Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model

Author(s): R. L. Prentice
Source: *Biometrika*, Vol. 69, No. 2 (Aug., 1982), pp. 331-342
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2335407
Accessed: 24-05-2018 04:35 UTC

Biometrika (1982), **69**, 2, pp. 331–42
Printed in Great Britain

Covariate measurement errors and parameter estimation in a failure time regression model

By R. L. PRENTICE

Fred Hutchinson Cancer Research Center, Seattle, Washington, U.S.A.

Summary

Estimation in Cox's failure time regression model is considered when the regression vector is subject to measurement error. A hazard function model is induced for the failure rate given the measured covariate and a partial likelihood function is derived for the relative risk parameters. This partial likelihood function may involve the baseline hazard function as well as the regression parameter, but useful inference techniques arise for testing whether the regression parameter equals zero and for more general inferences in important special cases. Explicit consideration is given to testing equality of survival curves when group membership is subject to misclassification and to relative risk estimation with normally distributed covariates. Approaches to regression estimation using the overall likelihood function, and a marginal likelihood function based on failure time ranks, are also indicated. Illustration of the possible effect of covariate errors on relative risk estimation is provided.

Some key words: Censoring; Classification error; Covariate error; Failure time; Hazard function; Log rank test; Marginal likelihood; Partial likelihood; Proportional hazards model; Regression; Relative risk.

1. Introduction

Suppose that a covariate vector z is related to a continuous failure time random variable $T \geqslant 0$ according to Cox's (1972) hazard function regression model

$$\lambda(t;\, z) = \lambda_0(t) \exp\{z^*(t)\,\beta\},$$

where $\lambda_0(\,.\,) \geqslant 0$ is an arbitrary baseline hazard function, $z^*(t) = \{z_1^*(t), \ldots, z_p^*(t)\}$ is a p vector consisting of functions of z or product terms between such functions and t, and β is a corresponding column p vector to be estimated. More generally a model of this type may be used to relate a covariate function $Z(t) = \{z(u);\, u \leqslant t\}$ to failure time via the hazard function

$$\lambda\{t;\, Z(t)\} = \lambda_0(t) \exp\{z^*(t)\,\beta\}, \tag{1}$$

where the p vector $z^*(t)$ may consist of functions of $Z(t)$, the covariate process up to time t, and interaction terms between $Z(t)$ and t. Note that the multiplicative factor $\exp\{z^*(t)\,\beta\}$ can be thought of as the relative risk of failure at t associated with $Z(t)$, as compared to a standard covariate function for which $z^*(t) \equiv 0$. The flexibility of (1), corresponding to the fact that $\lambda_0(\,.\,) \geqslant 0$ need not be restricted, the availability of convenient estimation techniques with good asymptotic properties (Cox, 1975; Efron, 1977; Oakes, 1977) and the existence of some important generalizations (Kalbfleisch & Prentice, 1980, Chapters 5, 7, 8) has led to widespread interest in the use of (1).

In many applications of (1) the covariate z, or covariate function $Z(t)$, will be subject to measurement error. For example, in the study that motivated this work, survivors of the atomic bombs in Hiroshima and Nagasaki were assigned radiation dose levels on the basis of location within their respective cities at the time of bombing and on the basis of individual shielding characteristics. Even if the physical assumptions concerning air dose as a function of distance from the epicentre are accurate it is acknowledged, by S. Jablon in an unpublished Atomic Bomb Casualty Commission report, that individual radiation dose estimates may differ from actual exposure levels by as much as 30% in an appreciable fraction of study subjects. It is then of interest to examine the possible effect of such measurement errors on estimated relative risks due to radiation, and to use information on or assumptions about the error distribution to develop improved risk estimators. The air dose assumptions have in fact been the subject of considerable recent controversy (Marshall, 1981).

2. A partial likelihood function with covariate errors

Suppose now that rather than the 'true' covariate, $Z(t)$, one has available only an 'estimate' $X(t) = \{x(u);\ u \leqslant t\}$. One would like to estimate the parameters β and $\lambda_0(\,.\,)$ in (1) but direct estimation of $\lambda\{t;\ Z(t)\}$ is prevented by measurement errors in the covariates. It is then natural to consider conditions under which inference on the hazard function $\lambda\{t;\ X(t)\}$, which is amenable to direct estimation, can lead to useful inferences on $\lambda\{t;\ Z(t)\}$.

The basic assumption that allows these two hazard functions to be related asserts a conditional independence, given $Z(t)$, of failure rate at t and $X(t)$; that is,

$$\lambda\{t;\ Z(t), X(t)\} = \lambda\{t;\ Z(t)\}. \tag{2}$$

This assumption requires the 'observed' covariate $X(t)$ to have no predictive value given the true covariate value. In general, one can write $\lambda\{t;\ X(t)\}$ as the conditional expectation of $\lambda\{t;\ Z(t), X(t)\}$, given $T \geqslant t$ and $X(t)$; that is,

$$\lambda\{t;\ X(t)\} = E_{\{T \geqslant t,\, X(t)\}} \lambda\{t;\ Z(t), X(t)\},$$

so that under (2)

$$\lambda\{t;\ X(t)\} = E_{\{T \geqslant t,\, X(t)\}} \lambda\{t;\ Z(t)\}.$$

The model (1) thereby induces a hazard function model

$$\lambda\{t;\ X(t)\} = \lambda_0(t)\, E_{\{T \geqslant t,\, X(t)\}} \exp\{z^*(t)\,\beta\}. \tag{3}$$

Note that the induced model is also a multiplicative hazards model, with a relative risk at time t,

$$E_{\{T \geqslant t,\, X(t)\}} \exp\{z^*(t)\,\beta\}, \tag{4}$$

that is a suitably weighted average of possible relative risks in (1), given the measured covariate $X(t)$ and given lack of failure prior to t. To this point no assumptions, except (2), have been made on the distribution of errors in the covariates. From (4) and uniqueness results for moment generating functions it is evident that one must minimally specify the marginal distributions for $z^*(t)$ given $\{T \geqslant t, X(t)\}$ at each t in order that the induced hazard function be specified up to the parameters β and $\lambda_0(\,.\,)$.

Even though (3) is of the same form as (1), the presence of $\{T \geqslant t\}$ in the conditioning event will usually imply some dependence of the relative risk function (4) on the baseline hazard function $\lambda_0(.)$. One can, nevertheless, derive a partial likelihood function for (4) using the argument of Cox (1975), and subsequently examine the dependence of inference techniques of interest on $\lambda_0(.)$. As usual some assumption on the censoring mechanism is necessary for likelihood specification. The most convenient censorship assumption asserts

$$\lambda\{t;\, X(t), \text{ no censorship in } [0,t)\} = \lambda\{t;\, X(t)\}, \tag{5}$$

so that a subject at risk at time t can be regarded as representative of the population without failure at time t and having measured covariate value $X(t)$.

Suppose now that $t_1 < \ldots < t_k$ are the ordered distinct failures in the sample and let $R(t_i)$ and $F(t_i)$ denote the risk set just prior to t_i and the set of subjects failing at t_i respectively, $i = 1, \ldots, k$. If we assume independent covariate errors on distinct study subjects and use an approximation to accommodate tied failure times (Breslow, 1974) the partial likelihood function can be written

$$L(\beta) = \prod_{i=1}^{k} \left[\prod_{l \in F(t_i)} E \exp\{z_l^*(t_i)\,\beta\} \right] \Big/ \left[\sum_{l \in R(t_i)} E \exp\{z_l^*(t_i)\,\beta\} \right]^{m_i}, \tag{6}$$

where m_i is the number of failures at t_i $(i = 1, \ldots, k)$ and the fact that the expectation over the distribution of $z_l^*(t_i)$ is conditional on $\{T \geqslant t_i, X_l(t_i)\}$ has been suppressed. Note that no approximation is involved in (6) if $m_i \equiv 1$ and that, otherwise, the approximation will be good if the ratios $m_i n_i^{-1}$ $(i = 1, \ldots, k)$ are small, where n_i is the size of the risk set $R(t_i)$.

Application of standard asymptotic likelihood methods to (6) for inference on the p vector β will require some conditions, presumably mild, on the censoring mechanism, on the covariate generating mechanism and on the form of the relative risk function (4); see, for example, Cox (1975) and Tsiatis (1981). Write $a = E[\exp\{z^*(t)\,\beta\}]$, $b = \partial a/\partial \beta$ and $c = \partial b/\partial \beta$, where the expectation is conditional on $\{T \geqslant t, X(t)\}$ and b and c are respectively $p \times 1$ and $p \times p$. The score statistic can be written

$$\partial \log L(\beta)/\partial \beta = \sum_{i=1}^{k} \left(\sum_{l \in F(t_i)} b_{il}\, a_{il}^{-1} - m_i \sum_{l \in R(t_i)} b_{il} \Big/ \sum_{l \in R(t_i)} a_{il} \right), \tag{7}$$

and the corresponding observed information matrix is

$$\begin{aligned} -\partial^2 \log L(\beta)/\partial \beta^2 = \sum_{i=1}^{k} \Bigg\{ & m_i \sum_{l \in R(t_i)} c_{il} \Big/ \sum_{l \in R(t_i)} a_{il} - m_i \sum_{l \in R(t_i)} b_{il} \sum_{l \in R(t_i)} b'_{il} \Big/ \Big(\sum_{l \in R(t_i)} a_{il} \Big)^2 \\ & - \sum_{l \in F(t_i)} (c_{il}\, a_{il}^{-1} - b_{il}\, b'_{il}\, a_{il}^{-2}) \Bigg\}, \end{aligned} \tag{8}$$

where, for example, $a_{il} = E[\exp\{z_l^*(t_i)\,\beta\}]$ and b' denotes the vector transpose of b.

These expressions can be used for testing and estimation on β provided that a, b and c can be specified as functions of β. As mentioned above, however, the relative risk function (4) will typically depend to some extent on the baseline hazard function $\lambda_0(.)$, as well as on β and the covariate error distribution. In fact, the expectation operation in (4) will usually involve both β and $\lambda_0(.)$. Since the primary reason for using the partial likelihood technique is avoidance of the 'nuisance' function $\lambda_0(.)$, the use of (6) will usually be restricted to situations in which the quantities a, b and c are exactly or approximately independent of $\lambda_0(.)$.

In order to illustrate the form of the relative risk function (4) and its dependence on $\lambda_0(.)$, consider the special case of a fixed covariate z. Let $f(z\,|\,x)$ denote the probability function, or probability density function, for z given the corresponding measured covariate x. Under (2) and (1) one can write

$$f(z \mid T \geqslant t, x) = K \operatorname{pr}(T \geqslant t \mid z) f(z \mid x) = K \exp\left\{-\int_0^t e^{z^*(u)\beta} \lambda_0(u)\, du\right\} f(z \mid x), \tag{9}$$

where $K = K(t, x)$ is an integration constant. The corresponding relative risk (4) is given by

$$\int e^{z^*(t)\beta} f(z \mid T \geqslant t, x)\, dz,$$

which may be a rather complicated function of β and $\lambda_0(\,.\,)$. The dependence on $\lambda_0(\,.\,)$ arises since, at any measured covariate x and specified t, subjects have been selectively eliminated from the risk set according to their underlying hazard rates $[\lambda_0(u) \exp\{z^*(u)\beta\};\ u \leqslant t]$. Given the measured covariate x, subjects with relatively large hazard rates (1) will then tend to be underrepresented in the risk sets $R(t)$ $(t > 0)$, with the degree of underrepresentation depending on both β and $\{\lambda_0(u);\ u \leqslant t\}$. That is, the covariate distribution in the risk set $R(t)$, given x, will usually depend to some extent on both β and $\lambda_0(\,.\,)$.

3. Score test for $\beta = 0$

An important exception occurs if $\beta = 0$, in which case the induced hazard function is merely $\lambda\{t;\ X(t)\} = \lambda_0(t)$ and there is no selective elimination from the risk set $R(t)$, $t > 0$ on the basis of underlying $Z(t)$ values. In general, one can write

$$a = a(t, \beta) = E[\exp\{z^*(t)\beta\}] = \int \exp\{z^*(t)\beta\} f\{z^*(t) \mid T \geqslant t, X(t)\}\, dz^*(t),$$

where the dependence of the conditional probability density function for $z^*(t)$ on β and $\lambda_0(\,.\,)$ has been suppressed. It follows that

$$\begin{aligned} b &= b(t, \beta) \\ &= \int z^*(t)' \exp\{z^*(t)\beta\} f\{z^*(t) \mid T \geqslant t, X(t)\}\, dz^*(t) \\ &\quad + \int \exp\{z^*(t)\beta\}\, \partial \log f\{z^*(t) \mid T \geqslant t, X(t)\}/\partial\beta\, f\{z^*(t) \mid T \geqslant t, X(t)\}\, dz^*(t). \end{aligned}$$

Since f is a probability function one has, at $\beta = 0$, $a = a(t, 0) \equiv 1$ and

$$b = b(t, 0) = \int z^*(t)' f\{z^*(t) \mid T \geqslant t, X(t)\}\, dz^*(t) = E\{z^*(t)'\}.$$

The score statistic (7) at $\beta = 0$ can therefore be written

$$v = \sum_{i=1}^{k} \left[\sum_{l \in F(t_i)} E\{z_l^*(t_i)\} - m_i n_i^{-1} \sum_{l \in R(t_i)} E\{z_l^*(t_i)\}\right]', \tag{10}$$

which does not depend on $\lambda_0(\,.\,)$. A finite population variance argument applied to $\Sigma E\{z_l^*(t_i)\}$, $l \in F(t_i)$, given $R(t_i)$ and $X_l(t_i)$, $l \in R(t_i)$, for each $i = 1, \ldots, k$ and the uncorrelatedness of the score statistic contributions $i = 1, \ldots, k$ (Cox, 1975) leads to

$$V = \sum_{i=1}^{k} (n_i - m_i)(n_i - 1)^{-1} m_i n_i^{-1}$$
$$\times \left[\sum_{l \in R(t_i)} E\{z_l^*(t_i)' z_l^*(t_i)\} - n_i^{-1} \sum_{l \in R(t_i)} E\{z_l^*(t_i)'\} \sum_{l \in R(t_i)} E\{z_l^*(t_i)\} \right]$$

as a corresponding variance estimator. This expression can also be derived by taking finite population expectations in (8) apart from the factors $(n_i - m_i)(n_i - 1)$, which are absent because of the tied data approximation used in the derivation of (6).

Evidently the hypothesis $\beta = 0$ in Cox's regression model (1) can be validly and efficiently tested in spite of covariate measurement errors, provided sufficient information or assumption is available on the covariate error distributions to ensure that the expectations

$$E\{z_l^*(t_i)\} = E\{z_l^*(t_i) \mid T \geqslant t_i, X_l(t_i)\}$$

can be specified for all $i = 1, \ldots, k$; $l \in R(t_i)$. The statistic $v' V^{-1} v$ will quite generally have an asymptotic χ_p^2 distribution under $\beta = 0$, when V is nonsingular. It is of interest to consider conditions on the error distribution under which v will reduce to the usual score statistic from (1) with $z^*(t)$ replaced by its estimated value $x^*(t)$. Alternatively, one can examine how information on the covariate error distribution can be used to derive an improved test.

The statistic (10) will reduce exactly, that is for all samples, to

$$D \sum_{i=1}^{k} \left\{ \sum_{l \in F(t_i)} x_l^*(t_i) - m_i n_i^{-1} \sum_{l \in R(t_i)} x_l^*(t_i) \right\}'$$

for some nonsingular $p \times p$ matrix D if and only if $E\{z^*(t)\} = Dx^*(t) + d(t)$ at all possible failure times, where $d(t)$ is some row p-vector function. Under these circumstances the usual score test ignoring covariate errors will be valid and can be expected to have good local power properties under (1).

As an important special case, consider the test of equality of survival curves among $p+1$ samples when sample membership is subject to misclassification. Suppose the samples are labelled $0, 1, \ldots, p$ and let $z^*(t) = z$ be a p vector with rth component equal to 1 if the subject's 'true' membership is sample r and equal to 0 otherwise. Let x be the corresponding p vector with sth component equal to 1 if the subject is assigned, possibly erroneously, to sample s, and equal to 0 otherwise. Let d_{rs} be the probability that a subject recorded as being in sample s is actually in sample r; that is, $d_{rs} = \mathrm{pr}(z_r = 1 \mid x_s = 1)$ for $r, s = 0, \ldots, p$. Let D be the $p \times p$ matrix with (r, s) entry $d_{rs} - d_{r0}$ for $r, s, = 1, \ldots, p$. From (9) one can easily show that, under $\beta = 0$, $f(z \mid T \geqslant t, x) = f(z \mid x)$, so that

$$E(z' \mid T \geqslant t, x) = E(z' \mid x) = Dx' + d_0, \tag{11}$$

where $d_0' = (d_{10}, \ldots, d_{p0})$. The score test for $\beta = 0$ from (1) in these circumstances is the so-called log rank test (Peto & Peto, 1972). The form of the expectation function (11) indicates the log rank test to be also the score statistic from (6) when the sample membership is subject to misclassification. As such one can expect the log rank statistic to retain good properties under proportional hazards alternatives, even in the presence of substantial misclassification.

The same type of robustness to misclassification will prevail for the time-dependent log rank test, for the comparison of $p+1$ samples when sample membership may change over follow-up time, provided the misclassification probabilities do not vary over time. More specifically suppose, in a notation similar to that given above, that

$$\begin{aligned} d_{rs}(t) &= \mathrm{pr}\{z_r(t) = 1 \mid x_s(t) = 1\} \\ &= \mathrm{pr}\{z_r(t) = 1 \mid x_s(t) = 1, x(u), u < t\} \end{aligned}$$

for $r, s = 0, 1, \ldots, p$. One then has, under $\beta = 0$,

$$E\{z(t)' \mid T \geqslant t, X(t)\} = E\{z(t)' \mid x(t)\} = D(t)\, x(t)' + d_0(t).$$

This expectation function is not of the form required for (9) to be equivalent to the usual time-dependent log rank test unless $D(t) \equiv D$; that is, unless the misclassification rates are independent of t. Rather (9) becomes

$$\sum_{i=1}^{k} D(t_i) \left\{ \sum_{l \in F(t_i)} x_l(t_i) - m_i n_i^{-1} \sum_{l \in R(t_i)} x_l(t_i) \right\}'$$

which is a weighted sum over the k failure times of observed minus conditionally expected numbers of failures in each sample with weight matrix, $D(t_i)$, that reflects the precision with which sample membership can be assigned at follow-up time t_i. Knowledge of misclassification probabilities can thereby be used to derive a test which is expected to have better local power properties than the ordinary time-dependent log rank test, particularly if the $D(t_i)$ matrices differ substantially over $i = 1, \ldots, k$.

4. Estimation of β

The partial likelihood function (6) can be used in a standard manner for β estimation merely by specifying the error distributions $f\{z^*(t) \mid T \geqslant t, X(t)\}$, so that the relative risk (4) is prescribed up to the regression parameter β. The fact that these error distributions, and hence the expectation operation in (4), typically depend on β and $\lambda_0(\,.\,)$ indicates, however, that one needs to be concerned with consistency of such specification with (1). In a variety of circumstances of practical importance, however, the dependence of the error distributions on β and $\lambda_0(\,.\,)$ will be negligible.

Consider first a fixed covariate z. From (9) it is evident that the distribution of z, given $\{T \geqslant t, x\}$ will be independent of β and $\lambda_0(\,.\,)$ whenever $\mathrm{pr}\,(T \geqslant t \mid z) = \mathrm{pr}\,(T \geqslant t \mid x)$. This equality will hold if $\beta = 0$, and approximately if $f(z \mid x)$ is concentrated, or if $p(T \geqslant t \mid z) \doteq 1$ over the follow-up period. This last condition is quite adequately fulfilled in the large cohort study mentioned in § 1.

Conditions under which the dependence of $f\{z^*(t) \mid T \geqslant t, X(t)\}$ on β and $\lambda_0(\,.\,)$ can be ignored will be more liberal for stochastic time-dependent covariates. For example, suppose components of $z^*(t)$ are all measured at time points $\tau_0 = 0, \tau_1, \ldots, \tau_s$, that may differ among study subjects. Knowledge of the measuring process may be used to specify $f\{z^*(\tau_j) \mid T \geqslant \tau_j, X(\tau_j)\}$ $(j = 0, \ldots, s)$. These distributions will not involve the parameters of (1). For $t \in [\tau_{j-1}, \tau_j)$ the error distribution $f\{z^*(t) \mid T \geqslant t, X(t)\}$ will depend on β and λ only to the extent that selective withdrawal from the risk set at specified $X(t)$, occurs prior to τ_j $(j = 0, \ldots, s)$. If components of $z^*(t)$ are measured at time points that are frequent, relative to the corresponding failure rates from (1), it will then usually be appropriate to specify covariate error distributions without regard to consistency with (1).

In other circumstances some *ad hoc* accommodation of the dependence of covariate error distributions on β and $\lambda_0(\,.\,)$ may be useful. Some more formal possibilities for accommodating (1) in β estimation are mentioned in §5.

In the remainder of this section it will be supposed that the covariate error distribution dependence on β and $\lambda_0(\,.\,)$ is negligible. Uniqueness results for moment generating functions indicate that $f\{z^*(t) \mid T \geqslant t, X(t)\}$ must be fully specified in order to determine (4) as a function of β. Any error distribution with convenient moment generating function can be readily utilized. Assuming the expectation and differentiation can be interchanged one can insert

$$a = E[\exp\{x^*(t)\,\beta\}], \quad b = E[z^*(t)'\exp\{z^*(t)\,\beta\}], \quad c = E[z^*(t)'\,z^*(t)\exp\{z^*(t)\,\beta\}]$$

in the log likelihood derivatives (7) and (8).

Consider the special case of normally distributed covariate errors. More precisely, suppose that $z^*(t)$ given $\{T \geqslant t, X(t)\}$ is normal with mean $\mu\{t, X(t)\}$ and variance $\Omega\{t, X(t)\}$. The relative risk function (4) is then

$$\exp[\mu\{t, X(t)\}\,\beta + \tfrac{1}{2}\beta'\,\Omega\{t, X(t)\}\,\beta]. \tag{12}$$

In the very special case in which $\mu\{t, X(t)\} - x^*(t)$ and $\Omega\{t, X(t)\}$ are independent of $X(t)$ the corresponding partial likelihood (6) will be identical to that based on (1) with $z^*(t)$ replaced by $x^*(t)$. A more interesting special case, illustrated in §6, presumes $\mu\{t, X(t)\} = x^*(t)$ and $\Omega\{t, X(t)\} = D\{t, X(t)\}\,C\,D\{t, X(t)\}$, where $D\{t, X(t)\} = \operatorname{diag}\{x^*(t)\}$ and C is positive-definite. Such a specification arises if ratios of components of $z^*(t)$ to the corresponding components of $x^*(t)$ have a fixed normal distribution with mean the 1-vector and variance matrix C. Then application of the partial likelihood (6) involves fitting a model in which the log relative risk is a quadratic function of $x^*(t)$. Application of (1) without acknowledgement of covariate errors, on the other hand, would fit a model in which the log relative risk is only linear in $x^*(t)$.

In order to apply asymptotic likelihood methods to (6) with relative risk function (12) one need only note that for all $i = 1, \ldots, k$; $l \in R(t_i)$

$$a_{il} = \exp(\mu_{il}\,\beta + \tfrac{1}{2}\beta'\,\Omega_{il}\,\beta)$$

from which, by differentiation,

$$b_{il} = (\mu_{il}' + \Omega_{il}\,\beta)\,a_{il}, \quad c_{il} = \{\mu_{il}' + \Omega_{il}\,\beta\}\,\{\mu_{il}' + \Omega_{il}\,\beta\}'\,a_{il} + \Omega_{il}\,a_{il},$$

where μ_{il} and Ω_{il} have been written for $\mu\{t_i, X_l(t_i)\}$ and $\Omega\{t_i, X_l(t_i)\}$, respectively.

In the notation of §3, the $p+1$ sample problem relative risk function can be written

$$E\{\exp(z\beta)\} = \sum_{r=0}^{p}\sum_{s=0}^{p} e^{\beta_r}\,d_{rs}\,x_s,$$

where $\beta_0 = 0$ and $x_0 = 1 - \Sigma_i\,x_i$, with the sum over $i = 1, \ldots, p$, provided that the dependence of $f(z \mid T \geqslant t, x)$ on β and $\lambda_0(\,.\,)$ can be ignored. Estimation of $\beta' = (\beta_1, \ldots, \beta_p)$ is readily carried out for specified classification probabilities d_{rs} $(r, s = 0, \ldots, p)$.

Before illustrating the possible effects of normally distributed covariate errors on β estimation some discussion of each of the assumptions (1), (2) and (5) is given. Some comments on covariate error distribution specification and on alternative possibilities for β estimation are also given.

5. Discussion and further results

Not surprisingly the possible effect of covariate errors on relative risk estimation can depend markedly on the parametric form of the underlying relative risk function. For example, suppose the exponential function in (1) is replaced by a linear function, so that

$$\lambda\{t;\, Z(t)\} = \lambda_0(t)\,\{1+z^*(t)\,\beta\}. \tag{13}$$

The induced hazard function, given $X(t)$, is then, under assumption (2),

$$\lambda\{t;\, Z(t)\} = \lambda_0(t)\,[1+E_{\{T\geqslant t,\, X(t)\}}\{z^*(t)\,\beta\}]. \tag{14}$$

Note that it will be necessary to restrict the range of β so that $1+E\{z^*(t)\,\beta\}\geqslant 0$ for all values occurring in the sample. It is then only necessary that $E\{z^*(t)\} = x^*(t)$ in order that the partial likelihood from (14) reduce to that based on (13) with $z^*(t)$ replaced by $x^*(t)$. This suggests that β estimation based on (13) will be less sensitive to covariate error distributional characteristics than is the case for inference based on (1).

Assumption (2), concerning the conditional independence of the failure rate on $X(t)$ given $Z(t)$, seems very natural indeed. On the other hand, it is possible to think of situations wherein

$$\lambda\{t;\, Z(t), X(t)\} \neq \lambda_0(t)\exp\{z^*(t)\,\beta\}$$

either because of inadequate modelling of the relationship between failure rate and $Z(t)$ or because $X(t)$ conveys some additional predictive information and is, in this sense, more than simply an estimator of $X(t)$. In these latter circumstances it would be reasonable to specify a model for $\lambda\{t, Z(t), X(t)\}$ from which an induced model $\lambda\{t, X(t)\}$ can be obtained.

Assumption (5) concerning censorship is convenient since the relative risk function (4) then directly enters the partial likelihood (6). In some situations, however, a more realistic censoring assumption will be

$$\lambda\{t, Z(t), \text{ no censoring in } [0,t)\} = \lambda\{t;\, Z(t)\} \tag{15}$$

in which case, under (1)

$$\lambda\{t;\, X(t), \text{ no censorship in } [0,t)\} = \lambda_0(t)\,E[\exp\{z^*(t)\,\beta\}],$$

where now the expectation is conditional on the absence of censorship in $[0,t)$ as well as on $\{T\geqslant t, X(t)\}$. A censorship mechanism that depends strongly on $Z(t)$ and the presence of substantial errors in covariate measurement could then induce, under (15), noticeably dependent censorship for failure time given $X(t)$. Such dependency, however, seems unlikely to be of practical importance in most applications.

The observations, given above, on the linear relative risk model (13) are consistent with measurement error results in the ordinary linear model, as reviewed by Cochran (1968). For example, in simple linear regression the distribution of x given z and the marginal distribution of z is usually specified in a way that leads to $E(z\,|\,x) = (1-c)\,\mu+cx$, where $0\leqslant c\leqslant 1$ is $\operatorname{cov}(z,x)$ divided by $\operatorname{var}(x)$ and μ is the mean of z. The coefficient of the regression on the measured covariate x is then 'deflated' by the factor c, relative to the regression on z. If the dependence of $f(z\,|\,T\geqslant t, x)$ on β and $\lambda_0(\,.\,)$ is negligible, a similar deflation by the factor $c' = c/\{1+(1-c)\,\mu\beta\}$ will also occur under these assumptions, with (13). Note that $0\leqslant c'\leqslant 1$ since $\mu\beta\geqslant -1$, and that $c'=0$ and $c'=1$ correspond to $c=0$ and $c=1$, respectively. In the so-called Berkson model $c=1$ and $E(z\,|\,x)=x$ and no deflation occurs. The condition mentioned above that

$E\{z^*(t) \mid T \geqslant t, X(t)\} = x^*(t)$ is a time-dependent generalization of the Berkson specification. With fixed covariates, for example, this generalized condition could not be expected to hold exactly for all t, even if $E(z \mid x) = x$, owing to the dependence of $f(z \mid T \geqslant t, x)$ on β and $\lambda_0(\,.\,)$.

It is evident under (13), or (1), or other parametric relative risk functions, that covariate error properties enter the partial likelihood (6) only through the conditional distributions of $z^*(t)$, given $\{T \geqslant t, X(t)\}$. It is, therefore, unnecessary to fully specify the joint distributions $\{X(t), Z(t) \mid T \geqslant t\}$ though it will often be natural to utilize, at least informally, distributional assumptions on the measured covariate given the true covariate and on the marginal distribution of the true covariate, in order to arrive at a suitable specification for $f\{z^*(t) \mid T \geqslant t, X(t)\}$.

An undesirable feature of the partial likelihood (6) is its dependence on the baseline hazard function $\lambda_0(\,.\,)$ in (1). Even though such dependence will often be negligible in practice, it is of interest to develop estimation techniques that are fully consistent with the basic hazard function (1): with a fixed covariate z, and with $z^*(t) \equiv z^*$ independent of t, the 'marginal likelihood' approach of Kalbfleisch & Prentice (1973) is readily adapted. More specifically, upon interchanging the expectation operators and the 'order-statistic' integrals in Kalbfleisch and Prentice's development, one obtains a marginal likelihood that can be written

$$E_{\{x_l,\, l \in R(t_1)\}} \prod_{i=1}^{k} \left\{ \prod_{l \in F(t_i)} e^{z_l^* \beta} \Big/ \left(\sum_{l \in R(t_i)} e^{z_l^* \beta} \right)^{m_i} \right\}. \tag{16}$$

where an approximation to accommodate tied data has again been made. This expression is entirely free of dependence on the baseline hazard function $\lambda_0(\,.\,)$. The score test results of §3 for a fixed covariate $z^*(t) = z^*$ could equally well have been derived from (16). More generally, it would be of interest to consider error distribution assumptions, $f(z \mid x)$, that give rise exactly or approximately to convenient estimation based on (16).

One could consider similarly the full likelihood function in order to carry out joint estimation on β and $\lambda_0(\,.\,)$. For example, with a fixed covariate z the overall likelihood function can be written

$$\prod_{i=1}^{n} E_{x_i} \left[\{\lambda_0(t_i)\, e^{z_i^*(t_i)\beta}\}^{\delta_i} \exp\left\{ -\int_0^{t_i} e^{z_i^*(u)\beta}\, \lambda_0(u)\, du \right\} \right], \tag{17}$$

where now $t_1, \ldots, t_n$ denote the observed failure or censoring times on the n study subjects, while δ_i and $z_i^*(u)$ are corresponding censoring indicators and covariate vectors at time u, respectively. One could apply ordinary likelihood methods to (17), upon specifying $f(z \mid x)$ and $\lambda_0(\,.\,)$ up to a finite set of parameters. One could also consider a Kaplan–Meier type estimation procedure for $\lambda_0(\,.\,)$ with β estimation from (6), or even an iterative procedure with a Kaplan–Meier type estimator from (17) for $\lambda_0(\,.\,)$, using a trial value of β, which in turn is used, in conjunction with (6), to obtain an updated β value. Such procedures have, however, not been pursued.

6. Illustration

Model (1) was applied to data on thyroid cancer incidence among survivors of the atomic bombs in Hiroshima and Nagasaki in order to relate thyroid cancer incidence to gamma radiation exposure level. The analyses presented here are for illustrative purposes only. A detailed analysis of these data, that takes account of neutron exposure level and other study subject characteristics, will be presented elsewhere.

A total of 105 clinically evident thyroid cancer cases, in a cohort of 98,611 persons, were identified by tumour registries in the two cities during the follow-up period for this study. Model (1) was applied with $z^*(t) = z$ being the true gamma dose in rads to the thyroid gland while $x^*(t) = x$ is the corresponding gamma dose estimate. These exposure estimates make use of the individuals' location and shielding characteristics at the time of bombing, as well as a number of physical assumptions on the air dose curves and on the fraction of whole body radiation attenuated by the thyroid gland. In the analyses that follow the baseline hazard function $\lambda_0(.)$ is allowed to depend arbitrarily on city and sex.

Assumption (2) concerning the conditional independence of thyroid cancer incidence on the estimated gamma dose x, given the true gamma dose z, seems appropriate. Assumption (5) on the censorship is probably mildly violated since most censorship in this sample corresponds to deaths without thyroid cancer. Such mortality, particularly for leukaemia and certain other types of cancer can be expected to depend on z, but not on x given z. This means that the set of study subjects without censorship at time t with an estimated exposure level x may not be entirely representative of the original distribution of z given x at $t = 0$ because of selective risk set exclusion via censorship. Furthermore, it means that $\lambda(t;\, x)$ should be interpreted as the thyroid cancer incidence rate at x in the presence of mortality without thyroid cancer. Since most study subjects survived the follow-up period for the study, and since most mortality is unrelated or is only moderately related to radiation dose, the impact of this dependency in the censorship should be slight.

The same type of consideration arises in relation to the dependence of the relative risk (4) on the thyroid cancer incidence rate function (1). Since the probability of not developing thyroid cancer $\text{pr}\,(T \geqslant t \mid z)$ is likely to exceed 0·98 for the entire study follow-up period, even for heavily exposed persons, it is quite adequate to approximate $f(z \mid T \geqslant t, x)$ by $f(z \mid x)$ on the basis of (9). In summary, the distribution of z, given $(T \geqslant t, x)$, will almost certainly not change in a practically important manner over the follow-up period, either on the basis of thyroid cancer occurrence or on the basis of mortality without thyroid cancer.

It is then necessary only to specify a distribution of z given x in order to carry out partial likelihood estimation of β. The first three columns on the left-hand side of Table 1 assume z given x to be normally distributed with mean x and variance $x^2\sigma^2$ according to

Table 1. *Maximum partial likelihood estimates for thyroid cancer relative risk as a function of gamma dose; values given are for covariate $\hat{\beta}$, with estimated standard errors in parentheses; σ indicates error scale*

	$\lambda_0(t)\,E\exp(z\beta)$ model				$\lambda_0(t)\,E(1+z\beta)$ model			
	$\sigma = 0$	$\sigma = 0{\cdot}2$	$\sigma = 0{\cdot}4$	$\sigma = 0$	$\sigma = 0$	$\sigma = 0$	$\sigma = 0{\cdot}2$	$\sigma = 0{\cdot}4$
Gamma dose in rads/100	0·525 (0·076)	0·410 (0·055)	0·353 (0·046)	1·370 (0·243)	2·069 (0·461)	2·738 (0·717)	2·738 (0·717)	2·738 (0·717)
(Gamma dose in rads/100)^2				−0·193 (0·062)		−0·314 (0·221)	−0·302 (0·213)	−0·271 (0·191)
Maximized log likelihood $+10^3$	−44·40	−49·34	−51·84	−37·83	−37·34	−36·61	−36·61	−36·61

the fixed distribution relative error assumption mentioned in §4. In these analyses the gamma dose has been truncated at 600 rads and has been standardized by dividing by 100. Maximum partial likelihood estimators for β and estimated standard errors are given for values of $\sigma = 0$, 0·2 and 0·4. These values correspond, respectively, to probabilities of 0%, 13% and 45% of a relative error in excess of 30% for the true exposure level given its estimate. The maximum partial likelihood estimates, $\hat{\beta}$, illustrate the fact that covariate error properties can affect relative risk estimation in a practically important manner. Note that the ratios of $\hat{\beta}$ to its asymptotic standard error are 6·91, 7·45 and 7·67 at $\sigma = 0$, 0·2 and 0·4, respectively. Near constancy of this ratio would be anticipated from the fact that the score test for $\beta = 0$ reduces under these error distribution assumptions to the regular score test from (1) with z replaced by x. The final column on the left-hand side of Table 1 provides evidence against a linear relationship between the log relative risk and gamma exposure, at least if covariate errors are ignored. A quadratic log relative risk model with normally distributed gamma dose levels z given x, could be applied with values $\sigma > 0$ upon inserting

$$E \exp(z\beta_1 + z^2\beta_2) = \exp\{x\beta_1 + x^2\beta_2 + \tfrac{1}{2}\theta_1^2(1-2\theta_2)^{-1} - \tfrac{1}{2}\log(1-2\theta_2)\}$$

into (6), where $\theta_1 = \sigma x(\beta_1 + 2x\beta_2)$ and $\theta_2 = \sigma^2 x^2 \beta$. Because of the rather complicated form of the resulting log likelihood derivatives, and because a linear relative risk model of the form (13) appeared to give a simpler representation of the data, such analyses were not carried out.

The right-hand side of Table 1 gives results of analyses based on (13) with $z^*(t) = z^*$ consisting of linear or linear and quadratic terms in gamma dose z. The analysis with only a linear term in z will apply to any covariate error distribution for which $E(z \mid T \geqslant t, x) = x$. Note that the maximized log likelihood is larger from this linear model than that from even the quadratic log relative risk analysis. More generally, one can consider estimation of the linear gamma dose coefficient under (13) and measurement error assumptions that give $E(z \mid T \geqslant t, x) = (1-c)\mu + cx$, as mentioned in §5. The desired regression coefficient β will be related to β' obtained by ignoring covariate measurement errors by $\beta = \beta'\{c - (1-c)\mu\beta'\}^{-1}$. For example, a value of 0·2 for the mean, truncated and standardized, gamma dose, and a value of 0·8 for the coefficient c along with $\hat{\beta}' = 2{\cdot}069$ from Table 1 gives $\hat{\beta} = 2{\cdot}732$. The corresponding asymptotic standard error estimate from Table 1 and the 'δ method' is $c\{c-(1-c)\mu\}^{-2}(0{\cdot}461) = 0{\cdot}639$.

If a quadratic term in z is added to the linear relative risk function the induced hazard function, assuming $f(z \mid T \geqslant t, x) = f(z \mid x)$ and $E(z \mid x) = x$, is

$$\lambda(t;\, x) = \lambda_0(t)\{1 + x\beta_1 + x^2(1+\sigma^2)\beta_2\},$$

where the variance of z given x has again been taken to be $\sigma^2 x^2$. The effect of varying the error scale parameter on the quadratic coefficient is illustrated in the final three columns of Table 1.

The work was done at the Radiation Effects Research Foundation in Hiroshima, while the author was on leave from the Fred Hutchinson Cancer Research Center and the Department of Biostatistics, University of Washington. He is grateful to Dr H. Kato for permission to use the thyroid cancer data in §6 and to Mark Mason for computer programming for Table 1. The work was partially supported by grants from the National Institutes of Health and by the Department of Energy via the National Academy of Sciences. Helpful suggestions from Dr M. Otake and a referee are also acknowledged.

REFERENCES

BRESLOW, N. E. (1974). Covariate analysis of censored survival data. *Biometrics* **30**, 89–99.
COCHRAN, W. G. (1968). Errors of measurement in statistics. *Technometrics* **10**, 637–66.
COX, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–220.
COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.
EFRON, B. (1977). Efficiency of Cox's likelihood function for censored data. *J. Am. Statist. Assoc.* **72**, 557–65.
KALBFLEISCH, J. D. & PRENTICE, R. L. (1973). Marginal likelihood based on Cox's regression and life model. *Biometrika* **60**, 267–78.
KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
MARSHALL, E. (1981). New A-bomb studies alter radiation estimates. *Science* **212**, 900–3.
OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64**, 441–8.
PETO, R. & PETO, J. (1972). Asymptotic efficient rank invariant test procedures (with discussion). *J. R. Statist. Soc.* A **135**, 185–206.
TSIATIS, A. A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.

[*Received December* 1980. *Revised September* 1981]