Published in final edited form as:

Lifetime Data Anal. 2009 March ; 15(1): 1–23. doi:10.1007/s10985-008-9088-y.

Analysis of quality-of-life adjusted failure time data in the presence of competing, possibly informative, censoring mechanisms

Andrea Rotnitzky, Andrea Bergesio, and Andres Farall

Summary

We derive estimators of the mean of a function of a quality-of-life adjusted failure time, in the presence of competing right censoring mechanisms. Our approach allows for the possibility that some or all of the competing censoring mechanisms are associated with the endpoint, even after adjustment for recorded prognostic factors, with the degree of residual association possibly different for distinct censoring processes. Our methods generalize from a single to many censoring processes and from ignorable to non-ignorable censoring processes.

Keywords

Cause-specific; Dependent censoring; Inverse weighted probability; Sensitivity analysis

1 Introduction

The goal of this paper is to provide methods for estimating the mean of a known function of a quality-of-life adjusted (QOL) failure time accounting for competing, possibly informative, censoring mechanisms. QOL adjusted failure time is defined by first assigning utilities to each of the patient's possible health conditions. A patient's QOL adjusted failure time is then calculated as the accumulation of the patient's utility from start of follow-up until the occurrence of a specific adverse event, such as death or disease relapse. For example, assigning a utility value of zero to time spent under toxicity or symptoms and a value of unity otherwise, gives a QOL adjusted failure time equal to the time spent without symptoms or toxicity -TWIST- (Gelber, Gelman and Goldhirsch, 1989). Using more complex utility functions, one can take into account the severity of toxicity and other patient conditions such as symptoms of distress and functional status.

Zhao and Tsiatis, in a series of papers (1997, 1999, 2000, 2001) resolved a major methodological problem in the analysis of QOL adjusted survival data. Specifically, Gelber et. al. (1989) and later Cox et. al. (1992) noted that censoring due to administrative end of follow-up is generally informative for a QOL adjusted time endpoint even if it is not informative for chronological event time. This is because, as indicated by Cox et. al., in two groups with identical censoring mechanisms, the group with poor QOL is accumulating QOL adjusted time slowly, and hence will tend to be censored earlier on the QOL scale. Zhao and Tsiatis applied the theory in Robins and Rotnitzky, 1992, to adjust for this induced dependent censoring. These authors, however, assumed that censoring and chronological event time are independent. van der Laan and Hubbard, 1999, relaxed this assumption and assumed that censoring may depend on the event time through recorded time dependent and independent patient covariates. However, no paper exists that contemplates the possibility that several competing censoring mechanisms exist, some that remain associated with chronological event time even after adjustment for recorded covariates, and others that do not. Yet, this is not a disparate possibility. In randomized studies administrative end of

follow-up typically induces non-informative censoring with chronological event time. In addition, loss to follow-up due to drop-out induces a competing censoring mechanism that is likely to be informative. Treatment discontinuation might induce yet another informative competing censoring process.

In this paper, we develop estimators of functionals of the QOL adjusted failure time distribution that can accommodate competing censoring mechanisms. Our estimators extend those of Scharfstein and Robins (2002) and Rotnitzky, Farall, Bergesio and Scharfstein (2007) throughout referred to as SR and RFBS. SR proposed an estimator of the survival distribution of chronological event time suitable for conducting sensitivity analysis under the assumption that some, but not all, joint prognostic factors for censoring and chronological event time are available. Their approach is to repeat inference under different values of a non-identifiable censoring bias parameter encoding the magnitude of the residual association between survival and censoring after adjusting for measured prognostic factors. RFBS extended this estimator to allow for the presence of competing censoring mechanisms. In this paper, we further extend the SR and RFBS methodology to estimation of the mean of a function of a QOL adjusted failure time endpoint. Our methods are suitable for conducting sensitivity analysis to different degrees of residual association with some, possibly all, competing censoring mechanisms after adjustment for measured prognostic factors.

2 A motivating example: the ACTG 175 study

The AIDS Clinical Trial Group ACTG 175 study (Hammer et. al., 1996) was a randomized, double-blind clinical trial comparing two nucleoside therapies, zidovudine (AZT) or didanosine (ddI), and two combination therapies, zidovudine/didanosine (AZT+ddI) or zidovudine/zalcitabine (AZT+ddC), in HIV-1 infected subjects with CD4 at baseline between 200 and 500 per cubic milliliter. Study enrolment was from December 1991 to October 1992, and follow-up ended in February 1995. Consequently, all subjects had a potential follow-up time of no less than 27 months and no more than 38 months. Baseline patient information included gender, age, ethnicity, CD4 count, Karnofsky score, prior use of antiretroviral therapy, whether or not the patient was asymptomatic, and whether or not the patient was homosexual, hemophiliac or was a drug user. The study design called for CD4 counts to be measured at baseline, week 8 and then every 12 weeks. In addition, whenever the patient visited the doctor's office because of adverse events (signs and symptoms) these events were recorded and scored according to the ACTG grading scheme (NIAID, 1996). The primary endpoint was time to a 50% decline in CD4 count from baseline, as confirmed by a second CD4 count within 3 to 21 days, progression to AIDS, or death. This study motivated the methods of SR and RFBS.

As reported in Hammer et al. (1996) a large fraction, roughly 50%, of the study participants prematurely discontinued treatment, but few, roughly 7% discontinued as mandated by the protocol for adverse events. Reasons for the majority of treatment discontinuations included low-grade toxic reactions, declining CD4 cell counts before the study end point was reached, the desire to seek other therapies and the demands of study participation. It is therefore of interest to investigate the arm-specific disease progression as if premature treatment discontinuation could have been prevented. To that end, in § 7 we will estimate, in the AZT arm, the mean of a quality-adjusted failure time endpoint which approximately quantifies the time spent without symptoms from the start of follow-up to the primary time endpoint which represents the expected time spent without symptoms that would be observed under full compliance. An exact definition of the QOL failure time endpoint is given in § 3.1.

To estimate the mean QOL failure time of interest in the AZT arm we could use the methods of van der Laan and Hubbard or of Zhao and Tsiatis which regard each subject as censored

at the time of treatment discontinuation. However, these methods make either the strong assumption that the variables recorded for data analysis are all the predictors of prognosis (i.e., of future health outcomes and failure time) that are used in the decision to discontinue treatment (van der Laan and Hubbard, 1999), or the even stronger assumption that the decision of treatment discontinuation is unrelated to the recorded time-dependent variables (Zhao and Tsiatis, 1997). However, an examination of the reasons listed above for treatment discontinuation suggests the possibility of residual dependence between treatment discontinuation and QOL failure time even after adjusting for all recorded variables. For instance, it is quite possible that a doctor's decision to remove a patient from protocol is based on clinical indicators of health deterioration that are not recorded in the analysis database. When residual dependence between treatment discontinuation and outcome exists, the aforementioned methods may fail to provide consistent estimators of the mean QOL failure time. As an example, if within levels of recorded covariates, a larger fraction of sicker subjects than healthier subjects discontinue treatment, then both the methods of van der Laan and Hubbard and those of Zhao and Tsiatis will overestimate the mean survival time in the AZT arm, as well as the mean TWIST.

We could try extending the SR-estimators of (chronological) failure time distributions to provide estimators of mean QOL failure time in the presence of informative censoring. However, this extension would not entirely meet the analytical needs in the ACTG 175 trial for the following reason. As indicated above, the ACTG trial had staggered entry during the period December 1991 to October 1992. Consequently, even though the maximum potential follow-up time was 38 months, several subjects were administratively censored before month 38. Thus, two competing censoring mechanisms were in operation in the ACTG trial, one due to treatment discontinuation and one due to reaching the administrative end of study. In the SR methodology the investigator specifies a range for the parameter encoding the residual dependence of the hazard of the minimum of competing censoring times on the censored outcome. However, this range might be rather difficult to specify if the reasons for each censoring type are quite different. The difficulty is magnified if some censoring processes are informative and some are not. To ameliorate this problem in their analysis of the ACTG trial, SR restricted the analytic time horizon from 38 months to 24 months because over this time period no subject was administratively censored. However, in doing so they discarded valuable information on the survival distribution from month 24 to 38. The magnitude of this information loss is seen in the following summary statistics. In the AZT arm, 609 subjects had complete baseline covariate information. Of these, 136 had reached the endpoint by month 38 and 104 had reached it by month 24, 65 had not reached the endpoint by month 38 and 253 had not reached it by month 24, 316 were censored due to drop-out by month 38 and 252 by month 24, and 92 were administratively censored prior to month 38 while no one was administratively censored prior to month 24. Thus SR discarded follow-up data collected after month 24 on the 253 subjects who remained at risk at that month. Note that valuable survival information for the interval [24, 28] was available from these 253 subjects because only 64 of them dropped out. The methods in this paper allow the incorporation of this valuable information to estimate the mean quality adjusted failure time over the larger time horizon of thirty eight months.

3 Study design, data and parameter of interest

We consider a follow-up study with staggered entry and maximum follow-up time τ^* . The study design calls for a vector of health variables $V(t)$ to be measured on each of n independent subjects at either random or predetermined times t ($t < \tau^*$) after entry to the study. Unless censored, an event time T^* is recorded in each subject. We assume that the exact date of the occurrence of the event is recorded on uncensored subjects, so that T^* is a non-negative continuous random variable. We assume that censoring can occur for one of

several reasons and that the cause of censoring is known. We also assume that two different causes cannot give rise simultaneously to the censoring event. This assumption is reasonable when, as in the ACTG 175 study, the censoring time can be regarded as the minimum of competing continuous censoring times. We let C^* denote the censoring time which is assumed to be a non-negative continuous random variable. Both T^* and C^* record time since the patient's entry to the study.

For technical reasons we will need to disregard data recorded after time $\tau = \tau^* - \epsilon$, where ϵ is a small positive number. We therefore define the event time $T = \min(T^*, \tau)$. This event time is potentially censored by $C = \min(C^*, \tau)$. We also define $V^H(t) \equiv \{V(u) : 0 \leq u \leq t\}$ where $V(0)$ is a vector of baseline covariates and for $u > 0$, $V(u)$ is defined as the vector of health variables recorded at time u if u is a measurement time; otherwise, it is defined as the vector of variables recorded at the last measurement time prior to u .

In the presence of censoring and with cause of censoring known, the observed data are comprised by n i.i.d. copies $O_1, \dots, O_n$ of $O = (\Delta, X, J, V^H(X))$ where $\Delta = I(T \leq C)$, $X = \min(T, C)$, $J = 0$ if $\Delta = 1$ and, otherwise, $J \in \{1, 2, \dots, j^*\}$ denotes the cause of censoring.

A quality-of-life adjusted failure time is a random variable $D = D(T)$ where for each $t \leq \tau$,

$$D(t) = \int_0^t s(u, V^H(u)) du \quad (1)$$

and where $s(u, V^H(u))$ is a user-specified utility function for a subject with health history $V^H(u)$ at time u . The function $D(t)$ measures the utility accumulated up to time t by a subject with covariate history $V^H(t)$. Note that D is equal to chronological event time T if we choose $s(u, V^H(u)) = 1$ for all u .

We seek to estimate

$$\mu^* = E\{g(D)\}$$

where $g(\cdot)$ is a specified function, such as $g(d) = d$ (yielding μ^* equal to the mean of D) or $g(d) = I(d \leq u)$ (yielding μ^* equal to the cumulative distribution function for D evaluated at u). Throughout we will assume that for all $u \in (0, \tau)$ and some constant K ,

$$\lambda_{c,j}(u|V^H(u), D, T > u) < K \text{ with probability 1.} \quad (2)$$

where $\lambda_{c,j}(u|\bullet, T > u) \equiv \lim_{h \rightarrow 0} \Pr(u < C \leq u + h, J = j | C \geq u, T > u, \bullet) / h$ is the conditional cause-specific hazard of censoring due to cause j given $\bullet$.

3.1. Definition of the QOL adjusted endpoint in our analysis of the ACTG 175 data

In our analysis of the ACTG 175 data we were interested in estimating the mean of TWIST, the time spent without symptoms, in the AZT arm. To that end, we considered the QOL failure time endpoint D defined as (1) where $s(t, V^H(t))$ was a dichotomous function taking the value 0 if at the visit to the doctor's office taking place at time t or, if no visit took place at t , at the closest time subsequent to t , the subject reported having had an adverse sign or symptom at time t (including, among others, fatigue, shortness of breath, pain swallowing and sore mouth), and the value 1 otherwise. The analytic time horizon was set to slightly less than 38 months, namely to 1120 days or equivalently 3.068 years, thus T was equal to $\min(T^*, \tau)$ with $\tau = 3.068$ years and T^* is time since entry to the study to AIDS, death or a 50% decline in CD4 count from baseline, as confirmed by a second CD4 count within 3 to

21 days. With this definition D was equal to the time spent without symptoms from start of follow-up until the minimum of the event time and 3.068 years and the parameter μ^* of interest was $E(D)$.

4 Sensitivity analysis

With observed data O , we can at most hope to estimate the observed data distribution P_O and functionals of it. However, it is intuitively clear, and indeed it can be proved formally, that the parameter of interest μ^* is not uniquely determined by P_O . To illustrate why this is the case, consider the ACTG 175 study. The mean of TWIST in the entire study population is the weighted average of the means of TWIST in i) the uncensored subjects, ii) the subjects censored administratively and iii) the subjects censored due to treatment discontinuation. With the observed data O we can hope to estimate i) and ii) (assuming, as we will do, that administrative censoring is non-informative). However, the observed data has no information about iii). To estimate this quantity we will need to make assumptions about the nature of the dependence of censoring with TWIST.

To that end, we will consider, in general, models that assume that the cause-specific hazard of censoring satisfies

$$\lambda_{c,j}(u|V^H(u), D, T > u) = \lambda_{0,j}(u, V^H(u)) \exp\{q_j(u, V^H(u), D)\} \quad (3)$$

for fixed, i.e. user specified, functions $q_j, j = 1, \dots, j^*$.

The function $q_j(u, V^H(u), D)$ encodes the residual dependence on the hazard ratio scale between the QOL adjusted failure time D and censoring at u due to cause j , after adjustment for measured prognostic factors $V^H(u)$. Following RFBS we call the functions q_j *cause-specific censoring bias functions*.

Model (3) can incorporate both informative and non-informative competing censoring mechanisms about QOL adjusted failure time, since a non-informative censoring mechanism j is modelled by setting $q_j \equiv 0$. For example, in the ACTG 175 trial consider the model that assumes two competing censoring times, i.e. $j^* = 2$, where the first one, $j = 1$, is time to administrative end of study, and the second one is minimum of time to treatment discontinuation and loss to follow-up, which for short, we refer to as drop-out time. To encode the assumption of independent administrative censoring, we set $q_1 = 0$. In contrast, to encode the possibility of dependent drop-out time even after adjustment for past CD4 count and baseline covariates, we set

$$q_2(u, V^H(u), D) = \alpha (D - D(u^-)) \quad (4)$$

where α is a scalar parameter. This model assumes that among subjects at risk and with the same recorded past at any given time u , the log-hazard of censoring due to drop out is proportional to the time $D - D(u^-)$ that they would spend without symptoms if, possibly contrary to fact, they were to remain in the study until the end of the analytic time horizon of 3.068 years. Negative values of α imply that subjects with shorter remaining times without symptoms would be more likely to drop-out than subjects with longer remaining times.

Arguing as in Scharfstein, Rotnitzky and Robins (1999) it is possible to show that for each choice of selection bias functions $q \equiv \{q_1, \dots, q_{j^*}\}$, model (3) will perfectly fit the observed data and will not be rejected by any statistical test. That is, the assumption that (3) holds for a given set q is not empirically verifiable. Since each choice q gives rise to a different distribution of D among the censored subjects, and since the observed data law determines the distribution of D in the uncensored units, then each q coupled with the observed data

distribution P_O gives rise to a different value of μ^* . Consequently, since the observed data cannot identify q , it cannot either identify μ^* . However, the parameter μ^* is identified by the observed data if we assume, as we do in model (3), that the functions q are fixed and known. Thus, the best we can hope for is to conduct a sensitivity analysis to various degrees of dependent censoring whereby we draw inferences about μ repeatedly under a range of censoring bias functions believed to be plausible. Substantive considerations may help restrict this range. For example, if as in our analysis of the ACTG 175 trial, cause $j=1$ stands for censoring due to administrative end of follow-up, and if it can be reasonably assumed that this gives rise to a non-informative censoring mechanism, we may restrict q_1 to be the null function. To help further restrict the range of the remaining selection bias functions $q_j, j=2, \dots, j^*$, we suggest modeling these functions parametrically, as in (4), and drawing repeated inferences, each time fixing the value of α at a specified value. Like model (4), the chosen model should be easily interpretable so that a plausible parameter range can be specified by subject matter experts, and be such that values of the parameters equal to zero correspond to the benchmark assumption of no residual association between the QOL adjusted endpoint and censoring.

Unfortunately, even though μ^* is identified under model (3), its estimation is not feasible under the model when, as will generally be the case when computing relevant QOL adjusted endpoints, $V^H(u)$ can take a large number of values. The reason for this limitation is that μ^* depends on the unknown functions $\Lambda_{0,j}(u, V^H(u)) \equiv \int_0^u \lambda_{0,j}(s, V^H(s)) ds, j=1, \dots, j^*$, of the high dimensional vector $V^H(u)$ and these functions cannot be well estimated using non-parametric smoothing techniques with the sample sizes found in practice due to the curse of dimensionality. One practical solution is to conduct inference about μ assuming that $\lambda_{0,j}(u, V^H(u))$ follows a lower dimension, yet still flexible, semiparametric model

$$\lambda_{0,j}(u, V^H(u)) = \lambda_{0,j}(u) \exp(\gamma_j^* w_j(u, V^H(u))), j=1, \dots, j^*. \quad (5)$$

where $\lambda_{0,j}(\cdot)$ is an unknown function, $w_j(u, V^H(u))$ is a specified vector function of $V^H(u)$ and γ_j^* is an unknown conformable parameter vector. Note that in a slight abuse of notation we use $\lambda_{0,j}$ to denote both the function of $(u, V^H(u))$ on the left hand side of (5) and the baseline hazard $\lambda_{0,j}(u)$ on the right hand side. This should not cause trouble throughout as the meaning of $\lambda_{0,j}$ should be obvious from the context. Of course, misspecification of the model for $\lambda_{0,j}(u, V^H(u))$ may yield inconsistent estimators of μ^* . It is precisely because of concerns about sensitivity to model misspecification that we propose using semiparametric model (5) rather than a fully parametric model for $\lambda_{0,j}(u, V^H(u))$, as the former is more flexible than the latter. Although ideally we would like not to make any assumptions on $\lambda_{0,j}(u, V^H(u))$, as indicated above in practice, this is not possible due to the curse of dimensionality.

5 Estimation

A key idea underlying the estimators of μ that we will develop in this section is that of implicit "non-uniform-pseudo-redistribution" to the right via inverse probability weighting of the uncensored subjects. The idea essentially generalizes a similar idea underlying the Kaplan-Meier estimator of a survivor function under independent censoring and is best understood by examining this simple case first. Without loss of generality suppose that the censored units are units $i=1, \dots, M$ and suppose that their censoring times satisfy $C_1 < C_2 < \dots, C_M$. Let n_m be the number of subjects censored at time C_m and let b_m be the number of subjects in the risk set at time $C_m, m=1, \dots, M$. The Kaplan-Meier estimator of $\mu^* = P(T > t^*)$ can be expressed as

$$n^{-1} \sum_{i=1}^n \frac{\Delta_i^{(t^*)}}{\hat{\pi}_i} I(T_i > t^*) \quad (6)$$

where $\Delta_i^{(t^*)} = 1$ if $C_i > T_i^{(t^*)}$ and $\Delta_i^{(t^*)} = 0$ otherwise, and

$\hat{\pi}_i = \prod_{m: C_m < T_i^{(t^*)}} \left(\frac{b_m - n_m}{b_m} \right)$ with $T_i^{(t^*)} = \min(T_i, t^*)$. The quantity $\hat{\pi}_i$ is the product of the estimated hazards $\left(\frac{b_m - n_m}{b_m} \right)$ of remaining uncensored at each of the observed censoring times C_m from time 0 to $T_i^{(t^*)}$ and as such it is itself a Kaplan-Meier estimator. Specifically, letting $\Lambda_c(t^*) = \int_0^{t^*} \lambda_c(t) dt$ denote the cumulative hazard of censoring at t^* and $\pi(t^*) = \prod_{0 \leq t \leq t^*} d\Lambda_c(t)$ denote the survivor function of C at t^* (where throughout $0 \leq t \leq t^*$ stands for the product integral from 0 to t^*), the quantity $\hat{\pi}_i$ is precisely the Kaplan-Meier of the survivor function $\pi(\cdot)$ evaluated at $T_i^{(t^*)}$.

We therefore see that the Kaplan-Meier estimator (6) at t^* is but a weighted average of the indicators of surviving past t^* among subjects uncensored in the interval $(0, t^*)$. In this weighted average, each subject i uncensored in the interval $(0, t^*)$ contributes to the estimation of μ^* and his contribution is weighted by the inverse of the estimated probability that a random person arrives uncensored and alive until subject's i observed failure time T_i or time t^* , whichever comes first. This weighting scheme is tantamount to uniformly redistributing to the right each censored subject at his time of censoring. Specifically, at time C_1 the n_1 censored subjects are redistributed among the $b_1 - n_1$ uncensored subjects still at risk at time C_1^- . Thus, of the subjects in the risk set at time C_1^- , the $b_1 - n_1$ subjects that were

not censored at time C_1 are weighted by $1/\{(b_1 - n_1)/b_1\} = 1 + \frac{n_1}{b_1 - n_1}$ so as to represent themselves (1) and their share $(n_1/(b_1 - n_1))$ of the n_1 censored subjects. At time C_2 , each of the n_2 censored subjects is redistributed to the right and represented by the remaining $b_2 - n_2$ in the risk set at time C_2^- that remain uncensored at time C_2 . Now, the weight of the $b_2 - n_2$ remaining subjects is updated from $1/\{(b_1 - n_1)/b_1\}$ to

$\left[1/\{(b_1 - n_1)/b_1\} \times \left\{ 1 + \frac{n_2}{b_2 - n_2} \right\} \right] = \left[1 / \prod_{m=1}^2 \{(b_m - n_m)/b_m\} \right]$ so as to account for their own weights and the weights that the b_2 censored subjects were already carrying at time t_2^- . Subjects with observed failure times in the interval (C_1, C_2) have weights just equal to $1/\{(b_1 - n_1)/b_1\}$ because their failure times cannot stand as surrogates for the failure times of subjects in risks sets at times C_2^- or later, since the latter are known to have survived at least until time C_2^- and the former are known to have failed prior to C_2^- . The computation continues in the same fashion to yield the expression (6).

We could have arrived at expression (6) after noticing that the parameter $\mu^* = E\{I(T > t^*)\}$

of interest can be expressed as $E\left\{ \frac{\Delta^{(t^*)}}{\pi(T^{(t^*)})} I(T > t^*) \right\}$, or equivalently that μ^* solves the

population equation $E\left[\frac{\Delta^{(t^*)}}{\pi(T^{(t^*)})} \{I(T > t^*) - \mu\} \right] = 0$. Expression (6) is the plug-in estimator of μ^* where the expectation is replaced by the empirical mean and $\pi(T^{(t^*)})$ is replaced by its non-parametric estimator.

Turn now to our problem. To estimate $\mu^* = E\{g(D)\}$ under our model (5) for dependent censoring we will use a similar inverse weighted probability representation for μ^* and proceed by computing the empirical weighted mean of the values of $g(D_i)$ among the uncensored subjects. Specifically, arguing as in Rotnitzky et. al. (2007) it can be shown that under (3) and (5), μ^* solves

$$E \left[\frac{\Delta}{\pi(T|V^H(T), D; \gamma, \Lambda_0)} \{g(D) - \mu\} \right] = 0 \quad (7)$$

when $\gamma = \gamma^*$ and where for any $\gamma_j, j = 1, \dots, j^*$, we define

$$\pi(t|V^H(t), D; \gamma, \Lambda_0) = \prod_{j=1}^{j^*} \prod_{0 \leq u \leq t} \{1 - r_j(u, V^H(u), D; \gamma_j) d\Lambda_{0,j}(u)\} \quad (8)$$

with

$$r_j(u, V^H(u), D; \gamma_j) \equiv \exp\{\gamma_j' w_j(u, V^H(u)) + q_j(u, V^H(u), D)\}$$

and $\Lambda_{0,j}(u) = \int_0^u \lambda_{0,j}(t) dt$. Under (5) and a slight strengthening of model (3), $\pi(t|V^H(t), D; \gamma^*, \Lambda_0)$ is the conditional probability of remaining uncensored at time t given the covariate history $V^H(t)$ and the QOL failure time D .

In analogy to the IPW representation of the Kaplan-Meier estimator, we propose estimating μ^* with the solution $\hat{\mu}$ of

$$\sum_{i=1}^n \frac{\Delta_i}{\pi(T_i|V_i^H(T_i), D_i; \hat{\gamma}, \hat{\Lambda}_0(\hat{\gamma}))} \{g(D_i) - \mu\} = 0 \quad (9)$$

where $\hat{\gamma}$ stands for a suitable estimator of $\gamma^* \equiv \{\gamma_1^*, \dots, \gamma_{j^*}^*\}$ and $\hat{\Lambda}_0(\hat{\gamma})$ for a suitable estimator of $\Lambda_0 \equiv \{\Lambda_{0,1}, \dots, \Lambda_{0,j^*}\}$ which depends on $\hat{\gamma}$. Computation of adequate estimators $\hat{\gamma}$ and $\hat{\Lambda}_0(\hat{\gamma})$ requires some thought. Specifically, even though the model defined by (3) and (5) is a relative risk model with time dependent covariates $w_j(u, V^H(u))$ and offset $q_j(u, V^H(u), D)$, we cannot estimate γ using an analogous of Cox's partial likelihood estimator. Doing so would require, in particular, evaluation of the offset $q_j(C_m, V_i^H(C_m), D_i)$ for subjects i in the risk set at each censoring time C_m but censored prior to their event time. However, this evaluation is unfeasible because D_i is unobserved for such subjects. To compute estimators of γ^* and Λ_0 we reason as follows.

For $j = 1, \dots, j^*$, define the j^{th} cause-specific counting process, $N_{C,j}(u) \equiv I(C \leq u, J = j)$ and define

$$dM_{*,C,j}(u; \Lambda_0, \gamma) \equiv dN_{C,j}(u) - d\Lambda_{0,j}(u) r_j(u, V^H(u), D; \gamma_j)$$

Under model (3) and (5), $E\{dM_{*,C,j}(u, \Lambda_0, \gamma^*) | T > u, X = u, V^H(u), D\} = 0$. Thus, the random quantity

$$m_*(O; \gamma, \Lambda_0, a) = \sum_{j=1}^{j^*} \int dM_{*,C,j}(u; \Lambda_0, \gamma) a_j(u, V^H(u)) I(T > u, X \geq u)$$

defined for any given collection of functions $a = \{a_j(\bullet, V^H(\bullet)) : j = 1, \dots, j^*\}$, has mean zero when γ is replaced by its true value γ^* . This suggests that if a consistent estimator $\hat{\Lambda}_{0,j}$ of $\Lambda_{0,j}$ were available we could contemplate estimating γ^* with the solution of an

estimating equation of the form $\sum_{i=1}^n m_*(O_i; \gamma, \hat{\Lambda}_0, a) = 0$. However, such procedure would also be unfeasible because, once again, to carry it out we would need to be able to compute, in particular, the values of $q_j(C_m, V_i^H(C_m), D_i)$ for subjects i in the risk set at C_m but censored prior to their event time. Fortunately, another inverse probability weighted trick resolves the problem.

Specifically, let

$$v(u, V^H(T), D; \gamma, \Lambda_0) = \frac{\pi(u^- | V^H(u^-), D; \gamma, \Lambda_0)}{\pi(T | V^H(T), D; \gamma, \Lambda_0)}$$

It can be verified that under (3) and (5), and assuming the process $V^H(u)$ cannot jump at u equal to the censoring time C ,

$$E\{\Delta v(u, V^H(T), D; \gamma^*, \Lambda_0) | T > u, X \geq u, V^H(u), D\} = 1 \text{ for all } u \in (0, T] \quad (10)$$

The formal proof of this result under the model defined by restrictions (3) and (5) follows as in Robins (1993). Under a slight strengthening of (3), $v(u, V^H(T), D; \gamma^*, \Lambda_0)$ is precisely the inverse of the conditional probability that $\Delta = 1$ given $V^H(u), D, T > u$ and $X \geq u$ and thus the result follows immediately.

One important implication of (10) is that

$$dM_{C,j}(u; \Lambda_0, \gamma) = dN_{C,j}(u) - \Delta v(u, V^H(T), D; \gamma, \Lambda_0) r_j(u, V^H(u), D; \gamma) d\Lambda_{0,j}(u)$$

is a computable, data dependent, function of Λ_0 and γ which satisfies

$$E\{dM_{C,j}(u; \Lambda_0, \gamma^*) | T > u, X \geq u, V^H(u), D\} = E\{dM_{*,C,j}(u; \Lambda_0, \gamma^*) | T > u, X \geq u, V^H(u), D\} = 0$$

Consequently, for any user-specified choice of functions $a_j(\cdot, \cdot)$, the data dependent function of γ, Λ_0

$$m(O; \gamma, \Lambda_0, a) = \sum_{j=1}^{j^*} \int dM_{C,j}(u; \Lambda_0, \gamma) a_j(u, V^H(u)) I(T > u, X \geq u)$$

has mean zero when γ is equal to γ^* and Λ_0 is the true cumulative baseline hazard of censoring. Thus, $m(O; \gamma, \Lambda_0, a)$ could, in principle, serve as an unbiased estimating function for γ so long as we could replace $\Lambda_{0,j}$ for each $j = 1, \dots, j^*$, with a suitable consistent estimator of it.

Unfortunately, such suitable estimator of $\Lambda_{0,j}$ can't be computed using Breslow's estimator of the cause-specific baseline hazard $\lambda_{0,j}$ for the following reason. If one were to apply Breslow's formula, one would estimate $\lambda_{0,j}(u)$ with 0 for $u < C_m$, $m = 1, \dots, M$ and one would estimate $\lambda_{0,j}(C_m)$, $m = 1, \dots, M$, with $\tilde{\lambda}_{0,j}(C_m; \hat{\gamma}_j)$ where

$$\tilde{\lambda}_{0,j}(C_m; \gamma) \equiv n_m^{(j)} \left\{ \sum_{i=1}^n I(X_i \geq C_m) r_j(C_m, V_i^H(C_m), D_i; \gamma_i) \right\}^{-1}, j=1, \dots, j^* \quad (11)$$

and $n_m^{(j)}$ is the number of subjects censored at time C_m due to cause j . However, $\tilde{\lambda}_{0,j}(C_m; \gamma)$ is not feasible because i) $\lambda_{0,j}(C_m; \hat{\gamma})$ depends on $\hat{\gamma}_j$ which is not available since, as we have seen, its computation requires solving an estimating function depending precisely on the unknown Λ_0 and, ii) $r_j(C_m, V_i^H(C_m), D_i; \gamma_j)$ depends on $q_j(C_m, V_i^H(C_m), D_i)$ which is not available for subjects i at risk at C_m but censored prior to their event time.

Difficulty i) can be bypassed if we can find, for each γ , a data dependent function $\hat{\Lambda}_{0,j}(\cdot; \gamma)$ such that $\Lambda_{0,j}(\cdot; \gamma^*)$ converges to $\Lambda_{0,j}(\cdot)$ at rate $\sqrt{n}$ because then, the estimating equation

$$\sum_{i=1}^n m(O_i; \gamma, \hat{\Lambda}_0(\gamma), a) = 0 \quad (12)$$

should have a solution $\hat{\gamma}$ consistent and asymptotically normal for γ^* . To find such $\hat{\Lambda}_{0,j}(\cdot; \gamma)$ we still need to resolve difficulty ii) as we do not have available the estimator (11) for any γ . We can indeed resolve difficulty ii) and produce the desired $\hat{\Lambda}_{0,j}(\cdot; \gamma)$ using an inverse probability weighted modification of the unfeasible quantity $\tilde{\lambda}_{0,j}(C_m; \gamma)$. Specifically, the identity (10) suggests that we can replace (11) with

$$\hat{\lambda}_{0,j}(C_m; \gamma) \equiv \frac{n_m^{(j)}}{\sum_{i=1}^n \Delta_i I(T_i > C_m) v(C_m, V_i^H(T_i), D_i; \gamma, \hat{\Lambda}_0(\gamma)) r_j(C_m, V_i^H(C_m), D_i; \gamma_j)}, j=1, \dots, j^* \quad (13)$$

However, the preceding display does not provide an explicit formula for $\hat{\lambda}_{0,j}(C_m; \gamma)$, because the denominator in the right hand side of (13) also depends on the unknowns $\hat{\lambda}_{0,j}(C_l; \gamma)$ for $l = M, M-1, \dots, m$ and $j = 1, \dots, j^*$. To compute $\hat{\lambda}_{0,j}(C_m; \gamma)$ we must solve (13) recursively, from the largest to the smallest censoring times. Specifically, as in Breslow's estimator, for $u < C_m$, $m = 1, \dots, M$, we set $\hat{\lambda}_{0,j}(u; \gamma) = 0$. This implies that

$v(C_m, V_i^H(T_i), D_i; \gamma, \hat{\Lambda}_0(\gamma))$ simplifies to $\prod_{k=1}^{j^*} \{1 - \hat{\lambda}_{0,k}(C_m; \gamma) r_k(C_m, V_i^H(C_m), D_i; \gamma_k)\}$ and for $m = 1, \dots, M-1$,

$$v(C_m, V_i^H(T_i), D_i; \gamma, \hat{\Lambda}_0(\gamma)) = \prod_{k=1}^{j^*} \prod_{C_l | C_m \leq C_l \leq T_i} \{1 - \hat{\lambda}_{0,k}(C_l; \gamma) r_k(C_l, V_i^H(C_l), D_i; \gamma_k)\}.$$

Then, when $m = M$, (13) is the same as

$$\hat{\lambda}_{0,j}(C_M; \gamma) = n_M^{(j)} \left\{ \sum_{i=1}^n \Delta_i \frac{I(T_i > C_M) r_j(C_M, V_i^H(C_M), D_i; \gamma_j)}{\prod_{k=1}^{j^*} \{1 - \hat{\lambda}_{0,k}(C_M; \gamma) r_k(C_M, V_i^H(C_M), D_i; \gamma_k)\}} \right\}^{-1}, j=1, \dots, j^* \quad (14)$$

and when $m < M$, (13) is the same as

$$\hat{\lambda}_{0,j}(C_m; \gamma) = n_m^{(j)} \left\{ \sum_{i=1}^n \Delta_i \frac{I(T_i > C_m) r_j(C_m, V_i^H(C_m), D_i; \gamma_j)}{\prod_{k=1}^{j^*} \prod_{C_l | C_m \leq C_l \leq T_i} \{1 - \hat{\lambda}_{0,k}(C_l; \gamma) r_k(C_l, V_i^H(C_l), D_i; \gamma_k)\}} \right\}^{-1}, j=1, \dots, j^* \quad (15)$$

We therefore compute $\hat{\lambda}_{0,j}(C_M; \gamma)$ for $j = 1, \dots, j^*$ as the solutions to the system (14) of j^* equations. Next, having computed $\hat{\lambda}_{0,j}(C_l; \gamma)$ for $l = m+1, \dots, M$ and $j = 1, \dots, j^*$, we compute $\hat{\lambda}_{0,j}(C_m; \gamma)$ for $j = 1, \dots, j^*$ as the solutions to the system of j^* equations (15).

In conclusion, to estimate μ^* we carry out the following three steps. First, we compute for each fixed γ , the estimators $\hat{\Lambda}_{0,j}(u, \gamma) = n_m^{-1} \sum_{C_m \leq u} \hat{\lambda}_{0,j}(C_m; \gamma)$ where $\lambda_{0,j}(C_m; \gamma)$ are calculated with the preceding recursive algorithm. Next, we estimate $\gamma = (\gamma_1, \dots, \gamma_{j^*})$ with $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_{j^*})$ solving (12). Finally, we estimate μ^* with $\hat{\mu}$ solving (9). However, in choosing the functions $a_j(\cdot, \cdot)$ that go in equations (12) we are not entirely free. In the Appendix we show that if $a_j(u, V^H(u))$ are functions $a_j(u)$ that depend on u only for $j = 1, \dots, j^*$, then (12) holds for all γ , i.e. it has infinite, uncountable, solutions. Consequently, the equations (12) must use functions $a_j(\cdot, \cdot)$ that depend on $V^H(u)$. One natural choice is $a_j(u, V^H(u)) = (\mathbf{0}^T, w_j(u, V^H(u))^T, \mathbf{0}^T)^T$ where the first and last $\mathbf{0}$ are zero vectors with

$\sum_{l=1}^{j-1} \dim(\gamma_l)$ and $\sum_{l=j+1}^{j^*} \dim(\gamma_l)$ rows respectively and $w_j(u, V^H(u))$ are the functions of the time dependent covariates that enter in model (5). This is the choice used in the analysis of the ACTG data presented in § 7 and in the simulation study in § 6.

In the Appendix we show that indeed, equations (14), (15), (12) and (9) are the empirical versions of population equations solved by the true values of the parameters $\Lambda_0(\cdot)$, μ and $\gamma = (\gamma_1, \dots, \gamma_{j^*})$. Under mild regularity conditions,

$\sqrt{n} \{ \hat{\Lambda}_0(\cdot; \hat{\gamma}) - \Lambda_0(\cdot) \}$ and $\sqrt{n} \{ (\hat{\mu}, \hat{\gamma}) - (\mu^*, \gamma^*) \}$ should converge to a Gaussian process and to a Normal distribution respectively. A proof of this result is beyond the scope of this paper.

Inverse probability weighting of uncensored subjects takes place in the computation of all our estimators, i.e. $\hat{\Lambda}_0(\cdot; \gamma)$, $\hat{\gamma}$ and $\hat{\mu}$. This weighting is tantamount to redistributing to the right the censored subjects, just as in the computation of the Kaplan-Meier estimator. However, two features distinguish the redistribution scheme of our estimators from that of the Kaplan-Meier estimators. First, in the K-M estimator, each censored subject is redistributed at the time of censoring among *all* subjects still uncensored and alive at that time. In contrast, in the computation of our estimators, a subject censored, say at time C , is approximately redistributed to just the subjects with the same recorded past $V^H(C)$ as him and still uncensored and alive at time C . Exact redistribution to just uncensored subjects with the same recorded past would occur if the conditional hazards of censoring $\lambda_{C,j}(u | V^H(u), D, T > u)$, $j = 1, \dots, j^*$ were estimated under a model that does not impose restrictions on the baseline hazards $\lambda_{0,j}(u | V^H(u))$ rather than under model (5) as we are forced to do

due to the curse of dimensionality. Second, in the K-M estimator, the redistribution to the right is *uniform* across all uncensored subjects, i.e. each uncensored subject represents the same "fraction" of a censored subject. In contrast, in the computation of our estimators, the redistribution (among subjects with the same recorded past) is non-uniform, since the weight carried by each uncensored subject depends on his value of the QOL failure time D through the selection bias functions $q_j(u, V^H(u), D)$. Exact uniform redistribution to the right among subjects with the same recorded past when the covariate history is low dimensional was applied in Malani (1995) to the estimation of a survival curve with informative, but ignorable, censoring. To illustrate our points, consider for example, the estimation of the baseline hazards $\lambda_{0,j}(\cdot; \gamma)$, $j = 1, \dots, j^*$. Consider first the estimation of $\lambda_{0,j}(C_M, \gamma)$. The only subjects that have T_i missing among those still at risk at C_M are those censored at C_M . These censored subjects, who would contribute to the denominator of (11), are represented in the denominator of (14) by those at risk and uncensored at C_M . This is carried out by implicitly assuming that the only values that D can take among subjects at risk at time C_M are the values of D taken by the subjects at risk but not censored at time C_M . Each subject i at risk and uncensored at time C_M is weighted so as to represent (approximately) himself

and $\left[\prod_{k=1}^{j^*} \{1 - \hat{\lambda}_{0,k}(C_M; \gamma) r_k(C_M, V_i^H(C_M), D_i; \gamma_k)\} \right]^{-1} - 1$ subjects censored at time C_M who have the same recorded past $V_i^H(C_M)$ and who would have had the same value of D as him/her if uncensored. Next, the same weighting strategy is followed for estimation of $\lambda_{0,j}(C_{M-1}; \gamma)$: the already weighted subjects who contributed to the estimation of $\lambda_{0,j}(C_M, \gamma)$, and the subjects with observed failure time between C_{M-1} and C_M are now weighted by $\left[\prod_{k=1}^{j^*} \{1 - \hat{\lambda}_{0,k}(C_{M-1}) r_k(C_{M-1}, V_i^H(C_{M-1}), D_i; \gamma_k)\} \right]^{-1}$, and so on.

We end this section with a note about variance estimation. The derivation of the variance of the limiting normal distribution of $\sqrt{n}((\hat{\mu}, \hat{\gamma}) - (\mu^*, \gamma^*))$ is subtle because $\hat{\mu}$ and $\hat{\gamma}$ depend on the estimator of the infinite dimensional parameters $\Lambda_{0,j}(\cdot)$, $j = 1, \dots, j^*$. In the Appendix we derive this variance, and a consistent estimator of it, with the help of ideas from semiparametric efficiency theory. Alternatively, the non-parametric bootstrap should provide a consistent variance estimator if, as we speculate, $\hat{\mu}$ is a regular, asymptotically, normal estimator of μ^* (Gill, 1989). However if, as we recommend here, one is to conduct repeated inference under different selection bias functions as a form of sensitivity analysis, bootstrap variance estimation may require impractically large computation time, and the computation of explicit analytic variance estimators may be preferable in practice, even if the calculations for the variance estimator are intricate.

6 Simulation studies

We conducted a simulation study to evaluate the performance of the estimator $\hat{\mu}$ of $\mu^* \equiv E(D)$ and of its estimated standard error in finite samples. The study consisted of 500 replications of samples of size $n = 500$. The maximum actual follow-up time was $\tau^* = 2.5$ and the maximum period for analysis τ was set to 2. We generated T^* according to an exponential distribution with mean equal to 2 and defined $T = \min(T^*, 2)$. Given T^* , we generated a 1×5 vector following a multivariate normal distribution $N(\eta(T^*), \Sigma)$, where $\Sigma_{ij} = 0.6^{|i-j|}$, $i, j = 1, \dots, 5$. This vector was meant to represent the values of $V(t)$ at times $t = t_0, \dots, t_4$ where $t_0 = 0$, $t_1 = 0.5$, $t_2 = 1$, $t_3 = 1.5$ and $t_4 = 2$. Then we set $V(u) = V(t_j)$, for t_j $u < t_{j+1}$. To create association between $V(\cdot)$ and T we set $\eta(T^*) = (T^*, T^*, T^*, T^*, T^*)^T$. To simulate staggered entry we generated a censoring variable C_1^* independent of T and $V^H(\cdot)$ following a Uniform(0, 2.5) distribution. With this set-up, in the absence of drop-out, 53%

of the patients are expected to experience the event in the interval $(0, \tau)$, i.e. $\Pr(T = 2) = 0.47$. Furthermore, the probability of not being administratively censored in the interval $(0, 2)$ is at least 0.2. We defined D using utilities $s(u, \mathbf{v}^H(u)) = \Pr(V(t_{j-1}) \leq \mathbf{v}(t_{j-1}))$ for $t_{j-1} < t_j$ to mimic a situation in which low values of $V(t)$ correspond to poor quality of life and thus are assigned lower scores. The score at a time u for a subject with covariate history $\mathbf{V}^H(u)$ is calculated as the proportion of subjects in the population with values as low or lower than $V(u)$.

In addition to C_1^* , given $(T^*, \mathbf{V}^H(T))$ we generated another random variable C_2^* , meant to represent an informative censoring process, with conditional hazard at t equal to $\exp\{\gamma_2 V(t_{j-1}) + q_2(t, \mathbf{V}^H(t), D)\}$ for $t \in [t_{j-1}, t_j]$, where $\gamma_2 = -0.25$. The function $q_2(t, \mathbf{V}^H(t), D)$ was defined as in (4) with $\alpha = -0.5$. We set $C = \min(C_1^*, C_2^*, \tau)$ and defined $J = 0$ if $C = \tau$, $J = 1$ if $C = C_1^*$ and $J = 2$ if $C = C_2^*$. Note that with these definitions, models (3) and (5) hold for $j = 1, 2$, with $\lambda_{C_1}(t, \mathbf{V}^H(t), D) = 1/(2.5 - t)$ independent of $(\mathbf{V}^H(t), D)$ (and therefore $\gamma_1 = 0$ and $q_1 = 0$), and $\lambda_{C_2}(t, \mathbf{V}^H(t), D) = \exp\{\gamma_2 V(t_{j-1}) + q_2(t, \mathbf{V}^H(t), D)\}$ for $t \in [t_{j-1}, t_j]$.

Table 1 reports the results of the simulations for the estimators of μ^* and γ^* and $\alpha = -0.5$ (the true value), $\alpha = 0$ and $\alpha = 1$. The first row and second rows of Table 1 report the Monte-Carlo mean and standard error of the estimators respectively, the third row reports the Monte Carlo mean of the estimator of standard errors given in the Appendix and the fourth row reports the coverage probability of the nominal 95% Wald intervals centered at the one-step estimators. The results show that the estimators and confidence intervals of μ^* , γ_1^* that use the true censoring bias function perform well. The estimator of γ_2^* is slightly biased downwards, possibly because the finite sample distribution of our estimator is not yet well approximated by its asymptotic normal distribution under our simulation setting with sample size $n = 500$. Note that in spite of the slight bias of the estimator of γ_2^* , the estimator of μ^* remains nearly unbiased. As expected, inference for μ^* and γ_2^* using the wrong censoring bias function is incorrect. Note that incorrectly assuming that $\alpha = 0$, i.e. that censoring is explainable by $\mathbf{V}^H(t)$ yields a biased estimator of μ and undercoverage of the Wald confidence interval. Note that even when $\alpha = 0$ our estimators should not necessarily converge in probability to the same limit as those of van der Laan and Hubbard (1999), even though $\alpha = 0$ is tantamount to assuming that all causes of censoring can be explained by the recorded variables $\mathbf{V}^H(t)$. The reason for this is that van der Laan and Hubbard (1999) do not model the competing, administrative and drop-out, censoring mechanisms separately, as we do here. Our estimators with $\alpha = 0$ do not necessarily converge to the same probability limit as those of Zhao and Tsiatis (2000), because these authors estimate μ^* under the stronger assumption that censoring is also independent of the recorded variables.

Quite surprisingly, inference about γ_1 was unaffected by misspecification of α . Note that this is not predicted by the theory, since estimation of γ_1 requires consistent estimation of γ_2 through the weights $\pi(u | \mathbf{V}^H(u), D; \gamma^*, \Lambda_0(\gamma^*))$.

7 Analysis of the ACTG 175 study

We estimated $\mu^* = E(D)$ where D is defined as in § 3.1 and τ equal to 1120 days (expressed using one year as a unit). We considered two censoring mechanisms. The first mechanism was administrative censoring and hence q_1 was set equal to 0. The second was defined as the minimum between the time to loss-to-follow-up and to treatment discontinuation. For this censoring process we considered the function $q_2(t, \mathbf{V}^H(t), D)$ defined in (4). We also assumed model (5) with

$$w_j(u, V^H(u)) = (CD4_u, AGE, IV, KAR, SYM_0, SYM_u), j=1, 2 \quad (16)$$

where $CD4_u$ is the last CD4 measurement prior to or at time u , AGE , IV , KAR and SYM_0 are baseline variables defined as age, the indicator of intravenous drug use, Karnovsky score and symptomatic HIV infection respectively, and SYM_u is a time dependent indicator of ever symptom over the last two months prior to u . We repeated the estimation for each α in the set $\{1, 0.7, 0.5, 0.3, 0, -0.3, -0.5, -0.7, -1\}$. As expected, the estimated mean values decrease as α decreases, i.e. as we assume that patients with worse prognosis have higher probability of dropping out. Nevertheless, the difference between the largest and lowest estimated means across values of α is roughly of 1/2 year. In addition, standard errors remain roughly equal to 0.05 for all α . From our analysis we conclude that the mean time spent without symptoms from entry to the study to the development of AIDS or death in the AZT arm, is not less than $1.88 - 2 \times 0.054 = 1.77$ years and not greater than $2.39 + 2 \times 0.04 = 2.47$ years.

8 Extensions and concluding remarks

The estimators of quantities associated with the distribution of a quality adjusted failure time endpoint available in the literature assume that the hazard of the minimum of competing censoring times is independent of the endpoint conditionally of measured prognostic factors as in van der Laan and Hubbard (1999) or completely independent of both the measured prognostic factors and the endpoint as in Zhao and Tsiatis (1997, 1999, 2000, 2001). Our approach gives the data analyst the possibility of conducting inference assuming informative censoring and specifying the non-identified degree of informativeness, thus rendering the possibility of conducting a sensitivity analysis. In addition, by allowing incorporation of refined cause-specific models for the censoring process, the methods provide more flexibility to approximate the actual censoring process than currently available methods that make assumptions directly on the minimum of the competing censoring times.

We have considered estimation of functionals of the form $\mu^* = E\{g(D)\}$ for an arbitrary, user-specified, function g . Adequate choices of g render estimands of practical relevance. For example, the choices $g(D) = D$ and $g(D) = I(D > d)$ give $\mu^* = E(D)$ and $\mu^* = \Pr(D > d)$ respectively. Our procedures can be extended in a straightforward manner to compute estimators of other estimands μ^* of interest so long as μ^* can be expressed as the solution of population moment equations $E\{U(D, \mu)\} = 0$ for some known function $U(\cdot; \cdot)$. This extension is accomplished by simply replacing $g(D) - \mu$ with $U(D, \mu)$ in every formula where $g(D) - \mu$ appears in our developments. In particular, we can estimate the median of D by replacing $U(D, \mu) = I(D > \mu) - 1/2$ for $g(D) - \mu$.

The efficiency of our estimators $(\hat{\mu}, \hat{\gamma})$ depends on the choice of functions $a_j(u, V^H(u))$, $j = 1, \dots, j^*$. It can be shown that there is an optimal choice $a_{j,opt}(u, V^H(u))$, $j = 1, \dots, j^*$, which yields a solution $(\hat{\mu}, \hat{\gamma})$ whose limiting distribution has variance equal to the semiparametric variance bound for (μ, γ) in the model that assumes (3) for specified q_j 's and (5). Unfortunately, the function $a_{j,opt}$ is not available for data analysis because it depends on the unknown distribution of the observed data. More importantly, the actual functional form of $a_{j,opt}$ is not available in closed form as $a_{j,opt}$ solves an exceedingly complicated integral equation with no closed form solution. Thus, in practice, we cannot hope to use our knowledge of the integral equations that $a_{j,opt}$ solves to compute estimators of μ^* with good efficiency properties. In our opinion, efficient estimation of μ^* is not a major practical issue. Improving the efficiency with which we estimate μ^* will generally have a minor impact in the overall conclusions of the sensitivity analysis. This is so because in general, the

uncertainty about μ^* due to lack of knowledge of the true censoring bias functions far exceeds the uncertainty due to sampling variability.

Acknowledgments

This work was supported by NIH grants R01 GM48704 (Rotnitzky, Bergesio and Farall) and R37 AI32475 (Rotnitzky).

Appendix

Population equations solved by $(\Lambda_0, \gamma^*, \mu^*)$. We have indicated in § 5 that γ^*, μ^* and Λ_0 are solutions of the population moment equations (7) and

$$E \{m(O; \gamma, \Lambda_0, a)\} = 0. \quad (17)$$

We will show below that under models (3) and (5), $\Lambda_0(\cdot)$ is actually equal to $\Lambda_0(\cdot; \gamma^*)$ where for each γ , $\Lambda_0(\cdot; \gamma)$ is defined by the (population) integral equation

$$\Lambda_{0,j}(u; \gamma) = \int_0^u \frac{\lambda_{c,j}(s|T>s)}{E \{ \Delta v(s, V^H(T), D; \gamma, \Lambda_0(\cdot, \gamma)) r_j(s, V^H(s), D; \gamma_j) | C \geq s, T > s \}} ds; u \geq 0 \quad (18)$$

We refer to (7), (17) and (18) as *population* equations because the expectations and cumulative hazards are computed under the true law of the observed data P_O . We note that the estimators $\hat{\Lambda}_{0,j}(u; \hat{\gamma})$, $\hat{\gamma}$ and $\hat{\mu}$ proposed in this paper are precisely solutions to the empirical versions of equations (7), (17) and (18), i.e. the equations obtained when the expectations and cumulative hazards are computed under the empirical law P_n that puts mass $1/n$ in each data point O_i , $i = 1, \dots, n$.

To show $\Lambda_0(\cdot)$ is equal to $\Lambda_0(\cdot; \gamma^*)$ satisfying (18) with γ equal to γ^* write

$$\begin{aligned} \Lambda_{0,j}(u) &= \int_0^u \lambda_{0,j}(s) ds = \int_0^u \frac{E \{ \lambda_{0,j}(s) r_j(s, V^H(s), D; \gamma_j^*) | C \geq s, T > s \}}{E \{ r_j(s, V^H(s), D; \gamma_j^*) | C \geq s, T > s \}} ds \\ &= \int_0^u \frac{\lambda_{c,j}(s|T>s)}{E \{ r_j(s, V^H(s), D; \gamma_j^*) | C \geq s, T > s \}} ds \\ &= \int_0^u \frac{\lambda_{c,j}(s|T>s)}{E \{ \Delta v(s, V^H(T), D; \gamma^*, \Lambda_0) r_j(s, V^H(s), D; \gamma_j^*) | C \geq s, T > s \}} ds \end{aligned}$$

where the third equality follows because $\lambda_{c,j}(s|T>s) = E \{ \lambda_{c,j}(s|T>s, V^H(s), D) | C \geq s, T > s \}$ and under model (3) and (5), $\lambda_{c,j}(s|T>s, V^H(s), D) = \lambda_0(s) r_j(s, V^H(s), D; \gamma_j^*)$ and the fourth equality follows by identity (10).

Proof that equation (12) is solved at any γ when $a_j(\cdot, \cdot)$ does not depend on $V^H(\cdot)$.

We will prove this by showing the following lemma which establishes an even stronger result.

Lemma A.1 If for all j , $a_j(u, V^H(u)) = a_j(u)$ is a function of u only, and if $\Lambda_0(\cdot, \gamma)$ is defined as in (18) where all expectations and cumulative hazards are computed under an arbitrary law P_O of O , then

$$E\{m(O; \gamma, \Lambda_0(\cdot, \gamma), a)\} = 0 \quad (19)$$

for all γ where the expectation is computed under the same P_O

The desired result that equation (12) is solved at any γ follows after noticing that (12) is precisely equation (19) with P_O replaced by P_n , the empirical law of the data, and $\Lambda_0(\cdot, \gamma)$ replaced by $\hat{\Lambda}_0(\cdot, \gamma)$ which is precisely the solution of (18) when the expectations and cumulative hazards are computed also under P_n .

To show Lemma A.1 consider the data configuration $O^* = (\Delta, X, J, \Delta V^H(T))$ which differs from the observed data configuration $O = (\Delta, X, J, V^H(X))$ in that, O^* does not include any part of the covariate process $V^H(\cdot)$ in censored subjects, while O includes the covariate process until C in censored subjects. The law P_{O^*} of the data O^* is determined by P_O because O^* is a function of O . Thus, in particular, for any function $g(O^*)$ of the data O^* , we have that

$$E_{P_{O^*}}\{g(O^*)\} = E_{P_O}\{g(O^*)\}. \quad (20)$$

Suppose that we specify a model A_q for the full data $(C, J, V^H(T), T)$ defined by the restriction

$$\lambda_{C,j}(u|V^H(T), T, T > u) = \lambda_{0,j}^*(u) \exp\{\gamma_j^{**'} w_j(u, V^H(u)) + q_j(u, V^H(u), D)\} \quad (21)$$

where γ_j^{**} and q_j are fixed and known and $\lambda_{0,j}^*(\cdot)$ is an unknown function. Here γ_j^{**} is an arbitrary conformable vector not necessarily equal to γ_j^* , $j=1, \dots, j^*$.

Arguing as in Scharfstein, Rotnitzky and Robins (1999) it can be shown that model A_q does not restrict the distribution of O^* . That is, given any fixed P_{O^*} , for each choice of γ_j^* and q_j , $j=1, \dots, j^*$, there exists a joint distribution of $(C, J, V^H(T), T)$ which satisfies (21) and whose marginal distribution of $(\Delta, X, J, \Delta V^H(T))$ is equal to P_{O^*} . In fact, also arguing as in Scharfstein et. al. (1999) it can be shown that there exists a unique such distribution. That is, the distribution of $(C, J, V^H(T), T)$ is identified under model A_q by P_{O^*} . Throughout we use P^* to denote the distribution of $(C, J, V^H(T), T)$ under A_q determined by P_{O^*} and we use the subscript $*$ to denote operations (i.e. expectations, conditional hazards, etc.) done with respect to P^* . With these definitions, under model A_q

$$E_{P_{O^*}}\{g(O^*)\} = E^*\{g(O^*)\} \quad (22)$$

for any function $g(O^*)$. Note that under model A_q it is not true that $E_{P_O}\{h(O)\} = E^*\{h(O)\}$ for an arbitrary function $h(O)$ of O because the marginal distribution of O under P^* is not necessarily equal to P_O .

For any non-negative function $\lambda_{0,j}^\dagger(\cdot)$ on the reals, let $\Lambda_{0,j}^\dagger(u) \equiv \int_0^u \lambda_{0,j}^\dagger(s) ds$. Define for any vector $\gamma^\dagger = (\gamma_1^\dagger, \dots, \gamma_{j^*}^\dagger)$ and any $\Lambda_{0,j}^\dagger(\cdot)$, $j=1, \dots, j^*$,

$$\pi(t|V^H(t), D; \gamma^\dagger, \Lambda_0^\dagger) \equiv \prod_{j=1}^{j^*} \prod_{0 \leq u \leq t} \left\{ 1 - \exp\{\gamma_j^\dagger w_j(u, V^H(u)) + q_j(u, V^H(u), D)\} d\Lambda_{0,j}^\dagger(u) \right\},$$

With this definition, under model A_q it holds that with $\pi(t|V^H(t), D, \gamma, \Lambda_0)$ defined as in (8), $\pi(t|V^H(t), D; \gamma^{**}, \Lambda_0^*) = \exp\left\{-\int_0^t \lambda_c^*(u|V^H(T), T, T > u) du\right\}$ and it is straightforward to

show that $E^* \left\{ \Delta \frac{\pi(u|V^H(u), D; \gamma^{**}, \Lambda_0^*)}{\pi(T|V^H(T), D; \gamma^{**}, \Lambda_0^*)} \middle| V^H(T), T, T > u, C \geq u \right\} = 1$ for all $u \in (0, T]$.

Consequently, defining $m(O; \gamma^\dagger, \Lambda_0^\dagger, a)$ and $dM_{C,j}(u; \Lambda_0^\dagger, \gamma^\dagger)$ for any $\Lambda_0^\dagger, \gamma^\dagger$ exactly like $m(O; \gamma, \Lambda_0, a)$ and $dM_{C,j}(u; \Lambda_0, \gamma)$ but with Λ_0 and γ replaced by $\Lambda_0^\dagger$ and $\gamma^\dagger$ we have that under model A_q , $E^*[dM_{C,j}(u; \Lambda_0^*, \gamma^{**})|V^H(T), T, T > u, X \geq u] = 0$ and, in particular, $E^*\{m(O; \gamma^{**}, \Lambda_0^*, a)\} = 0$.

The integral equation (18) depends on P_O only through P_{O^*} . Indeed, arguing exactly as in Scharfstein, Rotnitzky and Robins (1999) we can show that $\Lambda_{0,j}^*(u)$ is the unique $\Lambda_{0,j}(u; \gamma^{**})$ satisfying the integral equation (18). Thus, we have that $E^*\{m(O; \gamma^{**}, \Lambda_0(\gamma^{**}), a)\} = 0$. In the particular case in which $a_j(u, V^H(u))$ is a function $a_j(u)$ of u only for all $j = 1, \dots, j^*$, the function $m(O; \gamma^{**}, \Lambda_0(\gamma^{**}), a)$ depends on O through O^* only. Consequently, for any functions $a = \{a_j(u), j = 1, \dots, j^*\}$ that depend on u only we have

$$0 = E^*\{m(O; \gamma^{**}, \Lambda_0(\gamma^{**}), a)\} = E_{P_{O^*}}\{m(O; \gamma^{**}, \Lambda_0(\gamma^{**}), a)\} = E_{P_O}\{m(O; \gamma^{**}, \Lambda_0(\gamma^{**}), a)\}$$

where the second equality is by (22) and the third is by (20). Since we have taken γ^{**} arbitrary, this proves (19).

Variance estimation.

Define for any $1 \times (p+1)$ constant vector $\mathbf{k}$ and any set of $1 \times (p+1)$ vector-valued functions $\phi_j(u, V^H(u))$,

$$h(O; \mu, \gamma, \Lambda_0, \mathbf{k}, \phi) \equiv \frac{\Delta}{\pi(T|V^H(T), D; \gamma, \Lambda_0)} \mathbf{k} \{g(D) - \mu\} + m(O; \gamma, \Lambda_0, \phi)$$

The solution $(\hat{\mu}, \hat{\gamma})$ of the joint system (9), (12) also solves the equation in (μ, γ)

$$\sum_{i=1}^n h(O_i; \mu, \gamma, \hat{\Lambda}_0(\gamma), \mathbf{e}, a+d) = 0 \quad (23)$$

where $a+d = \{a_j\{u, V^H(u)\} + d_j(u), j = 1, \dots, j^*\}$ for $d_j(u)$ a $1 \times (p+1)$ vector function of u and $\mathbf{e}$ is the $1 \times (p+1)$ vector with 1 in its first entry and 0 elsewhere. This follows because, as proved above, when $a_j(u, V^H(u))$ does not depend on $V^H(u)$, equation (17) holds for all γ and, in particular, it holds for $\hat{\gamma}$.

Now, suppose momentarily that the solution $\Lambda_0(\cdot; \gamma)$ of the integral equation (18) were known for each γ . We could then consider estimating (μ^*, γ^*) with $(\tilde{\mu}_d, \tilde{\gamma}_d)$ solving the equation in (μ, γ)

$$\sum_{i=1}^n h(O_i; \mu, \gamma, \Lambda_0(\cdot; \gamma), \mathbf{e}, a+d) = 0 \quad (24)$$

Equation (24) differs from (23) in that the solution $\Lambda_0(\cdot; \gamma)$ to the population integral equation (18) replaces the solution $\hat{\Lambda}_0(\cdot, \gamma^*)$ of the empirical version of this equation. In contrast to the solutions of (23), the solutions $(\tilde{\mu}_d, \tilde{\gamma}_d)$ need not be the same for each choice of functions $d_j, j = 1, \dots, j^*$ because $\sum_{i=1}^n m(O_i; \gamma, \Lambda_0(\gamma), d)$ is not equal to 0 for an arbitrary γ . Using standard Taylor expansion arguments, we can easily derive that, under mild regularity conditions, equation (24) has a solution $(\tilde{\mu}_d, \tilde{\gamma}_d)$ such that $\sqrt{n}\{(\tilde{\mu}_d, \tilde{\gamma}_d) - (\mu^*, \gamma^*)\}$ converges to a normal distribution with variance equal to $\Gamma_d \Omega_d \Gamma_d^T$ where $\Omega_d = E\{h(O; \mu^*, \gamma^*, \Lambda_0, a + d)^{\otimes 2}\}$ and $\Gamma_d = \left\{ E \left[\partial h(O; \mu, \gamma, \Lambda_0, a + d) / \partial(\mu, \gamma) |_{(\mu^*, \gamma^*)}^T \right] \right\}^{-1}$.

The quantity $\Gamma_d \Omega_d \Gamma_d^T$ is not the variance of the limiting distribution of $(\hat{\mu}, \hat{\gamma})$ solving (23) because estimation of $\Lambda_0(\cdot; \gamma)$ impacts the asymptotic distribution of (μ, γ) . There exists, however, one specific choice of functions $d_j(\cdot)$, throughout denoted as $d_j^*(\cdot), j = 1, \dots, j^*$, such that $\Gamma_{d^*} \Omega_{d^*} \Gamma_{d^*}^T$ is indeed the variance of the limiting distribution of $(\hat{\mu}, \hat{\gamma})$. These functions are the unique, up to multiplying constants, choices $d_j^*(\cdot)$ such that $h(O; \mu^*, \gamma^*, \Lambda_0, \mathbf{e}, a + d^*)$, evaluated at the true values μ^*, γ^* and Λ_0 , is uncorrelated with the scores corresponding to all parameters, other than μ and γ , of all parametric sub-models of semiparametric model defined by (3) and (5). In the literature on semiparametric estimation, the closed linear span of the set of all such scores is called the nuisance tangent space. The collection of all random variables uncorrelated with all the elements of this set is called the orthogonal complement to the nuisance tangent space for (μ, γ) . Choosing an estimating function $h(O; \mu, \gamma, \Lambda_0(\gamma), \mathbf{e}, a + d^*)$ such that, when evaluated at the true parameter values μ^* and γ^* , is in the orthogonal complement of the nuisance tangent space ensures that the limiting distributions of the solution to (24) and of the solution to (23) (with d^* instead of d) are the same (Newey, 1990).

In a companion technical report we show that the function $d_j^*(\cdot)$ is the unique solution to the integral equation in $d_j(\cdot)$,

$$d_j(u) = Q_{1,j}(u; \gamma^*, \Lambda_0)^{-1} \left\{ -Q_{2,j}(u; \gamma^*, \Lambda_0, a_j) + Q_{3,j}(u; \mu^*, \gamma^*, \Lambda_0, \mathbf{e}) - \sum_{j'=1}^{j^*} \int_0^u K_{j',j}(s, u; \mu^*, \gamma^*, \Lambda_0, a_{j'} + d_{j'}) d\Lambda_{0,j'}(s) \right\} \quad (25)$$

where, using the shortcut notation $\pi(t) = \pi(t | V^H(t), D, \gamma^*, \Lambda_0)$ for all t ,

$$Q_{1,j}(u; \gamma^*, \Lambda_0) = E \left\{ \Delta I(T > u) \pi(T)^{-1} \pi(u) r_j(u, V^H(u), D; \gamma_j^*) \right\} \quad (26)$$

$$Q_{2,j}(u; \gamma^*, \Lambda_0, a_j) = E \left\{ \Delta I(T > u) \pi(T)^{-1} \pi(u) r_j(u, V^H(u), D; \gamma_j^*) a_j(u, V^H(u)) \right\} \quad (27)$$

$$Q_{3,j}(u; \mu^*, \gamma^*, \Lambda_0, \mathbf{e}) = E \left\{ \Delta I(T > u) \pi(T)^{-1} r_j(u, V^H(u), D; \gamma_j^*) \mathbf{e}[g(D) - \mu^*] \right\} \quad (28)$$

and

$$K_{j',j}(s, u; \mu^*, \gamma^*, \Lambda_0, a_{j'} + d_{j'}) = E \left\{ \Delta I(T > u) \pi(T)^{-1} \pi(s) [a_{j'}(s, V^H(s)) + d_{j'}(s)] \times r_{j'}(s, V^H(s), D; \gamma_{j'}^*) r_j(u, V^H(u), D; \gamma_j^*) \right\} \quad (29)$$

To derive a consistent estimator of $\Gamma_{d^*} \Omega_{d^*} \Gamma_{d^*}^T$, we define the functions $\hat{Q}_{1,j}(u, \gamma^*, \Lambda_0)$, $\hat{Q}_{2,j}(u, \gamma^*, \Lambda_0, a_j)$, $\hat{Q}_{3,j}(u, \mu^*, \gamma^*, \Lambda_0, \mathbf{e})$ and $\hat{K}_{j',j}(s, u, \mu^*, \gamma^*, \Lambda_0, a_{j'} + d_{j'})$ like the right

hand sides of (26), (27), (28) and (29) with expectations replaced by sample averages. Next, we compute $\hat{d}_j(u)$ solving (25) but with $Q_{1,j}(u; \gamma^*, \Lambda_0)$, $Q_{2,j}(u; \gamma^*, \Lambda_0, a_j)$, $Q_{3,j}(u; \mu^*, \gamma^*, \Lambda_0, \mathbf{e})$ and $K'_{j,j}(s, u; \mu^*, \gamma^*, \Lambda_0, a_j + d'_j)$ replaced by $\hat{Q}_{1,j}(u; \hat{\gamma}, \hat{\Lambda}_0(\hat{\gamma}))$, $\hat{Q}_{2,j}(u; \hat{\gamma}, \hat{\Lambda}_0(\hat{\gamma}), a_j)$, $\hat{Q}_{3,j}(u; \hat{\mu}, \hat{\gamma}, \hat{\Lambda}_0(\hat{\gamma}), \mathbf{e})$ and $\hat{K}'_{j,j}(s, u; \hat{\mu}, \hat{\gamma}, \hat{\Lambda}_0(\hat{\gamma}), a_j + d'_j)$. Finally, we compute

$$\hat{\Omega}_{d^*} \equiv n^{-1} \sum_{i=1}^n h(O_i; \hat{\mu}, \hat{\gamma}, \hat{\Lambda}_0(\hat{\gamma}), a + \hat{d})^{\otimes 2} \text{ and } \hat{\Gamma}_{d^*} \equiv \left\{ n^{-1} \sum_{i=1}^n \partial h(O_i; \mu, \gamma, \hat{\Lambda}_0(\hat{\gamma}), a + \hat{d}) / \partial (\mu, \gamma)^T \Big|_{(\mu, \gamma) = (\hat{\mu}, \hat{\gamma})} \right\}^{-1}$$

. Under regularity conditions, $\hat{\Gamma}_{d^*} \hat{\Omega}_{d^*} \hat{\Gamma}_{d^*}^T$ converges in probability to $\Gamma_{d^*} \Omega_{d^*} \Gamma_{d^*}^T$.

References

1. Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality of life assessment: can we keep it simple? *Journal of the Royal Statistical Society, Series B*. 1992; 3:353–393.
2. Gelber R, Gelman R, Goldhirsch A. A quality of life oriented endpoint for comparing treatment therapies. *Biometrics*. 1989; 45:781–795. [PubMed: 2790121]
3. Gill RD. Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method. *Scandinavian Journal of Statistics*. 1989; 16:97–128.
4. Hammer SM, Katsenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RJ, Henry WK, Lederman MM, Phair JP, Niu M, Hirsch MS, Merigan TC. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic milliliter. *New England Journal of Medicine*. 1996; 335:1081–1090. [PubMed: 8813038]
5. Malani HM. A modification of the redistribution to the right algorithm using disease markers. *Biometrika*. 1995; 82:515–526.
6. Newey W. Semiparametric efficiency bounds. *Journal of Applied Econometrics*. 1990;99–135.
7. NIAID, National Institute of Allergy and Infectious Diseases. Division of AIDS table for grading severity of adult adverse experiences. Rockville, MD: 1996.
8. Robins, J.; Rotnitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell, NP.; Dietz, K.; Farewell, B., editors. *AIDS Epidemiology*. Birkhauser, Boston: 1992.
9. Robins, J. Proceedings of the Biopharmaceutical Section, American Statistical Association. Alexandria, VA: American Statistical Association; 1993. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers; p. 24–33.
10. Robins, JM.; Rotnitzky, A.; Scharfstein, DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, E.; Berry, D., editors. *Statistical Models for Epidemiology, the Environment, and Clinical Trials*. 2000. p. 1–95.
11. Rotnitzky A, Farall A, Bergesio A, Scharfstein D. Analysis of failure time data under competing censoring mechanisms. *Journal of the Royal Statistical Society, Series B*. 2007 To appear.
12. Scharfstein D, Robins J. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*. 2002; 89:617–634.
13. Scharfstein DO, Rotnitzky A, Robins J. Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*. 1999; 94:1096–1120.
14. van der Laan M, Hubbard A. Locally efficient estimation of the quality adjusted lifetime distribution with right-censored data and covariates. *Biometrics*. 1999; 55:530–536. [PubMed: 11318210]
15. Zhao H, Tsiatis AA. A Consistent Estimator for the Distribution of Quality Adjusted Survival Time. *Biometrika*. 1997; 84:339–348.
16. Zhao H, Tsiatis AA. Efficient Estimation of the Distribution of Quality Adjusted Survival Time. *Biometrics*. 1999; 55:1101–1107. [PubMed: 11315054]
17. Zhao H, Tsiatis AA. Estimating Mean Quality Adjusted Lifetime With Censored Data. *Sankhya*. 2000; 62(1):175–188. **Series B**.

18. Zhao H, Tsiatis AA. Testing Equality of Survival Functions of Quality Adjusted Lifetime. *Biometrics*. 2001; 57:861–867. [PubMed: 11550938]

\$watermark-text

\$watermark-text

\$watermark-text

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Results of simulation study
True values $\mu = 0.788$, $\gamma_1 = 0$, $\gamma_2 = -0.25$

	$\alpha = -0.5$ (true value)			$\alpha = 0$ (wrong value)			$\alpha = 1$ (wrong value)		
	$\hat{\mu}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\mu}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\mu}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
MC mean	0.7906	0.0057	-0.2260	0.8171	0.0079	-0.4100	0.8777	0.0135	-0.8401
MC SE	0.0356	0.0416	0.0657	0.0362	0.0421	0.0720	0.0404	0.0433	0.1877
MC mean of estimated SE	0.0353	0.0383	0.0616	0.0358	0.0386	0.0665	0.0360	0.0387	0.1107
Coverage Probability	0.9400	0.9500	0.8980	0.8860	0.9440	0.3200	0.3160	0.9440	0.0260

\$watermark-text

\$watermark-text

\$watermark-text

Table 2

Sensitivity analysis for the mean of QOL adjusted survival in the AZT arm of the ACTG 175 study

α	1	0.7	0.5	0.3	0	-0.3	-0.5	-0.7	-1
$\hat{\mu}$	2.3913	2.3530	2.3166	2.2752	2.2000	2.1115	2.0474	1.9792	1.8753
Standard error of $\hat{\mu}$	0.0389	0.0413	0.0434	0.0452	0.0475	0.0492	0.0503	0.0518	0.0536