Biometrika (2002), **89,** 3, *pp.* 617–634

Printed in Great Britain

Estimation of the failure time distribution in the presence of informative censoring

By DANIEL O. SCHARFSTEIN

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205, U.S.A.

dscharf@jhsph.edu

and JAMES M. ROBINS

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

robins@hsph.harvard.edu

Summary

We present a method for estimating the survival curve of a continuous failure time random variable from right-censored data. Our method allows adjustment for informative censoring due to measured prognostic factors for time-to-event and censoring while simultaneously quantifying the sensitivity of the inference to residual dependence between failure and censoring due to unmeasured factors. We present the results of a simulation study and illustrate our approach using data from the AIDS Clinical Trial Group 175 study.

Some key words: Coarsening at random; Competing risks; Curse of dimensionality; Inverse probability of censoring weighted estimation; Kaplan–Meier estimator; Sequential ignorability of censoring.

1. Introduction

Consider estimation of the marginal survivor function $S_0(u) = \mathrm{pr}(T \geqslant u)$ of a continuous failure time random variable T from right-censored data. The nonparametric Kaplan & Meier (1958) estimator of the marginal survival function is inconsistent unless censoring is noninformative. When censoring is informative, two approaches have been proposed, those that do and do not use auxiliary information. Under the nonidentifiable assumption that data on all time-dependent and independent prognostic factors for failure that also predict censoring are available, Robins (1987) proved that $S_0(u)$ was nonparametrically identified and Robins & Rotnitzky (1992), Robins (1993), Satten et al. (2001) and Robins & Finkelstein (2000) proposed useful estimation techniques. When auxiliary prognostic factor data are not available, many authors have imposed nonidentifiable assumptions concerning the dependence between T and the censoring time C, and have then varied these assumptions to assess the sensitivity of the inference and to generate bounds for $S_0(u)$; see Fisher & Kanarek (1974), Slud & Rubinstein (1983), Klein & Moeschberger (1988), Klein et al. (1992), Zheng & Klein (1994, 1995) and Moeschberger & Klein (1995). In most studies, data are typically available on some but not all joint prognostic factors for censoring and survival. In the context of discrete time-to-event data, Scharfstein et al. (2001) developed a methodology that allows an analyst to adjust appropriately for

informative censoring due to the measured factors while simultaneously quantifying the sensitivity of the inference to nonidentifiable assumptions concerning the residual dependence between time-to-event and censoring due to unmeasured factors. Their approach relied on specification of classes of parametric models for the discrete-time cause-specific hazard of censoring, indexed by censoring bias functions quantifying the degree of residual dependence. For each member of the classes, Scharfstein et al. (2001) showed how to estimate the marginal survivor function for failure, and derived large-sample theory by appealing to standard finite parameter delta-method arguments.

In this paper, we consider the continuous-time setting. We proceed by specifying continuous time analogues of the models used in the discrete-time setting. That is, we specify classes of semiparametric models for the continuous-time cause-specific hazard of censoring, indexed by censoring bias functions. The semiparametric nature of these models makes large-sample theory for estimators of the marginal survivor function much more challenging, as the aforementioned standard delta-method arguments no longer apply. To handle this issue, we appeal to the general theory of estimation described in Appendix B of Scharfstein et al. (1999).

In § 2, we introduce the AIDS Clinical Trial Group (ACTG) 175 study, which provides context for the methodological developments. In § 3, we formalise the right-censored data structure, and introduce two models, differentiated by the dimension, low or high, of the available prognostic factors. In § 4, we describe our approach to estimation in these models. In § 5, we present the results of a simulation study. In § 6, we analyse the ACTG 175 data. Section 7 is devoted to a discussion.

2. THE ACTG 175 STUDY

The ACTG 175 study was a randomised, double-blind clinical trial designed to evaluate nucleoside monotherapy, i.e. zidovudine or didanosine, versus combination therapy, i.e. zidovudine/didanosine or zidovudine/zalcitabine, in HIV-1 infected individuals with CD4 cell counts of 200–500/mm^3. A total of 2467 subjects were randomised to one of four treatment arms: AZT, corresponding to zidovudine 200 mg three times daily; AZT + ddI, corresponding to zidovudine 200 mg three times daily plus didanosine 200 mg twice daily; AZT + ddC, corresponding to zidovudine 200 mg three times daily plus zalcitabine 0·75 mg three times daily; and ddI, corresponding to didanosine 200 mg twice daily. Enrolment began in December 1991 and was closed in October 1992. Patients were scheduled to be followed until November 1994. As a result, all subjects were scheduled for at least two years of follow-up. The CD4 counts were to be collected at baseline, week 8 and then every 12 weeks thereafter. The primary endpoint was time to 50% decline in CD4 from baseline, as confirmed by a second CD4 count within 3 to 21 days, AIDS or death. Baseline information was collected, including gender, age, race, ethnicity, CD4 count, Karnofsky score, prior use of anti-retroviral therapy and whether or not HIV was symptomatic, as well as risk factors such as homosexuality, injection drug use and haemophilia. Compliance information was also collected. The first published analysis of these data was based on an intent-to-treat model, i.e. compliance information was ignored, and Kaplan–Meier estimates of the treatment-specific distributions of the primary endpoint were presented (Hammer et al., 1996).

In contrast to the previous analysis, in this paper we are interested in estimating the treatment-specific distributions of time to primary endpoint if all subjects had remained on their assigned therapy. With this as our goal, for each subject, T is defined to be the

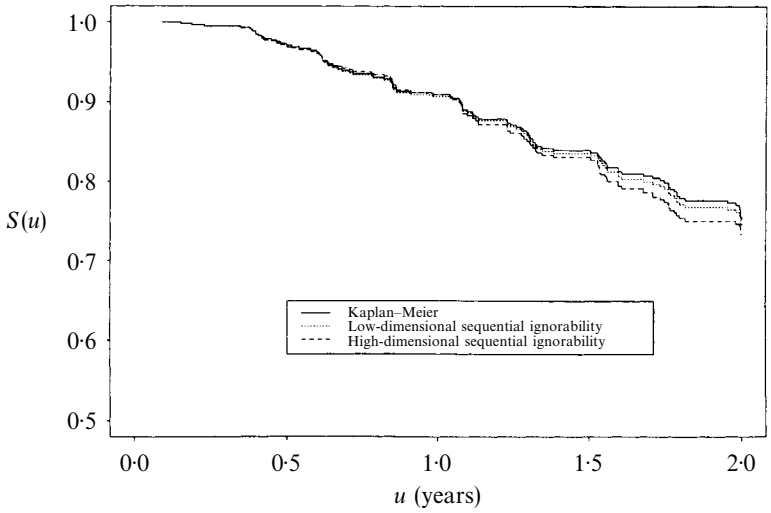

Fig. 1. Estimates of $S_0(u)$ under noninformative censoring, i.e. Kaplan–Meier, low-dimensional sequential ignorability and high-dimensional sequential ignorability.

time to the primary endpoint that would be observed under full compliance. Thus, we regard the censoring time, C, for a subject to be the smaller of time to loss to follow-up and time to discontinuation of assigned therapy. Censoring by discontinuation of therapy is thought to be potentially informative. Subjects could be lost to follow-up because they either dropped out of the study or reached the administrative date of end to follow-up. While the latter type of loss to follow-up would often be considered to be completely independent of the primary endpoint, it is plausible that the former type is informative. Since the focus of this paper is on informative censoring and no subject was administratively censored within the first two years of follow-up, we will consider a maximum of two years of follow-up on each individual and only seek to draw inference about $S_0(u)$ over this time horizon. By restricting the dataset in this fashion, we ensure that all censoring is potentially informative. Our models do not distinguish between the two competing causes of informative censoring. In future work, we plan to extend our methodology to accommodate not only cause of informative censoring, but also independent administrative censoring.

For ease of presentation, we will confine our analysis to the AZT arm. We analyse the 609 relevant subjects who had complete baseline information and who spent at least some time on treatment. In this cohort, 104 subjects experienced the primary events prior to two years, 253 subjects were censored prior to two years, and 252 subjects were known to have the primary event after two years. The solid line in Fig. 1 displays the Kaplan–Meier estimator for the survivor distribution of time to primary endpoint. Hammer et al. (1996) reported that 'younger patients, those reporting injection-drug use and those with lower CD4 cell counts, lower Karnofsky scores and symptoms of HIV infection at enrolment were significantly more likely to discontinue treatment before the study ended'. As a result, one would expect the true treatment-specific survival curves to lie below the Kaplan–Meier estimators; that is, the view presented by the Kaplan–Meier curve is likely to be too optimistic. If the recorded CD4 history plus the above baseline factors were the only joint prognostic factors for censoring and time to event, then the methods of Robins

and co-authors could provide consistent and asymptotically normal estimators of the treatment-specific survival curves. However, if there exist additional time-dependent or time-independent joint prognostic factors, then a residual dependence between time to censoring and primary event will exist, even after adjusting for the measured prognostic factors. As a result, existing methods fail. In § 6, we use our proposed methodology to perform a sensitivity analysis that examines how the estimated survival curves vary with assumed degrees of residual dependency.

3. DATA AND MODELS

3·1. *Data*

We assume that each subject has a maximum potential follow-up time of at least c^*. We seek to identify the distribution of the time-to-event on $[0, c^*]$. The following slight change in notation will be convenient. Let T^* and C^* be nonnegative time-to-event and censoring random variables which are absolutely continuous with respect to Lebesgue measure. Define $C = \min(C^*, c^*)$ and $T = \min(T^*, c^*)$. Note that the distributions of T and T^* are the same on $[0, c^*]$. In the context of the ACTG 175 analysis, c^* is two years, T^* is time to primary event, and C^* is the smaller of time to loss to follow-up and discontinuation of assigned treatment.

We consider each subject to be observed until time $X = \min(T, C)$. Let $\Delta = I(T \leqslant C)$ be the indicator of whether or not T was observed and let $\bar{V}_t = \{V_s : s < t\}$, where V_s is a vector of prognostic factors for T and C recorded at time s. The observable data for a subject is $O = (X, \Delta, \bar{V}_X)$. We assume that we observe n independent and identically distributed copies of O.

3·2. *Models*

Our approach is based on assuming a stratified proportional hazards model for the 'censoring mechanism'. We assume that

$$\lambda_C^\dagger(t \mid \bar{V}_t, T, T > t) = \lambda_0(t, \bar{V}_t) \exp\{q(t, \bar{V}_t, T)\}, \tag{1}$$

where $\lambda_C^\dagger(t|., T > t) = \lim_{h \to 0} \mathrm{pr}(t \leqslant C < t + h \mid C \geqslant t, ., T > t)/h$ is the conditional cause-specific hazard of censoring at time t given the information in ., $\lambda_0(t, \bar{V}_t)$ is an unknown function of t and $\bar{V}_t$, and $q(t, \bar{V}_t, T)$ is a known function of t, $\bar{V}_t$ and T. To ensure identifiability, we also impose the additional restriction that $\lambda_C^\dagger(t \mid \bar{V}_t, T, T > t) > 0$ with probability one for all $t \in [0, c^*]$. Specification of the function $q(t, \bar{V}_t, T)$ in (1) is equivalent to quantifying, for those who remain at risk at time t, the dependence, measured on a hazard ratio scale between T and censoring just after time t, after having adjusted for recorded prognostic factors, $\bar{V}_t$, up to time t. To emphasise this important property we refer to the function $q(t, \bar{V}_t, T)$ as a 'censoring bias function'. Note that the censoring bias function determines how T enters the proportional hazards regression model for the cause-specific hazard of censoring. Robins & Rotnitzky (1992), Robins (1993) and Robins & Finkelstein (2000) studied the special case in which $q(t, \bar{V}_t, T) = 0$ for all t. We can refer to this latter assumption as the assumption of sequential ignorability of censoring, which is a generalisation of the assumption of coarsening at random (Heitjan & Rubin, 1991); that is, coarsening at random implies sequential ignorability of censoring, but the converse does not hold (Robins & Rotnitzky, 1992).

The following three remarks, validation of which is analogous to the proof of Theorem 1 of Scharfstein et al. (1999), summarise the critical features of model (1).

Remark 1. We have that $S_0(u)$ is identified from the law, F_O, of the observed data. Specifically, let

$$\pi(t \mid \bar{V}_t, T; \Lambda_0) = \prod_{0 \leqslant s < t} [1 - \exp\{q(s, \bar{V}_s, T)\}\, d\Lambda_0(s, \bar{V}_s)],$$

where $\Lambda_0(t, \bar{V}_t)$ is the integral of $\lambda_0(s, \bar{V}_s)$ on $[0, t]$. Then $S_0(u)$ is the unique solution, $S(u)$, to

$$E\left(\frac{\Delta\{I(T \geqslant u) - S(u)\}}{\pi(T \mid \bar{V}_T, T; \Lambda_0)}\right) = 0, \tag{2}$$

where, from equation (1), $\Lambda_0(t, \bar{V}_t)$ itself is the unique solution to the identifiable, backward recursive equation

$$\Lambda_0(t, \bar{V}_t) = \int_0^t \frac{\lambda_C^\dagger(s \mid \bar{V}_s, T > s)\, ds}{E\left(\dfrac{\Delta \exp\{q(s, \bar{V}_s, T)\}\pi(s \mid \bar{V}_s, T; \Lambda_0)}{\pi(T \mid \bar{V}_T, T; \Lambda_0)} \,\middle|\, X \geqslant s, \bar{V}_s, T > s\right)}. \tag{3}$$

Remark 2. The function $q(t, \bar{V}_t, T)$ is not identified under equation (1) because any choice of $q(t, \bar{V}_t, T)$ is compatible with the law, F_O, of the observed data. Thus, no statistical test can reject any choice $q(t, \bar{V}_t, T)$.

Remark 3. By specifying the functions $q(t, \bar{V}_t, T)$, we do not place any restrictions on the distribution F_O of the observed data O. Thus, equation (1) with specified $q(t, \bar{V}_t, T)$ forms a nonparametric model for F_O.

We refer to model (1) with specified censoring bias function $q(t, \bar{V}_t, T)$ as model A_q. The above points tell us that model A_q, like the models proposed for nonparametric sensitivity analysis in the competing risks literature without auxiliaries, is a nonparametric, just, identified model for the distribution F_O of the observed data; that is, the censoring bias function is not identified, but, once specified, the model is nonparametric for F_O and $S_0(u)$ is identified. Thus, following the lead of the competing risks literature, we suggest drawing inference about $S_0(u)$ by varying the censoring bias function over a 'plausible range'. In § 3·3 below, we provide a useful parameterisation of the censoring bias function to facilitate such sensitivity analyses.

It is also interesting to note that, because model A_q does not place any restriction on F_O, it also places no a priori restriction on the marginal distribution of T. To see this, suppose that, under F_O, subjects die at time 0 with probability one and there is no censoring. Then, under A_q, the marginal distribution of T will be a single point mass at 0. Next, suppose that, under F_O, 30% of subjects are censored prior to c^* and all others are observed to survive past c^*. Then, under A_q, the marginal distribution of T will be a single point mass at c^*. This holds, whatever q might be. These are the most extreme distributions for the marginal distribution of T.

When data V_t are not recorded for data analysis, so that $O^- = (X, \Delta)$ are the observable data for a subject, our model becomes a new nonparametric, just, identified model for the competing risks setting without auxiliaries, characterised by the restriction

$$\lambda_C^\dagger(t \mid T, T > t) = \lambda_0(t) \exp\{q(t, T)\}, \tag{4}$$

where $q(t, T)$ is a known function and $\lambda_0(t)$ is an unknown nonnegative function. Remark 1 tells us that $S_0(u)$ will be identified when we specify censoring bias functions $q(t, T)$. Furthermore, equations (2) and (3), with $\bar{V}_t$ removed, provide a formula for computing

$S_0(u)$ from F_{O^-}. Remarks 2 and 3 tell us that $q(t, T)$ is not identified and that restriction (4) with specified $q(t, T)$ forms a nonparametric model for F_{O^-}. When $q(t, T) = 0$ for all t, restriction (4) is identical to the assumption of noninformative censoring or equivalently coarsening at random (Robins, 1987; Gill et al., 1997; Jacobsen & Keiding, 1995). As above, we suggest that sensitivity analyses be performed with respect to $q(t, T)$. The advantage of our model over the other nonparametric, just, identified models in the competing risks literature is that ours naturally generalises to settings in which data are collected on both time-independent and dependent prognostic factors for T and C.

Nonparametric estimation of $S_0(u)$ in model A_q with fixed censoring bias functions requires nonparametric estimation of the function $\Lambda_0(t, \bar{V}_t)$ in (1). From equation (3), we see that, when $\bar{V}_t$ is high-dimensional, nonparametric estimation of this function, in moderate-sized datasets, is infeasible because of the curse of dimensionality in the sense that the number of subjects at any given level of $\bar{V}_t$ is too small to yield a reasonable estimator (Huber, 1985; Robins & Ritov, 1997). Hence, estimation using model A_q is only feasible when, as in § 4·2 below, $\bar{V}_t$ is low-dimensional; for example, $\bar{V}_t = V$ for all t and V can be well approximated by a discrete random variable with a moderate number of levels.

When $\bar{V}_t$ is high-dimensional, it will be necessary to reduce the dimension of the unknown function $\lambda_0(t, \bar{V}_t)$. We will assume that the function $\Lambda_0(t, \bar{V}_t)$ follows a lower-dimensional model of the form

$$\lambda_0(t, \bar{V}_t) = \lambda_0(t) \exp\{\gamma_0' w(t, \bar{V}_t)\}, \tag{5}$$

where $\lambda_0(t)$ is an unknown function of t, $w(t, \bar{V}_t)$ is a specified q-dimensional function of t and $\bar{V}_t$, and γ_0 is an unknown parameter vector of the same dimension as $w(t, \bar{V}_t)$. Model A_q with additional restriction (5) is no longer a nonparametric model for F_O. The resulting model, which we refer to as B_q, is semiparametric; that is, it restricts the law of the observed data. Thus, we can in principle check our assumption (5) concerning the functional form of $\lambda_0(t, \bar{V}_t)$ as well as the functional form of the censoring bias function. However, because $q(t, \bar{V}_t, T)$ is not identified when the function $\lambda_0(t, \bar{V}_t)$ is left unrestricted, our ability to check the validity of the assumed functional form of the censoring bias function relies entirely upon the correct specification of $\lambda_0(t, \bar{V}_t)$. Therefore, rather than checking the specification of the assumed censoring bias function, we recommend that the analyst (a) should regard it as known when estimating $S_0(u)$ and then vary it in a sensitivity analysis, and (b) should choose the dimension of $w(t, \bar{V}_t)$ in (5) large enough so that any goodness-of-fit test will have little power to reject model B_q, but small enough so that the estimators described in § 4·3 have a nearly normal sampling distribution with a variance small enough to be of substantive use. We realise that it may be hard to meet the competing criteria in (b), as the choice of the dimension of $w(t, \bar{V}_t)$ will depend on the size of the dataset and the desired precision. In addition, since different choices for the functional form of $w(t, \bar{V}_t)$ cannot be easily distinguished based on goodness-of-fit tests and may lead to different inferences for $S_0(u)$, we also recommend that the analyst repeat the sensitivity analysis for various choices of $w(t, \bar{V}_t)$.

3·3. *Parameterisation of the censoring bias function*

To facilitate our sensitivity analysis approach in models A_q and B_q, it will be helpful to use a parameterised form $q(t, \bar{V}_t, T; \alpha)$ for the censoring bias function, where α is a parameter that we will vary, but not estimate, and satisfies the restriction that $\alpha = 0$ implies $q(t, \bar{V}_t, T; \alpha) = 0$. We impose this latter restriction so that $\alpha = 0$ corresponds to the assumption of sequential ignorability of censoring, or noninformative censoring when $\bar{V}_t$ is not

available. It is also essential that α be chosen to be interpretable, so that a plausible range can be specified by subject matter experts. The parameterisation that we use in the analysis of the ACTG 175 data is

$$q(t, \bar{V}_t, T; \alpha) = \alpha_1 \{I(T < c^*)(T - t) + I(T = c^*)(\alpha_2 - t)\}, \tag{6}$$

where $\alpha = (\alpha_1, \alpha_2)'$ and $\alpha_2 > c^*$. The censoring bias parameter α_2 is meant roughly to represent the mean time to event for subjects who do not experience the event prior to c^*. The parameter α_1 is interpreted as the log hazard ratio of dropping out at time t between subjects who are at risk at time t and have the same covariate history $\bar{V}_t$, but who differ by one unit in their ultimate time-to-event, with α_2 as the proxy time-to-event for subjects whose time-to-event is longer than c^*. Thus, $\alpha_1 > 0$ (< 0) implies that subjects with longer times to events are more (less) likely to be censored at any time t than subjects who would have shorter times to event. Note that this censoring bias function implies that a subject who would not experience the event prior to c^* has $\exp\{\alpha_1(\alpha_2 - t)\}$ times the hazard of censoring at time t as compared to a subject with the same covariate history but who drops out at time t. Thus, the choice of α_2 also indicates the degree to which subjects who would experience the event before and after time c^* are different from one another with respect to their risk of censoring. Furthermore, note that, regardless of the choice of α_2, $\alpha_1 = 0$ is equivalent to the assumption of sequential ignorability of censoring. When the censoring bias function is parameterised by α, we refer to models A_q and B_q as A_α and B_α, respectively.

4. Estimation

4·1. *Preliminaries*

Our estimation approach is a modification of the approach of Scharfstein et al. (1999) to right-censored data. Consider estimation of the parameter $\mu_0 = S_0(u)$ at a fixed time u. Using the general theory of estimation laid out in Appendix B of Scharfstein et al. (1999), we can construct the space of all 'estimating functions' for the finite-dimensional parameters, μ_0 in model A_q and $\psi_0 = (\mu_0, \gamma_0')'$ in model B_q, which are uncorrelated with the observed-data likelihood scores S_η of all correctly specified parametric submodels of the unknown infinite-dimensional components. This space is related to the set of influence functions for all regular and asymptotically linear estimators for the finite-dimensional parameters of the model. To be specific, each estimating function in the space is associated with an influence function obtained by normalising the estimating function so that its covariance matrix with the observed-data likelihood score for the finite-dimensional parameters is equal to the identity. In the literature on semiparametric estimation, our estimating function space is called the orthogonal complement of the nuisance tangent space (Newey, 1990; Bickel et al., 1993). Construction of these spaces for models A_q and B_q is available in a technical report from the authors.

While the estimating functions in the orthogonal complement of the nuisance tangent space depend on the unknown cumulative hazard function, the asymptotic variance of the estimator based on assuming that Λ_0 is known is the same as that of the estimator based on substitution of a $n^{1/4+\varepsilon}$-consistent estimator of Λ_0, where $\varepsilon > 0$. Hence, we will be able to derive the asymptotic properties for the latter estimator simply by studying the properties of the former estimator. This will allow us to derive analytical, as opposed to resampling-based, procedures for computing standard error estimates. This is critical as

resampling-based procedures become computationally intractable when one wishes, as we do, to calculate standard error estimates for a broad range of censoring bias parameters.

Before proceeding, it is important to highlight that, when $\alpha \neq 0$ in (6), estimation of the unknown baseline hazard functions or regression parameters in the proportional hazards models (4) and (1, 5) for the cause-specific censoring cannot proceed by using Breslow's estimator or the partial likelihood estimator, i.e. the estimators used when $\alpha = 0$. This is because the censoring bias function would appear in these estimators and this function depends on the failure time, which is potentially unobserved. Below, we propose alternative methods for estimating the unknown baseline hazard functions and regression parameters in the proportional hazards models for censoring.

4·2. *Model A_q with low-dimensional prognostic factors*

We assume that $\bar{V}_t$ is low-dimensional. In particular, we assume that $\bar{V}_t = V$ for all t, where V is a vector of time-independent factors which can be well approximated by a discrete random variable with a moderate number of levels. In model A_q, the orthogonal complement of the nuisance tangent space for μ_0 consists of all multiples of the single function

$$h(O;\, \mu_0, \Lambda_0;\, b^*) = \frac{\Delta}{\pi(T \mid V, T;\, \Lambda_0)} \{I(T \geqslant u) - \mu_0\} + A(O;\, \mu_0, \Lambda_0;\, b^*),$$

where

$$A(O;\, \mu_0, \Lambda_0;\, b^*) = (1-\Delta) b^*(V, C;\, \mu_0, \Lambda_0) + \frac{\Delta}{\pi(T \mid V, T;\, \Lambda_0)} \int_0^T b^*(V, s;\, \mu_0, \Lambda_0)\, d\pi(s \mid V, T;\, \Lambda_0)$$

and $b^*(v, t;\, \mu_0, \Lambda_0)$ is the unique solution to the integral equation

$$b(v, t;\, \mu_0, \Lambda_0) = \frac{J(v, t;\, \mu_0, \Lambda_0)}{K(v, t;\, \Lambda_0)} - \int_0^t b(v, s;\, \mu_0, \Lambda_0) \frac{f(v, s, t;\, \Lambda_0)}{K(v, t;\, \Lambda_0)}\, d\Lambda_0(s, v), \tag{7}$$

in which

$$J(v, t;\, \mu_0, \Lambda_0) = E\left(\frac{\Delta \{I(T \geqslant u) - \mu_0\} \exp\{q(t, v, T)\} I(T > t)}{\pi(T \mid V, T;\, \Lambda_0)} \middle| V = v \right),$$

$$K(v, t;\, \Lambda_0) = E\left(\frac{\Delta \pi(t \mid V, T;\, \Lambda_0) \exp\{q(t, V, T)\} I(T > t)}{\pi(T \mid V, T;\, \Lambda_0)} \middle| V = v \right),$$

$$f(v, s, t;\, \Lambda_0) = E\left(\frac{\Delta \pi(s \mid V, T;\, \Lambda_0) \exp\{q(s, V, T) + q(t, V, T)\} I(T > t)}{\pi(T \mid V, T;\, \Lambda_0)} \middle| V = v \right).$$

Suppose that $\Lambda_0(t, v)$, $J(v, t;\, \mu_0, \Lambda_0)$, $K(v, t;\, \Lambda_0)$ and $f(v, t;\, s, \Lambda_0)$ were known. Then $\tilde{\mu}_n(b^*)$, the solution to $E_n\{h(O;\, \mu_0, \Lambda_0;\, b^*)\} = 0$, would be regular and asymptotically linear with influence function $h(O;\, \mu_0, \Lambda_0;\, b^*)$. This implies that

$$n^{\frac{1}{2}} \{\tilde{\mu}_n(b^*) - \mu_0\} = n^{-\frac{1}{2}} \sum_{i=1}^{n} h(O_i;\, \mu_0, \Lambda_0;\, b^*) + o_P(1).$$

Thus, $n^{\frac{1}{2}} \{\tilde{\mu}_n(b^*) - \mu_0\}$ converges in distribution to a normal random variable with mean

zero and variance $E\{h(O;\,\mu_0,\Lambda_0;\,b^*)^2\}$. Since $J(v,t;\,\mu_0,\Lambda_0)$, $K(v,t;\,\Lambda_0)$ and $f(v,s,t;\,\Lambda_0)$ are unknown, we substitute

$$b(v,t;\,\mu_0,\Lambda_0)=\frac{J_n(v,t;\,\mu_0,\Lambda_0)}{K_n(v,t;\,\Lambda_0)}-\int_0^t b(v,s;\,\mu_0,\Lambda_0)\frac{f_n(v,s,t;\,\Lambda_0)}{K_n(v,t;\,\Lambda_0)}\,d\Lambda_0(s,v) \tag{8}$$

for (7), where $J_n(v,t;\,\mu_0,\Lambda_0)$, $K_n(v,t;\,\Lambda_0)$ and $f_n(v,s,t;\,\Lambda_0)$ are $J(v,t;\,\mu_0,\Lambda_0)$, $K(v,t;\,\Lambda_0)$ and $f(v,s,t;\,\Lambda_0)$, with expectations replaced by sample averages. The solution, $\tilde{b}_n^*$, to (8) will be a consistent estimator of the solution b^* to (7), whatever μ_0 might be. If we use the results of Robins et al. (1992), it can be shown that $\tilde{\mu}_n(b^*)$ and $\tilde{\mu}_n(\tilde{b}_n^*)$ have the same asymptotic properties.

Since Λ_0 is also unknown, we shall replace it by a $n^{\frac{1}{2}}$-consistent estimator Λ_n. Then we will estimate μ_0 as the solution $\mu_n(b_n^*)$ to $E_n\{h(O;\,\mu,\Lambda_n;\,b_n^*)\}=0$, where $b_n^*(v,t;\,\mu,\Lambda_n)$ is the solution to (8) with Λ_n substituted for Λ_0. This modified equation can be solved via forward recursion, since our estimator Λ_n is a step function; see below. Since $h(O;\,\mu_0,\Lambda_0;\,b^*)$ is orthogonal to the nuisance tangent space for Λ_0 and b_n^* is a consistent estimator of b^*, under mild conditions $\mu_n(b_n^*)$ will be regular and asymptotically linear with the same asymptotic properties as $\tilde{\mu}_n(b^*)$. Thus, $n^{\frac{1}{2}}\{\mu_n(b_n^*)-\mu_0\}$ converges to a normal random variable with mean zero and variance $E\{h(O;\,\mu_0,\Lambda_0;\,b^*)^2\}$. This latter variance can be estimated by $E_n\{h(O;\,\mu_n(b_n^*),\Lambda_n;\,b_n^*)^2\}$. It is important to note that, in finite samples, $\mu_n(b_n^*)$ is not guaranteed to be monotone in u. However, one can enforce monotonicity by using the pool adjacent violator algorithm (Ayer et al., 1955) without affecting the asymptotics.

To see how to estimate Λ_0, note that by simple algebra equation (3) can be rewritten as

$$\Lambda_0(t,v)=\int_0^t\frac{-\,dP(X\geqslant s,\Delta=0,V=v)}{E\left(\dfrac{\Delta I(T>s,V=v)\exp\{q(s,v,T)\}\pi(s\,|\,v,T;\,\Lambda_0)}{\pi(T|\,v,T;\,\Lambda_0)}\right)}. \tag{9}$$

We can estimate $\Lambda_0(t,v)$ by $\Lambda_n(t,v)$ as the unique solution to the empirical version of (9), namely

$$\Lambda_n(t,v)=\int_0^t\frac{-\,dP_n(X\geqslant s,\Delta=0,V=v)}{E_n\left(\dfrac{\Delta I(T>s,V=v)\exp\{q(s,v,T)\}\pi(s\,|\,v,T;\,\Lambda_n)}{\pi(T|\,v,T;\,\Lambda_n)}\right)}, \tag{10}$$

where P_n and E_n denote sample proportions and averages, respectively. Equation (10) has a solution, $\hat{\Lambda}_n(t,v)$, which is a step function with jumps at each of the unique censoring times in the group with $V=v$. The function can be found by computing the jump sizes, starting with a closed-form expression for the jump at the largest censoring time and then recursively solving a sequence of equations formed by working backwards in time. Under mild regularity conditions, we would expect $n^{\frac{1}{2}}\{\Lambda_n(.,v)-\Lambda_0(.,v)\}$ to converge to a Gaussian process. A rigorous proof of this result requires modern empirical process theory and lies outside the scope of this paper. An asymptotically equivalent and computationally simpler estimator of Λ_0 can be constructed by replacing $\pi(s\,|\,v,T;\,\Lambda_n)$ in (10) by $\pi(s^+|v,T;\,\Lambda_n)$. The solution to this modified equation is also a step function with jumps at the unique V-specific censoring times, but the jumps can be expressed in closed form. We denote this solution by $\hat{\Lambda}_n$.

The fact that the orthogonal complement of the nuisance tangent space for model A_q

is one-dimensional is not a coincidence: model A_q is nonparametric for the observed data O and Bickel et al. (1993) proved that, for any nonparametric model, all regular and asymptotically linear estimators of any functional of F_O have the same influence function. This tells us that all regular and asymptotically linear estimators have the same asymptotic properties and are therefore efficient. In fact, it is possible to construct a simpler regular and asymptotically linear estimator of μ_0, $\mu_n(0)$, which we take as the solution to $E_n\{h(O; \mu, \Lambda_n, 0)\} = 0$. This estimator is regular and asymptotically linear, since it is easy to see that $h(O; \mu, \Lambda_0; 0)$ is an unbiased estimating function. Now, $\mu_n(0)$ has the same asymptotic variance as $\mu_n(b_n^*)$, which can be estimated by $E_n\{h(O; \mu_n(0), \Lambda_n; b_n^*)^2\}$. This alternative estimator has some interesting properties. First, it is the solution to the empirical version of equation (3) with Λ_0 replaced by Λ_n. Secondly, it is guaranteed to be monotone in finite samples, whereas $\mu_n(b_n^*)$ is not. Thirdly, when $q(V, T) = 0$ for all t, $\mu_n(0)$ reduces to a weighted average of V-specific Kaplan–Meier estimators (Robins & Rotnitzky, 1992; Robins & Finkelstein, 2000). When V are missing or ignored and $q(t, T) = 0$ for all t, so that all subjects are assumed to belong to the same strata, then $\mu_n(0)$ is exactly Kaplan–Meier.

It is important to note that we could also estimate μ_0 by the asymptotically equivalent estimators $\hat{\mu}_n(b_n^*)$ or $\hat{\mu}_n(0)$ that replace Λ_n by $\hat{\Lambda}_n$ in calculating $\mu_n(b_n^*)$ or $\mu_n(0)$. The asymptotic variance of these estimators can be estimated as above but using the modified estimator, $\hat{\Lambda}_n$, of the cumulative baseline hazard function.

Finally, it is important to understand bounds on the estimated survivor curves when the censoring bias parameters in (6) are taken to extremes. It can be shown that, as α_1 goes to ∞, the resulting estimator converges to a weighted average of V-specific estimators in which each censored observation within the strata is assumed to fail at α_2. On $[0, c^*]$, this bound is the same for all α_2. As α_1 goes to $-\infty$, the resulting estimator converges to a weighted average of V-specific estimators in which each censored observation within a stratum of V is assumed to fail at the next observed failure time within the stratum. When V is ignored or missing, our bounds differ slightly from the bounds of Peterson (1976), in which the extremes correspond to the scenarios where all censored observations are assumed never to fail or to fail immediately after censoring.

4·3. *Model B_q with high-dimensional prognostic factors*

To ease notation, define $p_q(s, \bar{V}_s, T; \gamma_0) = \exp\{\gamma_0' w(s, \bar{V}_s) + q(s, \bar{V}_s, T)\}$. The orthogonal complement of the nuisance tangent space for $\psi_0 = (\mu_0, \gamma_0')'$ in model B_q contains an infinite number of elements, indexed by $(q+1)$-dimensional functions $\phi(\bar{V}_t, t)$. The elements are of the form

$$h(O; \psi_0, \Lambda_0; b_\phi^*) = e_1 \frac{\Delta}{\pi(T|\bar{V}_T, T; \gamma_0, \Lambda_0)}\{I(T \geqslant u) - \mu_0\} + A(O; \psi_0, \Lambda_0; b_\phi^*),$$

where e_1 is a $(q+1)$-dimensional vector whose first component is 1 and remaining components are zero, $\pi(t|\bar{V}_t, T; \gamma_0, \Lambda_0) = \prod_{0 \leqslant s < t}\{1 - p_q(s, \bar{V}_s, T; \gamma_0)\, d\Lambda_0(s)\}$,

$$A(O; \psi_0, \Lambda_0; b_\phi^*) = (1-\Delta) b_\phi^*(\bar{V}(C), C; \psi_0, \Lambda_0) + \frac{\Delta}{\pi(T|\bar{V}_T, T; \gamma_0, \Lambda_0)} \int_0^T b_\phi^*(\bar{V}_s, s; \psi_0, \Lambda_0)\, d\pi(s|\bar{V}_s, T; \gamma_0, \Lambda_0)$$

and $b_\phi^*(\bar{V}_t, t; \psi_0, \Lambda_0)$ is the unique solution to the Volterra-like integral equation

$$b_\phi^*(\bar{v}_t, t; \psi_0, \Lambda_0) = \phi(\bar{v}_t, t) - \frac{J_\phi(t; \psi_0, \Lambda_0) + \int_0^t f_{b_\phi^*}(s, t; \psi_0, \Lambda_0)\, d\Lambda_0(s) - L(t; \psi_0, \Lambda_0)}{D(t; \psi_0, \Lambda_0)}, \tag{11}$$

in which

$$D(t; \gamma_0, \Lambda_0) = E\left(\frac{\Delta\pi(t \mid \bar{V}_T, T; \gamma_0, \Lambda_0) p_q(t, \bar{V}_t, T; \gamma_0) I(T > t)}{\pi(T \mid \bar{V}_T, T; \gamma_0, \Lambda_0)}\right),$$

$$J_\phi(t; \gamma_0, \Lambda_0) = E\left(\phi(\bar{v}_t, t) \frac{\Delta\pi(t \mid \bar{V}_T, T; \gamma_0, \Lambda_0) p_q(t, \bar{V}_t, T; \gamma_0) I(T > t)}{\pi(T \mid \bar{V}_T, T; \gamma_0, \Lambda_0)}\right),$$

$$f_{b_\phi^*}(s, t; \gamma_0, \Lambda_0) = E\left(\frac{\Delta b_\phi^*(s; \psi_0, \Lambda_0)\pi(s \mid \bar{V}_T, T; \gamma_0, \Lambda_0) p_q(s, \bar{V}_s, T; \gamma_0) p_q(t, \bar{V}_t, T; \gamma_0) I(T > t)}{\pi(T \mid \bar{V}_T, T; \gamma_0, \Lambda_0)}\right),$$

$$L(t; \gamma_0, \Lambda_0) = E\left(\frac{\Delta e_1 \{I(T \geqslant u) - \mu_0\} p_q(t, \bar{V}_t, T; \gamma_0) I(T > t)}{\pi(T \mid \bar{V}_T, T; \gamma_0, \Lambda_0)}\right).$$

Suppose that $\Lambda_0(t)$ were known. If, in addition, (11) were directly available, then the solution, $\tilde{\psi}_n(b_\phi^*)$, to $E_n\{h(O; \psi_0, \Lambda_0; b_\phi^*)\} = 0$, would be regular and asymptotically linear with influence function $\Gamma_\phi^{-1} h(O; \psi_0, \Lambda_0; b_\phi^*)$, where $\Gamma_\phi = E\{\partial h(O; \psi_0, \Lambda_0; b_\phi^*)/\partial\psi\}$. This implies that

$$n^{\frac{1}{2}}\{\tilde{\psi}_n(b_\phi^*) - \psi_0\} = n^{\frac{1}{2}} \sum_{i=1}^{n} \Gamma_\phi^{-1} h(O_i; \psi_0, \Lambda_0; b_\phi^*) + o_P(1).$$

Thus, $n^{\frac{1}{2}}\{\tilde{\psi}_n(b_\phi^*) - \psi_0\}$ converges in distribution to a normal random vector with mean zero and covariance matrix $\Gamma_\phi^{-1} E\{h(O; \psi_0, \Lambda_0; b_\phi^*)^{\otimes 2}\}\Gamma_\phi^{-1\prime}$. Since (11) is not available even when Λ_0 is assumed known, we replace it by

$$b_\phi^*(\bar{v}_t, t; \psi_0, \Lambda_0) = \phi(\bar{v}_t, t) - \frac{J_{n,\phi}(t; \gamma_0, \Lambda_0) + \int_0^t f_{n,b_\phi^*}(s, t; \gamma_0, \Lambda_0)\, d\Lambda_0(s) - L_n(t; \psi_0, \Lambda_0)}{D_n(t; \gamma_0, \Lambda_0)}, \tag{12}$$

where $D_n(t; \gamma, \Lambda_0)$, $J_n(t; \gamma, \Lambda_0)$, $f_{n,b_\phi^*}(s, t; \gamma, \Lambda_0)$ and $L_n(t; \psi, \Lambda_0)$ are the empirical versions of $D(t; \gamma, \Lambda_0)$, $J_\phi(t; \gamma, \Lambda_0)$, $f_{b_\phi^*}(s, t; \gamma, \Lambda_0)$ and $L(t; \psi, \Lambda_0)$, respectively. The solution $b_{n,\phi}$ to (12) will be a consistent estimator of the solution b_ϕ^* to (11), whatever ψ_0 might be. As above, it can be shown that $\tilde{\psi}_n(b_\phi^*)$ and $\tilde{\psi}_n(b_{n,\phi}^*)$ have the same asymptotic properties.

Since $\Lambda_0(t)$ is unknown, it must be estimated. To motivate our approach, we define, for each γ, the following equation:

$$\Lambda_\gamma(t) = \int_0^t \frac{-dP(X \geqslant s, \Delta = 0)}{E\left(\frac{\Delta I(T \geqslant s) p_q(s, \bar{V}_s, T; \gamma)\pi(s \mid \bar{V}_s, T; \gamma, \Lambda_\gamma)}{\pi(T \mid \bar{V}_T, T; \gamma, \Lambda_\gamma)}\right)}. \tag{13}$$

When $\gamma = \gamma_0$, it can be shown that the unique solution, $\Lambda_{\gamma_0}(t)$, is equal to $\Lambda_0(t)$. This suggests that we define a profile estimator for cumulative baseline hazard function as the

solution to the empirical version of (13):

$$\Lambda_{n,\gamma}(t) = \int_0^t \frac{-dP_n(X \geqslant s, \Delta = 0)}{E_n\left(\dfrac{\Delta I(T \geqslant s) p_q(s, \bar{V}_s, T; \gamma) \pi(s \mid \bar{V}_s, T; \gamma, \Lambda_{n,\gamma})}{\pi(T \mid \bar{V}_T, T; \gamma, \Lambda_{n,\gamma})}\right)}. \tag{14}$$

For fixed γ, equation (14) can be uniquely solved via a backward recursive algorithm similar to the one described in § 4·2. Under mild regularity conditions, it can be shown that $n^{\frac{1}{2}}\{\Lambda_{n,\gamma}(.) - \Lambda_{\gamma}(.)\}$ converges to a Gaussian process, uniformly in γ.

With this profile estimator, we can now estimate ψ_0 by $\psi_n(b^*_{n,\phi})$, the solution to

$$E_n\{h(O; \psi, \Lambda_{n,\gamma}; b^*_{n,\phi})\} = 0, \tag{15}$$

where $b^*_{n,\phi}$ is the solution to (12) with Λ_0 replaced by $\Lambda_{n,\gamma}$. This modification of (12) is easily solved by forward recursion, as $\Lambda_{n,\gamma}$ is a step function. Since $h(O; \psi_0, \Lambda_0, b^*_{\phi})$ is orthogonal to the nuisance tangent space and $b^*_{n,\phi}$ is a consistent estimator of b^*_{ϕ}, it follows that $\psi_n(b^*_{n,\phi})$ is regular and asymptotically linear with influence function $\Gamma_{\phi}^{-1} h(O; \psi_0, \Lambda_0; b^*_{\phi})$. Thus, $n^{\frac{1}{2}}\{\psi_n(b^*_{n,\phi}) - \psi_0\}$ converges in distribution to a normal random vector with mean zero and covariance matrix $\Gamma_{\phi}^{-1} E\{h(O; \psi_0, \Lambda_0; b^*_{\phi})^{\otimes 2}\} \Gamma_{\phi}^{-1\prime}$. This asymptotic covariance matrix can be estimated by

$$\Gamma_{n,\phi}^{-1} E_n[h\{O; \psi_n(b^*_{n,\phi}), \Lambda_{n,\gamma_n(b^*_{n,\phi})}; b^*_{n,\phi}\}^{\otimes 2}] \Gamma_{n,\phi}^{-1\prime},$$

where $\Gamma_{n,\phi} = E_n[\partial h\{O; \psi_n(b^*_{n,\phi}), \Lambda_{n,\gamma_n(b^*_{n,\phi})}; b^*_{n,\phi}\}/\partial\psi]$ and can be obtained by numerical differentiation. In finite samples, $\mu_n(b^*_{n,\phi})$ is not guaranteed to be monotone in u. As above, one can make the estimator monotonic by using the pool adjacent violator algorithm without affecting the asymptotics.

It is important to note that the efficiency of the estimators of γ_0 depends on the choice of ϕ. The optimal choice of ϕ depends on F_O and requires the solution to an exceedingly complex integral equation. Thus, we forego trying to obtain a semiparametric efficient estimator. In the analysis of the ACTG 175 data and our simulation study, we chose

$$\phi(\bar{V}_t, t) = (0, w(t, \bar{V}_t)')'. \tag{16}$$

The solution to (15) can be found using the Newton–Raphson algorithm. However, each iteration of this algorithm requires K additional Newton–Raphson algorithms to compute the profile estimator for Λ_0. To circumvent this computationally intensive process, we propose two modifications of the algorithm. First, we replace $\pi(s \mid \bar{V}_s, T; \gamma, \Lambda_{n,\gamma})$ in (14) by $\pi(s + \mid \bar{V}_{s+}, T; \gamma, \Lambda_{n,\gamma})$. This serves to eliminate the K iterative algorithms in the computation of the profile estimator as the profile estimator can now be derived via a sequence of closed-form solutions. Secondly, we employ a one-step estimation procedure. We compute a consistent estimator of ψ_0 by simply solving (15) with $b^*_{n,\phi}$ equal to the right-hand side of (16). The resulting estimator, ψ_n^{init}, will be regular and asymptotically linear. Then we estimate ψ_0 by employing one step of the Newton–Raphson in solving (15) with ψ_n^{init} as the starting value. The estimator is

$$\psi_n^{\text{os}}(b^*_{n,\phi}) = \psi_n^{\text{init}} - \left(\frac{\partial E_n\{h_n(O; \psi_n^{\text{init}}, \Lambda_{n,\gamma_n^{\text{init}}}; b^*_{n,\phi})\}}{\partial\psi}\right)^{-1} E_n\{h(O; \psi_n^{\text{init}}, \Lambda_{n,\gamma_n^{\text{init}}}; b^*_{n,\phi})\},$$

where the partial derivative can be computed numerically. Our one-step estimator is regular and asymptotically linear and has the same influence function as $\psi_n(b^*_{n,\phi})$. Thus,

the asymptotic variance of $\psi_n^{\mathrm{os}}(b_{n,\phi}^*)$ can be estimated by

$$(\Gamma_{n,\phi}^{\mathrm{os}})^{-1} E_n[h(O;\, \psi_n^{\mathrm{os}}(b_{n,\phi}^*), \Lambda_{n,\gamma_n^{\mathrm{os}}}(b_{n,\phi}^*);\, b_{n,\phi}^*)^{\otimes 2}\}](\Gamma_{n,\phi}^{\mathrm{os}})^{-1\prime},$$

where $\Gamma_{n,\phi}^{\mathrm{os}} = E_n[\partial h\{O;\, \psi_n^{\mathrm{os}}(b_{n,\phi}^*), \Lambda_{n,\gamma_n^{\mathrm{os}}}(b_{n,\phi}^*);\, b_n^*\}/\partial\psi]$.

5. Simulation studies

5·1. *Preliminaries*

To evaluate the finite-sample performance of our estimation techniques under models A_α and B_α, we conducted two simulation studies. In each study, the maximum follow-up period was $c^* = 2$ years. In addition, the true values of α_1 and α_2 were $-0{\cdot}5$ and $2{\cdot}25$, respectively. We took the sample size to be 500 and generated 500 datasets. Since the true values of α_1 and α_2 are assumed unknown to the data analyst, for each dataset we fitted 10 models, corresponding to combinations of $\alpha_1 = -2{\cdot}0, -1{\cdot}0, -0{\cdot}5, 0{\cdot}0$ and $1{\cdot}0$ and $\alpha_2 = 2{\cdot}25$ and $3{\cdot}0$. For each dataset, we estimated $S_0(u)$ for 21 equally-spaced values of u on $[0, 2]$.

5·2. *Model A_α*

We generated data under the assumption that V was a Bernoulli random variable with mean $0{\cdot}4$ and the conditional law of T^* given V was exponential with mean $V + 1$. The true marginal distribution on $[0, 2]$ was therefore a mixture of exponentials with means 1 and 2. In addition, in (1), we assumed that $\lambda_0(1) = 0{\cdot}5$ and $\lambda_0(0) = 0{\cdot}75$, so that subjects with $V = 0$ have a higher hazard of being censored. The overall censoring rate under these conditions was 33%.

Figure 2 presents the results of this simulation. For each of the 10 models fitted, there are two associated diagrams. The first diagram compares the true survivor curve, solid line, to the average of the estimated survivor curves across simulations, dotted line. If we correctly guess α_1 and α_2, then our estimator is unbiased; otherwise, our estimator is biased. The bias is most heavily influenced by the degree of misspecification of α_1 as opposed to α_2. The second diagram assesses the coverage probability of 90% Wald-type confidence intervals, dashed line. The solid lines in these plots are fixed at $0{\cdot}9$. We see that, when α_1 and α_2 are correctly specified, our coverage rate is excellent. However, as expected, the coverage rate is sensitive to misspecification of both α_1 and α_2.

We also compared the Monte Carlo standard deviation of the parameter estimates to the mean of the standard error estimates not shown in Fig. 2. We found that our estimator of the standard error tends to underestimate slightly the true variability across all model fits; however, this should not be surprising as it is well known that Greenwood's formula suffers similarly (Andersen et al., 1993, pp. 259–60). It is interesting to note that our standard error estimator works well even if α_1 and α_2 are misspecified; note that our theory does not predict this.

5·3. *Model B_α*

We generated data under the assumption that T^* was exponentially distributed with mean 2. We assumed that repeated measurements of a prognostic factor for censoring and survival were scheduled to be collected at times 0, $0{\cdot}5$, $1{\cdot}0$, $1{\cdot}5$ and $2{\cdot}0$. Let Z_t $(t = 0, \ldots, 4)$ denote these five measurements. We then assumed that joint conditional distribution of $(Z_0, \ldots, Z_4)$ given T^* was multivariate normal, where the conditional mean and variance

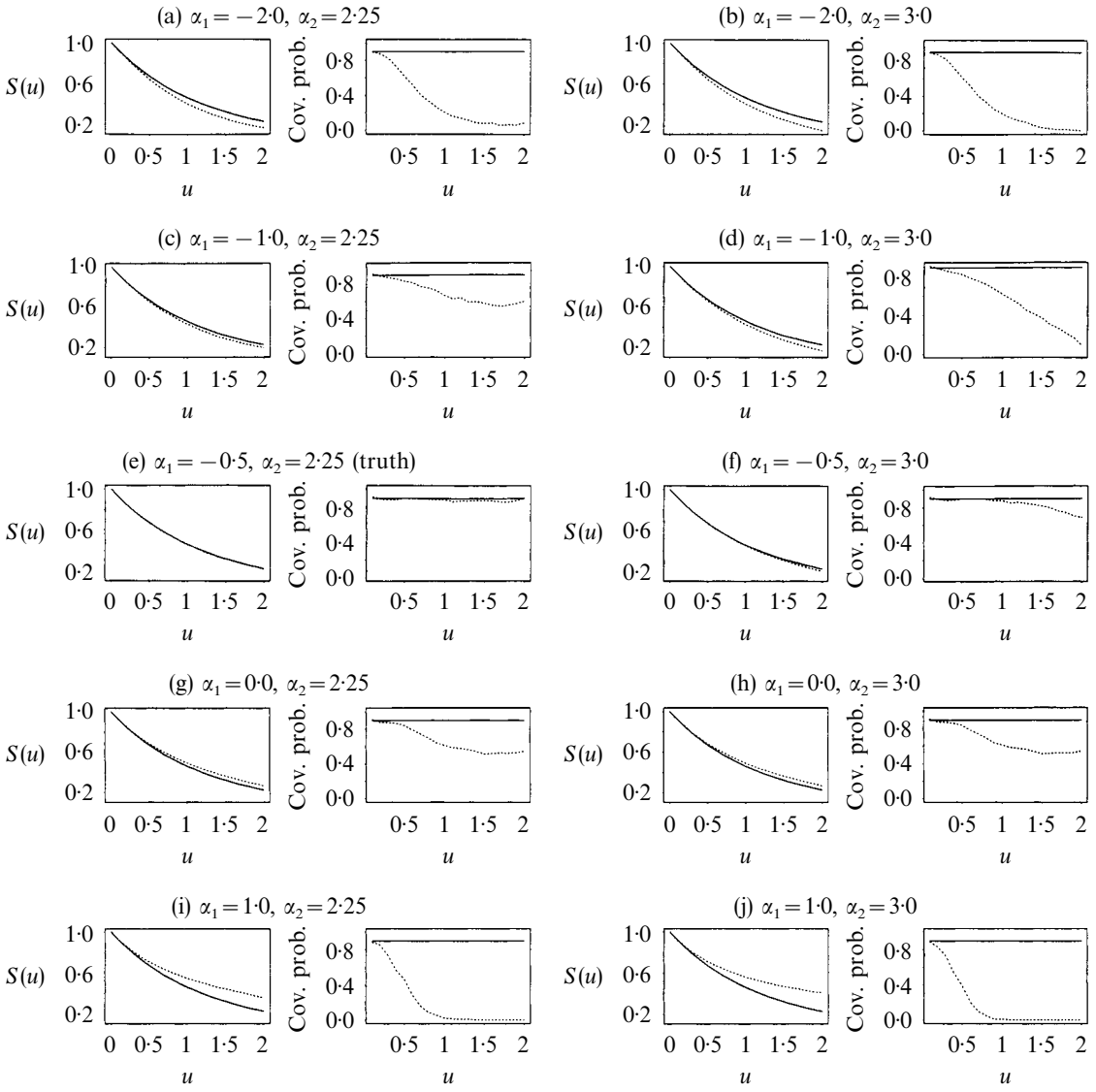

Fig. 2. Results of simulation for model A_α. For each of 10 pairs of α_1 and α_2, there are two associated diagrams. The first diagram compares the true survivor curve (solid line) to the average of the estimated survivor curves across simulations (dotted line). The second diagram assesses the coverage probability of 90% Wald-type confidence intervals (dotted line). The solid lines in these plots are fixed at 0·9.

of Z_t was T^* and 1, respectively, and the conditional covariance between Z_s and Z_t was $0{\cdot}6^{|s-t|}$. In addition, we assumed that the prognostic factor remains constant between measurements so that, for $0 \leqslant t \leqslant 2$,

$$V_t = \sum_{k=0}^{4} Z_k I\{0{\cdot}5k \leqslant t \leqslant 0{\cdot}5(k+1)\}.$$

Finally, in (5), we took $w(t, \bar{V}_t) = \bar{V}_t$, $\gamma_0 = -0{\cdot}25$ and $\lambda_0(t) = 1{\cdot}0$ for all t. These assumptions yielded an overall censoring rate of 35%.

We fitted the models using the one-step estimation routine, with the closed-form profile estimator for the baseline hazard. The results of this simulation are extremely similar to

those for model A_α in Fig. 2 except that the coverage results are slightly less sensitive to the choice of α_2. Secondly, results not shown indicate that our estimator of the standard error tends to have some positive bias, especially for larger values of u. This occurs even at the true values of α_1 and α_2. The degree of bias tends to increase with α_1.

6. Analysis of ACTG 175 data

In the analysis of the ACTG 175 data, we fitted three models:

Model 1. A_α without auxiliaries;
Model 2. A_α with the low-dimensional time-independent indicator of whether or not baseline CD4 is greater than the median CD4 level of 342 across all treatments;
Model 3: B_α with time-dependent CD4 count, note that CD4_t is the last CD4 measurement prior to t, and age, AGE, indicator of intravenous drug use, IV, Karnofsky score, KAR, and indicator of symptomatic HIV infection, SYM at baseline.

For Model 3, we took $w(t, \bar{V}_t) = (\text{CD4}_t, \text{AGE}, \text{IV}, \text{KAR}/100, \text{SYM})$. We varied α_1 from -1 to 1, and fixed $\alpha_2 = 3$, as our results were insensitive to choice of α_2.

Our results are displayed in Fig. 3. The solid curve represents our estimate of $S_0(u)$

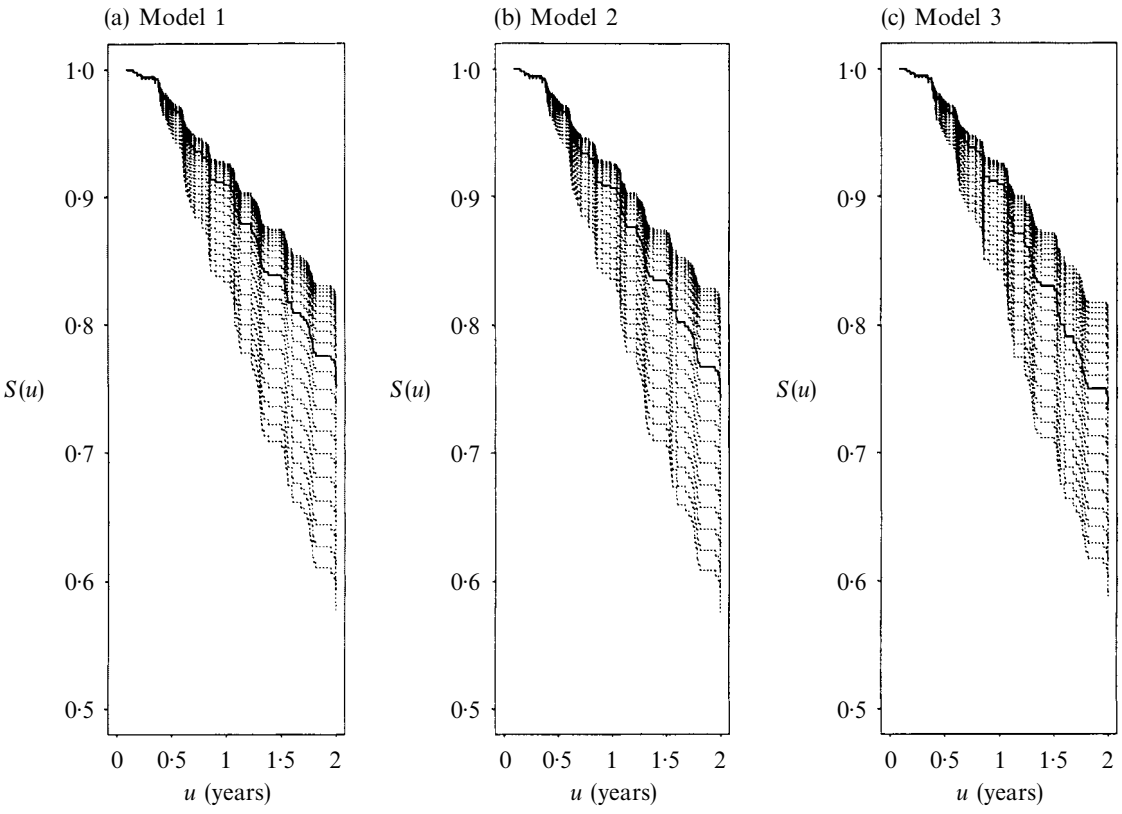

Fig. 3. Analysis of ACTG 175 data. Sensitivity analysis results for fitting model A_α without auxilliaries, Model 1, model A_α with a low-dimensional prognostic factor, Model 2, and model B_α with high-dimensional prognostic factors, Model 3. The solid curve represents our estimate of $S_0(u)$ under sequential ignorability of censoring, that is, $\alpha_1 = 0$. The dashed curves above the solid curve represent positive values of α_1, increasing in 0·1 increments. The curves lying below the solid curve correspond to negative values of α_1, decreasing in increments of 0·1.

under sequential ignorability of censoring, that is $\alpha_1 = 0$. To facilitate comparison, the solid curves are reproduced in Fig. 1. The dashed curves above the solid curve in Fig. 3 represent positive values of α_1, increasing in increments of 0·1. The curves lying below the solid curve correspond to negative values of α_1, decreasing in 0·1 increments.

In Model 1, $\alpha_1 = 0$ corresponds to the assumption of noninformative censoring and the solid line is the Kaplan–Meier estimate. In Model 2, we adjust for the indicator of high or low baseline CD4 count. In the AZT arm, 299 (310) subjects were classified with low (high) baseline CD4 count. Of these subjects, 70 (34) were observed to have a primary event prior to two years, 137 (116) were censored because of drop-out or non-compliance, and 92 (116) were known to have the primary event after two years. We see therefore that subjects with low baseline CD4 were more likely to be censored and experience the event prior to two years. If, after conditioning the cause-specific hazard of censoring on this baseline factor, there is no residual dependence on the time of primary event, $\alpha_1 = 0$, then we can expect a downward adjustment to the marginal survival curve. In fact, we do see a very mild correction, as depicted in Fig. 1. In Model 3, we adjusted for $CD4_t$, AGE, IV, KAR and SYM. When sequential ignorability is assumed to hold, that is $\alpha_1 = 0$, the regression coefficients, with standard errors, of these factors in the proportional hazards regression for the cause-specific hazard of censoring are $-0{\cdot}3001$ (0·0856), $-0{\cdot}0111$ (0·0154), 0·7448 (0·1922), $-2{\cdot}9231$ (1·1140) and $-2{\cdot}1290$ (1·7999), respectively. Thus, we see that subjects who have lower time-dependent CD4 are younger, are intravenous drug users and have lower baseline Karnofsky scores are significantly more likely to be censored. Since these factors are also associated with the time to primary event, we can expect that adjusting for them will also shift the marginal survivor curve to a greater extent than with Model 2. This is borne out in Fig. 1, but the adjustment is still quite small. The two-year probability of remaining free of the primary event under full compliance ranges from 0·734 in Model 3 to 0·753 in Model 1.

In other investigations of HIV-infected patients, time-dependent measures of clinical status such as severe weight loss and functional disability are strongly predictive of censoring and HIV-related morbidity and mortality, even after controlling for biological markers such as CD4 count and other baseline factors (Robins & Finkelstein, 2000). Since these variables were not available to us, it is plausible that the true value of α_1 in all of the above models is negative. If we consider a mild degree of residual dependence such as $\alpha_1 = -0{\cdot}5$, we see that the estimated two-year probabilities of remaining free of the primary event are 0·6704, 0·6626 and 0·6603 for Models 1, 2 and 3, respectively. This represents approximately an 8% change in the estimate of two-year survival probability, relative to the analyses which assumed $\alpha_1 = 0$.

7. DISCUSSION

We view our approach as a compromise between drawing a single inference based on sequential ignorability of censoring and presenting worst case bounds. We prefer our approach over reporting only bounds because the latter are often too wide to be of practical use and correspond to implausible scenarios. In the low-dimensional case, our estimates of the survivor function converge to well-established bounds as the censoring bias parameter $\alpha_1 \to \pm\infty$. We recommend that the investigator choose a parameterisation for which (i) experts have some feeling for the magnitude of the parameters, (ii) there exists parameter value(s) that correspond to sequential ignorability of censoring, and

(iii) at least in the low-dimensional case well-known bounds are achieved as the censoring bias parameters become large in absolute value.

Elicitation of plausible ranges for the censoring parameters from experts is an important issue. We suggest two approaches. In the first approach, one directly asks the expert to specify a range of relative risks, adjusted for recorded prognostic factors, of being censored at any time t for subjects whose time-to-event differs by one year. Potential problems with this approach are discussed in § 7.2.3 of Scharfstein et al. (1999). In the second approach, one uses the observed data to derive the imputed distribution of time-to-event for censored subjects for various values of the censoring bias parameters and asks the experts to assess the plausibility of these distributions relative to the actual censoring times and the distribution of time-to-events among subjects who were observed to fail. There are number of philosophical and inferential issues associated with this latter data-dependent approach, as discussed in § 8 of Scharfstein et al. (1999).

In future papers, we plan to extend our sensitivity analysis methodology to the k-sample and regression settings.

Acknowledgement

The authors are grateful to Mark van der Laan for fruitful discussions and to anonymous referees for suggestions that improved the quality of the paper. This research was partially supported by the National Institutes of Health.

References

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

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. & Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26**, 641–7.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.

Fisher, L. & Kanarek, P. (1974). Presenting censored survival data when censoring and survival times may not be independent. In *Reliability and Biometry: Statistical Analysis of Lifelength*, Ed. F. Proschan and R. Serfling, pp. 303–26. Philadelphia: SIAM.

Gill, R. D., Laan, M. J. van der & Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures and counterexamples. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, Ed. D. Lin and T. Fleming, pp. 255–94. New York: Springer-Verlag.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., Hirsch, M. S. & Merigan, T. C. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New Engl. J. Med.* **335**, 1081–90.

Heitjan, D. F. & Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19**, 2244–53.

Huber, P. J. (1985). Projection pursuit (with Discussion). *Ann. Statist.* **13**, 435–525.

Jacobsen, M. & Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time. *Ann. Statist.* **23**, 774–86.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457–81.

Klein, J. P. & Moeschberger, M. L. (1988). Bounds on net survival probabilities for dependent competing risks. *Biometrics* **44**, 529–38.

Klein, J. P., Moeschberger, M. L., Li, Y. H. & Wang, S. T. (1992). Estimating random effects in the Framingham Heart Study (with Discussion). In *Survival Analysis: State of the Art*, Ed. J. Klein and P. Goel, pp. 99–120. Dordrecht: Kluwer.

Moeschberger, M. L. & Klein, J. P. (1995). Statistical methods for dependent competing risks. *Lifetime Data Anal.* **1**, 195–204.

Newey, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Economet.* **5**, 99–135.

Peterson, A. V. (1976). Bounds for a joint distribution function with fixed subdistribution functions: Application to competing risks. *Proc. Nat. Acad. Sci.* **73**, 11–3.

ROBINS, J. M. (1987). A new approach to causal inference in mortality studies with sustained exposure periods — application to control of the healthy worker survivor effect (addendum). *Comp. Math. Applic.* **14**, 917–21.

ROBINS, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, Am. Statist. Assoc.*, pp. 24–33. Alexandria, VA: American Statistical Association.

ROBINS, J. M. & FINKELSTEIN, D. H. (2000). Correcting for non-compliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56**, 779–88.

ROBINS, J. M. & RITOV, Y. (1997). Toward a curse of dimensionality appropriate (coda) asymptotic theory for semiparametric models. *Statist. Med.* **16**, 285–319.

ROBINS, J. M. & ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology: Methodolgical Issues*, Ed. N. Jewell and K. Dietz, pp. 297–331. Boston: Birkhauser.

ROBINS, J. M., MARK, S. D. & NEWEY, W. K. (1992). Estimating exposure effects by modeling the expectation of exposure conditional on confounders. *Biometrics* **48**, 479–95.

SATTEN, G. A., DATTA, S. & ROBINS, J. M. (2001). An estimator for the surival function when data are subject to dependent censoring. *Statist. Prob. Lett.* **54**, 397–403.

SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with Discussion). *J. Am. Statist. Assoc.* **94**, 1096–146.

SCHARFSTEIN, D. O., ROBINS, J. M., EDDINGS, W. & ROTNITZKY, A. (2001). Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics* **57**, 404–13.

SLUD, E. V. & RUBINSTEIN, L. V. (1983). Dependent competing risks and summary survival curves. *Biometrika* **70**, 643–9.

ZHENG, M. & KLEIN, J. P. (1994). A self-consistent estimator of marginal survival functions based on dependent competing risk data and an assumed copula. *Commun. Statist.* A **23**, 2299–311.

ZHENG, M. & KLEIN, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82**, 127–38.

[*Received December* 2000. *Revised December* 2001]