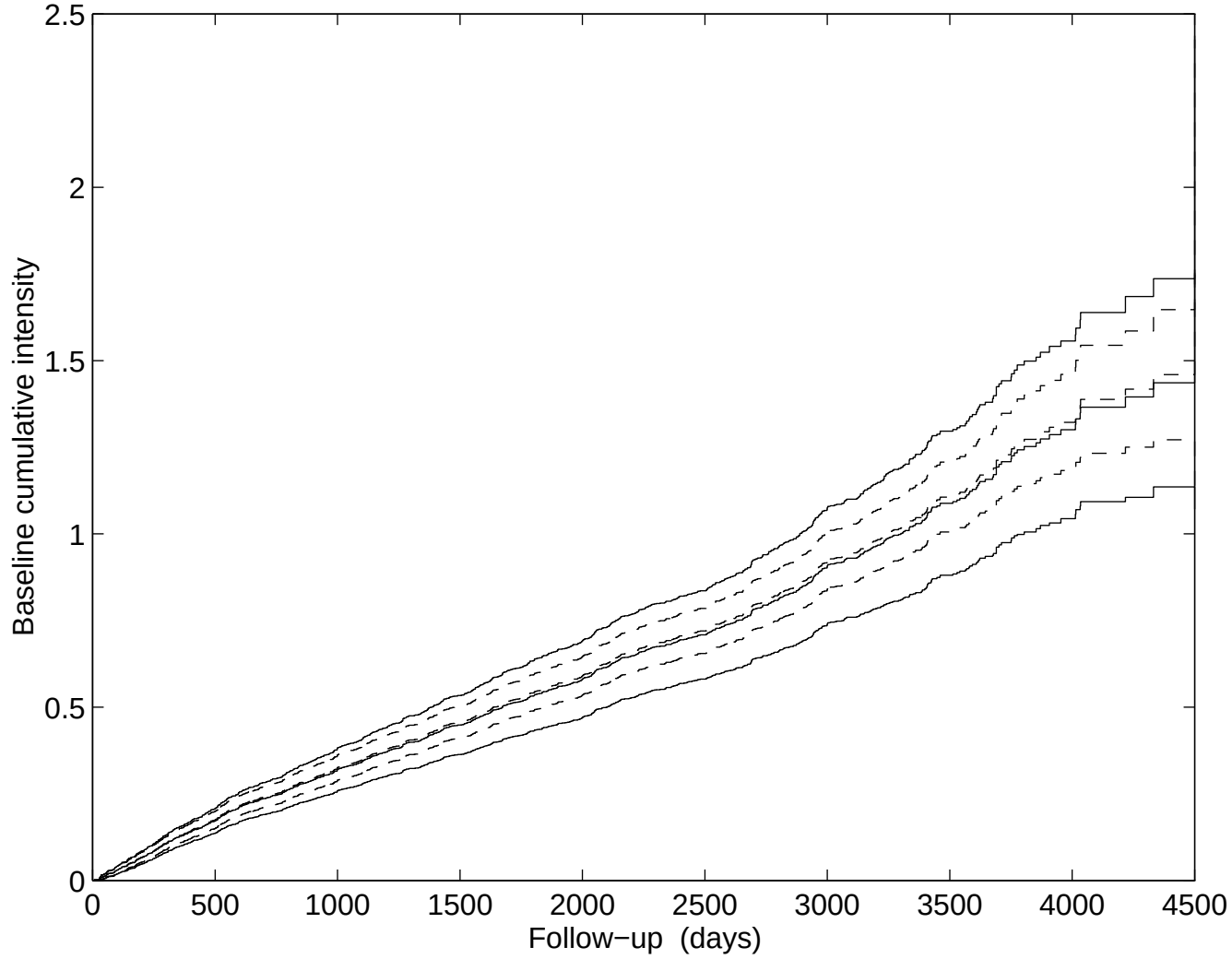

Figure 2: Estimate of baseline cumulative intensity function for SCC recurrence data for NPC trial. The dashed lines are the estimate and pointwise 95% confidence limits based on the naive analysis. The solid lines are the estimate and pointwise 95% confidence limits based on an robust analysis adjusted for random effects and covariate measurement error.