

# Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests

James M. Robins\* and Dianne M. Finkelstein\*\*

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

\*email: robins@epinet.harvard.edu

\*\*email: finkel@newton.mgh.harvard.edu

**SUMMARY.** AIDS Clinical Trial Group (ACTG) randomized trial 021 compared the effect of bacrim versus aerosolized pentamidine (AP) as prophylaxis therapy for pneumocystis pneumonia (PCP) in AIDS patients. Although patients randomized to the bacrim arm experienced a significant delay in time to PCP, the survival experience in the two arms was not significantly different ( $p = .32$ ). In this paper, we present evidence that bacrim therapy improves survival but that the standard intent-to-treat comparison failed to detect this survival advantage because a large fraction of the subjects either crossed over to the other therapy or stopped therapy altogether. We obtain our evidence of a beneficial bacrim effect on survival by artificially regarding the subjects as dependently censored at the first time the subject either stops or switches therapy; we then analyze the data with the inverse probability of censoring weighted Kaplan–Meier and Cox partial likelihood estimators of Robins (1993, *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pp. 24–33) that adjust for dependent censoring by utilizing data collected on time-dependent prognostic factors.

**KEY WORDS:** AIDS; Causal inference; Cox proportional hazards model; Informative censoring; Survival analysis; Time-dependent covariates.

## 1. Introduction

The AIDS Clinical Trial Group (ACTG) randomized trial 021 compared the effect of bacrim versus aerosolized pentamidine (AP) as prophylaxis therapy for recurrent pneumocystis pneumonia (PCP), a potentially fatal pulmonary infection, in AIDS patients (Hardy et al., 1992). The trial accrued 310 patients between July 1988 and November 1990. The primary endpoint was time to recurrent PCP, with survival as a secondary endpoint. Patients who were randomized to a given treatment were allowed to cross over to the other treatment if they developed PCP. Patients randomized to the bacrim arm experienced a significant delay in time to PCP compared to those in the AP arm (log-rank test  $p = .002$ ) with an estimated rate ratio of 3.25 based on a Cox proportional hazard regression analysis with assigned treatment arm as a single dichotomous covariate.

However, despite the clear advantage of bacrim for prevention of PCP, the survival experience in the two arms was not significantly different ( $p = .32$ ), although the median survival was 25.8 months in the bacrim arm compared with 22.8 months in the AP arm. There are at least two possible explanations for the failure to detect a significant survival advantage for bacrim. First, PCP is rarely a direct cause of death in these patients. Specifically, only 3 of the 14 PCP bouts in the bacrim arm proved fatal; only 2 of the 36 PCP bouts in the AP arm were fatal.

Second, due to the high rates of censoring (by end of or loss to follow-up), of treatment crossover, and of cessation of all prophylaxis therapy, the power to detect a beneficial effect of bacrim on survival, based on a standard intent-to-treat analysis, was poor. For example, of the 310 patients entered in the trial, 94 died and 216 were censored by loss to follow-up. Some subjects were lost to follow-up because they dropped out of the study. Others were lost to follow-up by virtue of reaching the administrative end to follow-up date of the trial while still alive. Of the 94 deaths, 21 occurred in subjects who had crossed over to the other treatment arm. Of the remaining 73 deaths, 37 occurred in subjects who had stopped all prophylactic therapy, 21 for nonmedical reasons and 16 for medical indications.

The first explanation is that bacrim has no biological effect on survival, while the second explanation is that there is a beneficial effect that we had inadequate power to detect using a standard intent-to-treat analysis. It is clearly important from a public health point of view to distinguish between these explanations and determine if bacrim prolongs survival when compared with aerosolized pentamidine. Therefore, in this paper, we will reanalyze the trial under additional assumptions which, if true, provide (a) greater power than the assumptions underlying the use of the standard intent-to-treat log-rank test for detecting an effect of bacrim on survival and (b) allow us to estimate the magnitude of that

effect had subjects neither stopped therapy nor crossed over to the alternative therapy. Our approach is to (artificially) regard subjects as dependently censored at the first time a subject stops therapy, switches therapy, or is lost to follow-up. We then apply the inverse probability of censoring weighted (IPCW) Kaplan–Meier estimator and log-rank test of Robins (1993) that adjusts for dependent censoring by utilizing data collected on time-dependent risk factors for failure and censoring. The IPCW versions of these statistics differ from their ordinary versions in that, in calculating the contribution of a subject at risk at time  $t$ , (i) the subject is given a weight inversely proportional to an estimate of the conditional probability of having remained uncensored until time  $t$  and (ii) this estimate is based on the fit of a Cox proportional hazards model for censoring in which time-dependent prognostic factors for failure and censoring are entered as covariates. If a subject stops or switches therapy due to toxicity, it would be inappropriate to regard the subject as noncompliant unless an appropriate palliative therapy is available to combat the toxicity. Fortunately, in the 021 data, information was collected regarding whether cessation or change of therapy was due to toxicity.

More precisely, we will report four different causal analyses. The first analysis is similar to a standard intent-to-treat analysis with mortality as the outcome; i.e., we compare mortality experience in the two treatment arms with death censored only by loss to or end of follow-up. Thus, all 94 deaths are regarded as failures. The novelty in the first analysis is that we allow for the possibility of dependent censoring by loss to follow-up.

In our second analysis, we compare mortality in the two treatment arms while regarding a subject as dependently censored by the minimum of time to loss to follow-up and time to treatment crossover. Thus, only the 73 of the 94 deaths that occurred before a subject crossed over to the other arm are included as failures; the other 21 deaths are regarded as censored. This analysis attempts to estimate what the survival curves would have been if the possibility of crossover to the other treatment arm after the development of PCP had been eliminated from the treatment protocol.

In our third analysis, we compare the mortality in the two treatment arms while regarding subjects as dependently censored at the minimum of time to loss to follow-up, time to treatment crossover, and time to voluntarily stopping therapy (for nonmedically related reasons). In this analysis, only the 52 of the 94 deaths in which subjects had neither crossed over nor voluntarily stopped therapy are regarded as failures. For public health purposes, this analysis may be preferred since it attempts to estimate the survival benefit of bactrim compared to aerosolized pentamidine had no subject voluntarily stopped their assigned therapy without medical indication.

Our final analysis compares mortality in the two treatment arms while regarding subjects as censored at the minimum of time to loss to follow-up, crossover, or stopping therapy for any reason. In this analysis, we are attempting to compare the effect of bactrim versus aerosolized pentamidine on survival if all subjects were forced to stay on their assigned therapy. Even if we could successfully estimate this parameter, it would only be of public health relevance if, as discussed above, the toxicities that led to medically indicated termination of therapy could be ameliorated with appropriate palliative therapy.

Since one would expect that the sickest patients would tend to cross over or to stop therapy, it is clear that we must regard censoring by crossover or stopping therapy as dependent. However, Robins (1993) and Robins and Rotnitzky (1992) showed that, if we have available data on all time-dependent prognostic factors for mortality that independently predict censoring, then one can correct for the dependence between the censoring and failure by replacing the Kaplan–Meier estimator, log-rank test, and Cox partial likelihood estimator of the ratio of the treatment-arm-specific mortality rates by their inverse probability of censoring weighted versions. Extensive data on time-dependent prognostic factors were available in the data file. Specifically, at approximately 8-week intervals, data were recorded on the following six potential prognostic factors: total white blood count (WBC), hemoglobin, Karnovsky score (a clinical measure of functional status), asthenia score (a measure of weight loss and lean body mass), episodes of pneumocystis pneumonia (PCP), and CD4 lymphocyte count.

Our fundamental assumption is that, conditional on the treatment arm  $Z$  and on the recorded history  $\bar{\mathbf{V}}(t)$  of the above six vector of time-dependent covariates  $\mathbf{V}(t)$ , the cause-specific hazard of censoring  $C$  at time  $t$  does not further depend on the possibly unobserved failure time  $T$ , i.e.,

$$\lambda_C(t | \bar{\mathbf{V}}(t), Z, T > t) = \lambda_C(t | \bar{\mathbf{V}}(t), Z, T > t), \quad (1)$$

where overbars are used to denote histories, so  $\bar{\mathbf{V}}(t) \equiv \{\mathbf{V}(x); 0 \leq x \leq t\}$ . Robins (1997) has previously referred to this assumption as no unmeasured confounders for censoring or, equivalently, the assumption of sequential ignorability of censoring. Throughout,  $\lambda_C(t | (\cdot), T > t)$  is the cause-specific hazard of censoring at time  $t$  given that both  $X = \min(T, C)$  exceeds  $t$  and the information in  $(\cdot)$ . Robins (1987) shows that (1) is sufficient to identify the treatment-arm-specific distributions of  $T$ . Note that (1) would be false if there were an important prognostic factor for both failure  $T$  and censoring  $C$  that was not included in  $\mathbf{V}(t)$ . Therefore, it should be a primary goal of the investigators conducting the study to collect data on a sufficient number of covariates in  $\mathbf{V}(t)$  to make (1) plausible. However, in the presence of censoring, whether (1) is or is not true is not subject to empirical test. The usual assumption of independent censoring is that, given the treatment arm, the cause-specific hazard of  $T$  among subjects at risk is the marginal hazard of  $T$ , i.e.,

$$\lambda_T(t | Z, C \geq t) = \lambda_T(t | Z). \quad (2)$$

Given (1), (2) will be true if (and essentially only if), in each treatment arm,  $\bar{\mathbf{V}}(t)$  does not predict censoring at  $t$ , i.e.,

$$\lambda_C(t | \bar{\mathbf{V}}(t), Z, T > t) = \lambda_C(t | Z, T > t) \quad (3)$$

since, as noted by Robins (1987), Jacobsen and Keiding (1995), and Gill, van der Laan, and Robins (1997), (2) is equivalent to

$$\lambda_C(t | Z, T, T > t) = \lambda_C(t | Z, T > t). \quad (4)$$

Note that, in contrast to (1), (3) can be empirically tested by modeling  $\lambda_C(t | \bar{\mathbf{V}}(t), Z, T > t)$  using a time-dependent Cox proportional hazards model for censoring. It follows that, if prognostic factors recorded in  $\bar{\mathbf{V}}(t)$  determine censoring (i.e., (3) is false), then the usual assumption (2) of independent

censoring given treatment arm  $Z$  that justifies the use of the Kaplan–Meier (KM) and Cox partial likelihood estimators and the log-rank test will likely be false. However, if (1) holds, the IPCW versions of these statistics can fully correct for bias due to the dependent censoring attributable to  $\bar{\mathbf{V}}(t)$ . In practice, we would not expect (1) to be precisely true but, given a rich collection of prognostic factors recorded in  $\mathbf{V}(t)$ , it may well be approximately true. When (1) is approximately true, our methods should reduce, even if not totally eliminate, the bias due to dependent censoring.

The remainder of the paper is organized as follows. In Section 2, we reduce the dimension of  $\bar{\mathbf{V}}(t)$  by retaining only those covariates that are significant prognostic factors for mortality. We then introduce the IPCW versions of the KM and Cox partial likelihood estimators and the log-rank test and describe why they are effective in removing bias due to dependent censoring. We also discuss the efficiency advantage of the IPCW KM and Cox partial likelihood estimators compared with the standard KM and Cox partial likelihood estimators when censoring is independent. In Section 3, we present the results of our four causal analyses. In Section 4, we compare our IPCW methods with alternative approaches based on the G-computation algorithm of Robins (1986). We conclude with a discussion.

## 2. IPCW Methods

### 2.1 Dimension Reduction

To construct IPCW estimators, it is necessary to estimate the treatment-arm-specific hazards of censoring conditional on time-dependent prognostic factors for failure, which we shall do using a time-dependent Cox proportional hazards model for censoring. However, because of the large number of potential prognostic factors included in  $\bar{\mathbf{V}}(t)$ , it is first useful to reduce these to a relevant subset. For a time-dependent prognostic factor to cause selection bias or confounding, it must generally be a prognostic factor both for failure and for censoring. We first fit a stratified time-dependent Cox model for failure,

$$\lambda_T \{t | \bar{\mathbf{V}}(t), Z, C > t\} = \gamma_{0Z}(t) \exp \left\{ \sum_{j=1}^6 \psi_j V_j(t) \right\}, \quad (5)$$

where  $V_j(t)$  were the most recent recorded values of the six individual prognostic factors included in the vector  $\mathbf{V}(t)$ ,  $\gamma_{0Z}(t)$  is a treatment-specific baseline hazard, and subjects are censored only by loss to follow-up. Note the model specification given in (5) entails that, conditional on their most recent values, past values of  $\mathbf{V}(u)$  do not predict failure at  $t$ . We kept only those prognostic factors significant at the  $p = .12$  level. These were Karnovsky score at  $t$  ( $Karn(t)$ ), coded as one if the Karnovsky score was 75 or less and zero otherwise; asthenia score ( $Ast(t)$ ), coded as one if the clinical asthenia score was two or greater at  $t$  and zero otherwise; and hemoglobin at  $t$  ( $Hgb(t)$ ), coded as one if less than or equal to eight and zero otherwise. The estimated hazard ratios  $\exp(\hat{\psi}_j)$  in (5) associated with  $\mathbf{V}^*(t) \equiv \{Karn(t), Ast(t), Hgb(t)\}$  were  $\{4.2, 2.2, 4.5\}$ . Conditional on these three variables, the prognostic factors  $WBC(t)$ ,  $PCP(t)$ , and  $CD4(t)$  count did not significantly predict the hazard of death. The reader may be surprised, given the well-known prognostic value of CD4

and PCP in AIDS patients, that they were not independent predictors of mortality. However, standard analyses do not usually adjust for factors such as Karnovsky score and asthenia, which are quite sensitive measures of a patient's current clinical condition and thus may render CD4 count and PCP history nonpredictive conditional on the values of these factors. We refit model (5) under the three alternative definitions of censoring, corresponding to our other three causal analyses. The variables selected as significant prognostic factors were quite consistent across these alternative analyses.

For the purposes of constructing our IPCW estimators, we can reduce the six vector  $\mathbf{V}(t)$  to the three vector  $\mathbf{V}^*(t)$  provided that the analog of (1) holds with  $\bar{\mathbf{V}}^*(t)$  in place of  $\bar{\mathbf{V}}(t)$ , i.e.,

$$\lambda_C \{t | \bar{\mathbf{V}}^*(t), Z, T, T > t\} = \lambda_C \{t | \bar{\mathbf{V}}^*(t), Z, T > t\}. \quad (6)$$

Now one might suppose that, given (1) holds, (6) will hold provided the variables in  $\mathbf{V}^*(t)$  are the only prognostic variables for failure; i.e., one might suppose that (6) would hold if (as suggested by our fit of model (5)) the cause-specific hazard of failure at  $t$  does not depend on the covariate history  $\bar{\mathbf{V}}(t)$  except through  $\bar{\mathbf{V}}^*(t)$ , i.e.,

$$\lambda_T \{t | \bar{\mathbf{V}}(t), Z, C > t\} = \lambda_T \{t | \bar{\mathbf{V}}^*(t), Z, C > t\}. \quad (7)$$

However, as noted by Robins (1986), this supposition is incorrect, i.e., (1) and (7) do not imply (6). But if, in addition to (1) and (7), the remaining variables  $\mathbf{V}^\dagger(u) = (CD4(u), PCP(u), WBC(u))$  in  $\mathbf{V}(u)$  do not predict future values of the prognostic factors  $\mathbf{V}^*(u)$ , then (6) follows. More precisely, Robins (1986) shows that (1), (7), and

$$\begin{aligned} f \{ \mathbf{V}^*(u) | \bar{\mathbf{V}}^\dagger(u^-), \bar{\mathbf{V}}^*(u^-), Z, T \geq u, C \geq u \} \\ = f \{ \mathbf{V}^*(u) | \bar{\mathbf{V}}^*(u^-), Z, T \geq u, C \geq u \} \end{aligned} \quad (8)$$

together imply (6). Equation (8) states that the history  $\bar{\mathbf{V}}^\dagger(u^-)$  of the other time-dependent covariates do not predict jumps in the  $\mathbf{V}^*(u)$  process at time  $u$  given past  $\mathbf{V}^*(u^-)$  history. This result is a special case of the following condition for lack of confounding due to Robins (1986). A sufficient condition for a time-varying covariate vector  $\mathbf{V}^\dagger(u)$  to be a nonconfounder in the presence of data on a second time-varying covariate vector  $\mathbf{V}^*(u)$  is that  $\mathbf{V}^\dagger(u)$  is not an independent predictor of both outcome (i.e., equation (7) holds) and subsequent  $\mathbf{V}^*(u)$  history (i.e., equation (8) holds). On the other hand,  $\mathbf{V}^\dagger(u)$  not being a significant predictor of the outcome is not, by itself, a sufficient condition for  $\mathbf{V}^\dagger(u)$  to be a nonconfounder in the presence of data on  $\mathbf{V}^*(u)$ .

Thus, before we could justify eliminating white blood count, CD4 count, and PCP history from the analysis, we needed to empirically check whether (8) is true. To do so, we fit three separate models. We first modeled the logit of the probability of change in  $Karn$  at each of the approximately eight weekly visits as a linear function of treatment arm and the values of the covariates in  $\mathbf{V}$  at the previous visit. When we did so, none of the components  $\mathbf{V}^\dagger$  were significant predictors at the  $p = .15$  level. We then modeled the logit of the probability of change in  $Ast$  at each visit as a linear function of the value of  $Karn$  at that visit and values of the covariates recorded in  $\mathbf{V}$  at the previous visit. Again, none of

the covariates in  $\mathbf{V}^\dagger$  significantly predicted current *Ast*. Finally, we specified a linear logistic model for the probability of change in *Hgb* at each visit, conditional on the values of *Karn* and *Ast* at that visit and the values of  $\mathbf{V}$  at the previous visit. Once again, the covariates  $\mathbf{V}^\dagger$  were not significant predictors. Thus, we concluded, based on this empirical analysis, that we could accept (8) as (approximately) true. Having accepted (7) as (approximately) true based on our fit of (5) and having assumed (1), we can conclude that (6) is (approximately) true, and we eliminate the variables *CD4*, *WBC*, and *PCP* recorded in  $\mathbf{V}^\dagger$  from further consideration. We note that alternative approaches to dimension reduction might be considered. We have focused on the approach reported here because it illustrates the condition for lack of confounding described above, which we believe to be relatively unfamiliar to the biometric community. We wish to emphasize that the choice of the significance level  $p = .15$  for variable inclusion or exclusion is somewhat ad hoc. As discussed in detail by Robins and Greenland (1986), a more formal, less ad hoc approach might be to elicit a subjective prior distribution for the joint distribution of all the unknown parameters in the model and carry out a full Bayesian analysis. However, such an approach lies outside the scope of the present paper.

## 2.2 Construction of Estimators

Given (6), we are now ready to construct our IPCW estimators based on the time-dependent prognostic factors  $\mathbf{V}^*(t)$ . The first step is to specify a model for the right-hand side of equation (6),

$$\begin{aligned} \lambda_C \{t \mid \bar{\mathbf{V}}^*(t), Z, T > t\} \\ = \lambda_{0Z}(t) \exp \{ \alpha_{Z1} \text{Karn}(t) + \alpha_{Z2} \text{Ast}(t) + \alpha_{Z3} \text{Hgb}(t) \} \\ \equiv \lambda_{0Z}(t) \exp \{ \alpha'_Z \mathbf{V}^*(t) \}, \end{aligned} \quad (9)$$

where  $\mathbf{V}^*(t) = \{\text{Karn}(t), \text{Ast}(t), \text{Hgb}(t)\}'$ . Since both the baseline hazards  $\lambda_{0Z}(t)$  and  $\alpha_Z$  may depend on treatment arm, model (9) constitutes separate treatment-arm-specific models for censoring and therefore was fit to data from the two arms separately. This flexibility is substantively important because the reasons for censoring are quite different in the two arms. For example, in those analyses in which treatment crossover was regarded as censoring, not only was the overall rate of censoring quite different in the two treatment arms but also, as we shall see in Section 3.3, the clinical status of censored patients differed in the two arms.

The IPCW KM estimator for failure is given in equation (10). As mentioned in the Introduction, the IPCW KM estimator differs from the ordinary KM estimator by weighting the contribution of a subject at risk at time  $t$  by the inverse of an estimate of the conditional probability of having remained uncensored until time  $t$ , based on the fit of model (9). Specifically, let  $\hat{\alpha}_Z$  be the Cox partial likelihood estimate of  $\alpha_Z$  of model (9) in treatment arm  $Z$ . Further, let  $X = \min(T, C)$ . Let  $Y(u) = I(X \geq u)$  be the at-risk indicator and let  $\tau = I(T = X)$  be the failure indicator that takes the value one if the subject is a failure and zero if the subject is censored. Now we say the data  $(\bar{\mathbf{V}}^*(T), Z, T)$  are coarsened at random (CAR) (Heitjan and Rubin, 1991; Robins and Rotnitzky, 1992) if  $\lambda_C \{t \mid \bar{\mathbf{V}}^*(T), Z, T, T > t\} = \lambda_C \{t \mid \bar{\mathbf{V}}^*(t),$

$Z, T > t\}$ . Note CAR differs from (6) in that  $\bar{\mathbf{V}}^*(T)$  replaces  $\mathbf{V}^*(t)$  on the left-hand side of the equality. Hence, CAR implies (6), but the converse is false. Under CAR and model (9), a consistent estimate of the conditional probability that subject  $i$  is uncensored through time  $t$  given  $(\bar{\mathbf{V}}^*(T), Z, T)$  is provided by the following time-dependent extension of the Kaplan–Meier product limit estimator for censoring:

$$\begin{aligned} \hat{K}_i^{\mathbf{V}^*}(t) \\ = \prod_{\{j: X_j < t, \tau_j = 0, Z_j = Z_i\}} [1 - \hat{\lambda}_{Z_i}(X_j) \exp \{ \hat{\alpha}'_{Z_i} \mathbf{V}_i^*(X_j) \}], \end{aligned}$$

where

$$\hat{\lambda}_Z(X_j) = \frac{(1 - \tau_j)}{\left\{ \sum_{i=1}^n \exp(\alpha'_Z \mathbf{V}_i^*(X_j)) Y_i(X_j) I(Z_i = Z) \right\}}$$

is the Cox estimator of the baseline hazard function for censoring  $\lambda_{0Z}$  in arm  $Z$  and, for any proposition  $B$ ,  $I(B) = 1$  if  $B$  is true and is zero otherwise. We write the estimate as  $\hat{K}_i^{\mathbf{V}^*}(t)$  to emphasize its dependence on  $\mathbf{V}_i^*(t)$ . Let  $\hat{K}_i^0(t)$  be the usual treatment-arm-specific Kaplan–Meier estimator of the probability of being uncensored by time  $t$  in treatment arm  $Z_i$ . Note that  $\hat{K}_i^{\mathbf{V}^*}$  and  $\hat{K}_i^0$  are equal whenever the estimator  $\hat{\alpha}_Z$  is the zero vector. Now define the subject-specific weight,

$$\hat{W}_i(t) = \hat{K}_i^0(t) / \hat{K}_i^{\mathbf{V}^*}(t),$$

so that  $\hat{W}_i(t)$  will converge to one for all  $t$  if and only if the history  $\mathbf{V}^*(t)$  of prognostic factors for failure does not predict the hazard of censoring at  $t$  given  $Z$  (i.e., within each treatment arm, there is no dependent censoring). In the presence of dependent censoring,  $\hat{W}_i(t)$  will not converge to one.

We are now ready to define the IPCW estimators of Robins (1993). The IPCW Kaplan–Meier estimator for failure in treatment arm  $z$ ,  $z \in \{0, 1\}$ , differs from the ordinary Kaplan–Meier estimator for failure only in that the contribution of a subject at risk at any time  $X_i$  is weighted by the subject-specific weight  $\hat{W}_i(X_i)$ . Thus, our IPCW KM estimate of the treatment-arm-specific marginal probability of remaining alive through time  $t$  is

$$\hat{S}_T(t \mid z) = \prod_{\{i: X_i < t\}} \frac{1 - \tau_i \hat{W}_i(X_i) I(Z_i = z)}{\left\{ \sum_{k=1}^n Y_k(X_i) \hat{W}_k(X_i) I(Z_k = z) \right\}}. \quad (10)$$

Since, in (10),  $\hat{K}^0(t)$  cancels from the numerator and denominator, we could have replaced  $\hat{W}(X_i)$  by  $1/\hat{K}^{\mathbf{V}^*}(X_i)$  in (10). The key to the consistency of  $\hat{S}_T(t \mid z)$  when (6) and (9) are true is the fact that  $\text{fail}(z, X_i) = I(Z_i = z) \tau_i / \hat{K}_i^{\mathbf{V}^*}(X_i)$  estimates the number of subjects in treatment arm  $z$  who would have been observed to fail at time  $X_i$  in the absence of any censoring and  $\text{risk}(z, X_i) = \sum_{k=1}^n Y_k(X_i) I(Z_k = z) / \hat{K}_k^{\mathbf{V}^*}(X_i)$  estimates the number of subjects who would have been alive and at risk at time  $X_i$  in the absence of any censoring. Thus, the ratio  $\text{fail}(z, X_i) / \text{risk}(z, X_i)$  estimates

the hazard of death at  $X_i$  in the absence of censoring; it follows that  $\hat{S}_T(t | z)$  estimates the probability  $S_T(t | z)$  of surviving without failure (i.e., of remaining alive) until time  $t$  in the absence of censoring.

The reason  $risk(z, X_i)$  consistently estimates the number of persons who would be at risk at time  $X_i$  in the absence of censoring under CAR and (9) is as follows. For each person  $k$  who is observed to survive uncensored to time  $X_i$  (i.e.,  $Y_k(X_i) = 1$ ) and who has an estimated conditional probability of, say,  $\hat{K}_k^{V^*}(X_i) = .25$  of having avoided censoring until time  $X_i$ , there would, on average, have been three other prognostically similar persons (i.e., ghosts with similar values of  $(\bar{V}^*(T), Z, T)$ ) who (i) were censored before time  $X_i$  and who (ii) would have, like subject  $k$ , survived to at least time  $X_i$  had censoring been prevented. We therefore assign person  $k$  a weight of  $4 = 1/.25$  in our estimate  $risk(z, X_i)$  of the number of subjects in treatment arm  $z$  who would have been at risk at  $X_i$  in the absence of censoring. Similarly, a subject in treatment arm  $z$  who fails at  $X_i$  with  $\hat{K}_i^{V^*}(X_i) = .25$  must ghost for three other similar subjects who would have had a similar failure time and similar prognostic factor history had censoring been prevented. Thus, in estimating the number of subjects who would have failed at time  $X_i$  in the absence of censoring, we assign subject  $i$ , the observed failure, a weight of  $4 = 1/.25$  in computing our estimate  $fail(z, X_i)$ . When (6) and (9) hold but CAR does not, the above argument is not strictly valid because  $\hat{K}_k^{V^*}(X_i)$  is no longer a consistent estimate of the probability that subject  $k$  remains uncensored through  $X_i$  given  $(V_k^*(T_k), Z_k, T_k)$ ; nonetheless, it can be shown using other more mathematical arguments that  $\hat{S}_T(t | z)$  remains consistent.

Suppose we wish to compare the marginal survival experience in the two treatment arms using the Cox proportional hazards model

$$\lambda_T(t | Z) = \lambda_0(t) e^{\beta Z}, \quad (11)$$

where  $Z = 1$  if a subject were randomized to the aerosolized pentamidine arm and  $Z = 0$  if a subject were randomized to the bacrim arm, so that  $\beta$  is positive if bacrim has a beneficial effect on survival compared to aerosolized pentamidine. Robins' (1993) IPCW Cox partial likelihood score  $U(\beta)$  for  $\beta$  differs from the ordinary Cox partial likelihood score only in that the contribution of the subject  $j$  at risk at time  $X_i$  is weighted by  $\hat{W}_j(X_i)$ , i.e.,

$$U(\beta) = \sum_i \tau_i \hat{W}_i(X_i) \times \left[ Z_i - \frac{\sum_{j=1}^n Y_j(X_i) \hat{W}_j(X_i) Z_j e^{\beta Z_j}}{\sum_{j=1}^n Y_j(X_i) \hat{W}_j(X_i) e^{\beta Z_j}} \right]. \quad (12)$$

If (6) and (9) are correct, Robins (1993) proves that  $\hat{S}_T(t | z)$  is consistent for  $S_T(t | z) = \text{pr}(T > t | Z = z)$  and the solution  $\hat{\beta}$  to the IPCW score equation  $0 = U(\beta)$  is a consistent asymptotically normal estimator of the parameter  $\beta$  of the Cox model (11).

## 2.3 Efficiency

If in (12) we had replaced  $\hat{W}(X_i)$  by  $1/\hat{K}^{V^*}(X_i)$ , the solution  $\hat{\beta}$  to  $0 = U(\beta)$  would still be consistent and asymptotically normal. However, Robins (1993) shows that using the weight  $\hat{W}(X_i)$  in (12) rather than  $1/\hat{K}^{V^*}(X_i)$  has important efficiency advantages. Indeed, by using  $\hat{W}(X_i)$  in (12), Robins (1993) shows that our IPCW estimator  $\hat{\beta}$  is guaranteed to be asymptotically more efficient than the usual Cox partial likelihood estimator of  $\beta$  whenever the partial likelihood estimator is consistent (i.e., whenever there is independent censoring). This is because our estimator  $\hat{\beta}$  is specifically designed to recover information about the parameter  $\beta$  of (11) from the data on the prognostic factors  $V^*(u)$ . Similarly, Robins (1993) shows that our IPCW estimate  $\hat{S}_T(t | z)$  is guaranteed to be asymptotically more efficient than the standard Kaplan-Meier estimator for failure in treatment arm  $z$  whenever the latter estimator is consistent (i.e., whenever censoring is independent). Robins (1993) discusses how to further modify  $U(\beta)$  and  $\hat{S}_T(t | z)$  to further increase efficiency and indeed to obtain doubly robust locally semiparametric efficient estimators of  $\beta$  and  $S_T(t | z)$ .

## 2.4 Confidence Intervals and Log-Rank Tests

Asymptotic 95% Wald confidence intervals for the cumulative hazard  $\Lambda(t | z) = -\ln S(t | z)$  in the one-sample problem and  $\beta$  in the two-sample problem are  $\hat{\Lambda}(t | z) \pm 1.96\{\hat{\Sigma}^*(t, z)\}^{1/2}$  and  $\hat{\beta} \pm 1.96\{\hat{I}^{-1}\hat{\Sigma}(\hat{\beta})\hat{I}^{-1}\}^{1/2}$ , where  $\hat{\Lambda}(t | z) = -\ln \hat{S}(t | z)$ ,  $\hat{I} = \partial U(\hat{\beta})/\partial \beta$ , and  $\hat{\Sigma}^*(t, z)$  and  $\hat{\Sigma}(\hat{\beta})$  are the estimators of the asymptotic variances of  $\hat{\Lambda}(t | z)$  and  $U(\beta)$  given in the Appendix. The IPCW log-rank test of independence between  $T$  and  $Z$  is the score test  $\chi(0) = U(0)/\{\hat{\Sigma}(0)\}^{1/2}$  of the hypothesis that  $\beta = 0$  in model (11).

## 3. Results of Analyses

### 3.1 Evidence for Dependent Censoring

Table 1 records the log-rank  $z$ -score (standard normal deviate) and 95% Wald confidence interval for the Cox model parameter  $\beta$  for our first analysis, in which the only censoring is by loss to follow-up and all 94 deaths are regarded as failures. We report four subsidiary analyses, 1a-1d. Analysis 1d is the analysis described in the previous subsection; i.e., we adjusted for dependent censoring due to Karnovsky score, asthenia, and hemoglobin by fitting model (9) as described above. Analysis 1a is a standard intent-to-treat analysis; i.e., the log-rank  $z$ -score and confidence interval for  $\hat{\beta}$  are those obtained using the standard log-rank test and Cox partial likelihood. Formally, they can be computed by fitting model (9) with the parameters  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  fixed at zero. Analyses 1b and 1c are intermediate between 1a and 1d. Specifically, in analysis 1b, we appropriately adjust for dependent censoring due to Karnovsky score but do not adjust for dependent censoring due to asthenia or hemoglobin. Formally, in analysis 1b, we fit model (9) with the parameters  $\alpha_2$  and  $\alpha_3$  for asthenia and hemoglobin fixed at zero. In analysis 1c, we adjust for dependent censoring due to Karnovsky score and asthenia but not due to hemoglobin by fitting model (9) with the parameter  $\alpha_3$  for hemoglobin fixed at zero. From Table 1, we see that there is little evidence of dependent censoring

**Table 1**  
Treatment comparison; death censored by loss to follow-up only; 94 deaths

Analysis	Covariates included	z-Value of IPCW log-rank test	IPCW partial likelihood estimate $\hat{\beta}$	IPCW 95% confidence interval for $\beta$	Length of IPCW 95% confidence intervals
1a	None	.78	.15	[-.24, .55]	.79
1b	Add Karnovsky	.89	.17	[-.21, .58]	.79
1c	Add asthenia	.92	.17	[-.20, .56]	.76
1d	Add Hgb < 8	.87	.16	[-.21, .54]	.75

**Table 2**  
Treatment comparison; death censored by minimum of treatment crossover and loss to follow-up only; 73 deaths

Analysis	Covariates included	z-Value of IPCW log-rank test	IPCW partial likelihood estimate $\hat{\beta}$	IPCW 95% confidence interval for $\beta$	Length of IPCW 95% confidence intervals
2a	None	1.33	.30	[-.15, .77]	.92
2b	Add Karnovsky	1.47	.32	[-.11, .78]	.89
2c	Add asthenia	1.50	.32	[-.10, .78]	.88
2d	Add Hgb < 8	1.56	.33	[-.09, .78]	.87

since our estimate of  $\hat{\beta}$  remains essentially unchanged as we successively adjust for potential dependent censoring due to Karnovsky score, asthenia, and hemoglobin. As mentioned above, asymptotic theory implies that, in the presence of independent censoring, the IPCW Wald intervals are narrower than the standard partial likelihood Wald intervals of analysis 1a. The results presented in the final column of Table 1 are consistent with the asymptotic theory.

Table 2 records the results of our second analysis, in which we censor subjects at the minimum of time to treatment crossover and loss to follow-up. From examining the  $\hat{\beta}$  column in Table 2, we find there is a hint of dependent censoring, with the partial likelihood estimator of  $\hat{\beta}$  in the analysis of 2a slightly biased toward the null. Again, there is an indication

of slight improvement in efficiency as we adjust for additional time-dependent prognostic factors in analyses 2b–2d.

Table 3 records the results of our third analysis, in which subjects are censored by the minimum of time to treatment crossover, voluntary cessation of all treatment, and loss to follow-up. In Table 3, we see slightly stronger evidence of dependent censoring, with the Cox partial likelihood estimator  $\hat{\beta}$  of analysis 3a biased toward the null.

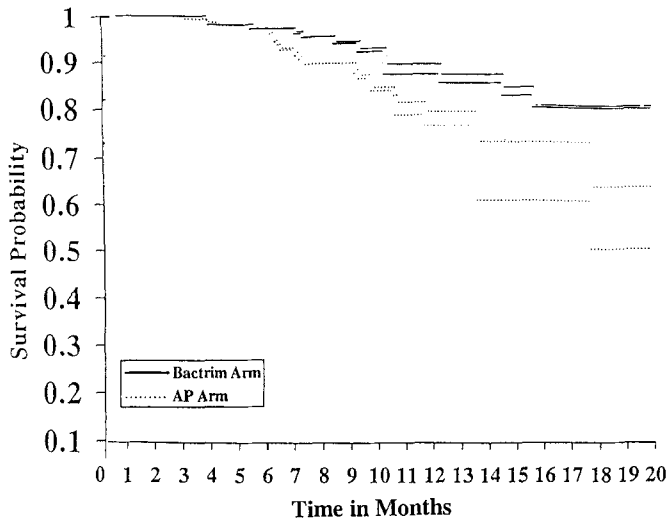
Table 4 records the results of our fourth analysis, in which we censor by the minimum of time to crossover, cessation of therapy for any reason, and loss to follow-up. In Table 4, there is fairly strong evidence of dependent censoring, with the ordinary Cox partial likelihood estimator biased toward the null.

**Table 3**  
Treatment comparison; death censored by minimum of crossover, voluntary stop, and loss to follow-up only; 52 deaths

Analysis	Covariates included	z-Value of IPCW log-rank test	IPCW partial likelihood estimate $\hat{\beta}$	IPCW 95% confidence interval for $\beta$	Length of IPCW 95% confidence intervals
3a	None	2.02	.56	[.02, 1.20]	1.18
3b	Add Karnovsky	2.24	.61	[.08, 1.21]	1.13
3c	Add asthenia	2.30	.62	[.09, 1.21]	1.12
3d	Add Hgb < 8	2.34	.62	[.10, 1.21]	1.11

**Table 4**  
Treatment comparison; death censored by minimum of crossover, stop for any reason, and loss to follow-up; 36 deaths

Analysis	Covariates included	z-Value of IPCW log-rank test	IPCW partial likelihood estimate $\hat{\beta}$	IPCW 95% confidence interval for $\beta$	Length of IPCW 95% confidence intervals
4a	None	1.70	.63	[-.10, 1.50]	1.60
4b	Add Karnovsky	2.30	.80	[.13, 1.55]	1.42
4c	Add asthenia	2.36	.78	[.14, 1.55]	1.41
4d	Add Hgb < 8	2.55	.84	[.21, 1.57]	1.36



**Figure 1.** In each arm, the upper curve is the ordinary KM curve and the lower curve is the IPW KM curve adjusted for Karnovsky score, asthenia, and hemoglobin. Censoring is the minimum of time to stopping therapy, crossover, or loss to follow-up.

### 3.2 Causal Interpretation of Results

Recall that the number of deaths regarded as failures decreases from 94 in analysis 1 to 36 in analysis 4. Thus, as expected, the length of the 95% Wald interval for the Cox parameter  $\hat{\beta}$  increases from .75 in model 1d to 1.78 in model 4d, tracking, in reverse, the decline in the number of failures. However, the point estimate  $\hat{\beta}$  of  $\beta$  increases from .16 in model 1d to .33 in model 2d to .62 in model 3d and, finally, to .84 in model 4d. Assuming, in each of our four analyses, the subsidiary analysis d appropriately corrects for any bias due to dependent censoring, the increasing values of  $\hat{\beta}$  as we go from analysis 1d to analysis 4d reflect the fact that the parameter  $\beta$  of the Cox model (11) being estimated represents a different causal parameter in the four different analyses. In analysis 4,  $\exp(\beta)$  represents the mortality hazard ratio in the aerosolized pentamidine arm compared to the bactrim arm had, contrary to fact, all subjects remained on their assigned treatment. In analysis 3, it represents this hazard ratio in a hypothetical study in which all subjects had remained on their assigned treatment unless discontinuation of treatment was medically indicated. In analysis 2,  $\exp(\beta)$  represents the hazard ratio that would be observed had, contrary to fact, treatment crossover been prohibited (although voluntary and involuntary, i.e., medically indicated, cessation of therapy are allowed). In analysis 1,  $\exp(\beta)$  represents the usual intent-to-treat hazard ratio comparing mortality rates by treatment assignment.

The fact that the IPCW log-rank  $z$ -score exceeds 1.96 and the IPCW 95% confidence interval for  $\beta$  excludes zero in analyses 3d and 4d implies that, given (6) and (9) are correct, we can conclude that there is strong evidence that bactrim therapy offers a survival advantage over aerosolized pentamidine therapy if (i) subjects remain on their assigned therapy until discontinuation of therapy becomes medically indicated (analysis 3) or (ii) remain on it always (analysis 4),

assuming that appropriate palliative therapy to ameliorate any toxicity were available.

### 3.3 Source of the Dependent Censoring

We have seen that, in analysis 4 reported in Table 4, there is strong evidence that the usual Cox partial likelihood estimate of .63 is biased downward due to dependent censoring, with an estimated bias of  $.21 = .84 - .63$ . The usual and IPCW Kaplan-Meier plots shown in Figure 1 clarify the source of this bias. Specifically, in the AP arm, we see that the usual KM estimator of the survival curve lies significantly above the IPCW KM estimator, while, in the bactrim arm, there is only a small difference between the usual KM estimator and the IPCW KM estimator. These curves suggest that, in the AP arm, subjects who are censored by crossing over or stopping treatment were subjects with poor prognosis (as indicated by their Karnovsky, asthenia, and hemoglobin measurements) compared with subjects who remained uncensored. In contrast, in the bactrim arm, it appears that the censored and uncensored subjects were roughly comparable in terms of the above prognostic factors. Such results are exactly what would be expected based on our understanding of the underlying biology. Bactrim is a highly effective drug in preventing PCP. Over 60% of the bactrim patients who stopped therapy or crossed over did so either because they had developed abdominal pain, elevated hepatic or pancreatic enzymes, or rash. None of these complications of bactrim therapy are important prognostic factors for survival. Thus, in the bactrim arm, one would expect, as is observed, little dependent censoring. In contrast, aerosolized pentamidine has little direct toxicity. People who cross over or stop therapy do so (i) because PCP has developed or (ii) because they become so ill that either they are unable to continue on AP therapy (which requires one be fit enough to use an inhalant machine) or their physician decides to cross them over to the other treatment. Thus, one would expect that subjects stopping AP therapy or switching from AP therapy are quite ill compared to subjects who continue on AP therapy, resulting in, as observed, a significant degree of dependent censoring in the AP arm.

## 4. Adjusting for Dependent Censoring Using the G-Computation Algorithm Estimator

In this section, we consider an alternative method for estimating the marginal survival curve  $S_T(t | z)$  based on the G-computation algorithm estimator of Robins (1987) and argue that it is inferior to that based on the IPCW KM estimator. Robins (1987) showed that, when (6) holds, the marginal survival curve  $S_T(t | z)$  in arm  $Z = z$  can be written as the sum (i.e., integral) over all possible prognostic factor histories  $\bar{\mathbf{V}}^*(t)$  of the joint probability of surviving to time  $T$  with prognostic factor history  $\bar{\mathbf{V}}^*(t)$  in the absence of censoring, i.e.,  $S_T(t | z)$  is given by

$$\begin{aligned} & \int \cdots \int \exp \left\{ - \int_0^t \lambda_T(u | C > u, \bar{\mathbf{v}}^*(u), Z = z) \right\} \\ & \times \prod_{m=0}^{\text{int}(t)} dF \{ \mathbf{v}^*(m) | \bar{\mathbf{v}}^*(m^-), Z = z, T > m, C > m \}, \end{aligned} \quad (13)$$

where  $\bar{\mathbf{v}}^*(t)$  is a realization of  $\bar{\mathbf{V}}^*(t)$ ,  $\text{int}(t)$  is the smallest integer less than  $t$ , and we have, in the notation, assumed that the  $\mathbf{V}^*(t)$  process jumps only at discrete times  $k = 0, 1, 2, \dots$  days from randomization. Robins (1986) refers to (13) as the G-computation algorithm formula for  $S_T(t | z)$ . The form of (13) suggests estimating  $S_T(t | z)$  by substituting estimates for the unknown quantities  $\lambda_T\{u | C > u, \bar{\mathbf{v}}^*(u), Z = z\}$  and  $f\{\mathbf{v}^*(m) | \bar{\mathbf{v}}^*(m^-), Z = z, T > m, C > m\}$  in (13) and then evaluating (13) by Monte Carlo integration. In carrying out this program, it would be natural to estimate  $\lambda_T\{u | C > u, \bar{\mathbf{v}}^*(u), Z = z\}$  using a time-dependent Cox model for the hazard of failure. The unknown density  $f\{\mathbf{v}^*(m) | \bar{\mathbf{v}}^*(m^-), Z = z, T > m, C > m\}$  could be estimated using linear logistic models similar to those used to test equation (8). Versions of the G-computation algorithm estimator have been used by Lagakos (1977), Finkelstein and Schoenfeld (1992), Murray and Tsiatis (1996), Gray (1992), and others with the intention of increasing efficiency by recovering information from surrogate marker data in the presence of independent censoring. However, as just noted, the G-computation algorithm estimator can also adjust for dependent censoring due to  $\mathbf{V}^*(u)$ .

The above Monte Carlo G-computation algorithm estimator of  $S_T(t | z)$  has efficiency advantages compared with the IPCW KM estimator of  $S_T(t | z)$  in the sense that, if the models for the conditional hazard of  $T$  and for jumps in the  $\mathbf{V}^*(m)$  process in (13) are correctly specified, the G-computation algorithm estimator will always be more efficient than any IPCW KM estimator (Robins and Rotnitzky, 1992). Nevertheless, we suggest that, in adjusting for dependent censoring due to time-dependent prognostic factors, the IPCW KM estimator be used rather than the Monte Carlo G-computation algorithm estimator because the IPCW KM estimator has several advantages over the G-computation algorithm estimator that outweigh the potential efficiency advantage of the latter. Specifically, we would not wish to recommend a method to correct for dependent censoring attributable to the prognostic factors  $\mathbf{V}^*(t)$  that can produce inconsistent estimates of the survival curve  $S_T(t | z)$  when, in fact, censoring is independent and thus ordinary KM estimator would be consistent. Now, under independent censoring with six true, the IPCW KM estimator is guaranteed to be consistent. In contrast, even under independent censoring, the Monte Carlo G-computation algorithm estimator will be consistent only if the Cox model for  $\lambda_T\{u | C > u, \bar{\mathbf{V}}^*(u), Z = z\}$  and the model for  $f\{\mathbf{V}^*(m) | \bar{\mathbf{V}}^*(m^-), Z = z, T > m, C > m\}$  are correctly specified. In high-dimensional problems, it is inevitable that these models will be misspecified. Finally, in contrast to the G-computation algorithm estimator, the IPCW KM estimator can, as we have seen above, be easily and naturally generalized to estimate the parameter  $\beta$  of the Cox regression model (11) comparing the effect of treatment arm on the marginal hazard of failure.

## 5. Discussion

Malani (1995) has also proposed a weighted version of the KM and Cox partial likelihood estimator. However, her weights are not proportional to the inverse of the conditional probability of surviving uncensored unless censoring is independent of failure (i.e., equation (2) holds). Thus, when our fundamental

assumption (6) of sequentially ignorable censoring is true but the assumption (2) of independent censoring is false, the Malani weighted estimator, in contrast to the IPCW weighted estimators discussed in this paper, will be inconsistent. In summary, the IPCW Kaplan-Meier and Cox partial likelihood estimators can be used to correct for bias due to noncompliance and dependent censoring when (most of) the noncompliance and dependent censoring can be explained by measured time-dependent prognostic factors and a (nearly) correct time-dependent Cox model for the hazard of censoring given these prognostic factors can be specified.

Finally Robins, Rotnitzky, and Scharfstein (1999) have developed sensitivity analysis methodology to investigate the sensitivity of our inferences to the fundamental assumption (6) of no unmeasured confounders. In a later report, we will apply these methods to the 021 data set.

## RÉSUMÉ

L'essai randomisé 021 ACTG a comparé l'effet du bactrim par rapport aux aérosols de pentamidine (AP) comme prophylaxie de la pneumocystose pulmonaire (PCP) chez des patients atteints du SIDA. Bien que les patients randomisés dans le bras bactrim aient connu un délai significatif à la survenue de la PCP, les survies dans les deux bras de l'essai n'étaient pas significativement différentes ( $p = .32$ ). Dans ce papier, nous montrons que la thérapie par bactrim améliore la survie, mais que la comparaison standard en "intention de traiter" échoue dans la détection de cette amélioration parce qu'une large proportion des sujets change de thérapie ou arrête complètement le traitement. Nous obtenons cet effet bénéfique du bactrim sur la survie en considérant artificiellement les sujets comme non indépendamment censurés à la première survenue soit de l'arrêt soit du changement de thérapie; nous avons alors analysé les données avec les estimateurs IPCW Kaplan-Meier, et par vraisemblance partielle de Cox proposés par Robins (1993) qui ajustent en cas de censures non indépendantes en utilisant les données recueillies sur des facteurs pronostiques dépendant du temps.

## REFERENCES

- Finkelstein, D. and Schoenfeld, D. A. (1994). Analyzing survival in the presence of an auxiliary variable. *Statistics in Medicine* **13**, 1747-1754.
- Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random: Characterizations, conjectures and counterexamples. In *Proceedings of the First Seattle Symposium on Survival Analysis*, 255-294.
- Gray, R. (1992). A kernel method for incorporating information on disease progression and the analysis of survival. *Biometrika* **81**, 527-539.
- Hardy, W. D., Feinberg, J., Finkelstein, D. M., et al. (1992). A controlled trial of trimethoprim-sulfamethoxazole or aerosolized pentamidine for secondary prophylaxis of *Pneumocystis carinii* pneumonia in patients with the acquired immunodeficiency syndrome: ACTG 021. *New England Journal of Medicine* **327**, 1842-1848.
- Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *The Annals of Statistics* **19**, 2244-2253.
- Jacobsen, N. and Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time. *The Annals of Statistics* **23**, 774-786.



- Lagakos, S. W. (1977). Using auxiliary variables for improved estimates of survival time. *Biometrics* **33**, 399–404.
- Malani, H. M. (1995). Modification of the redistribution to the right algorithm using disease markers. *Biometrika* **82**, 515–526.
- Murray, S. and Tsiatis, A. (1996). Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrics* **52**, 137–151.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512.
- Robins, J. M. (1987). Addendum to “A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect.” *Computers and Mathematics with Applications* **14**, 923–945.
- Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24–33. Alexandria, Virginia: American Statistical Association.
- Robins, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality*. Lecture Notes in Statistics (120), M. Berkane (ed), 69–117. New York: Springer-Verlag.
- Robins, J. M. and Greenland, S. (1986). The role of model selection in causal inference from non-experimental data. *American Journal of Epidemiology* **123**, 392–502.
- Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology—Methodological Issues*, N. Jewell, K. Dietz, and B. Farewell (eds), 297–331. Boston: Birkhäuser.
- Robins, J. M., Rotnitzky, A., and Scharfstein, D. O. (1991). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment and Clinical Trials*, M. E. Halloran and D. Berry (eds), 1–94. New York: Springer-Verlag.

Received December 1997. Revised January 2000.

Accepted January 2000.

## APPENDIX

### Calculation of Standard Errors

Robins (1993) showed that, under (6), (9), and (11), the asymptotic variances of  $n^{1/2}\{\hat{\Lambda}(t|z) - \Lambda(t|z)\}$  and  $n^{-1/2} \times U(\beta)$  are consistently estimated by  $n\hat{\Sigma}^*(t, z) = \hat{\Omega}_1^*(t, z) - \hat{\Omega}_2^*(t, z) - \hat{\Omega}_3^*(t, z)$  and  $n^{-1}\hat{\Sigma}(\hat{\beta}) = \hat{\Omega}_1(\hat{\beta}) - \hat{\Omega}_2(\hat{\beta}) - \hat{\Omega}_3(\hat{\beta})$ . Here

$$\hat{\Omega}_1^*(t, z) = n^{-1} \sum_i \{\hat{Q}_i^*(t, z, 0)\}^2,$$

$$\hat{Q}^*(t, z, x) = \int_x^t \frac{I(Z=z)d\hat{M}_T^*(u)\hat{W}(u)}{\sum_{j=1}^n Y_j(u)\hat{W}_j(u)I(Z_j=z)/n},$$

$\hat{W}(u)$  is as defined in the paragraph before equation (10), and

$$\hat{M}_T^*(u) = N_T(u) - \int_0^u d\hat{\Lambda}(x|z)Y(x),$$

$$N_T(u) = I[X \leq u, \tau = 1],$$

$$d\hat{\Lambda}(u|z) = \hat{\lambda}_0(u|z) = \frac{\sum_{j=1}^n dN_{Tj}(u)\hat{W}_j(u)I(Z_j=z)}{\sum_{j=1}^n Y_j(u)\hat{W}_j(u)I(Z_j=z)}.$$

Also,

$$\tilde{\Omega}_1(\beta) = n^{-1} \sum_i \{\hat{Q}_i(\beta, 0)\}^2,$$

$$\hat{Q}(\beta, x) = \int_x^\infty d\hat{M}_T(u, \beta)\hat{W}(u) \times \left[ Z - \frac{\sum_{j=1}^n Y_j(u)\hat{W}_j(u)Z_j e^{\beta Z_j}}{\sum_{j=1}^n Y_j(u)\hat{W}_j(u) e^{\beta Z_j}} \right],$$

$$\hat{M}_T(u, \beta) = N_T(u) - \int_0^u d\hat{\Lambda}_0(x, \beta)Y(x)e^{\beta Z},$$

$$d\hat{\Lambda}_0(x, \beta) = \hat{\lambda}_0(x, \beta) = \frac{\tilde{E}\{dN_T(x)\hat{W}(x)\}}{\tilde{E}\{\hat{W}(x)e^{\beta Z}Y(x)\}},$$

and

$$\tilde{E}(\mathbf{H}) \equiv n^{-1} \sum_{i=1}^n \mathbf{H}_i.$$

We shall need the following notation. Given model (9), for a random  $\mathbf{H}(u)$ , define

$$\tilde{\mathcal{L}}\{\mathbf{H}(u), \hat{\alpha}_z, z\} = \frac{\tilde{E}\{\mathbf{H}(u)Y(u)e^{\hat{\alpha}_z' \mathbf{V}^*(u)}I(Z=z)\}}{\tilde{E}\{Y(u)e^{\hat{\alpha}_z' \mathbf{V}^*(u)}I(Z=z)\}},$$

where  $\hat{\alpha}_z$  is the maximum partial likelihood estimator of  $\alpha_z$  in model (9). For any random vector functions  $\mathbf{H}_1(u), \mathbf{H}_2(u)$ , define  $\mathbf{H}_{1 \circ 2}(u) = \mathbf{H}_1(u)\mathbf{H}_2'(u)$  and set

$$\begin{aligned} & \tilde{\Phi}\{\mathbf{H}_1(u), \mathbf{H}_2(u)\} \\ &= \tilde{E} \left[ \int_0^\infty dN_C(u) \right. \\ & \quad \times \left\{ \tilde{\mathcal{L}}(\mathbf{H}_{1 \circ 2}(u), \hat{\alpha}_Z, Z) \right. \\ & \quad \left. \left. - \tilde{\mathcal{L}}(\mathbf{H}_1(u), \hat{\alpha}_Z, Z) \tilde{\mathcal{L}}(\mathbf{H}_2(u), \hat{\alpha}_Z, Z)' \right\} \right], \end{aligned}$$

$$\tilde{\Phi}_z \{ \mathbf{H}_1(u), \mathbf{H}_2(u) \} = I(Z = z),$$

where  $N_C(u) = I(X \leq u, \tau = 0)$ . Then

$$\tilde{\Omega}_3^*(t, z) = \tilde{E} \left[ \int_0^\infty dN_C(u) I(Z = z) \times \left\{ \tilde{L}(\hat{Q}^*(t, z, u), \hat{\alpha}_z, z) \right\}^2 \right],$$

$$\tilde{\Omega}_3(\beta) = \tilde{E} \left[ \int_0^\infty dN_C(u) \left\{ \tilde{L}(\hat{Q}(\beta, u), \hat{\alpha}_Z, Z) \right\}^2 \right],$$

$$\tilde{\Omega}_2^*(t, z) = \tilde{\Phi}_z(\hat{Q}^*(t, z, u), \mathbf{V}^*(u)) [\tilde{\Phi}_z \{ \mathbf{V}^*(u), \mathbf{V}^*(u) \}]^{-1} \times \tilde{\Phi}_z \{ \mathbf{V}^*(u), \hat{Q}^*(t, z, u) \},$$

and

$$\tilde{\Omega}_2(\beta) = \tilde{\Phi}(\hat{Q}(\beta, u), \mathbf{V}^*(u)) [\tilde{\Phi} \{ \mathbf{V}^*(u), \mathbf{V}^*(u) \}]^{-1} \times \tilde{\Phi} \{ \mathbf{V}^*(u), \hat{Q}(\beta, u) \}.$$

The above is a consequence of the following theorem of Robins (1993).

**THEOREM:** Let  $\beta_0$  be the true value of  $\beta$ . Then given (6), (9), and (11),  $n^{1/2}(\hat{\beta} - \beta_0)$  and  $n^{-1/2}U(\beta_0)$  are asymptotically normal with mean zero and variances  $I^{-1}\Sigma(I^{-1})'$  and  $\Sigma$ , where  $I = E\{n^{-1}\partial U(\beta_0)/\partial \beta\}$  and  $\Sigma = \Omega_1 - \Omega_2 - \Omega_3$ . Here  $\Omega_1 = E\{Q^2(\beta_0, 0)\}$ , where

$$Q(\beta, x) = \int_x^\infty dM_T(u, \beta) W(u) \times \left[ Z - \frac{E\{Y(u)W(u)Ze^{\beta Z}\}}{E\{Y(u)W(u)e^{\beta Z}\}} \right],$$

$W(u) = K^0(u)/K(u)$ , and  $K^0(u)$  and  $K(u)$  are the probability limits of  $\hat{K}^0(u)$  and  $\hat{K}^{\mathbf{V}^*}(u)$ ,

$$M_T(u, \beta) = N_T(u) - \int_0^u \lambda_0(t) Y(t) e^{\beta Z} dt.$$

Also

$$\Omega_2 = \Phi \{ Q(\beta_0, u), \mathbf{V}^*(u) \} [\Phi \{ \mathbf{V}^*(u), \mathbf{V}^*(u) \}]^{-1} \times \Phi \{ \mathbf{V}^*(u), Q(\beta_0, u) \}$$

and

$$\Omega_3 = E \left[ \int_0^\infty dN_C(u) [\mathcal{L} \{ Q(\beta_0, u), \alpha_Z, Z \}]^2 \right],$$

where

$$\begin{aligned} & \Phi \{ \mathbf{H}_1(u), \mathbf{H}_2(u) \} \\ &= E \left[ \int_0^\infty dN_C(u) \right. \end{aligned}$$

$$\begin{aligned} & \times [\mathcal{L} \{ \mathbf{H}_{1 \circ 2}(u), \alpha_Z, Z \} \\ & \quad \left. - \mathcal{L} \{ \mathbf{H}_1(u), \alpha_Z, Z \} \mathcal{L} \{ \mathbf{H}_2(u), \alpha_Z, Z \}] \right] \end{aligned}$$

and

$$\mathcal{L} \{ \mathbf{H}(u), \alpha_z, z \} = \frac{E \{ \mathbf{H}(u) Y(u) e^{\alpha_z' \mathbf{V}^*(u)} I(Z = z) \}}{E \{ Y(u) e^{\alpha_z' \mathbf{V}^*(u)} I(Z = z) \}}.$$

We will not reprove the entire theorem here but will only sketch the proof that  $\hat{\beta}$  is consistent for  $\beta_0$ . To do so, we require the following lemma.

**LEMMA A.1:** Given (6),

$$E\{Y(u)/K(u) \mid T, Z\} = I(T > u).$$

*Proof of lemma.* Since

$$K(u) = \exp \left[ - \int_0^u \lambda_C \{ x \mid Z, \bar{\mathbf{V}}^*(x), T > x \} dx \right],$$

straightforward calculation shows that

$$\begin{aligned} Y(u)/K(u) &\equiv I(T > u) I(C > u)/K(u) \\ &= I(T > u) \left\{ 1 - \int_0^u dM_C(x)/K(x) \right\}, \end{aligned}$$

where

$$dM_C(x) = dN_C(x) - \lambda_C \{ x \mid Z, \bar{\mathbf{V}}^*(x), T \} Y(x) dx.$$

This is fundamental identity (10) of Robins (1993). The lemma now follows from (6) and the mean zero property of stochastic integrals of martingales.

*Sketch of proof of consistency of  $\hat{\beta}$ .* First,  $n^{-1}U(\beta)$  is algebraically equal to  $n^{-1}\Sigma_i \hat{Q}_i(\beta, 0)$ , which clearly converges uniformly in probability to  $E[Q(\beta, 0)]$ . Under weak regularity conditions,  $\hat{\beta}$  will be consistent if we can show  $E[Q(\beta, 0)] = 0$  if and only if  $\beta = \beta_0$ . Now, by Lemma A.1,  $E\{Q(\beta, 0)\} = E\{Q_{\text{ful}}(\beta, 0)\}$ , where

$$\begin{aligned} Q_{\text{ful}}(\beta, x) &= \int_x^\infty dM_{\text{ful}, T}(u, \beta) K^0(u) \\ &\quad \times \left[ Z - \frac{E \{ K^0(u) I(T > u) Z e^{\beta Z} \}}{E \{ K^0(u) I(T > u) e^{\beta Z} \}} \right] \end{aligned}$$

and

$$M_{\text{ful}, T}(u, \beta) = I(T \leq u) - \int_0^u \lambda_0(t) I(T > t) e^{\beta Z} dt.$$

But under (11),  $E\{Q_{\text{ful}}(\beta, 0)\} = 0$  if and only if  $\beta = \beta_0$ .