Additive Hazards Regression with Current Status Data

D. Y. Lin; David Oakes; Zhiliang Ying


Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199806%2985%3A2%3C289%3AAHRWC%3E2.0.CO%3B2-1

*Biometrika* is currently published by Biometrika Trust.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/bio.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Additive hazards regression with current status data

BY D. Y. LIN

Department of Biostatistics, Box 357232, University of Washington, Seattle, Washington 98195, U.S.A.
danyu@biostat.washington.edu

DAVID OAKES

Department of Biostatistics, University of Rochester Medical Center, 610 Elmwood Avenue, Box 630, Rochester, New York 14642, U.S.A.
oakes@metro.bst.rochester.edu

AND ZHILIANG YING

Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, New Jersey 08855, U.S.A.
zying@stat.rutgers.edu

SUMMARY

Current status data arise when the only knowledge about the failure time of interest is whether the failure occurs before or after a random monitoring time. We propose to analyse such data by the semiparametric additive hazards model, which specifies that the hazard function for the failure time associated with a set of possibly time-dependent covariates is the sum of an arbitrary baseline hazard function and a regression function of covariates. Under certain conditions on the monitoring time, one can make inferences about the regression parameters of the additive hazards model by using the familiar asymptotic theory and software for the proportional hazards model with right censored data. An application to a carcinogenicity experiment is provided.

Some key words: Additive risk model; Counting process; Failure time; Interval censoring; Martingale; Partial likelihood; Proportional hazards; Time-dependent covariate.

1. Introduction

Current status data arise when the failure time of interest \( T \) cannot directly be observed, but can only be determined to lie below or above a random monitoring time \( C \). Such data are commonly encountered in biomedicine, economics, sociology and other scientific areas. For example, in carcinogenicity experiments, animals are randomly assigned to various doses of a suspected carcinogen and are examined at sacrifice or death time for evidence of a malignancy. The time to tumour onset is of interest, but not directly observable. Rather, one knows only the age at death and whether or not the tumour is present at that time.

There is much less information about the failure time \( T \) in current status data than in the familiar right-censored data. In the latter, one observes \( \min(T, C) \) and \( I(C \leq T) \), where
$I(.)$ is the indicator function; the probability of observing $T$ exactly is positive. In the former, however, one only observes $C$ and $I(C \leq T)$; the exact value of $T$ is never observed.

A number of authors, including Ayer et al. (1955), Peto (1973), Turnbull (1976) and Groeneboom & Wellner (1992, § 2.3), have introduced algorithms for the nonparametric maximum likelihood estimation of the distribution function of $T$ with current status data. Groeneboom & Wellner (1992, §§ 4.1, 5.1) showed that the estimator is consistent and converges at $n^{1/3}$-rate to a complicated limiting distribution.

Semiparametric regression methods have also been studied. Finkelstein (1986) developed a method for fitting the proportional hazards model (Cox, 1972) to current status data, but the properties of her estimator are still unknown. Diamond & McDonald (1991) and Shiboski & Jewell (1992) translated the proportional hazards model to a semiparametric binary regression model, and proposed some ad hoc estimation procedures. Huang (1996) recently showed that, profiled over the cumulative baseline hazard function, the maximum likelihood estimator for the regression parameter under the proportional hazards model is asymptotically normal with the $n^2$-convergence rate. Klein & Spady (1993) and Rabinowitz, Tsiatis & Aragon (1995) considered the linear regression model. All the aforementioned regression methods are fairly complicated, requiring the estimation of the distribution function, or even the density, of $T$. Furthermore, none of them handles time-dependent covariates.

In the present paper, we study semiparametric methods for analysing current status data under the additive hazards regression model. This model specifies that the hazard function of $T$ at time $t$, given the history of a $p$-dimensional covariate process $Z(.)$ up to $t$, has the form

$$\lambda(t|Z) = \lambda_0(t) + \beta_0 Z(t), \quad (1.1)$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, and $\beta_0$ is a $p$-vector of unknown regression parameters. Some related models are discussed in § 5. Normally, $Z(.)$ is restricted to external covariates (Kalbfleisch & Prentice, 1980, p. 123).

There are two major motivations for using model (1.1). First, the additive hazards model describes a different aspect of the association between the failure time and covariates than the proportional hazards model, and is more plausible than the latter in many applications (Aranda-Ordaz, 1983; Buckley, 1984; Cox & Oakes, 1984, p. 74; Thomas, 1986; Breslow & Day, 1987, pp. 122–31, 182; Lin & Ying, 1994). Secondly, we have discovered that there exist surprisingly simple methods for making inferences about $\beta_0$ under model (1.1) with current status data.

In the next section, we present the new methodology for the special case in which the monitoring time $C$ is independent of $T$ and $Z$. In § 3, we extend the results of § 2 to allow the dependence of $C$ on $Z$ through the proportional hazards model. In § 4, we report the results of our simulation studies and provide an illustration with real data taken from a carcinogenicity experiment. In § 5, we discuss the extension of our approach to more general interval censored data and give some concluding remarks.

2. Inference Procedures with Independent Monitoring

In this section, we assume that $C$ is independent of $T$ and $Z$. Let $\{T_i, C_i, Z_i(.)\}$ ($i = 1, \ldots, n$) be independent replicates of $\{T, C, Z(.)\}$. The observations consist of $C_i$, $\delta_i := I(C_i \leq T_i)$ and $Z_i(t)$ ($t \leq C_i$).

For $i = 1, \ldots, n$, define the counting process $N_i(t) = \delta_i I(C_i \leq t)$, which jumps by unity
whenever subject \( i \) is monitored at time \( t \) and found still to be failure-free. The process \( N_i(t) \) is regarded as censored when subject \( i \) is monitored and found to have experienced failure. Subject \( i \) is at risk at \( t \) if and only if \( N_i(t) \) is still uncensored and has not taken its jump by time \( t \), which happens if and only if \( C_i \geq t \). Consider now the hazard \( dH_i(t) \), say, of a jump of \( N_i(t) \) for a subject that has not yet been monitored. Two things have to happen for \( dN_i(t) = 1 \), namely (i) \( C_i = t \), and (ii) the resulting monitoring reveals that subject \( i \) has been failure-free up to \( t \). Let us denote the hazard of event (i) by \( d\Lambda_i(t) \). Under model (1.1), event (ii) has the conditional probability
\[
pr \{ T_i \geq t \mid Z_i(s), s \leq t \} = e^{-\Lambda_0(t) - \beta_0 Z_i^*(t)},
\]
where
\[
\Lambda_0(t) = \int_0^t \lambda_0(s) ds, \quad Z_i^*(t) = \int_0^t Z_i(s) ds.
\]
The product of the two probabilities for events (i) and (ii) is of the form
\[
dH_i(t) = e^{-\beta_0 Z_i^*(t)} dH_0(t), \tag{2.1}
\]
where \( dH_0(t) = e^{-\Lambda_0(t)} d\Lambda_0(t) \).

Equation (2.1) is of course the well-known Cox proportional hazards model. Letting \( Y_i(t) = I(C_i \geq t) \), we show in the Appendix that the compensated counting processes
\[
M_i(t) := N_i(t) - \int_0^t Y_i(s)e^{-\beta_0 Z_i^*(s)} dH_0(s) \quad (i = 1, \ldots, n)
\]
are martingales with respect to the \( \sigma \)-filtration
\[
\mathcal{F}_t := \sigma \{ N_i(s), Y_i(s), Z_i^*(s) : s \leq t, i = 1, \ldots, n \}.
\]
Thus, one can make inferences about \( \beta_0 \) by applying the partial likelihood principle to model (2.1) with the data \( \{ N_i(\cdot), Y_i(\cdot), Z_i^*(\cdot) \} \) \( (i = 1, \ldots, n) \). Specifically, the partial likelihood score function and observed information matrix for \( \beta_0 \) are
\[
U(\beta) := \sum_{i=1}^n \int_0^\infty \left\{ Z_i^*(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right\} dN_i(t), \tag{2.2}
\]
\[
\mathcal{J}(\beta) := \sum_{i=1}^n \int_0^\infty \left\{ \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \frac{S^{(1)}(\beta, t^{\otimes 2})}{S^{(0)}(\beta, t^2)} \right\} dN_i(t),
\]
where
\[
S^{(k)}(\beta, t) = \sum Y_j(t)e^{-\beta Z_j^*(t)Z_i^*(t)^{\otimes k}} \quad (k = 0, 1, 2),
\]
and \( a^{\otimes 1} = a, a^{\otimes 2} = aa' \). The maximum partial likelihood estimator \( \hat{\beta} \) is the solution to \( U(\beta) = 0 \). By the counting-process martingale arguments (Andersen & Gill, 1982), the random vectors \( n^{-\frac{1}{2}} U(\beta_0) \) and \( n^{-\frac{1}{2}}(\hat{\beta} - \beta_0) \) converge in distribution to zero-mean normal random vectors with covariance matrices \( \Omega \) and \( \Omega^{-1} \), respectively, where \( \Omega \) is the limit of \( n^{-1} \mathcal{J}(\beta_0) \) and can be consistently estimated by \( n^{-1} \mathcal{J}(\beta_0) \) or \( n^{-1} \mathcal{J}(\hat{\beta}) \). As usual, we assume that \( \Omega \) is nonsingular.

The above asymptotic results can be used to test nonparametrically the null hypothesis that \( \beta_0 = 0 \) and to make semiparametric inferences about the regression parameters. These procedures are much simpler than the existing regression methods for current status data,
not involving any complicated nonparametric estimation of the distribution of $T$. No special software is required to implement the proposed methods. In fact, one can just input the data \( \{C_i, \delta_i, Z_i^\tau(.)\} \) \((i = 1, \ldots, n)\) into standard software for fitting the proportional hazards model with right-censored data. Incidentally, if model (1.1) were replaced by a proportional hazards model, then \( dH_i(t) \) would not be of the proportional hazards form so that partial likelihood methods would not be applicable.

Interestingly, the distribution of the censoring time for \( N_i(t) \) also involves \( \beta_0 \) so that the censoring times are informative about \( \beta_0 \). This implies that \( \hat{\beta} \) may not be fully efficient. It is possible to improve efficiency by utilizing the extra information about \( \beta_0 \) provided by the censoring times, as will be discussed at the end of § 3.

When the monitoring time distribution is degenerate, \( U(\beta) \) has a close connection with the parametric likelihood score function. To see this, suppose that \( C_i = c \), a positive constant. We then have a simple regression problem with binary responses \( I(T_i \geq c) \) \((i = 1, \ldots, n)\). Under the model (1.1), \( \Pr(T_i \geq c \mid Z_i) = e^{-\alpha_0 - \beta_0 Z_i} \), where \( \alpha_0 = \Lambda_0(c) \) and \( \bar{Z}_i = Z_i^\tau(c) \). The likelihood score equations for the unknown parameters \( \alpha_0 \) and \( \beta_0 \) are

\[
\sum_{i=1}^n \left( 1 - e^{-\alpha_0 - \beta_0 \bar{Z}_i} \right)^{-1} \{I(T_i \geq c) - e^{-\alpha_0 - \beta_0 \bar{Z}_i}\} = 0.
\]

Removing the weights \( \left( 1 - e^{-\alpha_0 - \beta_0 \bar{Z}_i} \right)^{-1} \) from the above equations and eliminating the nuisance parameter \( \alpha \), we obtain the following equation for \( \beta_0 \):

\[
\sum_{i=1}^n \left( \bar{Z}_i - \frac{\sum_{j=1}^n e^{-\beta_0 \bar{Z}_j} \bar{Z}_j}{\sum_{j=1}^n e^{-\beta_0 \bar{Z}_j}} \right) I(T_i \geq c) = 0.
\]

The left-hand side of the above equation is exactly what (2.2) reduces to in this special case. Since the maximum likelihood estimator is asymptotically efficient, we expect \( \hat{\beta} \) to have high efficiency at least when \( C \) is not too dispersed.

3. INFERENCES WITH DEPENDENT MONITORING

When the monitoring time \( C \) is not independent of the covariate vector \( Z \), we formulated their relationship through the proportional hazards model. Specifically, the hazard function of \( C \) at time \( t \) conditional on the covariate history up to \( t \) is given by

\[
d\Lambda_c(t \mid Z) = e^{\gamma_0 Z(t)} d\Lambda_{c,0}(t), \quad (3.1)
\]

where \( \Lambda_{c,0}(t) \) is an unspecified baseline cumulative hazard function, and \( \gamma_0 \) is a \( p \)-vector of unknown regression parameters. We require that \( C \) and \( T \) be independent conditional on \( Z \).

By the arguments leading to model (2.1), we can show that, under models (1.1) and (3.1),

\[
dH_i(t) = e^{-\beta_0 \bar{Z}_i(t)} + \gamma_0 Z_i(t) \ dH_0(t), \quad (3.2)
\]

where \( dH_0(t) = e^{-\Lambda_0(t)} d\Lambda_{c,0}(t) \). This is also a proportional hazards model. Furthermore, as shown in the Appendix, the compensated counting processes

\[
M_i(t) := N_i(t) - \int_0^t Y_i(s) e^{-\beta_0 \bar{Z}_i(s)} + \gamma_0 Z_i(s) \ dH_0(s) \quad (i = 1, \ldots, n) \quad (3.3)
\]

are martingales with respect to the \( \sigma \)-filtration \( \mathcal{F}_t \). Here and in the sequel, we adopt the
notation $N_i, M_i, H_i, H_0$, etc. used in § 2. This should not create any ambiguity since § 2 may be regarded as a special case of the present section with $\gamma_0 = 0$.

As in § 2, one could apply the partial likelihood principle to model (3-2) with the augmented data $\{N_i(\cdot), Y_i(\cdot), Z^*_i(\cdot), Z_i(\cdot)\}$ ($i = 1, \ldots, n$) to make inferences about $\beta_0$ and $\gamma_0$. However, it will be more efficient to estimate $\gamma_0$ by applying the partial likelihood theory directly to model (3-1) since the $C_i$ are always observed. Let $\hat{\gamma}$ be the maximum partial likelihood estimator for $\gamma_0$ based on the data $\{C_i, Z_i(\cdot)\}$ ($i = 1, \ldots, n$), which is the root of the score function

$$U_{\gamma}(\gamma) = \sum_{i=1}^{n} \left\{ Z_i - \frac{\sum_{j=1}^{n} Y_j(C_i) e^{\gamma Z_j(C_i)} Z_j(C_i)}{\sum_{j=1}^{n} Y_j(C_i) e^{\gamma Z_j(C_i)}} \right\}. $$

Given $\hat{\gamma}$, we estimate $\beta_0$ by the estimating function $U_{\beta}(\beta; \hat{\gamma})$, where

$$U_{\beta}(\beta; \gamma) = \sum_{i=1}^{n} \int_{0}^{\infty} \left\{ Z_i(t) - \frac{S^{(1)}(\beta, \gamma, t)}{S^{(0)}(\beta, \gamma, t)} \right\} dN_i(t)$$

is the partial likelihood score function for $\beta_0$ under model (3-2), and

$$S^{(k)}(\beta, \gamma, t) = \sum_{j=1}^{n} Y_j(t) e^{-\beta Z_j(t) + \gamma Z_j(t)} Z_j(t)^{\otimes k} \quad (k = 0, 1).$$

Denote the resulting estimator by $\hat{\beta}$.

Note that $S^{(k)}(\beta, \gamma, t) (k = 0, 1)$ differ from $S^{(k)}(\beta, t)$ only in that the former involve an extra factor $e^{\beta Z_j(t)}$, which is free of $\beta$. Thus, the numerical properties of $U_{\beta}(\beta; \hat{\gamma})$ as a function of $\beta$ are very similar to those of (2-2), and $\hat{\beta}$ can be obtained by the usual Newton-Raphson algorithm.

Let

$$\hat{\Omega}_{\beta}(\beta; \gamma) = n^{-1} \frac{\partial U_{\beta}(\beta; \gamma)}{\partial \beta}, \quad \hat{\Omega}_{\gamma}(\beta; \gamma) = n^{-1} \frac{\partial U_{\beta}(\beta; \gamma)}{\partial \gamma}, \quad \hat{D}_{\beta}(\gamma) = n^{-1} \frac{\partial U_{\beta}(\gamma)}{\partial \gamma},$$

and let $\Omega_{\beta}, \Omega_{\gamma},$ and $D_\gamma$ denote their limits at $\beta = \beta_0$ and $\gamma = \gamma_0$. Assume that $\Omega_{\beta}$ and $D_\gamma$ are nonsingular. We show in the Appendix that the random vectors $n^{-1} U_{\beta}(\beta_0; \hat{\gamma})$ and $n^{-1} (\hat{\beta} - \beta_0)$ converge in distribution to zero-mean normal random vectors with covariance matrices $\Omega_{\beta} - \Omega_{\beta} D_{\gamma}^{-1} \Omega_{\beta}^{\prime} D_{\gamma}^{-1}$ and $V_{\beta} := \Omega_{\beta}^{\prime} - \Omega_{\beta} D_{\gamma}^{-1} \Omega_{\beta}^{\prime} D_{\gamma}^{-1}$, respectively, and the limiting covariance matrices can be consistently estimated by replacing $\Omega_{\beta}, \Omega_{\beta}^{\prime},$ and $D_\gamma$ by their respective sample estimates.

If $\gamma_0$ is known, then $\Omega_{\gamma}^{\prime}$ will be the limiting covariance matrix for the maximum partial likelihood estimator of $\beta_0$ under model (3-2). Since $\Omega_{\beta}^{\prime} V_{\beta} \Omega_{\gamma}^{\prime} = \Omega_{\beta}^{\prime} (\Omega_{\beta}^{\prime} - \Omega_{\beta} D_{\gamma}^{-1} \Omega_{\beta}^{\prime} D_{\gamma}^{-1}) \Omega_{\gamma}^{\prime}$, $V_{\beta}$ is always smaller than $\Omega_{\beta}^{\prime} V_{\beta} \Omega_{\gamma}^{\prime}$ in that $\Omega_{\beta}^{\prime} V_{\beta} \Omega_{\gamma}^{\prime}$ is positive semidefinite, $V_{\beta}$ is always smaller than $\Omega_{\beta}^{\prime}$ in that $\Omega_{\beta}^{\prime} V_{\beta} \Omega_{\gamma}^{\prime}$ is positive semidefinite; i.e. it will be more efficient to estimate $\gamma_0$ from the data even if $\gamma_0$ is known. Thus, we can gain efficiency by using the two-stage method even when we know that the monitoring time is independent of the covariates, that is, $\gamma_0 = 0$. On the other hand, standard likelihood theory indicates that the covariance matrix for the maximum partial likelihood estimator of $\beta_0$ under model (3-2) with unknown $\gamma_0$ is always larger than $\Omega_{\beta}^{\prime}$. These results confirm that the proposed two-step estimation procedure is indeed more efficient than the simultaneous partial-likelihood estimation of $\beta_0$ and $\gamma_0$ under model (3-2) based on $\{N_i(\cdot), Y_i(\cdot), Z^*_i(\cdot), Z_i(\cdot)\}$ ($i = 1, \ldots, n$). If for some reason one component of $Z(t)$ is the integral of another component of $Z(t)$, so that $Z(t)$ and $Z^*(t)$ share a common component, the parameters $\beta_0$ and $\gamma_0$ in model (3-2) will not be identifiable. In this case, only the two-step procedure can be used.
4. Numerical results

4.1. Simulation studies

A series of simulation studies were conducted to assess the performance of the methods developed in the previous two sections. The failure times were generated from model (1.1) with \( \lambda_0(.) = 1, \beta_0 = 0.5 \) and \( Z \) being a uniform random variable on \( (0, \sqrt{12}) \). The monitoring times were generated from the exponential distribution with hazard rate \( \lambda_{c,0}e^{\gamma_0Z} \), where \( \lambda_{c,0} = 0.5, 1.0 \) or 1.5, and \( \gamma_0 = 0 \) and 0.5 for the cases of independent and dependent monitoring times, respectively. Sample sizes of 100 and 200 were considered. There were 10,000 simulation samples for each combination of the simulation parameters.

Table 1 displays the Monte Carlo estimates for the sampling means and standard errors of the estimators \( \hat{\beta} \) and \( \tilde{\beta} \), the sampling means of their standard error estimators \( \mathcal{S}^{-1/2}(\hat{\beta}) \) and \( n^{-1/2}\hat{\mathcal{V}}_{\hat{\beta}}(\tilde{\beta}; \gamma) \), and the empirical coverage probabilities of the 95% Wald confidence intervals. These results suggest that the biases of the parameter estimators and their standard error estimators are negligible, at least for \( n > 100 \). In addition, the 95% confidence intervals have proper coverage probabilities. As expected, \( \hat{\beta} \) is more efficient than \( \tilde{\beta} \) in the case of independent monitoring time.

Table 1. Summary statistics for the simulation studies

(a) Independent monitoring time

<table>
<thead>
<tr>
<th>( \lambda_{c,0} = 0.5 )</th>
<th>1.0</th>
<th>1.5</th>
<th>( \lambda_{c,0} = 0.5 )</th>
<th>1.0</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of ( \hat{\beta} )</td>
<td>0.56</td>
<td>0.54</td>
<td>0.53</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td>Stand. error of ( \hat{\beta} )</td>
<td>0.45</td>
<td>0.40</td>
<td>0.40</td>
<td>0.29</td>
<td>0.26</td>
</tr>
<tr>
<td>Mean of ( \mathcal{S}^{-1/2}(\hat{\beta}) )</td>
<td>0.42</td>
<td>0.38</td>
<td>0.39</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>Cov. prob. of 95% CI</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>Mean of ( \tilde{\beta} )</td>
<td>0.56</td>
<td>0.54</td>
<td>0.53</td>
<td>0.53</td>
<td>0.52</td>
</tr>
<tr>
<td>Stand. error of ( \tilde{\beta} )</td>
<td>0.42</td>
<td>0.36</td>
<td>0.35</td>
<td>0.27</td>
<td>0.24</td>
</tr>
<tr>
<td>Mean of ( n^{-1/2}\hat{\mathcal{V}}_{\hat{\beta}}(\tilde{\beta}; \gamma) )</td>
<td>0.40</td>
<td>0.35</td>
<td>0.34</td>
<td>0.26</td>
<td>0.23</td>
</tr>
<tr>
<td>Cov. prob. of 95% CI</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

(b) Dependent monitoring time

<table>
<thead>
<tr>
<th>( \lambda_{c,0} = 0.5 )</th>
<th>1.0</th>
<th>1.5</th>
<th>( \lambda_{c,0} = 0.5 )</th>
<th>1.0</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of ( \hat{\beta} )</td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.52</td>
<td>0.51</td>
</tr>
<tr>
<td>Stand. error of ( \hat{\beta} )</td>
<td>0.38</td>
<td>0.42</td>
<td>0.50</td>
<td>0.25</td>
<td>0.29</td>
</tr>
<tr>
<td>Mean of ( n^{-1/2}\hat{\mathcal{V}}_{\hat{\beta}}(\tilde{\beta}; \gamma) )</td>
<td>0.38</td>
<td>0.41</td>
<td>0.49</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>Cov. prob. of 95% CI</td>
<td>0.95</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

4.2. A real example

We now provide an illustration with real data taken from a carcinogenicity experiment. The study was originally reported by Hoel & Walburg (1972); the data were reproduced by Finkelstein & Wolfe (1985). The purpose of the study was to compare the time until the onset of lung tumour for untreated mice in a germ-free versus conventional environment. A total of 144 RFM mice were involved in the study. At the times of death/sacrifice, lung tumours were found in 27 out of the 96 mice assigned to the conventional environment, as compared with 35 out of the 48 mice assigned to the germ-free environment. Since lung tumours are nonlethal in RFM mice, it is reasonable to regard the tumour
onset time $T$ to be independent of the death time $C$. However, the survival experiences may be different between the conventional and germ-free environments. Therefore, the methods of § 3 seem to be the most appropriate.

Let $Z = 0$ for conventional mice and $Z = 1$ for germ-free mice. The maximum partial likelihood estimate $\hat{\gamma}$ for the death time is $-1.96$, with estimated standard error of 0.24, which indicates that germ-free mice live much longer than conventional mice. The risk difference estimate $\hat{\beta}$ for the onset of lung tumour is 0.00071, with estimated standard error of 0.00041. The normalised test statistic is 1.73, resulting in a two-sided $P$-value of 0.084. These results suggest that lung tumours tend to occur earlier in the germ-free environment than in the conventional environment.

The above data were previously analysed by Finkelstein (1986) and Huang (1996) under the proportional hazards model. Finkelstein reported a standard-normal statistic of 1.64 for testing no group difference in tumour incidence, with an associated $P$-value of 0.10. Huang obtained a log relative risk estimate of 0.55 with an estimated standard error of 0.29, corresponding to a $P$-value of 0.054. Incidentally, our estimate $\hat{\beta}$ is not directly comparable to Huang’s estimate since the former pertains to $\{\lambda(t|Z = 1) - \lambda(t|Z = 0)\}$ and the latter to $\log\{\lambda(t|Z = 1)/\lambda(t|Z = 0)\}$.

5. Remarks

Given $\beta_0$, the values of $Z(t)$ are constrained so that the right-hand side of (1.1) is nonnegative. This constraint may sometimes be achieved by replacing $Z(t)$ in (1.1) by $Z(t)q(t)$, where $q$ is a prespecified function. One may also remove the constraint by using a nonnegative link function, such as $\lambda_0(t) + e^{\beta_0 Z(t)}$. The basic results presented in this paper hold for any link functions, although the linear form of (1.1) gives rise to the simplest formulae.

As with the case of right-censored data, it is essential to assume the conditional independence of $T$ and $C$ given $Z$, without which the problem is non-identifiable. The proportional hazards modelling of $C$ is somewhat restrictive. However, model (3.1) is fully testable since the $C_i$ are always observed. In many applications, especially in controlled experiments, the structures for the monitoring time imposed in §§ 2 and 3 are appropriate. In such cases, the simplicity of the inference procedures seems to justify their restrictive assumptions.

In the case of independent monitoring, that is $\gamma_0 = 0$, any functional of the covariate history may be used in model (3.1), and the resulting two-stage estimator is more efficient than $\hat{\beta}$. It can be shown by the Cauchy–Schwarz inequality that the most efficient functional is $Z^* e^{-\beta_0 Z(t)} - \Lambda_0(t)$. One may replace $\beta_0$ and $\Lambda_0(t)$ involved in this functional by any preliminary estimators or guesses. Note that $\Lambda_0(t)$ can be estimated by solving the integral equation $\hat{H}_0(t) = \int_0^t e^{-\Lambda(u)} d\hat{\Lambda}_{c,o}(u)$, where $\hat{H}_0$ and $\hat{\Lambda}_{c,o}$ are the Aalen–Breslow type estimators of $H_0$ and $\Lambda_{c,o}$ given at the end of the Appendix. The estimator $\hat{\beta}$ corresponding to the most efficient functional does not achieve the semiparametric information bound. One may derive semiparametric efficient estimators of $\beta_0$ for both independent and dependent monitoring along the lines of Huang (1996), but the resulting inference procedures will be much more complicated than those of §§ 2 and 3.

In some applications including cancer screening trials, there is more than one monitoring time for each study subject, which gives rise to general interval censored data. Suppose that there are $K$ monitoring times. For each of the $K$ monitoring times, we may construct a modification of $U(\beta)$ or $U_0(\beta; \gamma)$. Then a linear combination of these $K$ functions will
yield a single estimating function for $\beta_0$. Inference procedures similar to those of §§ 2 and 3 can be obtained. These new theoretical developments, along with their applications, will be reported elsewhere.

ACKNOWLEDGEMENT

We are grateful to the Editor and a referee for useful comments and to Professor Cun-Hui Zhang for helpful discussions. This research was supported by the U.S. National Institutes of Health, National Science Foundation and National Security Agency.

APPENDIX

Some technical results

It is helpful to introduce some notation in connection with the estimation of model (3-1). Let $N_i(t) = I(C_i \leq t)$ and

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)e^{\varphi Z_i(s)} \, d\Lambda_{c,0}(s) \quad (i = 1, \ldots, n).$$

It is well known that $M_i(t) (i = 1, \ldots, n)$ are martingales with respect to the $\sigma$-filtration

$$\sigma\{N_i(s), Z_i(s): s \leq t, i = 1, \ldots, n\}.$$

Furthermore,

$$n^{\frac{1}{2}}(\hat{\gamma} - \gamma_0) = D^{-1} n^{\frac{1}{2}} \sum_{i=1}^n \int_0^\infty \{Z_i(t) - e(\gamma_0, t)\} \, dM_i(t) + o_p(1), \quad (A.1)$$

where $e(\gamma, t) = E[1(t) e^{Z_t(0)} Z_t(1)] / E[1(t) e^{Z_t(0)}]$ (Andersen & Gill, 1982).

We now prove that $M_i(t)$ defined in (3-3) are martingales with respect to the $\sigma$-filtration $\mathcal{F}_t$. It suffices to show that, for any $\Delta t > 0$, $E[M_i(t + \Delta t)|\mathcal{F}_t] = M_i(t)$ or

$$E(N_i(t + \Delta t) - N_i(t)|\mathcal{F}_t) = \int_{(t,t+\Delta t]} \text{pr}(C_i \geq s|\mathcal{F}_t) e^{-\beta_0 Z_i(t) + \gamma Z_i(0)} \, dH_0(s). \quad (A.2)$$

Clearly, both sides of (A-2) are zero if $C_i \leq t$. Thus, we only need to establish the equality on the set $\{C_i > t\}$. However,

$$E(N_i(t + \Delta t) - N_i(t)|\mathcal{F}_t, C_i > t) = \text{pr}(t < C_i \leq t + \Delta t, T_i \geq C_i, C_i > t, Z_i)$$

$$= \int_{(t,t+\Delta t]} e^{-\lambda_0 s - \beta_0 Z_i(t)} \text{pr}(C_i \geq s|Z_i) e^{\varphi Z_i(0)} \, d\Lambda_{c,0}(s)$$

$$= \int_{(t,t+\Delta t]} \text{pr}(C_i \geq s|\mathcal{F}_t, C_i > t) e^{-\beta_0 Z_i(t) + \gamma Z_i(0)} \, dH_0(s).$$

Hence, (A-2) holds.

Simple algebraic manipulation shows that

$$U_p(\beta_0; \gamma_0) = \sum_{i=1}^n \int_0^\infty \left\{Z_i^*(t) - \frac{S^{(1)}(\beta_0, \gamma_0, t)}{S^{(0)}(\beta_0, \gamma_0, t)}\right\} \, dM_i(t),$$

which is a martingale integral. A standard calculation of predictable variation yields

$$n^{-\frac{1}{2}} U_p(\beta_0; \gamma_0) = n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^\infty \{Z_i^*(t) - e(\beta_0, \gamma_0, t)\} \, dM_i(t) + o_p(1), \quad (A.3)$$

where $e(\beta, \gamma, t) = E[S^{(1)}(\beta, \gamma, t)] / E[S^{(0)}(\beta, \gamma, t)].$
By Taylor series expansion,
\[ n^{-1} U_p(\beta_0, \hat{\gamma}) = n^{-1} U_p(\beta_0, \gamma_0) - \hat{O}_p(\beta_0, \gamma^*) n^{-1} (\hat{\gamma} - \gamma_0), \]
where \( \gamma^* \) is on the line segment between \( \hat{\gamma} \) and \( \gamma_0 \). It then follows from (A·1), (A·3), the law of large numbers and the consistency of \( \hat{\gamma} \) that
\[
n^{-1} U_p(\beta_0, \hat{\gamma}) = n^{-1} \sum_{i=1}^n \int_0^\infty \{ Z_i^*(t) - e(\beta_0, \gamma_0, t) \} dM_i(t) \]
\[ - \Omega_p^* D_p^{-1} n^{-1} \sum_{i=1}^n \int_0^\infty \{ Z_i(t) - e(\gamma_0, t) \} dM_i(t) + o_p(1). \]
\[ \text{(A·4)} \]

Since the right-hand side of (A·4) is a sum of \( n \) independent and identically distributed zero-mean random vectors, a direct application of the Lindeberg central limit theorem shows that \( n^{-1} U_p(\beta_0, \hat{\gamma}) \) is asymptotically normal. Standard calculations of predictable variation for martingale integrals yield \( \Omega_p^* \) and \( \Omega_p^* D_p^{-1} \Omega_p^* \) as the limiting variances for the first and second terms on the right-hand side of (A·4), respectively. Note now that
\[
dM_i(t) = I(T_i \geq t) dM_i(t) + \{ I(T_i \geq t) - \text{pr}(T_i \geq t | Z_i) \} Y_i(t) e^{\gamma Z_i(t)} d\Lambda_{c,0}(t). \]
Since \( \{ Z_i(t) - e(\gamma_0, t) \} dM_i(t) \) does not involve \( T_i \), its covariance with \( \{ I(T_i \geq t) - \text{pr}(T_i \geq t | Z_i) \} \) must be zero. Thus,
\[
cov \left[ \int_0^\infty \{ Z_i^*(t) - e(\beta_0, \gamma_0, t) \} dM_i(t), \int_0^\infty \{ Z_i(t) - e(\gamma_0, t) \} dM_i(t) \right]
= E \left[ \int_0^\infty \{ Z_i^*(t) - e(\beta_0, \gamma_0, t) \} dM_i(t), \int_0^\infty \{ Z_i(t) - e(\gamma_0, t) \} dM_i(t) \right]
= E \left[ \int_0^\infty \{ Z_i^*(t) - e(\beta_0, \gamma_0, t) \} \{ Z_i(t) - e(\gamma_0, t) \} I(T_i \geq t) dN_i(t) \right]
= E \left[ \int_0^\infty \{ Z_i^*(t) - e(\beta_0, \gamma_0, t) \} \{ Z_i(t) - e(\gamma_0, t) \} dN_i(t) \right],
\]
which is exactly \( \Omega_p^* \). Thus, the covariance between the two terms on the right-hand side of (A·4) equals the variance of the second term. Consequently, the limiting covariance matrix for \( n^{-1} U_p(\beta_0, \hat{\gamma}) \) is \( \Omega_p^* - \Omega_p^* D_p^{-1} \Omega_p^* \).

The proof for the consistency of \( \hat{\beta} \) is similar to that of the maximum partial likelihood estimator since the slope matrix \( \hat{\Omega}(\beta; \hat{\gamma}) \) is positive semidefinite and its limit is assumed to be positive definite at \( \beta_0 \). Given the consistency, the desired asymptotic normality for \( \hat{\beta} \) follows from the Taylor series expansion of \( U_p(\hat{\beta}; \hat{\gamma}) \) at \( U_p(\beta_0; \hat{\gamma}) \), together with the asymptotic normality of \( n^{-1} U_p(\beta_0; \hat{\gamma}) \). The consistency of the limiting covariance matrix estimators follows from the arguments for establishing the consistency of the covariance matrix estimator for the maximum partial likelihood estimator (Andersen & Gill, 1982).

The variance–covariance formulae given in §§ 2 and 3, as well as their derivations given above, implicitly assumed that \( \Lambda_c \) is continuous. In the presence of discontinuities, a correction is needed for the predictable variation (Fleming & Harrington, 1991, Theorem 2.6.2). Consequently, the limiting covariance matrix for \( n^{-1} U_p(\beta_0, \hat{\gamma}) \) is \( \Psi_p := \Omega_p^* - \Omega_p^* D_p^{-1} D_p^* \Omega_p^* \), and that of \( n^{-1} (\hat{\beta} - \beta_0) \) is \( \Omega_p^* D_p^{-1} \Psi_p \Omega_p^* \), where
\[
\Omega_p^* = E \left[ \int_0^\infty \{ Z_i^*(t) - e(\beta_0, \gamma_0, t) \} \otimes^2 \{ 1 - e^{-\beta_0 Z_i(t) + \gamma_0 Z_i(t)} \Delta H_0(t) \} Y_i(t) e^{\gamma Z_i(t)} dH_0(t) \right],
\]
\[
D_p^* = E \left[ \int_0^\infty \{ Z_i(t) - e(\gamma_0, t) \} \otimes^2 \{ 1 - e^{\gamma Z_i(t)} \} Y_i(t) e^{\gamma Z_i(t)} d\Lambda_{c,0}(t) \right].
\]
Consistent estimators can again be obtained by inserting proper sample estimators for the unknown quantities in the covariance matrix expressions. Note that $H_0(t)$ and $\Lambda_{c,0}(t)$ can be estimated by the Aalen–Breslow type estimators

$$
\hat{H}_0(t) = \sum_{i=1}^{n} \int_{0}^{t} \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)e^{-\sum_{j=1}^{n} \frac{\hat{r}_j(s)+\hat{g}_j(s)}}},
\hat{\Lambda}_{c,0}(t) = \sum_{i=1}^{n} \int_{0}^{t} \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)e^{\sum_{j=1}^{n} \frac{\hat{r}_j(s)+\hat{g}_j(s)}}}.
$$

REFERENCES


[Received September 1996. Revised September 1997]