STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:3383–3393 (DOI: 10.1002/sim.1279)

Choice of stratification in Poisson process analysis of recurrent event data with environmental covariates

Anup Dewanji[1] and Suresh H. Moolgavkar[2,*,†]

[1]*Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta, India*
[2]*Fred Hutchinson Cancer Research Center, Seattle, Washington 98109-1024, U.S.A.*

SUMMARY

The Poisson process approach for studying the association between environmental covariates and recurrent events depends on the stratification of study period into intervals within which the baseline intensities are assumed constant. In this work we investigate the problem of bias and variance due to misspecification of this stratification. We suggest a cross-validation approach to choosing a stratification model to balance the trade-off between bias and variance. We also establish a connection between the Poisson process approach and case cross-over studies. Copyright

KEY WORDS: bias–variance trade-off; stratification model; cross-validation; chronic respiratory disease; case cross-over design

1. INTRODUCTION

A point process formulation is commonly used for regression analysis with recurrent event data, in which either the intensity is modelled as a function of the covariates or covariate processes [1], or some other 'marginal' quantities are modelled to avoid strong assumptions on the recurrent event process [2–4]. These approaches focus on subject specific covariates, but fail for environmental covariates as they are the same for all the subjects at any event time. However, environmental covariates are also important, especially in the study of air pollution. Daily measures of air pollution indices, average temperature and humidity etc., may affect relapse of, say, respiratory diseases, which can be identified with recurrent event data. In Section 5 we consider one such example where the recurrent events are hospital admissions for chronic respiratory diseases. Although one could think of subject-specific covariates like age, sex, some health index etc., we do not consider them for our illustration to keep it simple.

In a recent article, Dewanji and Moolgavkar [5] propose a Poisson process approach for analysing such data, in which the recurrent events in an individual subject are assumed to follow a non-homogeneous Poisson process with intensity depending on environmental

* Correspondence to: Suresh H. Moolgavkar, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, MP-665, P.O. Box 19024, Seattle, Washington 98109-1024, U.S.A.
† E-mail: smoolgav@fhcrc.org

Received August 2001
Accepted March 2002

covariates. Although the corresponding regression coefficients (the relative risk parameters) are assumed to be the same for all the individuals, the baseline intensities can be different for different individuals. In addition, the baseline intensity for an individual is assumed to be piecewise constant in time having different values in different time intervals (called 'strata'). Therefore, different choices of stratification represent different models, with finer stratification giving more flexible models. Our proposed method considers the conditional distribution of events in a stratum given the number of events in that stratum so that the baseline intensities within a stratum cancel out and the resulting likelihood becomes a function of the regression coefficients only. With finer stratification, there may be no event in some strata leading to loss of information and, hence, larger variance. We shall see later in Section 3 that expected information depends on within-strata variation in the covariates, which is less for smaller strata.

A model with a small number of big strata is misspecified if the true model corresponds to a finer stratification. In such a case, the estimates obtained from the assumed model are biased. A model with a large number of small strata may be able to avoid this bias, but, because of the loss of information, the corresponding estimates are less efficient having large variance. The purpose of this work is to investigate this bias–variance trade-off. We derive the expressions for asymptotic bias and variance. We also conduct a simulation study to investigate the small sample behaviour of these quantities. At the end, we try to come up with guidelines for choosing a stratification model that is 'optimal' in some sense.

In the next section, we briefly review the Poisson process approach of reference [5]. We derive the expressions for asymptotic bias and variance in Sections 3, and also discuss our simulation results. Section 4 presents our method for choice of stratification given the covariates. Some properties of this method are investigated by simulation. In Section 5 we present the analysis of the data considered in reference [5]. Section 6 discusses the connection between this Poisson process approach and case cross-over studies and Section 7 ends with some remarks.

2. THE POISSON PROCESS APPROACH

Let x_t denote the vector of covariates at time t consisting of all the subject and environment specific covariates. For this approach, we consider only the time dependent covariates, and, specifically, the environmental covariates. For each subject (say, ith) under study, we consider a non-homogeneous Poisson process to describe the occurrences of events in the subject with intensity given by

$$\lambda_i(t, x_t) = \lambda_i \exp[x_t^{\mathrm{T}} \beta] \tag{1}$$

The baseline intensity λ_i varies from subject to subject but is independent of time and the relative risk parameter β, which is of primary interest, remains the same over all the subjects. It is possible to allow some limited amount of time dependence in the λ_i's because of the Poisson process formulation. Assume the λ_i's, as function of time t, to be piecewise constant over the study period $(0, \tau]$, say, as in the following:

$$\lambda_i(t) = \lambda_{il} \quad \text{for } t \in I_l = (\tau_{l-1}, \tau_l], \quad \text{for } l = 1, \ldots, K \tag{2}$$

with $0 = \tau_0 < \tau_i < \cdots < \tau_K = \tau$ being prespecified. We call these I_l's 'strata' from now on.

It is expected that, for a particular study at hand, there will be some natural candidate stratifications. For our example in Section 5, the natural candidates correspond to years, seasons, months etc. In the absence of such logical choices, one has to use judgement on when the baseline intensities may be constant, which is a difficult task. Let d_{il} denote the number of events for the ith subject in the stratum I_l. Since the events in disjoint strata I_l's are independent (because of the Poisson process assumption), the contribution of the ith individual to the conditional (on d_{il}) likelihood is given by

$$L_i(\beta) = \prod_{l=1}^{K} \left\{ \frac{\prod_{j=1}^{d_{il}} \exp[x_{t_{ilj}}^{\mathrm{T}}\beta]}{(\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t)^{d_{il}}} \right\} \tag{3}$$

where t_{ilj}'s denote the d_{il} event times for the ith subject in I_l with $x_{t_{ilj}}$ being the value of x_t at time $t = t_{ilj}$. Taking the product of $L_i(\beta)$'s over all i, we get the total conditional likelihood as given in equation (9) of reference [5]. Note that this likelihood can also be derived by considering the superimposition of all the individual Poisson processes as a single Poisson process with intensity $(\sum_{i=1}^{n} \lambda_{il}) \exp[x_t^{\mathrm{T}}\beta]$, for $t \in I_l$, and conditioning the likelihood on $d_l = \sum_{i=1}^{n} d_{il}$. The likelihood for β can also be viewed as a profile likelihood by substituting the λ_{il}'s (or $(\sum_{i=1}^{n} \lambda_{il})$) by the corresponding maximum likelihood estimates under fixed β in the original likelihood (see reference [5] for details).

3. ASYMPTOTIC BIAS AND VARIANCE

Let us, for the sake of notational convenience, assume that there is only one environmental covariate of interest. Using the likelihood contribution (3), one can easily write down the score function as

$$u(\beta) = \sum_{i=1}^{n} \sum_{l=1}^{K} \sum_{j=1}^{d_{il}} \left\{ x_{t_{ilj}} - \frac{\int_{I_l} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}{\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t} \right\} \tag{4}$$

Since x_t is the regressor process, we hold it to be non-stochastic in further calculation of expectation and variance. Also, this being a conditional analysis given the d_{il}'s, we consider the corresponding conditional expectations. Note that the only random quantity in (4) is $x_{t_{ilj}}$. Because of the Poisson process assumption, we have

$$E[x_{t_{ilj}} | d_{il}] = \frac{\int_{I_l} x_t \lambda_i(t, x_t)\,\mathrm{d}t}{\int_{I_l} \lambda_i(t, x_t)\,\mathrm{d}t} \tag{5}$$

where $\lambda_i(t, x_t)$ denotes the Poisson intensity, for the ith subject, at time t with covariate value x_t. If the model (or stratification) is true, that is $\lambda_i(t, x_t) = \lambda_{il} \exp[x_t\beta]$ for $t \in I_l$, then this expectation (5) becomes equal to

$$\frac{\int_{I_l} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}{\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t} \tag{6}$$

for all i and l, so that the expectation of the score function (4) is zero. Therefore, the estimating equation for the coefficient β is unbiased. Hence, from the first-order Taylor series

covariates. Although the corresponding regression coefficients (the relative risk parameters) are assumed to be the same for all the individuals, the baseline intensities can be different for different individuals. In addition, the baseline intensity for an individual is assumed to be piecewise constant in time having different values in different time intervals (called 'strata'). Therefore, different choices of stratification represent different models, with finer stratification giving more flexible models. Our proposed method considers the conditional distribution of events in a stratum given the number of events in that stratum so that the baseline intensities within a stratum cancel out and the resulting likelihood becomes a function of the regression coefficients only. With finer stratification, there may be no event in some strata leading to loss of information and, hence, larger variance. We shall see later in Section 3 that expected information depends on within-strata variation in the covariates, which is less for smaller strata.

A model with a small number of big strata is misspecified if the true model corresponds to a finer stratification. In such a case, the estimates obtained from the assumed model are biased. A model with a large number of small strata may be able to avoid this bias, but, because of the loss of information, the corresponding estimates are less efficient having large variance. The purpose of this work is to investigate this bias–variance trade-off. We derive the expressions for asymptotic bias and variance. We also conduct a simulation study to investigate the small sample behaviour of these quantities. At the end, we try to come up with guidelines for choosing a stratification model that is 'optimal' in some sense.

In the next section, we briefly review the Poisson process approach of reference [5]. We derive the expressions for asymptotic bias and variance in Sections 3, and also discuss our simulation results. Section 4 presents our method for choice of stratification given the covariates. Some properties of this method are investigated by simulation. In Section 5 we present the analysis of the data considered in reference [5]. Section 6 discusses the connection between this Poisson process approach and case cross-over studies and Section 7 ends with some remarks.

2. THE POISSON PROCESS APPROACH

Let x_t denote the vector of covariates at time t consisting of all the subject and environment specific covariates. For this approach, we consider only the time dependent covariates, and, specifically, the environmental covariates. For each subject (say, ith) under study, we consider a non-homogeneous Poisson process to describe the occurrences of events in the subject with intensity given by

$$\lambda_i(t, x_t) = \lambda_i \exp[x_t^{\mathrm{T}} \beta] \tag{1}$$

The baseline intensity λ_i varies from subject to subject but is independent of time and the relative risk parameter β, which is of primary interest, remains the same over all the subjects. It is possible to allow some limited amount of time dependence in the λ_i's because of the Poisson process formulation. Assume the λ_i's, as function of time t, to be piecewise constant over the study period $(0, \tau]$, say, as in the following:

$$\lambda_i(t) = \lambda_{il} \quad \text{for } t \in I_l = (\tau_{l-1}, \tau_l], \quad \text{for } l = 1, \ldots, K \tag{2}$$

with $0 = \tau_0 < \tau_i < \cdots < \tau_K = \tau$ being prespecified. We call these I_l's 'strata' from now on.

It is expected that, for a particular study at hand, there will be some natural candidate stratifications. For our example in Section 5, the natural candidates correspond to years, seasons, months etc. In the absence of such logical choices, one has to use judgement on when the baseline intensities may be constant, which is a difficult task. Let d_{il} denote the number of events for the ith subject in the stratum I_l. Since the events in disjoint strata I_l's are independent (because of the Poisson process assumption), the contribution of the ith individual to the conditional (on d_{il}) likelihood is given by

$$L_i(\beta) = \prod_{l=1}^{K} \left\{ \frac{\prod_{j=1}^{d_{il}} \exp[x_{t_{ilj}}^{\mathrm{T}}\beta]}{(\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t)^{d_{il}}} \right\} \tag{3}$$

where t_{ilj}'s denote the d_{il} event times for the ith subject in I_l with $x_{t_{ilj}}$ being the value of x_t at time $t = t_{ilj}$. Taking the product of $L_i(\beta)$'s over all i, we get the total conditional likelihood as given in equation (9) of reference [5]. Note that this likelihood can also be derived by considering the superimposition of all the individual Poisson processes as a single Poisson process with intensity $(\sum_{i=1}^{n} \lambda_{il}) \exp[x_t^{\mathrm{T}}\beta]$, for $t \in I_l$, and conditioning the likelihood on $d_l = \sum_{i=1}^{n} d_{il}$. The likelihood for β can also be viewed as a profile likelihood by substituting the λ_{il}'s (or $(\sum_{i=1}^{n} \lambda_{il})$) by the corresponding maximum likelihood estimates under fixed β in the original likelihood (see reference [5] for details).

3. ASYMPTOTIC BIAS AND VARIANCE

Let us, for the sake of notational convenience, assume that there is only one environmental covariate of interest. Using the likelihood contribution (3), one can easily write down the score function as

$$u(\beta) = \sum_{i=1}^{n} \sum_{l=1}^{K} \sum_{j=1}^{d_{il}} \left\{ x_{t_{ilj}} - \frac{\int_{I_l} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}{\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t} \right\} \tag{4}$$

Since x_t is the regressor process, we hold it to be non-stochastic in further calculation of expectation and variance. Also, this being a conditional analysis given the d_{il}'s, we consider the corresponding conditional expectations. Note that the only random quantity in (4) is $x_{t_{ilj}}$. Because of the Poisson process assumption, we have

$$E[x_{t_{ilj}} | d_{il}] = \frac{\int_{I_l} x_t \lambda_i(t, x_t)\,\mathrm{d}t}{\int_{I_l} \lambda_i(t, x_t)\,\mathrm{d}t} \tag{5}$$

where $\lambda_i(t, x_t)$ denotes the Poisson intensity, for the ith subject, at time t with covariate value x_t. If the model (or stratification) is true, that is $\lambda_i(t, x_t) = \lambda_{il} \exp[x_t\beta]$ for $t \in I_l$, then this expectation (5) becomes equal to

$$\frac{\int_{I_l} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}{\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t} \tag{6}$$

for all i and l, so that the expectation of the score function (4) is zero. Therefore, the estimating equation for the coefficient β is unbiased. Hence, from the first-order Taylor series

expansion of $u(\hat{\beta})$ as

$$0 = u(\hat{\beta}) \approx u(\beta) + (\hat{\beta} - \beta)u'(\beta) \tag{7}$$

we get $0 = E[u(\beta)] \approx E[-(\hat{\beta} - \beta)u'(\beta)]$. Here $u'(\beta)$ denotes the derivative of $u(\beta)$ with respect to β. Since $-u'(\beta)$ is a positive constant (as will be seen later in (10)), this implies that $\hat{\beta}$ is asymptotically unbiased for β under the true model specification. Note that no condition on the process x_t is needed for this result to hold except that x_t be an external process (that is, the marginal distribution of x_t does not involve the parameters of the recurrent event process). Therefore, even if the process x_t is autocorrelated, or there are time trends in x_t, $\hat{\beta}$ is still asymptotically unbiased for β under the true model specification.

If the assumed model is more general than the true model (that is, the assumed stratification is finer than the true stratification and each assumed stratum is completely contained in exactly one true stratum), the expectation (5) is still equal to (6), for all i, l and j, leading to an unbiased estimating equation, as before. If, however, the true model corresponds to a different stratification than the assumed one and the kinds mentioned above (that is, there is at least one stratum I_l and at least one subject having an event in that I_l, for whom the baseline intensity is not the same throughout I_l), then the cancellation of the baseline intensity in (5) does not take place; hence, the estimating equation is not unbiased. The magnitude and direction of bias depend on the values of x_t and the true λ_{il}'s and the 'distance' between the assumed and the true stratifications in a complicated way. In order to illustrate this, let us consider the following simple example.

Suppose the true stratification is different from the assumed one in only one stratum, namely I_l for a fixed l, in a way that $I_l = I_{l1} \cup I_{l2}$ and $I_{l1} \cap I_{l2} = \phi$, with true baseline intensities λ_{il1} and λ_{il2} in I_{l1} and I_{l2}, respectively, instead of being the assumed λ_{il} in I_l, for the ith subject only having an event in I_l. Then, from (5)

$$\begin{aligned} E[x_{t_{ij}} | d_l] &= \frac{\lambda_{il1} \int_{I_{l1}} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t + \lambda_{il2} \int_{I_{l2}} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}{\lambda_{il1} \int_{I_{l1}} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t + \lambda_{il2} \int_{I_{l2}} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t} \\ &= \frac{\int_{I_{l1}} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t + r_i \int_{I_{l2}} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}{\int_{I_{l1}} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t + r_i \int_{I_{l2}} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t} \end{aligned} \tag{8}$$

where $r_i = \lambda_{il2}/\lambda_{il1}$. As this is not equal to (6), this makes the only contribution to the expectation of $u(\beta)$, which is the difference (8) *minus* (6). A simple calculation shows that, if $r_i > 1$, or $\lambda_{il2} > \lambda_{il1}$, then (8) $\geqslant$ ($\leqslant$) (6) if and only if

$$\left(\int_{I_{l1}} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t\right)\left(\int_{I_{l2}} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t\right) \leqslant (\geqslant) \left(\int_{I_{l1}} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t\right)\left(\int_{I_{l2}} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t\right) \tag{9}$$

If the x_t's in I_{l1} are, in general, smaller than those in I_{l2}, then the LHS in (9) is smaller than the RHS and, hence, the bias, given by (8) *minus* (6), is positive. Since $-u'(\beta)$ is positive, we have, from (7), positive bias also in $\hat{\beta}$. Note that, if $r_i < 1$, then we have the reverse inequality in (9). This simple example illustrates how bias gets in $\hat{\beta}$ due to misspecification of model or stratification. Extending this line of derivation for more general type of misspecification, one can see that if both x_t's and the baseline intensities are generally increasing (or, decreasing)

within an assumed stratum, then the bias in $\hat{\beta}$ is positive, whenever the true stratification is contained in the assumed one in the sense that each true stratum is completely contained in exactly one assumed stratum. Even if the true and assumed stratifications deviate a little from the above mentioned relationship, $\hat{\beta}$ can still be positively biased because of the dominance of the positive bias contributions. However, on the other hand, if the x_t's and the baseline intensities change monotonically within an assumed stratum but in opposite directions, the bias will be negative. Once it is seen how bias develops, through the expectation of (4), one can think of other scenarios for either positive or negative bias. As a matter of fact, bias could be zero or negligible because of cancellation of different bias contribution terms of opposite signs depending on the values of the x_t's and the baseline intensities.

The asymptotic variance of $u(\beta)$ is given by the expected information $i(\beta) = E[-u'(\beta)]$. From (4), we have $-u'(\beta) = E[-u'(\beta)]$ and is given by

$$i(\beta) = \sum_{l=1}^{K} d_l \left\{ \frac{(\int_{I_l} x_t^2 \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t)(\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t) - (\int_{I_l} x_t \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t)^2}{(\int_{I_l} \exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t)^2} \right\} \tag{10}$$

This asymptotic variance depends only on the assumed stratification, not on the true one. Also, note that the asymptotic variance of $\hat{\beta}$ is the inverse $i^{-1}(\beta)$ of (10).

By the Cauchy–Schwarz inequality, the numerator of each of the terms in (10) is non-negative. Also, if the x_t's are constant within a stratum, then the corresponding numerator is zero (the score function (10) is identically equal to zero). As expected, if there is no variation in the x_t values within a stratum, there is no information on the regression coefficient β in that stratum. If there is more (less) within-strata variation in the x_t values, there is more (less) information, given by $i(\beta)$, on the regression coefficient β leading to smaller (larger) variance of $\hat{\beta}$. This fact can be used, by noting the pattern in the x_t values, to obtain a suitable stratification model so that the variance of $\hat{\beta}$ is small (or the estimate is precise). Although the results on bias suggest finer stratification, the x_t values within strata, in general, tend to be similar in such a case. Therefore, as argued above, the estimate becomes less precise. This is evident in the example of reference [5]. This is the bias–variance trade-off as mentioned in the Introduction and discussed further in Section 6. A simulation study also supports the above findings on bias and variance in finite samples.

4. MODEL CHOICE VIA CROSS-VALIDATION

Any available knowledge on time dependence of the baseline intensities and temporal pattern of the x_t values can be used in restricting the choice to a few suitable stratification models. However, although the x_t values can be known before the analysis, the baseline intensities are unknown nuisance parameters. Hence, this choice of a few models has to be done in rather an *ad hoc* way. In this section, we suggest a method of cross-validation to choose a stratification model out of these few (say, S in number), given the covariates (x_t's).

Various authors have used cross-validation techniques for different purposes including choice of prediction model [6], classification tree [7] and smoothing parameter in non-parametric estimation [8]. Likelihood-based cross-validation techniques have also been used [9, 10] for adaptive choice of smoothing parameter. Stone [11] has established asymptotic equivalence between likelihood based cross-validation and Akaike's criterion for choice of model.

Table I. Proportions of chosen models when model $\mathscr{P}_3$ is true.

Baseline intensity				Pattern of x_t	Proportions				
(a_1,b_1)	(a_2,b_2)	(a_3,b_3)	(a_4,b_4)		$\mathscr{P}_1$	$\mathscr{P}_2$	$\mathscr{P}_3$	$\mathscr{P}_4$	$\mathscr{P}_5$
(0.01,0.015)	(0.015,0.02)	(0.02,0.025)	(0.025,0.03)	I	0.02	0.27	0.56	0.12	0.03
				II	0.21	0.33	0.26	0.12	0.08
(0.01,0.02)	(0.02,0.03)	(0.03,0.04)	(0.04,0.05)	I	0.00	0.14	0.61	0.21	0.04
				II	0.18	0.29	0.33	0.11	0.09
(0.01,0.03)	(0.03,0.05)	(0.05,0.07)	(0.07,0.09)	I	0.00	0.01	0.74	0.15	0.10
				II	0.11	0.20	0.42	0.12	0.15
(0.00,0.04)	(0.04,0.08)	(0.08,0.12)	(0.12,0.16)	I	0.00	0.00	0.79	0.12	0.09
				II	0.13	0.16	0.47	0.19	0.05
(0.01,0.03)	(0.01,0.03)	(0.01,0.03)	(0.01,0.03)	I	0.74	0.14	0.06	0.04	0.02
				II	0.52	0.16	0.16	0.08	0.08
(0.00,0.16)	(0.00,0.16)	(0.00,0.16)	(0.00,0.16)	I	0.73	0.14	0.07	0.03	0.03
				II	0.45	0.18	0.20	0.12	0.05

For a fixed stratification model indexed by s (for $s=1,\ldots,S$), we define cross-validated log-likelihood (CVLL(s)) by

$$\mathrm{CVLL}(s)=\sum_{i=1}^{n}\log L_i^{(s_f)}(\hat{\beta}_{-i}^{(s)}) \tag{11}$$

where $L_i^{(s_f)}(\cdot)$ is the likelihood contribution from the ith subject, as in (3), corresponding to the finest stratification indexed by s_f and $\hat{\beta}_{-i}^{(s)}$ is the estimate of β obtained by removing the ith individual under the model indexed by s. This corresponds to the 'leave-one-out' cross-validation and one chooses the model indexed by s_0 so that $\mathrm{CVLL}(s_0)=\max_s \mathrm{CVLL}(s)$.

We carry out a small simulation study in order to investigate how good the above cross-validation technique is for choosing the true model. We take the observation period as $[0,100]$ and consider five stratification models given by $\mathscr{P}_1:[0,100]$, $\mathscr{P}_2:\{[0,50],(50,100]\}$, $\mathscr{P}_3:\{[0,20],(20,50],(50,70],(70,100]\}$, $\mathscr{P}_4:\{[0,20],(20,40],(40,50],(50,70],(70,90],(90,100]\}$, $\mathscr{P}_5:\{[0,10],(10,20],\ldots,(90,100]\}$. We take $n=100$ and $\beta=0.05$. For each simulation, data on n subjects (that is, n Poisson processes) are generated using $\mathscr{P}_3$ as the true model with $\lambda_1,\lambda_2,\lambda_3$ and λ_4 being the four vectors of baseline intensities for the n subjects in the four respective strata of $\mathscr{P}_3$ and x_t, as before, being the single covariate. The cross-validated log-likelihood (CVLL) for this simulated data is calculated (note that s_f here corresponds to the stratification $\mathscr{P}_5$) for all five models and the one giving the maximum CVLL is chosen, as prescribed before. Based on 100 such simulations, we report in Table I the proportions of times the different models are chosen. Here, we report only the cases when we assume $\mathscr{U}[a_i,b_i]$ distribution for λ_i, for $i=1,\ldots,4$, for different values of (a_i,b_i)'s. As before, we assume that the covariate takes a constant value x_t in $(t-1,t]$, for $t=1,\ldots,100$, and consider only two patterns of x_t, in which it follows (I) $\mathscr{U}[6+0.05t,10+0.05t]$ and (II) $\mathscr{U}[6,15]$ distribution, respectively, so that the x_t's have an increasing trend and no trend at all.

We notice that the results depend on the patterns of x_t. When x_t's are increasing (pattern I), their values are less heterogeneous within strata for, say, $\mathscr{P}_5$ than for the other coarser models. As a result, the finer stratification models give less precise estimates, and so the models $\mathscr{P}_4$ and $\mathscr{P}_5$ are chosen less often. However, as difference between the λ_i's increases, the true model

$\mathscr{P}_3$ is chosen more often, as expected, and since the bias becomes more prominent for $\mathscr{P}_1$ and $\mathscr{P}_2$, there is a shift in choosing the finer models. When the λ_i's follow the same distribution (the last two rows), then the bias is negligible and as a result the coarser models (in particular the model $\mathscr{P}_1$) are chosen more often to give more precise estimate. For the second pattern of x_t having the same distribution for all t, x_t values are equally homogeneous within strata and, therefore, the estimates are equally precise for all the models. Note that, from (5), bias is also expected to be small. Hence, the choices are somewhat evenly distributed with $\mathscr{P}_3$ being chosen more often as the λ_i's become more disparate.

We also worked with exponential distributions for the λ_i's and normal distributions for the x_t's leading to similar results. We also generated data from other models (than $\mathscr{P}_3$) and found similar results. The objective of this cross-validation method is to select a model that maintains a 'balance' between bias and variance. It may not be choosing the 'true' model all the time, but it addresses the problem of bias–variance trade-off in an adaptive way. In the next section, we propose a strategy based on paired two-sample tests to choose a set of models having statistically 'similar' CVLL values, out of which the one with coarsest stratification will be selected.

For a study with a large number of individuals, this 'leave-one-out' cross-validation becomes computationally prohibitive. One can then use m-fold cross-validation with a suitable m, in which the set of individuals is divided into m random parts (referred to as cross-validation samples) with roughly equal sizes and then the above cross-validation method is applied on the m parts. In the next section, we illustrate application of this cross-validation method with real data.

5. AN ILLUSTRATION

Dewanji and Moolgavkar [5] presented analyses of a data set on multiple hospital admissions for chronic respiratory disease in King County over the period 1990–1995. The environmental covariates of interest were daily indices of air pollution and temperature. There were 5362 admissions for 2801 individuals over this period. In the analyses, three air pollution indices were considered: carbon monoxide (CO) on the same day; particulate matter less then 10 microns in diameter (PM_{10}) lagged by three days, and an index of light scattering (LS), which is a surrogate measure of fine particles, on the previous day. In addition, the analyses also included temperature (TMP) and day of week. To keep our illustration simple, we exclude day of week from our analysis here. As in reference [5], we consider five different stratification models: a single stratum over the entire period; 6 strata corresponding to each of the years; 25 strata representing different seasons; 72 strata corresponding to each month, and 144 strata with two for each month.

We consider 28-fold cross-validation for our illustration dividing the 2801 individuals in 27 sets of size 100 and one of 101. Here, the last stratification with 144 strata is the finest, and we use that for our calculation of CVLL (see (11)) for all the above five models. Although the cross-validation method of the previous section chooses the model having the largest CVLL value, our experience is that many models lead to similar values of CVLL. The one-standard-error rule of reference [7] has not been satisfactory for differentiating between models, as there is large variation between different cross-validation samples, but for a fixed cross-validation sample there is little difference between different models.

For any two models under comparison, one can, intuitively, consider a paired two-sample test based on, in general, the m pairs of individual contributions to the two corresponding

Table II. CVLL values and the chosen model in bold face.

Covariates	Stratification Models				
	One stratum	6 strata	25 strata	72 strata	144 strata
CO	−14625.86	−14588.96	**−14586.31**	−14586.44	−14586.50
PM_{10}	−12140.46	**−12123.92**	−12122.12	−12122.00	−12122.13
LS	−12861.02	**−12845.04**	−12843.05	−12842.98	−12843.06
CO + TMP	−14611.03	**−14588.06**	−14590.34	−14588.32	−14588.38
PM_{10} + TMP	−12129.65	**−12121.48**	−12124.89	−12122.22	−12122.42
LS + TMP	−12854.31	**−12843.98**	−12847.34	−12846.13	−12846.40
CO + PM_{10} + TMP	−12133.92	**−12120.23**	−12124.71	−12121.13	−12121.23
CO + LS + TMP	−12855.59	**−12842.98**	−12846.27	−12845.18	−12845.40

CVLL's and conclude if one set of m contributions is statistically larger than the other. We suggest the following strategy. Consider the model s_0 with the largest CVLL value and compare it with the other models using a paired test with one-sided alternative as described above. This gives a set of models, including s_0, which have statistically homogeneous CVLL values. The coarsest one in this set is the chosen model, since that leads to more precise estimates (as argued in Section 3).

In this illustration with $m=28$, we consider a parametric test (paired t-test) and a non-parametric test (Wilcoxon signed rank test) for comparing models. In order to be conservative, we use 0.1 as the size of the two tests and the null hypothesis of homogeneity is rejected if at least one test rejects it. The results (the CVLL values) for different covariates are reported in Table II, in which the one in bold face indicates the chosen model by the above strategy. Besides considering the three air pollution indices (covariates) separately in the first three rows, we also consider them one by one after effect of temperature (TMP) is controlled linearly with a distinct slope for each season (see reference [5]). The corresponding results are presented in the second three rows. The results for two pollutants and temperature are in the last two rows.

As more covariates are included in the model, we expect less time dependence in the baseline intensities and the coarser models to be chosen. However, the model with a single stratum is strongly rejected. The model with six strata seems to be the choice, even with a single covariate (except for CO). We notice, from the results of reference [5], that the estimates are similar over the models with six and more strata and, among them, the standard errors are smallest with six strata.

6. CONNECTION WITH CASE CROSS-OVER STUDIES

Note that the conditional likelihood of Section 2 is derived by viewing the data as arising from a follow-up study. Those individuals not having any event in the study period do not contribute anything to the likelihood because of the conditioning. The likelihood (3) reminds one of the case cross-over likelihood, in which the covariate at the time of an event is compared with the covariates at some suitably chosen 'control' time periods (see reference [12, 13]). Here, the covariates at the event times in a stratum are compared, instead, with those at all the time

points in that stratum. This likelihood, therefore, can be viewed as that from a case cross-over study possibly with multiple events and the set of control periods being the whole time interval in a particular stratum. This may be thought of as a continuous version of the usual case cross-over study. If for an individual there is at most one event in each stratum and the integral in the denominator of (3) is replaced by a finite sum (since, usually, x_t is piecewise constant), the resulting likelihood is similar to that of reference [12] for a case cross-over study. However, for multiple events, our likelihood is much simpler than the one derived in reference [12].

One may recall that the original idea behind the case cross-over design was to study subject specific transient covariates [14]. Since, in principle, a case cross-over design chooses controls from the same subject but at different times, all subject specific non-transient covariates are automatically matched, thus precluding the analysis of relative risks associated with them. We observe the same phenomenon in our study of recurrent event also. This approach is, therefore, useful for studying subject specific transient covariates, or, in particular, environmental covariates. Navidi [12] addressed the bias problem in case cross-over design due to time trends (non-stationarity, in general) in covariates; our Poisson process approach leads to a solution similar to his. Note that, as in the case cross-over design comparing hazards for the same person at different times, our approach also requires the assumption of constant baseline intensity within a stratum which may vary arbitrarily between individuals thus incorporating the effects of any time-constant covariates. The stratification, however, allows some limited time-dependence in the baseline intensity.

The likelihood for a case cross-over study is derived by mimicking the derivation of likelihood for a matched case-control study. However, the assumptions of case-control studies do not always apply to the case cross-over studies, as discussed in reference [13]. For example, possible autocorrelation between covariates over time makes the observed covariates for case and control periods dependent; also, for environmental covariates, two cases occurring on the same day must have the same covariate value, a between-stratum constraint not present in case-control studies. Because the Poisson process approach is based on different assumptions, these problems do not arise.

The issue of bias–variance trade-off in the case cross-over design was first discussed by Mittleman *et al.* [15]. As we have noticed already, a similar problem arises in the Poisson process approach. Note that the different strata in a particular stratification model in the Poisson process approach are the different reference or comparison sets from the case cross-over viewpoint. For a model with large strata, the comparison sets are large, leading to bias due to trend or seasonality. This trend or seasonality component, if it exists, makes the baseline intensity within a stratum time-dependent, which violates the model assumption and leads to bias. With smaller strata or comparison sets, the covariate values within a stratum tend to be similar leading to the problem of overmatching with virtually no information on the covariate effect and, hence, loss of precision. In the previous section, we discussed some ways to resolve this trade-off issue.

7. CONCLUDING REMARKS

As noted in reference [5], one can allow some deviations from the Poisson process assumption. For example, the baseline intensity of a subject may be allowed to change after an event, depending possibly on the number of accumulated events. Then, the corresponding profile

likelihood, following the same technique as in reference [5], turns out to be

$$\prod_{i=1}^{n}\prod_{j=0}^{d_i-1}\left\{\frac{\exp[x_{t_{i,j+1}}^{\mathrm{T}}\beta]}{\int_{t_{i,j}}^{t_{i,j+1}}\exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}\right\}$$

where d_i now denotes the number of events in the ith subject at times $t_{i,1}<\cdots<t_{i,d_i}$ with $t_{i,0}=0$. This is similar to the stratification modelling of reference [1]. Note that this likelihood also looks like a case cross-over likelihood with each event being a case time and the corresponding control period being the time interval since the preceding event time (or time zero for the first event) until the current event time in the same subject. This is similar to the original case cross-over likelihood of reference [14]. More generally, one can also introduce time dependence in the baseline intensities in a piecewise manner, as in (2). Then, the resulting profile likelihood is again a case cross-over likelihood as before but with the control period being the time interval since the beginning of the current stratum until the current event time. One can think of employing a cross-validation technique (similar to that in Sections 4 and 5) to choose an 'optimum' stratification model.

Instead of recurrent events, if we have fatal events leading to censored survival data, analysis of environmental covariates can be carried out in similar manner. With piecewise constant baseline hazard rates (different for different subjects), the profile likelihood for β is

$$\prod_{i\in D}\left\{\frac{\exp[x_{t_i}^{\mathrm{T}}\beta]}{\int_{\tau_{l_i}}^{t_i}\exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}\right\}$$

where D denotes the set of failed subjects, t_i is the failure time for the ith subject in D and τ_{l_i} is the starting time of the stratum in which t_i falls. This also looks like a case cross-over likelihood with control period for each event being the time interval since the beginning of the current stratum until the event time. If we allow the baseline hazard rates to be the same (piecewise constant) for all the subjects, the corresponding profile likelihood becomes

$$\prod_{l=1}^{K}\prod_{i\in D_l}\left\{\frac{\exp[x_{t_i}^{\mathrm{T}}\beta]}{\sum_{j\in\mathscr{R}(\tau_{l-1})}\int_{\tau_{l-1}}^{\min(\tau_l,t_i)}\exp[x_t^{\mathrm{T}}\beta]\,\mathrm{d}t}\right\}$$

where K, D_l and the τ_l's are the number of strata, the set of subjects having the event in the lth stratum, and the times defining the strata, respectively, t_i's denote the observation times (having the event or not) for the ith subject and $\mathscr{R}(\tau_{l-1})$ is the set of subjects at risk at time $\tau_{l-1}-$. This is similar to both the case cross-over likelihood and the Cox likelihood. Therefore, this can be used for analysis of both environmental and subject-specific (transient or not) covariates. As commented in the previous paragraph, a cross-validation technique can be employed to choose an 'optimum' stratification model. However, the properties of the profile likelihood based estimates need to be investigated.

REFERENCES

1. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.
2. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989; **84**:1065–1073.

3. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association* 1993; **88**:811–820.
4. Lawless JF, Nadeau JC. Some simple robust methods for the analysis of recurrent events. *Technometrics* 1995; **37**:158–168.
5. Dewanji A, Moolgavkar SH. A Poisson process approach for recurrent event data with environmental covariates. *Environmentrics* 2000; **11**:665–673.
6. Stone M. Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B* 1974; **36**:111–147.
7. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984; 75–80.
8. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman and Hall: London, 1990; 42–43, 159.
9. Staniswalis J. The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association* 1989; **84**:276–283.
10. Chaudhuri P, Dewanji A. On a likelihood based approach in nonparametric smoothing and cross-validation. *Statistics and Probability Letters* 1995; **22**:7–15.
11. Stone M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, Series B* 1977; **39**:44–47.
12. Navidi W. Bidirectional case cross-over designs for exposures with time trends. *Biometrics* 1998; **54**:596–605.
13. Lumley T, Levy D. Bias in the case cross-over design: implications for studies of air pollution. *Environmentrics* 2000; **11**:689–704.
14. Maclure M. The case cross-over design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* 1991; **133**:144–153.
15. Mittleman MA, Maclure M, Robins JM. Control sampling strategies for case cross-over studies: an assessment of relative efficiency. *American Journal of Epidemiology* 1995; **142**:91–98.