

Estimation of the time-dependent accelerated failure time model in the presence of confounding factors

BY JAMES ROBINS

Occupational Health Program and Department of Biostatistics, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

SUMMARY

Cox & Oakes (1984, p. 66) introduced the 'strong version' of the accelerated failure time model with time-dependent exposures. We provide conditions under which this model could be used to estimate, from observational data, the causal effect of a time-varying exposure or treatment on time to an event of interest in the presence of time-dependent confounding variables. We propose a class of semiparametric tests and estimators for the model parameters. This class contains an estimator that is semiparametric efficient in the sense of Begun et al. (1983).

Some key words: AIDS; Causal; Observational study; Semiparametric analysis; Survival data; Time-dependent covariate.

1. INTRODUCTION

The purpose of this paper is first to provide conditions under which it would be possible to estimate, from observational data, the causal effect of a time-dependent treatment or exposure on time to an event of interest in the presence of time-dependent confounding covariates and, then, to construct estimators for the treatment effect under these conditions. Our approach will be based on the semiparametric estimation of the parameters of the 'strong version' of the accelerated failure time model with time-dependent exposures or treatments introduced by Cox & Oakes (1984, p. 66).

The usual approach to the estimation of the effect of a time-varying treatment on survival is to model the hazard of failure at t as a function of past treatment history using a time-dependent proportional hazards model. We show that the usual approach may be biased, whether or not one further adjusts for past confounder history in the analysis, when (a) there exists a time-dependent risk factor for, or predictor of, the event of interest that also predicts subsequent treatment, and (b) past treatment history predicts subsequent risk factor level. The following four examples demonstrate conditions (a) and (b) will often be true in an observational study in which there is 'treatment by indication'.

The drug zidovudine, formerly AZT, used in the treatment of the acquired immunodeficiency syndrome, is a direct red-blood cell toxin that is contra-indicated in anaemic subjects, i.e. subjects with depressed red-cell counts, since the toxic effects of zidovudine can worsen the anaemia. Further, anaemic patients are at increased risk of death. Thus in an observational study of the effect of zidovudine on survival of patients with the acquired immunodeficiency syndrome, anaemia is both a risk factor for death

and a predictor of subsequent treatment with zidovudine. Further, as a red-blood-cell toxin, past zidovudine treatment is a risk factor for the development of anaemia. As a second example, many physicians withdraw women from exogenous oestrogens at the time they develop an elevated blood cholesterol. Therefore, in a study of the effect of postmenopausal oestrogen on cardiac mortality, the time-dependent cardiac risk factor cholesterol is a predictor of subsequent exposure. Further, past oestrogen treatment influences subsequent cholesterol levels. As a third example, in observational studies of the efficacy of cervical cancer screening on mortality, women who have had operative removal of their cervix due to invasive disease are no longer at risk for further screening, i.e. exposure, but are at increased risk for death. Therefore, the covariate, 'operative removal of the cervix', is an independent risk factor for death and a predictor of subsequent exposure. Further, past screening can prevent the need for operative removal of the cervix by allowing early removal of superficial lesions in the pre-invasive stage. As a final epidemiologic example, in occupational mortality studies, unhealthy workers who terminate employment early are at increased risk of death compared to other workers and receive no further exposure to the chemical agent under study. Therefore, the time-dependent covariate, employment status at time t , is an independent risk factor for death and a predictor of exposure to the study agent. In addition, previous exposure to the study agent may lead to early termination of employment if the agent causes a disabling illness. Epidemiologists refer to these covariates as time-dependent confounders.

In § 2, we formally state the fundamental assumption that, if true, will allow us to draw causal inferences from observational data, and show that, under this assumption, standard approaches based on the Cox proportional hazard model cannot be used to test for a treatment effect. We introduce the strong version of the accelerated failure time model with time-dependent treatments in § 3, and consider fully parametric likelihood-based inference for the model parameters in § 4. In § 5, we consider semiparametric G -estimation of the model parameters in the simple setting of a dichotomous treatment. Section 6 generalizes the model of § 3. Section 7 considers non-dichotomous treatments. In § 8, we show that the optimal estimator in our class is semiparametric efficient. Section 9 provides computational formulae that should be useful for applications. Section 10 extends our results to allow for censoring. We conclude with a discussion.

2. CAUSAL INFERENCE FROM OBSERVATIONAL DATA

2.1. *The data*

For pedagogic purposes we shall consider a study of the effect of zidovudine treatment on the survival of patients with the acquired immunodeficiency syndrome. Let T_i be a continuous random variable recording the survival time for the i th study subject, for $i = 1, \dots, n$, with time measured from study enrollment. Let $A_i(t)$ record subject i 's zidovudine dosage rate at t and $L_i(t)$ record the value at time t of a vector of various time-dependent and time-independent covariates such as CD4-lymphocyte count, red blood cell count, and gender. For any time-dependent random variable $Z_i(t)$, let $\bar{Z}_i(t) = \{Z_i(u); 0 \leq u \leq t\}$ be the history of the Z -process through t . Note that $Z_i(t)$ is defined only for $t \leq T_i$. In the absence of censoring the observable random variables are then $\{T_i, \bar{A}_i(T_i), \bar{L}_i(T_i)\}$, which we assume are independent and identically distributed.

Following Cox & Oakes (1984), Robins (1986, 1987) and Rubin (1978), we shall also assume there exists a latent 'baseline' failure time random variable U_i representing subject i 's survival time had zidovudine always been withheld.

2.2. The fundamental assumption of no unmeasured confounders

Our fundamental assumption of no unmeasured confounders is

$$U_i \perp\!\!\!\perp A_i(t^+) \mid \bar{L}_i(t), \bar{A}_i(t), T_i > t, \quad (2.1)$$

where $A \perp\!\!\!\perp B \mid C$ means A is independent of B given C (Dawid, 1979). This assumption implies that, conditional on zidovudine history and the history of all recorded time-independent and time-dependent covariates through t , increments in zidovudine dosage rate at t^+ are independent of the baseline failure time random variable U_i . This assumption will be true if all risk factors for, i.e. predictors of, baseline failure time U_i that are used by patients and physicians to determine the dosage of zidovudine at t^+ are recorded in $\bar{L}_i(t)$ and $\bar{A}_i(t)$. For example, since physicians tend to withhold zidovudine from anaemic subjects, and in untreated subjects, anaemia is an independent predictor of survival, our fundamental assumption would be false if $\bar{L}_i(t)$ does not contain red cell count history. It is the primary goal of the epidemiologists conducting an observational study to collect data on a sufficient number of covariates to ensure that our assumption (2.1) will be true. In this paper, we assume this goal has been realized, while recognizing that, in practice, this would never be precisely true and may, on occasion, not even be approximately true.

Assumption (2.1) is the fundamental condition that will allow us to draw causal inferences from observational data. It is precisely because (2.1) cannot be guaranteed to hold in an observational study that it is so very hazardous to draw causal inferences from observational data. Note that if, at each time t^+ , the dose of zidovudine was chosen at random by the flip of a coin, then (2.1) would be true even if the probability that the coin landed heads depended on past covariate and zidovudine-history. It is because physical randomization guarantees (2.1) that most people accept that valid causal inferences can be obtained from a randomized trial. See Rubin (1978), Robins (1986) and Holland (1986) for further discussion.

Henceforth, it will be convenient to discretize the times at which data on covariates in $L_i(t)$ were recorded. Suppose that measurements on time-dependent covariates were recorded only at times $k = 0, 1, 2, \dots$. Let $L_{k,i}$ be the vector consisting of all time-dependent covariate measurements made on subject i at time k except that $L_{0,i}$ also includes data on any time-independent covariates such as race, sex, etc. We suppose that each subject's zidovudine dosage rate can be assumed to be constant in each interval $(k, k+1]$ and denote by $A_{k,i}$ the zidovudine dosage rate in that interval. Thus, jumps in the zidovudine process occur only at times k^+ while jumps in the L -process measurements occur only at times k . Write $\bar{A}_{k,i} = (A_{0,i}, \dots, A_{k,i})$. Using these conventions, L -history and zidovudine history through k can be written as $\bar{L}_{k,i}$ and $\bar{A}_{k-1,i}$ respectively. Therefore, (2.1) is equivalent to the assumption that the zidovudine dose rate $A_{k,i}$ at k^+ is independent of U_i conditional on L -history $\bar{L}_{k,i}$ and treatment history $\bar{A}_{k-1,i}$ through k . That is, for $k = 0, 1, \dots$,

$$A_{k,i} \perp\!\!\!\perp U_i \mid \bar{L}_{k,i}, \bar{A}_{k-1,i}, T_i > k, \quad (2.2)$$

where, for notational convenience, we define $\bar{A}_{-1,i}$ and $\bar{L}_{-1,i}$ to be random variables that take the value zero with probability one.

2.3. Bias of standard methods

The sharp causal null hypothesis of no treatment effect on survival is that each subject's observed and baseline lifetimes are the same. That is, for all i ,

$$U_i = T_i. \quad (2.3)$$

Given our assumption (2.1), the restriction on the distribution of the observables implied by (2.3) is

$$T_i \sqcup A_{k,i} | \bar{L}_{k,i}, \bar{A}_{k-1,i}, T_i > k. \quad (2.4)$$

Thus, under (2.1), an α -level test of (2.4) is an α -level test of (2.3). One standard approach to testing for a causal effect of zidovudine is to test the hypothesis

$$\lambda\{t | \bar{A}_i(t)\} = \lambda(t), \quad (2.5)$$

where $\lambda(t | \cdot)$ is the hazard at t of T_i conditional on \cdot , by testing the hypothesis $\theta = 0$ in a time-dependent Cox proportional hazards model

$$\lambda\{t | \bar{A}_i(t)\} = \lambda_0(t) \exp\{\theta W_i(t)\}, \quad (2.6)$$

where $W_i(t)$ is a real-valued function of $\bar{A}_i(t)$ such as

$$W_i(t) = \int_0^t A_i(u) du$$

or $W_i(t) = A_i(t)$.

A second standard approach to testing for the causal effect of zidovudine is to test the hypothesis

$$\lambda\{t | \bar{A}_i(t), \bar{L}_i(t)\} = \lambda\{t | \bar{L}_i(t)\} \quad (2.7)$$

by testing whether $\theta = 0$ in a correctly specified time-dependent Cox model

$$\lambda\{t | \bar{A}_i(t), \bar{L}_i(t)\} = \lambda_0(t) \exp\left\{\theta W_i(t) + \sum_{m=1}^M \gamma_m Z_{m,i}(t)\right\}, \quad (2.8)$$

where $\{Z_{1,i}(t), \dots, Z_{M,i}(t)\}$ are real-valued functions of $\bar{L}_i(t)$ and the γ_m are unknown parameters.

The test of $\theta = 0$ in model (2.6) can be an α -level test of (2.4), and thus of (2.3), only if (2.4) implies (2.5). The test of $\theta = 0$ in model (2.8) can be an α -level test of (2.4) only if (2.4) implies (2.7). Since (2.4) neither implies (2.5) nor (2.7), inference concerning the causal null hypothesis (2.3) cannot be based on (2.6) or (2.8) except under special circumstances. For example, it follows from Theorem (A.2) and Corollary (A.2) of Robins (1989a) that (2.4) implies (2.5) if either

$$\lambda\{t | \bar{A}_i(t), \bar{L}_i(t)\} = \lambda\{t | \bar{A}_i(t)\} \quad (2.9)$$

or

$$f(A_{k,i} | \bar{A}_{k-1,i}, \bar{L}_{k,i}, T_i \geq k) = f(A_{k,i} | \bar{A}_{k-1,i}, T_i \geq k). \quad (2.10)$$

To show this, replace $d_{G=(\bar{E}_i)}$ by T in the statement and proof of Theorem (A.2) and Corollary (A.2) of Robins (1989a) and follow the proof of Theorem 6.1 of Robins (1986).

If (2.9) holds, we say L -history is not an independent risk factor for death. If (2.10) holds, we say L -history is not a predictor of subsequent exposure. If (2.9) and (2.10) are both false, we say L -history is a confounder for the causal effect of treatment history on survival (Robins, 1986, 1989a).

As a second example, straightforward calculations show that if

$$f(L_{k,i} | \bar{A}_{k-1,i}, \bar{L}_{k-1,i}, T_i \geq k) = f(L_{k,i} | \bar{L}_{k-1,i}, T_i \geq k) \quad (2.11)$$

then (2.7) implies (2.4), although (2.4) still does not imply (2.7). If (2.11) holds, we say that past treatment history is not a predictor of subsequent covariates status. Thus, even

when (2.11) is true, we do not obtain α -level test of (2.4) from an α -level test of (2.7). Nonetheless, if (2.11) is true and a α -level test of the hypothesis (2.7) accepts, we can accept (2.4), although, when viewed as a test of (2.4), the true α -level of this test may differ from its nominal level.

Importantly, in each of the four examples provided in § 1, equations (2.9), (2.10) and (2.11) were all false, so that inferences concerning the causal null hypothesis (2.3) under assumption (2.1) could not be based on proportional hazards models (2.6) or (2.8). However, we shall see that inferences concerning the null hypothesis (2.3) can be based on the strong version of the accelerated failure-time model.

3. STRONG VERSION OF THE ACCELERATED FAILURE TIME MODEL

In this section following Cox & Oakes (1984, § 5.2) we shall assume that the latent baseline variable U_i is related to the observables $\{T_i, \bar{A}_i(T_i)\}$ by the relationship

$$U_i = \int_0^{T_i} \exp\{\psi_0 A_i(u)\} du, \quad (3.1)$$

where $\psi_0 \in R^1$ is an unknown parameter. This model is a particular example of the strong version of the accelerated failure time model with time-dependent covariates proposed by Cox & Oakes (1984, § 5.2). Note this model is a structural or counterfactual model in the sense that it links the observable random variables with the counterfactual latent random variable U_i . The hypothesis $\psi_0 = 0$ is equivalent to the causal null hypothesis (2.3). Thus the model (3.1) is correctly specified under the null hypothesis (2.3), and an α -level test of $\psi_0 = 0$ is an α -level test of (2.3).

Suppose that $A_i(t)$ is dichotomous with $A_i(t) = 1$ if subject i is being treated at t and $A_i(t) = 0$ otherwise. Then (3.1) implies that, for any subject i who received continuous treatment, $U_i = \exp(\psi_0) T_i$. Thus, $\psi_0 < 0$ implies continuous treatment extends life by a factor $\exp(-\psi_0)$ and $\psi_0 > 0$ implies that continuous treatment decreases life by a factor $\exp(-\psi_0)$.

To better understand (3.1), it is convenient to follow the suggestion of Cox & Oakes (1984) and consider two time scales. Namely, let u correspond to the baseline time scale and t the actual or real time scale. Let the mapping from t to u for the i th individual be given by

$$u = \int_0^t \exp\{\psi_0 A_i(x)\} dx.$$

Then $du/dt = \exp\{\psi_0 A_i(t)\}$ and $dt/du = \exp\{-\psi_0 A_i(t)\}$. dt/du corresponds to the relative rate at which real time is being used up compared to baseline time. Thus, if individual i has U_i years of baseline time to be used up, the actual time T_i at which these U_i years of baseline time will have been used is given by (3.1).

4. FULLY PARAMETRIC LIKELIHOOD-BASED INFERENCE

Cox & Oakes (1984, § 5.2) and Robins & Tsiatis (1992) considered inference for the parameters ψ_0 of (3.1) in the absence of time-dependent confounding. In this section, we consider fully parametric likelihood-based inference for ψ_0 in the presence of the time-dependent confounders $L_i(t)$. In order to write the likelihood as a function of the

parameter ψ we define

$$H_i(\psi) = \int_0^{T_i} \exp\{\psi A_i(u)\} du.$$

Thus, under model (3.1), $U_i = H_i(\psi_0)$. Since the map from $\{T_i, \bar{L}_i(T_i), \bar{A}_i(T_i)\}$ to $\{H_i(\psi), \bar{L}_i(T_i), \bar{A}_i(T_i)\}$ is one to one with strictly positive Jacobian determinant $\partial H_i(\psi)/\partial T_i = \exp\{\psi A_i(T_i)\}$, we have

$$f_{\{T, \bar{L}(T), \bar{A}(T)\}}\{T_i, \bar{L}_i(T_i), \bar{A}_i(T_i)\} = \left\{ \frac{\partial H_i(\psi)}{\partial T_i} \right\} f_{\{H(\psi), \bar{L}(T), \bar{A}(T)\}}\{H_i(\psi), \bar{L}_i(T_i), \bar{A}_i(T_i)\}. \quad (4.1)$$

Define $\text{int}(t)$ to be the largest integer less than t . Then, by a decomposition into a product of conditional probabilities, the right-hand side of (4.1) can be written

$$\left[\frac{\partial H_i(\psi)}{\partial T_i} \right] f\{H_i(\psi)\} \prod_{m=0}^{m=\text{int}(T_i)} f\{L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi)\} f\{A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}, H_i(\psi)\}, \quad (4.2)$$

which, for $\psi = \psi_0$, can be written

$$\left[\frac{\partial H_i(\psi_0)}{\partial T_i} \right] f\{H_i(\psi_0)\} \prod_{m=0}^{m=\text{int}(T_i)} f\{L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\} \prod_{m=0}^{m=\text{int}(T_i)} f\{A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}\}, \quad (4.3)$$

since (2.2) and (3.1) together imply

$$f\{A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\} = f\{A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}\}. \quad (4.4)$$

The condition (4.4) is the sole restriction on the joint distribution of the observables implied by (2.2) and (3.1). In the interest of clarity, we shall describe how one would generate simulated data from our model. To draw R independent realizations $\{t_r, \bar{a}_r(t_r), \bar{l}_r(t_r)\}$ ($r = 1, \dots, R$), from the law of $\{T, \bar{A}(T), \bar{L}(T)\}$, first initialize r to 1. Then proceed as follows.

Step 1. Draw h_r from $f_{H(\psi_0)}(h)$ and set $m = 0$.

Step 2. Draw $l_{m,r}$ from $f(l_m | \bar{l}_{m-1,r}, \bar{a}_{m-1,r}, h_r)$ and $a_{m,r}$ from $f(a_m | \bar{l}_{m,r}, \bar{a}_{m-1,r})$.

Step 3. If

$$h_r > \int_0^{m+1} \exp\{\psi_0 a_r(t)\} dt \quad (4.5)$$

increase m by one and return to Step (2); otherwise, let $x, x \in (m, m+1]$, solve

$$h_r = \int_0^x \exp\{\psi_0 a_r(t)\} dt,$$

set t_r equal to x , $\bar{a}_r(t_r)$ equal to $\bar{a}_{m,r}$, and $\bar{l}_r(t_r)$ equal to $\bar{l}_{m,r}$. Finally, if $r < R$, increase r by 1 and return to step 1.

That is, if a simulated subject r with baseline time h_r manages to survive to real time m , we randomly draw $l_{m,r}$ and $a_{m,r}$ and then use (4.5) to determine whether subject r has survived to real time $m+1$ or whether the subject has died at a time t_r in the interval $(m, m+1]$.

We can specify a fully parametric model for the distribution of $\{T_i, \bar{L}_i(T_i), \bar{A}_i(T_i)\}$ by specifying parametric models $f(L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi_0); \phi)$, $f(H_i(\psi_0); \theta)$ and $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \alpha)$. From (4.3), the maximum likelihood estimators of the true values

$(\phi_0, \theta_0, \psi_0)$ are the values of $(\hat{\phi}, \hat{\theta}, \hat{\psi})$ that maximize

$$\prod_i \left\{ \frac{\partial H_i(\psi)}{\partial T_i} \right\} f\{H_i(\psi); \theta\} \prod_{m=0}^{m=\text{int}(T_i)} f(L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi); \phi). \quad (4.6)$$

5. SEMIPARAMETRIC ESTIMATION OF ψ_0

It follows from (4.6) that fully parametric likelihood-based inference for ψ_0 is completely robust to misspecification of models for $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$. In contrast, fully parametric likelihood based inference for ψ_0 will not be robust to misspecification of the models for $f\{H_i(\psi_0)\}$ and $f\{L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\}$. Unfortunately, since the baseline failure time $H_i(\psi_0)$ is not directly observed, one would expect that it would be difficult if not impossible for an investigator to specify a parametric model for $f\{L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\}$ believed to approximate the truth. In contrast, one might expect that an investigator might be much more secure specifying a parametric model for the densities $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$, since such models require only that the investigator have a general understanding of the behavioural process by which patients and physicians chose the zidovudine dose $A_{m,i}$ at m^+ given their knowledge of past zidovudine and covariate history. Thus, we now develop semiparametric tests and estimators of ψ_0 under the assumption that we have correctly specified models for the densities $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ while allowing the densities $f\{H_i(\psi_0)\}$ and

$$f\{L_{m,i} | \bar{L}_{m-1,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\}$$

to remain completely unrestricted. In the setting of a single time-independent dichotomous treatment, Rosenbaum & Rubin (1983), Rosenbaum (1984) and Robins, Mark & Newey (1992) have previously considered estimating treatment effects in observational studies by modelling the probability of treatment given covariates. In this section we outline the main results of the paper in the simple setting of a dichotomous time-dependent treatment variable and model (3.1). In subsequent sections these results are generalized.

We shall construct an asymptotic α -level test of the hypothesis that a particular test value ψ^* of the parameter ψ equals the true value ψ_0 , based on a correctly specified logistic model

$$f(A_{m,i} = 1 | \bar{L}_{m,i}, \bar{A}_{m-1,i}) = \frac{\exp(\alpha'_0 W_{m,i})}{1 + \exp(\alpha'_0 W_{m,i})} \quad (m = 1, 2, \dots), \quad (5.1)$$

where $\alpha_0 \in R^p$ is an unknown parameter and $W_{m,i}$ is a p -dimensional function of $(\bar{L}_{m,i}, \bar{A}_{m-1,i})$. As a simple example, $W_{m,i}$ might be the eleven vector consisting of the constant one, $A_{m-1,i}$, $A_{m-2,i}$, m , $(mA_{m-1,i})$, $(A_{m-1,i} A_{m-2,i})$, red cell count at m , average red cell count prior to m , age, race and gender. The key idea is to note that if ψ^* equals ψ_0 , then, under (5.1) and (4.4), for subjects with $T_i > m$,

$$P_{m,i}(\alpha, \theta) = \frac{\exp[\alpha' W_{m,i} + \theta\{H_i(\psi^*)\}]}{1 + \exp[\alpha' W_{m,i} + \theta\{H_i(\psi^*)\}]} \quad (5.2)$$

is a correctly specified logistic model for $f\{A_{m,i} = 1 | \bar{L}_{m,i}, \bar{A}_{m-1,i}, H_i(\psi^*)\}$ with true value of α equal to α_0 and the true value θ_0 of θ equal to 0, since (4.4) implies $\theta_0 = 0$. In model (5.2) $H_i(\psi^*)$ is simply treated as a fixed covariate. Since $\psi^* = \psi_0$ implies $\theta_0 = 0$, we obtain, for each choice of m , an asymptotic α -level test of the hypothesis $\psi^* = \psi_0$ by performing an α -level score test of the hypothesis that $\theta_0 = 0$ in the logistic model (5.2). In particular,

we obtain a test of the causal null hypothesis (2.3), i.e. the hypothesis $\psi_0 = 0$, by choosing ψ^* to be 0 and thus $H_i(\psi^*)$ to be T_i in (5.2).

A single asymptotic α -level G -test of the hypothesis $\psi^* = \psi_0$ that combines information over all times m is obtained by performing any standard likelihood based test of the hypothesis $\theta_0 = 0$ based on the 'likelihood' of the extended logistic model

$$\mathcal{L}(\alpha, \theta) = \prod_{i=1}^n \prod_{m=0}^{\text{int}(T_i)} \mathcal{L}_{m,i}(\alpha, \theta), \quad (5.3)$$

where

$$\mathcal{L}_{m,i}(\alpha, \theta) = \{P_{m,i}(\alpha, \theta)\}^{A_{m,i}} \{1 - P_{m,i}(\alpha, \theta)\}^{1 - A_{m,i}}. \quad (5.4)$$

That is, the likelihood ratio test

$$-2 \log \{ \mathcal{L}(\tilde{\alpha}, 0) / \mathcal{L}(\hat{\alpha}, \hat{\theta}) \} \quad (5.5)$$

and the score test

$$\Gamma^2 \equiv S_{\theta}^2(\tilde{\alpha}, 0) \hat{I}^{\theta\theta} \quad (5.6)$$

of the hypothesis $\theta_0 = 0$ are both asymptotically chi-squared distributed on one degree of freedom when $\psi^* = \psi_0$, where $(\hat{\alpha}, \hat{\theta})$ maximizes (5.3), $\tilde{\alpha}$ maximizes (5.3) when θ is constrained to be 0, $S_{\theta}(\alpha, \theta) = \partial \log \mathcal{L}(\alpha, \theta) / \partial \theta$, $\hat{I}^{\theta\theta} \equiv (\hat{I}_{\theta\theta} - \hat{I}_{\alpha\theta} \hat{I}_{\alpha\alpha}^{-1} \hat{I}_{\theta\alpha})^{-1}$ and, for example, $\hat{I}_{\theta\alpha} = -\partial^2 \log \{ \mathcal{L}(\tilde{\alpha}, 0) \} / \partial \theta \partial \alpha$. Here ψ^* is held fixed in each maximization.

The likelihood ratio test (5.5) and the score test (5.6) can be computed using standard logistic regression software by regarding the data as $\sum \text{int}(T_i + 1)$ independent Bernoulli responses, where each subject i contributes $\text{int}(T_i + 1)$ responses $A_{m,i}$ with associated covariates $\{W_{m,i}, H_i(\psi^*)\}$ for $m = 0, \dots, \text{int}(T_i)$.

Since $\mathcal{L}_{m,i}(\alpha, \theta)$ is fixed given $\{\bar{L}_{k,i}, \bar{A}_{k-1,i}, H_i(\psi^*)\}$ ($k > m$), (5.3) is a correctly specified partial likelihood under (5.1) and (4.4) when $\psi^* = \psi_0$. It follows that (5.5) and (5.6) are partial likelihood ratio and score tests of the true hypothesis $\theta_0 = 0$ when $\psi^* = \psi_0$ and thus are asymptotically distributed χ_1^2 .

Write the score statistic Γ^2 as $\Gamma^2(\psi^*)$ to make explicit its dependence on ψ^* . The set of ψ^* for which $\Gamma^2(\psi^*)$ is less than the $1 - \alpha$ percentile of a χ_1^2 random variable is an asymptotic $(1 - \alpha)$ confidence interval or set for ψ_0 . The solution $\tilde{\psi}$ of $\Gamma^2(\psi^*) = 0$ is a point estimate of ψ_0 , whose properties we study in § 7. If we had defined (5.2) by

$$P_{m,i}(\alpha, \theta) = \frac{\exp[\alpha' W_{m,i} + \theta G_{m,i}\{H_i(\psi^*)\}]}{1 + \exp[\alpha' W_{m,i} + \theta G_{m,i}\{H_i(\psi^*)\}]}, \quad (5.7)$$

where $G_{m,i}\{H_i(\psi^*)\} = g\{H_i(\psi^*), \bar{A}_{m-1,i}, \bar{L}_{m,i}\}$ and $g(\cdot, \cdot, \cdot)$ is a fixed real valued function, for example

$$g\{H_i(\psi^*), \bar{A}_{m-1,i}, \bar{L}_{m,i}\} = \{\log H_i(\psi^*)\} A_{m-1,i},$$

then, when $\psi^* = \psi_0$ and (5.1) and (4.4) hold, (5.3) remains a correctly specified partial likelihood with true values α_0 and $\theta_0 = 0$. Therefore the set of ψ^* for which $\Gamma^2(\psi^*)$ is less than the $1 - \alpha$ percentile of a χ_1^2 random variable remains a $(1 - \alpha)$ confidence interval or set for ψ_0 . In § 8, we give the function $g(\cdot, \cdot, \cdot)$ that minimizes confidence interval length.

6. THE GENERAL STRONG VERSION OF THE ACCELERATED FAILURE TIME MODEL

In this section, we generalize model (3.1). Model (3.1) is a special case of the general strong version of the accelerated failure time model with time-dependent covariates that

links U_i to $\{T_i, \bar{A}_i(T_i)\}$ by assuming

$$U_i = h\{T_i, \bar{A}_i(T_i), \psi_0\}, \tag{6.1}$$

where $\psi_0 \in R^v$ is an unknown vector of parameters to be estimated, and $h(., ., .)$ is a known smooth function, satisfying

- (i) monotonicity: $h\{t, \bar{A}_i(t), \psi\} > h\{u, \bar{A}_i(u), \psi\}$ if $t > u$;
- (ii) identity: $h\{t, \bar{A}_i(t), \psi\} = t$ if $\bar{A}_i(t)$ is identically zero on $(0, t)$;
- (iii) $h\{t, \bar{A}_i(t), 0\} = t$ so that $\psi_0 = 0$ represents the null hypothesis of no causal effect of treatment on time to failure.

Define $H_i(\psi) = h\{T_i, \bar{A}_i(T_i), \psi\}$ so that $H_i(\psi_0) = U_i$. Equation (4.4) remains the sole restriction on the distribution of the observables implied by (2.1) and (6.1). A concrete example of model (6.1) with $\psi_0 \in R^2$ is discussed in § 11.

7. GENERALIZATIONS

We now generalize the results of § 5 by allowing $A_{m,i}$ to be continuous, ordinal or discrete and ψ_0 in model (6.1) to be of dimension $v > 1$. Given a correctly specified model

$$f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}) = f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \alpha_0), \tag{7.1}$$

where $\alpha_0 \in R^p$ is an unknown parameter and where, for each $\alpha \in R^p$, $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \alpha)$ is a density with respect to a measure μ , we define

$$\begin{aligned} \mathcal{L}(\alpha, \theta, \psi) &= \prod_{i=1}^n \mathcal{L}_i(\alpha, \theta, \psi), \quad \mathcal{L}_i(\alpha, \theta, \psi) = \prod_{m=0}^{\text{int}(T_i)} \mathcal{L}_{m,i}(\alpha, \theta, \psi), \\ \mathcal{L}_{m,i}(\alpha, \theta, \psi) &= \frac{f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \alpha) \exp[\theta' Q_{m,i}\{A_{m,i}, H_i(\psi)\}]}{\int f(a_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \alpha) \exp[\theta' Q_{m,i}\{a_{m,i}, H_i(\psi)\}] d\mu(a_{m,i})}, \end{aligned} \tag{7.2}$$

$\theta \in R^v$ and

$$Q_{m,i}\{A_{m,i}, H_i(\psi)\} \equiv q(A_{m,i}, H_i(\psi), \bar{A}_{m-1,i}, \bar{L}_{m,i}) \in R^v,$$

where $q(., ., ., .)$ is a fixed function. Note that if $\psi_0 \in R^1$, $A_{m,i}$ is dichotomous, (7.1) is given by the logistic model (5.1), and

$$Q_{m,i}\{A_{m,i}, H_i(\psi_0)\} \equiv [G_{m,i}\{H_i(\psi)\}] A_{m,i},$$

one can calculate that $\mathcal{L}_{m,i}(\alpha, \theta, \psi)$ is given by (5.4) with $P_{m,i}(\alpha, \theta)$ as defined in (5.7). We have modified the notation of § 5 in that the dependence of $\mathcal{L}_{m,i}(\alpha, \theta, \psi)$ on ψ has now been made explicit.

Equations (4.4) and (7.1) imply that $\mathcal{L}_{m,i}(\alpha, \theta, \psi_0)$ is a correctly specified model for $f\{A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\}$ with true values α_0 and $\theta_0 = 0$ so that $\mathcal{L}(\alpha, \theta, \psi_0)$ is a correctly specified partial likelihood. We shall use the following notation. Let $\tilde{\alpha} = \tilde{\alpha}(\psi)$ maximize $\mathcal{L}(\alpha, 0, \psi)$ and $(\hat{\alpha}, \hat{\theta})$ maximize $\mathcal{L}(\alpha, \theta, \psi_0)$. Define

$$S_\theta(\alpha, \theta, \psi) \equiv \frac{\partial \log \mathcal{L}(\alpha, \theta, \psi)}{\partial \theta}, \quad S_\theta(\psi) \equiv S_\theta(\alpha_0, 0, \psi), \quad S_{\theta,i}(\alpha, \theta, \psi) \equiv \frac{\partial \log \mathcal{L}_i(\alpha, \theta, \psi)}{\partial \theta},$$

$$S_{\theta,i}(\psi) \equiv S_{\theta,i}(\alpha_0, 0, \psi), \quad S_{\theta,i} \equiv S_{\theta,i}(\alpha_0, 0, \psi_0).$$

Also $S_\alpha(\alpha, \theta, \psi)$, $S_{\alpha,i}(\alpha, \theta, \psi)$, and $S_{\alpha,i}$, are defined analogously. Define

$$S_i \equiv (S'_{\alpha,i}, S'_{\theta,i})', \quad i \equiv E(S_i S_i'), \quad S(\alpha, \theta, \psi) \equiv \{S'_\alpha(\alpha, \theta, \psi), S'_\theta(\alpha, \theta, \psi)\}',$$

$$I(\alpha, \theta, \psi) \equiv -\frac{\partial S(\alpha, \theta, \psi)}{\partial (\alpha, \theta)'}, \quad \hat{I}(\psi) \equiv I(\tilde{\alpha}, 0, \psi).$$

For any square $p + v$ dimensional square matrix a , write

$$a = \begin{bmatrix} a_{\alpha\alpha} & a_{\theta\alpha} \\ a_{\alpha\theta} & a_{\theta\theta} \end{bmatrix},$$

where $a_{\alpha\alpha}$ and $a_{\theta\theta}$ are square matrices of dimension p and v respectively. Define $a^{\theta\theta} \equiv (a^{-1})_{\theta\theta} = (a_{\theta\theta} - a_{\alpha\theta} a_{\alpha\alpha}^{-1} a_{\theta\alpha})^{-1}$. Note $\hat{I}_{\alpha\alpha} \equiv \hat{I}_{\alpha\alpha}(\psi)$ does not depend on ψ .

Since $\mathcal{L}(\alpha, \theta, \psi_0)$ is a partial likelihood and $\theta_0 = 0$, it follows that $n^{-1}\hat{I}(\psi_0)$ converges in probability to i ; $n^{1/2}\hat{\theta}$ is asymptotically normal with mean 0 and asymptotic variance $i^{\theta\theta}$; and $n^{-1/2}S_\theta(\tilde{\alpha}, 0, \psi_0)$ is asymptotically normal with mean 0 and asymptotic variance $(i^{\theta\theta})^{-1}$. Therefore the score statistic

$$\Gamma^2(\psi) \equiv S'_\theta(\tilde{\alpha}, 0, \psi) \hat{I}^{\theta\theta}(\psi) S_\theta(\tilde{\alpha}, 0, \psi)$$

is asymptotically distributed χ^2_v when $\psi = \psi_0$ so that the set of ψ for which $\Gamma^2(\psi)$ is less than the upper α -percentage point of χ^2_v random variable is an asymptotic $(1 - \alpha)$ confidence set for ψ_0 . Under regularity conditions, $S_\theta(\tilde{\alpha}, 0, \psi) = 0$ will have a solution $\tilde{\psi}$, which we shall call a G -estimate, that is consistent for ψ_0 . A Taylor expansion gives

$$0 = n^{-1/2}S_\theta(\tilde{\alpha}, 0, \tilde{\psi}) = n^{-1/2}S_\theta(\tilde{\alpha}, 0, \psi_0) + \left\{ n^{-1} \frac{\partial S_\theta(\tilde{\alpha}, 0, \psi_0)}{\partial \psi'} \right\} n^{1/2}(\tilde{\psi} - \psi_0) + o_p(1).$$

Thus, $n^{1/2}(\tilde{\psi} - \psi_0)$ will be asymptotically normal with mean 0 and asymptotic variance $B(i^{\theta\theta})^{-1}B'$ which can be consistently estimated by $n\hat{B}(\hat{I}^{\theta\theta})^{-1}\hat{B}'$, where

$$B^{-1} = E \left[\frac{\partial S_{\theta,i}(\alpha_0, 0, \psi_0)}{\partial \psi'} \right] \equiv E \left[\frac{\partial S_{\theta,i}(\psi_0)}{\partial \psi'} \right], \quad \hat{I}^{\theta\theta} \equiv \hat{I}^{\theta\theta}(\tilde{\psi}), \quad \hat{B}^{-1} \equiv n^{-1} \frac{\partial S_\theta(\tilde{\alpha}, 0, \tilde{\psi})}{\partial \psi'}.$$

8. EFFICIENCY PROPERTIES OF $\tilde{\psi}$

In this section, we show that increasing the number of covariates in our model (7.1) never decreases and often increases the efficiency with which ψ_0 is estimated. Let $\alpha^{(j)}$ ($j = 1, \dots, J$) represent the parameter vector in a finite sequence of correctly specified nested models for $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ ordered by increasing dimension $p^{(j)}$ of $\alpha^{(j)}$. Since the models are correctly specified, all but the first $p^{(1)}$ components of each $p^{(j)}$ -vector of the true values $\alpha_0^{(j)}$ are zero. Let $\tilde{\psi}^{(j)}$ be the estimator of ψ_0 based on solving $S_\theta(\tilde{\alpha}^{(j)}, 0, \psi) = 0$. Then

$$\text{var}^A \{ n^{1/2}(\tilde{\psi}^{(j)} - \psi_0) \} = B(i^{(j),\theta\theta})^{-1}B'$$

is nonincreasing with increasing j since B does not depend on the model (7.1) and $\{i^{(j),\theta\theta}\}^{-1} = \{\text{var}^A(n^{1/2}\hat{\theta}^{(j)})\}^{-1}$ is known from standard likelihood and partial likelihood theory to be nonincreasing with j . Thus, it is advantageous to use richly parameterized models for $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ both to guard against misspecification bias and to increase efficiency. However, our results assume that $\tilde{\alpha}^{(j)}$ is $n^{1/2}$ -consistent for $\alpha_0^{(j)}$. This limits the number of free parameters in a model for $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ as a function of sample size although results of Newey (1990) suggest that $n^{1/4}$ -consistency is sufficient.

Write $\tilde{\psi}$ as $\tilde{\psi}(q)$ to emphasize the dependence of $\tilde{\psi}$ on the choice of q . Define

$$q_{\text{opt}}\{A_{m,i}, H_i(\psi_0), \bar{A}_{m-1,i}, \bar{L}_{m,i}\} = E\{S_\psi(\psi_0) | \bar{L}_{m,i}, \bar{A}_{m-1,i}, H_i(\psi_0)\},$$

where $S_\psi(\psi)$ is the score for ψ for a single observation obtained by replacing ψ_0 by ψ in (4.3) and then differentiating the natural logarithm of (4.3) with respect to ψ . Let

$q_{opt}\{A_{m,i}, H_i(\psi), \bar{A}_{m-1,i}, \bar{L}_{m,i}\}$ be the function obtained by replacing $H_i(\psi_0)$ by $H_i(\psi)$ in $q_{opt}\{A_{m,i}, H_i(\psi_0), \bar{A}_{m-1,i}, \bar{L}_{m,i}\}$. Robins (1992) proves that $\text{var}^A[n^{1/2}\{\psi(q_{opt}) - \psi_0\}]$ does not depend on the model (7.1) and attains the semiparametric variance bound for ψ_0 under the sole restriction (4.4). That is, there exists no regular estimator of ψ_0 with asymptotic variance less than of $\tilde{\psi}(q_{opt})$ that is guaranteed to be asymptotically normal and unbiased whatever be the densities

$$f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i}), f\{H_i(\psi_0)\}, f\{L_{m,i} | \bar{L}_{m-1,i}, A_{m-1,i}, H_i(\psi_0)\}.$$

Further, the semiparametric variance bound under the sole restriction (4.4) is the same as that under the joint restriction that both (4.4) hold and (7.1) hold. The estimator $\tilde{\psi}(q_{opt})$ is not feasible since q_{opt} depends on the unknown joint distribution of $\{T_i, \bar{L}_i(T_i), \bar{A}_i(T_i)\}$. Robins (1992) discusses adaptive estimation of q_{opt} .

9. COMPUTATIONAL FORMULAE

In this section, we provide explicit formulae necessary to compute the score test $\Gamma^2(\psi)$ and the estimate $\hat{B}(\hat{I}^{\theta\theta})^{-1}\hat{B}'$ of the asymptotic variance of $\tilde{\psi}$. We assume that $\tilde{\alpha}$ and $\hat{I}_{\alpha\alpha}$ are available from the output of a standard software package used to perform likelihood based inference on model (7.1). Define $A^{[2]} = AA'$. Then straightforward differentiation gives

$$S_{\theta}(\tilde{\alpha}, 0, \psi) = \sum_{i=1}^n \sum_{m=0}^{\text{int}(T_i)} Q_{m,i}\{A_{m,i}, H_i(\psi)\} - \int Q_{m,i}\{a_m, H_i(\psi)\}f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \tilde{\alpha}) d\mu(a_m), \tag{9.1}$$

$$\hat{I}_{\theta\theta}(\psi) = \sum_{i=1}^n \sum_{m=0}^{\text{int}(T_i)} \left\{ \int Q_{m,i}^{[2]}\{a_m, H_i(\psi)\}f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \tilde{\alpha}) d\mu(a_m) - \left[\int Q_{m,i}\{a_m, H_i(\psi)\}f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \tilde{\alpha}) d\mu(a_m) \right]^{[2]} \right\}, \tag{9.2}$$

$$\hat{I}_{\theta\alpha}(\psi) = \sum_{i=1}^n \sum_{m=0}^{\text{int}(T_i)} \int Q'_{m,i}\{a_m, H_i(\psi)\}S_{\alpha,m,i}(a_m, \tilde{\alpha}, \psi)f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \tilde{\alpha}) d\mu(a_m), \tag{9.3}$$

where

$$S_{\alpha,m,i}(A_{m,i}, \tilde{\alpha}, \psi) = \frac{\partial \log \mathcal{L}_{m,i}(\tilde{\alpha}, 0, \psi)}{\partial \alpha},$$

$$nB^{-1} = \sum_{i=1}^n \sum_{m=0}^{\text{int}(T_i)} \frac{\partial Q_{m,i}\{A_{m,i}, H_i(\tilde{\psi})\}}{\partial \psi'} - \int \frac{\partial Q_{m,i}\{a_m, H_i(\tilde{\psi})\}}{\partial \psi'} f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \tilde{\alpha}) d\mu(a_m). \tag{9.4}$$

10. CENSORING

We give a brief sketch of how our approach can be generalized to allow for censoring by end of follow-up when, as in § 5, the treatment $A_{m,i}$ is dichotomous and model (3.1) holds. Generalizations of our approach that allow for nondichotomous treatments,

censoring by competing risks prior to end of follow-up, and missing treatment and covariate data are discussed by Robins (1989b, 1992) and Robins et al. (1992). We assume a common end of follow-up date for all subjects, and let C_i record the known potential censoring time defined as the difference between the end of follow-up date and the i th individual's date of entry into the study. Since C_i is known, we can and do include it in each $\bar{L}_{m,i}$. We observe $\{X_i = \min(T_i, C_i), \bar{L}_i(X_i), \bar{A}_i(X_i), \tau_i = I(T_i < C_i)\}$. In this setting it might be natural to replace $H_i(\psi)$ by

$$X_i^*(\psi) = \int_0^{x_i} \exp\{\psi A_i(t)\} dt.$$

Unfortunately, if $\psi_0 \neq 0$, $X_i^*(\psi_0)$ is not independent of $A_{m,i}$ given $(\bar{L}_{m,i}, \bar{A}_{m-1,i})$ and, thus, an alternative approach is necessary. The key idea is to note that our assumption (4.4) implies that any function of $(H_i(\psi_0), \bar{L}_{m,i}, \bar{A}_{m-1,i})$ will be independent of $A_{m,i}$ given $(\bar{L}_{m,i}, \bar{A}_{m-1,i})$. Thus, we shall define observable random variables $X_{m,i}(\psi)$ and $\Delta_{m,i}(\psi)$ that are functions of $(H_i(\psi), \bar{L}_{m,i}, \bar{A}_{m-1,i})$ to be used as a basis for inference concerning ψ_0 . Specifically $X_{m,i}(\psi) \equiv \min\{H_i(\psi), C_{m,i}(\psi)\}$, where

$$C_{m,i}(\psi) \equiv C_i - m + \int_0^m \exp\{\psi A_i(t)\} dt \quad (\psi \geq 0),$$

$$C_{m,i}(\psi) \equiv \{\exp(\psi)\}(C_i - m) + \int_0^m \exp\{\psi A_i(t)\} dt \quad (\psi < 0),$$

$X_{m,i}(\psi)$ and $\Delta_{m,i}(\psi) \equiv I\{H_i(\psi) < C_{m,i}(\psi)\}$ are observables since $T_i > C_i$ implies $H_i(\psi) > C_{m,i}(\psi)$ and furthermore

$$\{\Delta_{m,i}(\psi_0), X_{m,i}(\psi_0)\} \perp\!\!\!\perp A_{m,i} \mid \bar{L}_{m,i}, \bar{A}_{m-1,i}. \quad (10.1)$$

Define

$$Q_{m,i}\{A_{m,i}, X_{m,i}(\psi), \Delta_{m,i}(\psi)\} = q(A_{m,i}, X_{m,i}(\psi), \Delta_{m,i}(\psi), \bar{A}_{m-1,i}, \bar{L}_{m,i}),$$

where $q(\cdot, \cdot, \cdot, \cdot, \cdot)$ is a known real-valued function. Three simple examples of functions $Q_{m,i}\{A_{m,i}, X_{m,i}(\psi), \Delta_{m,i}(\psi)\}$ are

$$A_{m,i}X_{m,i}(\psi), \quad A_{m,i}\Delta_{m,i}(\psi)X_{m,i}(\psi), \quad A_{m,i}\{1 - \Delta_{m,i}(\psi)\}X_{m,i}(\psi).$$

Expression (10.1) implies that the results of § 7 concerning the estimator $\tilde{\psi}$ remain true with $Q_{m,i}\{A_{m,i}, X_{m,i}(\psi), \Delta_{m,i}(\psi)\}$ replacing $Q_{m,i}\{A_{m,i}, H_i(\psi)\}$ except now, since (Huber, 1981, p. 133; Robins & Tsiatis, 1991), $\Delta_{m,i}(\psi)$ is discontinuous in ψ , B^{-1} equals $\partial E\{S_{\theta,i}(\psi)\}/\partial\psi'$ evaluated at ψ_0 and \hat{B}^{-1} is the matrix of numerical partial derivatives of $n^{-1}S(\tilde{\alpha}, 0, \psi)$ with respect to ψ evaluated at $\tilde{\psi}$ based on a step size of $O(n^{-1/2})$.

11. DISCUSSION

We have developed semiparametric $n^{1/2}$ -consistent G -estimators of the parameter ψ_0 of the strong version of the accelerated failure time model under assumption (4.4) when we have available a correctly-specified richly parameterized model for the densities $f(A_{m,i} \mid \bar{L}_{m,i}, \bar{A}_{m-1,i})$. The fact that the semiparametric variance bound for the parameter ψ_0 is finite under the sole restriction (4.4) implies that, if the distribution of the data is sufficiently smooth, with appropriate choice of bandwidth, we could obviate the need to specify a model for $f(A_{m,i} \mid \bar{L}_{m,i}, \bar{A}_{m-1,i})$ and yet still construct asymptotically normal

distribution-free tests of the hypothesis $\psi_0 = 0$ and asymptotically normal and unbiased estimates of ψ_0 , by substituting a nonparametric multivariate kernel density estimate of $f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ for $f(a_m | \bar{L}_{m,i}, \bar{A}_{m-1,i}; \tilde{\alpha})$ in the estimating function (9.1) (Newey, 1990; Robins, 1992). This implies that, in principle, given a sufficiently large sample size, we can obtain an asymptotically normal distribution-free test of the strict causal null hypothesis (2.3) provided only that we have collected data on a sufficient number of covariates to ensure assumption (2.1) holds. In practice, the actual sample size required for the finite sample distribution of the test statistic to be well-approximated by the standard normal may be unobtainable due to the curse of dimensionality (Huber, 1985) and some model for $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ will be required.

In closing, we note again that the 'strong version' of the accelerated failure time model makes a strong noninteraction assumption. That is, if two subjects i and j have identical observed failure times and observed zidovudine treatment histories then, according to the model, they would have had identical failure times if zidovudine had always been withheld. Our model (6.1) can be generalized to allow the effect of treatment to depend on measured covariates by assuming $U_i = h(T_i, \bar{A}_i(T_i), \bar{L}_i(T_i), \psi_0)$ where $h(\cdot, \cdot, \cdot, \cdot)$ satisfies conditions (i), (ii) and (iii) of § 6. For example, we might specify

$$U_i = \int_0^{T_i} \exp \{ \psi_{0,1} A_i(u) + \psi_{0,2} A_i(u) Z_i(u) \} du,$$

when $Z_i(u)$ is red-blood cell count at u and $\psi_0 = (\psi_{0,1}, \psi_{0,2})'$. Robins (1989b, 1992) and Robins et al. (1992) provide a class of failure time models, the structural nested failure time models, which includes the strong version of the accelerated failure time model as a sub-class, but allows the magnitude of the treatment effect to depend on unmeasured factors. That is, failure time in the absence of treatment, U_i , need not be a deterministic function of the observables $\{T_i, \bar{A}_i(T_i), \bar{L}_i(T_i)\}$.

In closing, we note that the parameter ψ_0 of model (3.1) can be estimated consistently under weaker assumptions than assumption (2.2). For example, suppose $A_{k,i} = (A_{k,i}^{(1)}, A_{k,i}^{(2)})'$ with $A_{k,i}^{(1)}$ recording the prescribed treatment and $A_{k,i}^{(2)}$ the actual treatment taken in $(k, k+1]$. One might then regard

$$A_{k,i}^{(1)} \sqcup U_i | \bar{L}_{k,i}, \bar{A}_{k-1,i}, T_i > k \quad (11.1)$$

as true, but (2.2) as false, if an independent predictor of both U_i and actual treatment had not been included in $\bar{L}_{k,i}$. Under assumption (11.1), semiparametric estimation of ψ_0 can proceed as above except that parametric models for $f(A_{m,i}^{(1)} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$ would replace the models for $f(A_{m,i} | \bar{L}_{m,i}, \bar{A}_{m-1,i})$.

REFERENCES

- BEGUN, J. M., HALL, W. J., HUANG, W. M. & WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432-52.
- COX, D. R. & OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc. B* **41**, 1-31.
- HOLLAND, P. (1986). Statistics and causal inference. *J. Am. Statist. Assoc.* **81**, 945-61.
- HUBER, P. (1981). *Robust Statistics*. New York: Wiley.
- HUBER, P. (1985). Projection pursuit. *Ann. Statist.* **13**, 435-74.
- NEWAY, W. K. (1990). Semiparametric efficiency bounds. *J. Appl. Econ.* **5**, 99-135.
- ROBINS, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—applications to control of the healthy worker survivor effect. *Math. Model.* **7**, 1393-512.

- ROBINS, J. M. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J. Chron. Dis.* **40**, 139S-161S (Suppl. 2).
- ROBINS, J. M. (1989a). The control of confounding by intermediate variables. *Statist. Med.* **8**, 679-701.
- ROBINS, J. M. (1989b). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, Ed L. Sechrest, H. Freeman and A. Mulley, pp. 113-59. Washington, DC: NCHSR, U.S. Public Health Service.
- ROBINS, J. M. (1992). Estimating the causal effect of a time-varying treatment on survival using a new class of failure time models. *Comm. Statist. A*. To appear.
- ROBINS, J. M., BLEVINS, D., RITTER, G. & WULFSOHN, M. (1992). G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia (PCP) on the survival of AIDS patients. *Epidemiology*. To appear.
- ROBINS, J. M., MARK, S. & NEWAY, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders *Biometrics*. To appear.
- ROBINS, J. M. & TSIATIS, A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Comm. Statist. A* **20** (8), 2609-31.
- ROBINS, J. & TSIATIS, A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* **79**, 311-9.
- ROSENBAUM, P. R. (1984). Conditional permutation tests and the propensity score in observational studies. *J. Am. Statist. Assoc.* **79**, 565-74.
- ROSENBAUM, P. R. & RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.
- RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6**, 34-58.

[Received July 1990. Revised December 1991]