HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2017 June 15.

Published in final edited form as:

Stat Med. 2016 June 15; 35(13): 2183–2194. doi:10.1002/sim.6852.

Evaluating predictors of competing risk outcomes when censoring depends on time-dependent covariates, with application to safety and efficacy of HIV treatment

Judith J. Lok* and **Michael D. Hughes**

Department of Biostatistics, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA

Abstract

We propose a prediction model for the cumulative incidence functions of competing risks, based on a logit link. Because of a concern about censoring potentially depending on time-varying covariates in our motivating HIV application, we describe an approach for estimating the parameters in the prediction models using inverse probability of censoring weighting under a missingness at random assumption. We then illustrate the application of this methodology to identify predictors of the competing outcomes of virologic failure (VF), an efficacy outcome, and treatment limiting adverse event (TLAE), a safety outcome, among HIV-infected patients first starting antiretroviral treatment.

Keywords

Competing risks; HIV/AIDS; Missing at random; Logit link; Personalized medicine; Safety and efficacy

1. Introduction

Competing risks occur when there is only one event time for each patient, but the event can be of several (competing) types. For example, for HIV-infected patients followed while receiving a specific treatment, time might be the time from initiation of that treatment to regimen failure, and the competing event types that reflect regimen failure might be virologic failure (VF), an efficacy outcome, or a treatment limiting adverse event (TLAE), a safety outcome mandating treatment discontinuation. These two event types are competing in that a TLAE leads to discontinuation of treatment and hence prevents observation of VF on that treatment, and vice versa. Because of the time-dependent nature of competing risks data, these data are often subject to censoring due to loss-to-follow-up, and to administrative censoring such as that arising at the end of a study. The analysis of competing risks data has been studied extensively (see e.g. [1] or [2]) for the situation in which censoring is noninformative (as is likely with administrative censoring), but not for the case where censoring may depend on time-varying covariates, which may arise when censoring is due to

*Correspondence to: Department of Biostatistics, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA. jlok@hsph.harvard.edu.

loss-to-follow-up. In this article, we consider the specific problem of evaluating predictors of the cumulative incidence of each competing event type at a given time when censoring may depend on time-dependent covariates.

Our motivating application concerns the analysis of data from a randomized clinical trial that evaluated the effects of different initial antiretroviral treatments for HIV-infected patients. Our aim is to evaluate what pre-treatment patient characteristics might predict each of the competing efficacy (VF) and safety (TLAE) outcomes, and whether the differences between randomized treatments for each of the competing outcomes might vary according to these characteristics. In doing so, we also allow for a third type of competing clinically significant event which we call “treatment limiting other event” (TLOE). This third event type includes other events which mandate discontinuation of the initial treatment such as a patient's need for a concomitant medication which might adversely interact with the initial antiretroviral treatment. However, some participants in the clinical trial discontinued their initial treatment for other reasons which, from a clinical perspective, do not mandate discontinuation, notably loss-to-follow-up from the trial. We are therefore interested in evaluating patient characteristics that predict the cumulative incidence of each of VF and TLAE in the counterfactual scenario “had no-one discontinued treatment for reasons other than VF, TLAE or TLOE”. To predict what might happen under this counterfactual scenario, one would want to censor a patient's follow-up when he or she deviated from this scenario, i.e. he or she discontinued the randomized treatment for reasons other than VF, TLAE or TLOE. Such censoring might be informative because it may depend on factors that are prognostic for the outcomes of interest.

This article takes informative censoring into account using inverse probability of censoring weighting (IPCW), see e.g. [3]. In order to apply IPCW, we assume that the unobserved outcomes due to censoring are missing at random (MAR), i.e. all factors that are prognostic for both censoring and the (counterfactual) outcome of interest are available for analysis. These factors include both baseline and time-dependent patient information. With IPCW under MAR, when a patient's follow-up is censored, his or her weight is distributed among “similar” patients, with “similar” based on a patient's covariate and treatment history measured before censoring (see e.g. [4]). In the absence of predictors of censoring, the Kaplan Meier estimator can be used to calculate the IPCW weights; then, IPCW reduces to Efron's redistribute-to-the-right algorithm [5], used to calculate the Kaplan Meier estimator [6].

IPCW for time-to-event outcomes under MAR has been studied by, e.g., [7]. We extend these ideas to the competing risk setting, assuming discrete-time censoring and event time processes. The focus on discrete time is motivated by our application, in which the event times are often determined by measurements made at pre-specified visit times, such as the HIV RNA measurements used to define VF and the safety laboratory tests that often define TLAE. If the intervals are small, it is also a good approximation of a continuous-time setting. The current paper therefore focuses on the use of IPCW in a discrete-time setting, evaluating pre-treatment predictors of the cumulative incidence function for each of several competing risks.

In this article we focus on evaluating predictors of the cumulative incidence function at one time point. In our application, this time point was at 3 years after treatment initiation, chosen because the trial followed many patients for this duration. The benefit of focusing on one time point is that we only need to specify a link function for the cumulative probabilities of the different types of events at one time point. This approach allows for effects of baseline predictors that are time-dependent, without explicitly modeling how these effects depend on time. This is in contrast to [8] and [9], who specify how the effect of baseline predictors varies over time. They then do extensive model checking to verify which effects depend on time. In particular, [9] assume that the effect of baseline covariates is constant over time on the (possibly transformed) absolute risk scale, whereas [8] assume censoring does not depend on time-varying covariates which predict the outcome of interest. The advantage of the methods used by [8] and [9] is that the conclusions obtained from these analyses are stronger, in the sense that they estimate the effects at each time point in one overall analysis; the advantage of our method is the weaker assumptions. If interest lies in how effects depend on time, a cause-specific hazards approach ([10]) could also be used. Other differences between our approach and [8] and [9] lie in the different choice of link function and in the fact that we allow for time-dependent predictors of both censoring and outcome.

Section 2 introduces the setting and notation. Section 3 describes IPCW for the prediction of competing risks outcomes under censoring that is MAR. Section 4 applies this method to the data from an HIV clinical trial to estimate how pre-treatment covariates predict efficacy and safety of the initial treatment regimens. A discussion concludes this paper in Section 5.

2. Setting and notation

For ease of reading, we define notation in the context of the above motivating HIV application, but the concepts are easily adapted to other competing risks settings. Similarly, we also suppress a subscript which labels individual patients in this article. Define T as the time until VF, TLAE or TLOE, whichever comes first. Let J indicate the type of event, with $J = 1$, $J = 2$ and $J = 3$ indicating VF, TLAE and TLOE, respectively. Let X be a vector of baseline (pre-treatment) covariates. We are interested in the counterfactual scenario in which no-one discontinues initial treatment for reasons other than VF, TLAE and TLOE.

We consider a discrete-time setting defined by $k = 1,\ldots, K$ time periods, with times $0 = \tau_1 < \tau_2 < \ldots < \tau_{K+1} = \tau$, such that τ_k and τ_{k+1} are the beginning and end of the kth period. We assume that visits for measuring covariates and outcomes such as VF are scheduled at each of the τ_k. Let L_k be the covariates at τ_k in the counterfactual scenario, with $\bar{L}_k$ their history. We define several indicator variables, taking the values 1 or 0 according to whether or not something occurred, as follows. Let $LTFU_k$ be an indicator of whether the person was lost to follow-up or discontinued treatment for reasons other than VF, TLAE or TLOE at or before the beginning of period k, so at or before τ_k. Let $ADMIN_k$ be an indicator of whether the person was administratively censored at or before the beginning of period k, so at or before τ_k. Finally, let C_k be an indicator of whether loss to follow-up or administrative censoring occurred at or before τ_k. Then, if $C_k = 0$, L_k and the outcome of period k, $(1_{T \le \tau_{k+1}}, J1_{T \le \tau_{k+1}})$, with 1_B being the indicator of whether event B happened, are observed. The full data, under the counterfactual scenario, is $(\bar{L}_K, T, J)$, which is observed until censoring. The

full data and the censoring information, in the order in which it could potentially be observed, is

$$(L_1, 1_{T\leq\tau_2}, J1_{T\leq\tau_2}, \mathrm{ADMIN}_2, \mathrm{LTFU}_2, L_2, 1_{T\leq}\tau_3, \ldots, \mathrm{ADMIN}_K, \mathrm{LTFU}_K, , 1_{T\leq\tau_{K+1}} J1_{T\leq\tau_{K+1}}).$$

After experiencing an event while on initial treatment, a patient can no longer be censored. The reason is that all the relevant data for such a patient has been observed. Thus, we only consider discontinuation of treatment for LTFU/ADMIN before the first event.

3. Evaluating predictors of competing risks outcomes under Missing At Random using Inverse Probability of Censoring Weights

The quantities of interest are the probability (cumulative incidence) of VF and the probability of TLAE between treatment initiation (baseline) and time τ since treatment initiation, had no-one discontinued their initial treatment for reasons other than VF, TLAE or TLOE. To predict these probabilities based on a vector of baseline characteristics, $X \in \mathbb{R}^m$, we estimate the cumulative incidence functions $P(T < \tau, J = j|X)$ for $j = 1, 2, 3$ using a logit link in a generalized linear model, similar to logistic regression:

$$P(T \leq \tau, J=j|X=x) = \frac{1}{1+e^{-(\beta_{0,j}+\beta_{1,j}x)}}; \quad (1)$$

with $\beta_{0,j} \in \mathbb{R}$ and $\beta_{1,j}^{\top} \in \mathbb{R}^m$. This is a type-specific logistic regression model. The method is easily generalizable to other link functions.

If (X, T, J) was observed for each patient, one could simply do maximum likelihood estimation. The maximum likelihood estimating equations based on the full data are

$$P_n \begin{pmatrix} 1 \\ X \end{pmatrix} \left(1_{\{T\leq\tau, J=j\}} - \frac{1}{1+e^{-(\beta_{0,j}+\beta_{1,j}X)}} \right) = 0, \quad (2)$$

where P_n denotes the empirical average over all patients. It is easy to see that these estimating equations are unbiased, by taking expectations and conditioning on X. In the absence of censoring, one could fit model (1) by solving the estimating equations (2) using any standard logistic regression software for each type of event separately. If censoring does not depend on patient characteristics, one could use Efron's redistribute-to-the right algorithm [5] to weight the observations and then fit a weighted logistic regression model.

When censoring depends on time-dependent patient characteristics, model (1) can be fitted using IPCW. IPCW has been described in, e.g., [3] for binary outcomes, and for multiple censoring mechanisms in, e.g., [11] and [12]. We review this for the current setting. In the following, write $Y_k = 1_{\{T\leq\tau_k\}}$, with $\bar{Y}_k$ its history until τ_k. Write

$$Q(\overline{Y}_{K+1}, J, X;\boldsymbol{\beta}) = \begin{pmatrix} 1 \\ X \end{pmatrix} \left(1_{\{T \le \tau, J=j\}} - \frac{1}{1+e^{-(\boldsymbol{\beta}_{0,j}+\boldsymbol{\beta}_{1,j}X)}} \right), \quad (3)$$

where we chose the notation Q to refer to the estimating equation. As is usual in the causal inference literature, we assume Consistency:

Assumption 3.1

(Consistency). For every k = 1,…, K, if $C_k = 0$ then $\overline{Y}_{K+1}$, $Y_{k+1}J$ and $\bar{L}_k$ are equal to their observed counterparts.

This states that as long as the patients follow the counterfactual scenario, their observed outcomes are the same as their outcomes under the counterfactual scenario. Unbiased estimating equations can then be generated based on the observed data by noting that

$$E\left(\frac{1_{C_k=0}}{P(C_k=0|\overline{L}_k, \overline{Y}_{K+1}, Y_{K+1}J)} Q(\overline{Y}_{K+1}, J, X;\boldsymbol{\beta}) \right) = 0. \quad (4)$$

Equation (4) can be proven by conditioning on $\bar{L}_K$, $\overline{Y}_{K+1}$, $Y_{K+1}J$:

$$\begin{aligned} & E\left(\frac{1_{C_K=0}}{P(C_K=0|\overline{L}_K, \overline{Y}_{K+1}J)} Q(\overline{Y}_{K+1}, J, X;\boldsymbol{\beta}) \right) \\ =& E\left(E\left[\frac{1_{C_K=0}}{P(C_K=0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J)} Q(\overline{Y}_{K+1}, J, X;\boldsymbol{\beta}) | \overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J \right] \right) \\ =& E\left(\frac{1}{P(C_K=0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J)} Q(\overline{Y}_{K+1}, J, X;\boldsymbol{\beta}) E[1_{C_K=0}|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J] \right) \\ =& E(Q(\overline{Y}_{K+1}, J, X;\boldsymbol{\beta})), \end{aligned}$$

which is zero at the truth. In order for this to work, one needs Positivity (e.g., [3] or [13]):

Assumption 3.2

(Positivity). P ($C_K = 0|\bar{L}_K, \overline{Y}_{K+1}, Y_{K+1}J$) > 0 for all possible values of ($\bar{L}_K, \overline{Y}_{K+1}, Y_{K+1}J$).

This assumption states that no matter what a patient's full data are, there is a positive probability of observing the full data on this patient. Without this assumption, one may be dividing by zero in equation (4). In the context of using IPCW, if a patient is censored, then his or her weight is re-distributed to similar patients who are uncensored; but for this to work, there have to be similar patients who are still uncensored.

The probability in equation (4) can be re-written as

$$P(C_K=0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J) = \prod_{k=1}^{K} P(C_K=0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J, C_{K-1}=0). \quad (5)$$

Now consider Missing At Random, see, e.g., [3] for the current situation:

Assumption 3.3

(Missing At Random (MAR)). *For every* $k = 2,\ldots,K$,

$$(\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J) \perp \text{ADMIN}_k | \overline{L}_{k-1}, \overline{Y}_k{=}0, C_{k-1}{=}0,$$

and

$$(\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J) \perp \text{LTFU}_k | \overline{L}_{k-1}, \overline{Y}_k{=}0, |_{k-1}{=}0, \text{ADMIN}_K{=}0.$$

Here, $\perp$ indicates conditional independence [14].

Assumption 3.3 says that loss to follow-up and administrative censoring can depend on the data collected so far, but not further on the prognosis of the patients under the counterfactual scenario. Under MAR, the IPCW probability in (4) and (5) can be re-written since censoring only depends on past observed values:

$$\begin{aligned}
&P(C_K{=}0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J) \\
&= \prod_{k=2}^{K} P(C_K{=}0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J, C_{K-1}{=}0) \\
&= \prod_{k=2}^{K} P(\text{ADMIN}_k{=}0, \text{LTFU}_K{=}0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J, C_{k-1}{=}0) \\
&= \prod_{k=2}^{K} P(\text{ADMIN}_k{=}0, |\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J, C_{k-1}{=}0) \cdot p(\text{LTFU}_k{=}0|\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J, C_{k-1}{=}0, \text{ADMIN}_k{=}0) \\
&= \prod_{k=2}^{K} P(\text{ADMIN}_k{=}0|\overline{L}_{K-1}, \overline{Y}_K, C_{k-1}{=}0) \cdot P(\text{LTFU}_K{=}0|\overline{L}_{K-1}, \overline{Y}_K, C_{k-1}{=}0, \text{ADMIN}_k{=}0).
\end{aligned} \tag{6}$$

Under Consistency Assumption 3.1, all quantities in this expression can be replaced by their observed counterparts. Note that if $\overline{Y}_k$ indicates that an event (VF, TLAE or TLOE) took place in (τ_{k-1}, τ_k], a patient can no longer be censored, and the product effectively runs until $k-1$.

In many cases, analysts assume that administrative censoring, or equivalently date of randomization in the study, is independent of $(\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1}J, \overline{\text{LTFU}}_{k-1})$. In that case, no covariates are needed in the model for administrative censoring, because

$$\begin{aligned}
&P(\text{ADMIN}_k{=}0|\overline{L}_{k-1}, \overline{Y}_k, C_{k-1}{=}0) \\
&= \frac{P(\text{ADMIN}_k{=}0 \text{and} \text{LTFU}_{k-1}{=}0|\overline{L}_{k-1}, \overline{Y}_k)}{P(\text{ADMIN}_{k-1}{=}0 \text{and} \text{LTFU}_{k-1}{=}0|\overline{L}_{k-1}, \overline{Y}_k)} \\
&= \frac{P(\text{ADMIN}_k{=}0|\text{LTFU}_{k-1}{=}0, \overline{L}_{k-1}, \overline{Y}_k)}{P(\text{ADMIN}_{k-1}{=}0|\text{LTFU}_{k-1}{=}0, \overline{L}_{k-1}, \overline{Y}_k)} \\
&= \frac{P(\text{ADMIN}_k{=}0)}{P(\text{ADMIN}_{k-1}{=}0)},
\end{aligned} \tag{7}$$

independent of ($\bar{L}_{k-1}$, $\overline{Y}_k$). One may however want to include some covariates in the prediction model for ADMIN to increase precision [15].

In our application, it is possible that simplification (7) does not necessarily hold. For example, it is possible that patients with characteristics which suggest that they might be harder to follow, may be under-represented in the patient population enrolling early, if clinical sites target enrollment first on patients who might be easier to follow. In that case, it may still be reasonable to assume that administrative censoring, or, equivalently, date of randomization in the study, depends on some selection of baseline covariates but not further on the prognosis of patients under the scenario of interest, or on their LTFU pattern. We could formalize this assumption as

$$(\overline{L}_K, \overline{Y}_{K+1}, Y_{K+1} J, \overline{\mathrm{LTFU}}_{k-1}) \perp \overline{\mathrm{ADMIN}}_k | V, \quad (8)$$

where V is a subset of the baseline covariates thought to predict date of enrollment. Under equation (8), since V is part of L_0,

$$\overline{\mathrm{ADMIN}}_k | V, \overline{\mathrm{LTFU}}_{k-1} \sim \overline{\mathrm{ADMIN}}_k | V, \overline{L}_K, \overline{Y}_{K+1}, Y_{K+1} J, \overline{\mathrm{LTFU}}_{k-1},$$

where $\sim$ indicates that the left hand side has the same distribution as the right hand side, so that the following simplification holds in equation (6):

$$P(\mathrm{ADMIN}_k = 0 | \overline{L}_{k-1}, \overline{Y}_k, C_{k-1} = 0) = P(\mathrm{ADMIN}_k = 0 | V, C_{k-1} = 0). \quad (9)$$

Thus, only covariates that potentially predict randomization date need to be included in the model for administrative censoring.

For confidence intervals and p-values, we employed a non-parametric bootstrap using Efron's percentile method (e.g., [16] page 327). All analyses were carried out using SAS version 9.2.

4. Application to predictors of VF and TLAE in HIV-infected patients

Study A5095 conducted by the AIDS Clinical Trials Group (ACTG) was a blinded randomized clinical trial comparing three combination drug regimens in previously untreated HIV-infected patients [17, 18]. One arm was discontinued early because of inferiority after review by a data and safety monitoring board [17]. We focus on the two remaining arms: zidovudine/lamivudine plus efavirenz (3-drug regimen) and zidovudine/lamivudine/abacavir plus efavirenz (4-drug regimen). A total of 765 patients were randomized to these regimens. We excluded seven patients who never initiated randomized treatment and were not followed. This resulted in 382 patients who received the 4-drug regimen and 376 patients who received the 3-drug regimen. In this study, the time of VF was defined as the time of the first of two successive HIV–1 RNA values of 200 or more copies per milliliter of plasma at least 16 weeks after randomization. TLAEs and TLOEs were assessed by the site investigators [17]. TLOEs included required discontinuation of study treatment because of the need for medications which could not be taken with study treatment, clinical events, pregnancy, or death.

We focused on the first 144 weeks of treatment, as relatively few patients were followed much longer at the time of study closure. By this time, 144 patients had experienced VF, 57 TLAE and 23 TLOE; 89 had been lost to follow-up on randomized treatment, and 138 were administratively censored prior to completing 144 weeks of treatment (because the study closed to follow-up).

We investigated how the probabilities of VF and TLAE during the first 144 weeks of treatment depend on pre-treatment (baseline) patient characteristics and the treatment to which a patient was randomized, in the situation in which the initial randomized treatment is taken as long as feasible; that is, until a VF, TLAE or TLOE occurs. Thus, interest lies in the counterfactual scenario in which no-one discontinues treatment for reasons other than a VF, TLAE or TLOE.

Univariate and multivariate models were evaluated. A specific focus of the analysis concerned interactions between patient characteristics and randomized treatment, because this would facilitate understanding as to whether there are patient groups for whom the risk (higher probability of TLAE) to benefit (lower probability of VF) profile might favor one treatment more than another. The method described in Section 3 was applied to allow for the possibility of informative censoring due to discontinuation of randomized treatment for reasons other than VF, TLAE or TLOE. Visits in ACTG A5095 were planned at weeks 2, 4, then every 4 weeks through to week 24, and every 8 weeks after week 24. For simplicity of the analysis, because the study's design incorporated evaluations at least every 8 weeks, we considered 8-week periods throughout follow-up, including the first 24 weeks. Thus, we only considered the values of time-dependent variables at the beginning of each of the 18 8-week periods that covered the 144 weeks of follow-up, so at weeks 0, 8, 16, 24, …, 136. Intermittent missing values were imputed using last observation carried forward.

To estimate the IPCW weights, we applied pooled (over the periods) logistic regression. A simple conditioning argument shows that pooled logistic regression leads to consistent estimates of the censoring probabilities if the logistic model is correctly specified. We based the model for LTFU on a literature review of variables that might predict LTFU in other studies of HIV-infected patients, see [19, 20, 21, 22, 23, 24, 25]. The variables included in the IPCW models for LTFU were randomized treatment, age (indicator for ≤30 years old), sex, injection drug use (ever versus never), race/ethnicity (black non-Hispanic, Hispanic, white non-Hispanic/other), baseline $\log_{10}$ viral load, and a time-dependent indicator of CD4 count ≤200. In addition, to allow the odds of LFTU to vary over time periods, we included indicator variables for“period 3”, “period 4”, “period 5-8”, “period 9-13”, and “period 14-18”, with period 2 representing the reference group for these variables (in period 1, only one patient had no visit after baseline and thus was LTFU, and we modeled that separately without including covariates). The grouping of periods in these variables was chosen to reflect the frequency of losses to follow-up over time. As the study was designed to follow all patients for a minimum of about two years, there was no administrative censoring in the first 14 periods, so we set the administrative censoring probabilities to 0 for periods 1-14. In period 15, there was only one case of administrative censoring, so we included no covariates in the administrative censoring model for period 15. We then used a pooled logistic regression model for ADMIN in periods 16 to 18, including the same predictor variables as

used in the model for LTFU except that indicator variables were included for "period 17" and "period 18", with period 16 being the reference period. Table 1 summarizes the results of the two pooled logistic regression models for remaining in follow-up, with end of follow-up due either to LTFU or to administrative censoring. Injection drug users were less likely to remain in follow-up ($OR = 0.41$, $p = 0.002$), and so were patients of Hispanic ethnicity versus white non-Hispanic/other race and ethnicity ($OR = 0.60$, $p = 0.07$). Hispanic patients were also less likely to remain in follow-up during periods 16 to 18, indicating that they tended to be administratively censored earlier. Using these models, the weights ranged from 1.0 to 2.9, and the sum of the weights was 759. Note that from the IPCW theory described in Section 3, the expected value of the sum of the weights divided by the baseline sample size equals one, and so the sum of the weights should be about the same as the sample size at time zero. In using the bootstrap method to obtain confidence intervals, we truncated the weights at 10 as suggested by [26], thus assuming that no patient characteristics lead to a probability of censoring that is greater than 9 out of 10. The need for such truncation was, however, rare: in the 5000 bootstrap samples, only three samples had a patient for whom the weight was truncated.

We investigated the following baseline covariates as predictors of VF and TLAE: randomized treatment, CD4 count (square root-transformed), viral load ($\log_{10}$-transformed), age (30 years or younger versus not, and as a continuous covariate), sex, injection drug use (ever versus never), and race (black non-Hispanic versus Hispanic versus white non-Hispanic/other). Table 2 describes the results of the univariate analyses for both VF and TLAE during the first 144 weeks of treatment. In addition, Table 2 shows results for the composite outcome of VF or TLAE or TLOE; this composite outcome is akin to a "regimen failure" type of outcome often analyzed in HIV studies. For randomized treatment, although not statistically significant, the odds ratios reflect what might be anticipated: lower odds of VF but higher odds of TLAE for the 4-drug versus 3-drug regimen (a pattern of effects which is lost when analyzing the composite outcome). Significant predictors of increased odds of VF at the 0.05 level were injection drug use and black non-Hispanic race/ethnicity; higher baseline viral load was marginally significant (p=0.10). For TLAE, no variables were significant at the 0.05 level though older age was marginally significantly associated with an increased odds of TLAE (p=0.06). Of note, injection drug use was significantly associated both with censoring due to LTFU and with VF, so indicative of potentially informative censoring and hence the potential need for use of methods such as those proposed in this paper (there may still be such a need even if the associations are not statistically significant). Note that results for the composite outcome VF/TLAE/TLOE tend to reflect averages of the associations found for VF and TLAE whereas the competing risks analysis provides more useful information about characteristics that might separately predict efficacy and safety outcomes.

We also fitted models including one baseline covariate, treatment, and a treatment-covariate interaction term, for each covariate that was significant at the 0.20 level in the univariate analysis for VF, TLAE, or the composite outcome. In this analysis of univariate interactions with treatment, we only found a significant interaction of treatment with injection drug use when considering the composite outcome (VF/TLAE/TLOE) ($p = 0.04$), and a marginally

significant interaction of treatment with injection drug use when considering VF ($p = 0.07$). All other interactions were non-significant at the 0.20 level.

Table 3 describes the multivariate analysis, including randomized treatment, all covariates which had a univariate p-value less than 0.20 for either the VF or TLAE event types or the composite outcome VF/TLAE/TLOE, and interaction terms with treatment if they were significant at the 0.20 level in the corresponding univariate analysis. The p-values and confidence intervals for injection drug use and the interaction of this covariate with treatment for TLAE may not be reliable, because the histograms of the bootstrap estimates for these parameters did not appear to be approximately normal; this is probably due to the fact that only 5 of the 82 patients who reported ever being an injection drug user experienced a TLAE. In the multivariate model for VF, the interaction of injection drug use and randomized treatment suggests that among patients who never injected drugs, adding a fourth drug to the antiviral regimen has little impact on the odds of VF ($OR = 0.94$, $p = 0.78$), whereas among those who reported ever injecting drugs, adding a fourth drug to the ART regimen significantly decreases the odds of VF ($OR = 0.29$, $p = 0.03$) (and the odds of the composite outcome, $OR = 0.29$, $p = 0.03$). As with any subgroup analysis, such analyses need to be interpreted cautiously and ideally replicated, but if real it might suggest a subpopulation in which the 4-drug regimen might have greater virologic benefit. Considering the other main effects in the models, analysis of the competing events rather than a composite outcome might help identify predictors which are important for specific outcomes, and might be relevant for patient management. For example, from Table 3, it appears that older patients might be at increased odds of TLAE but not of VF during the first 144 weeks of treatment, and black non-Hispanic patients and patients with higher pre-treatment viral load might have an increased odds of VF but not TLAE.

We conducted a number of sensitivity analyses to explore issues that are pertinent to the use of inverse probability of censoring methods. Our main approach was to develop a model for the censoring due to LTFU based on the form of associations described in other studies in the literature as use of model selection methods may affect coverage of confidence intervals. There is a balance in terms of bias versus variance tradeoff in evaluating the associations of interest between using more flexible versus more parsimonious functional forms for continuous variables in the censoring models ([15]). For the effect of time in the model for censoring due to loss to follow-up, we allowed such flexibility by including multiple indicator variables for groups of time periods. However, a simpler model involving fewer parameters for the time effect (allowing for an initial increased odds of LTFU and then a linear trend in log odds over time), there was very little change in either the parameter estimates or confidence intervals from those shown in Table 3. There was also very little change when we replaced the continuous log viral load variable by indicator variables for tertiles of viral load.

Another sensitivity analysis was motivated by the fact that including variables in the censoring model which predict censoring but not the outcome might decrease precision, whereas including variables that predict the outcome but not the censoring might increase precision [15]. We therefore re-fitted the univariate models (for those variables that had a p-value < 0.30 in the initial univariate analysis) and multivariate models for the competing

risks of interest using different nuisance parameter models for LFTU and ADMIN censoring. Specifically, we added into the censoring models, those variables that predicted VF or TLAE at the modest level of $p < 0.30$ in the original analysis. The results were very similar to those presented in Tables 2 and 3, so in this application there did not appear to be any notable change in precision in estimated associations.

5. Discussion

We fitted the probability of two types of competing events, VF and TLAE, and the composite outcome VF/TLAE/TLOE, using a logit link. These models could all be well-specified if the probability of TLOEs complements TLAEs and the VFs within the probability of events, resulting in a logit link for all three models. However, these models are probably not all correctly specified. If they are not correctly specified, the resulting estimates provide the best possible summary measure based on the logit link, minimizing the expected logistic loss function. For example, for the probability of an event, the model can be interpreted as the best possible prediction of the form $P(\text{event before time } \tau | X = x) = 1/(1 + e^{-(\beta_0+\beta_1 x)})$ based on the expected logistic loss. In the case of no censoring and correct model specification, minimizing the expected logistic loss leads to the maximum likelihood estimate. For the competing risks, our models also estimate the best possible prediction of this form based on the logistic loss function.

We conclude that IPCW can be adapted to the competing risk setting when censoring depends on time-dependent patient characteristics, to identify predictors of the different types of events.

6. Software

The programs, available upon request to jlok@hsph.harvard.edu, can handle different predictors of administrative censoring, loss to follow-up, and the outcomes (VF and TLAEs). The number of periods concerned can also be varied. The data are available upon request as well.

Acknowledgments

The authors are indebted to the participants who volunteered for ACTG A5095, and to the ACTG A5095 study team. The authors would also like to thank Christina Lalama for sharing her thoughts and code.

Contract/grant sponsor: This work was supported by the National Institutes of Health [grant numbers AI024643, AI68634]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. John Wiley & Sons, Inc; Hoboken, New Jersey: 2002.
2. Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, N. Statistical models based on counting processes. Springer series in statistics, Springer–Verlag; New York: 1993.
3. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association. 1995; 90(429):106–121.

4. Hernan MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. Pharmacoepidemiology and drug safety. 2005; 14:477–491. [PubMed: 15660442]
5. Efron, B. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 4. Prentice Hall; New York: 1967. The two sample problem with censored data; p. 831-853.
6. Malani HM. A Modification of the Redistribution to the Right Algorithm Using Disease Markers. Biometrika. 1995; 82(3):515–526.
7. Robins JM, Finkelstein D. Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics. 2000; 56(3):779–788. [PubMed: 10985216]
8. Scheike TH, Zhang MJ, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. Biometrika. 2008; 95(1):205–220.
9. Gerds TA, Scheike TH, Andersen PK. Absolute risk regression for competing risks: interpretation, link functions, and prediction. Statistics in Medicine. 2012; 31:3921–3930. [PubMed: 22865706]
10. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association. 1999; 94(446):496–509.
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 (Statistical Methodology). 2007; 69:307–327.
12. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. Statistics in Medicine. 2009; 28:1725–1738. [PubMed: 19347843]
13. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. American Journal of Epidemiology. 2011; 173:569–577. [PubMed: 21289029]
14. Dawid AP. Conditional independence in statistical theory (with discussion). Journal of the Royal Statistical Society. 1979; B41:1–31.
15. Rotnitzky A, Li L, Li X. A note on overadjustment in inverse probability weighted estimation. Biometrika. 2010; 97(4):997–1001. [PubMed: 22822256]
16. Van der Vaart, AW. Asymptotic statistics. Cambridge series in statistical and probabilistic mathematics, Cambridge University Press; Cambridge: 1998.
17. Gulick R, Ribaudo H, Shikuma C, Lustgarten S, Squires K, M W III, Acosta E, Schackman B, Pilcher C, Murphy R, et al. Triple-nucleoside regimens versus Efavirenz-containing regimens for the initial treatment of HIV-1 infection. New England Journal of Medicine. 2004; 350:1850–1861. [PubMed: 15115831]
18. Gulick R, Ribaudo H, Shikuma C, Lalama C, Schackman B, M W III, Acosta E, Schouten J, Squires K, Pilcher C, et al. Three- vs four-drug antiretroviral regimens for the initial treatment of HIV-1 infection. Journal of the American Medical Association. 2006; 296(7):769–781. [PubMed: 16905783]
19. Dudley J, Jin S, Hoover D, Metz S, Thackeray R, Chmiel J. The Multicenter AIDS Cohort Study: Retention after 9 1/2 years. American Journal of Epidemiology. 1995; 142(3):323–330. [PubMed: 7631636]
20. Ioannidis JPA, Bassett R, Hughes MD, Volberding PA, Sacks HS, Lau J. Predictors and impact of patients lost to follow-up in a long-term randomized trial of immediate versus deferred antiretroviral treatment. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology. 1997; 16(1):22–30. [PubMed: 9377121]
21. Arici C, Ripamonti D, Maggiolo F, Rizzi M, Finazzi MG, Pezzotti P, Suter F. Factors associated with the failure of HIV-positive persons to return for scheduled medical visits. HIV Clinical Trials. 2002; 3(1):52–57. [PubMed: 11819186]
22. Lanoy E, Mary-Krause M, Tattevin P, Dray-Spira R, Duvidier C, Fischer P, Obadia Y, Lert F, Costagliola D. the Clinical Epidemiology Group of the French Hospital Database on HIV infection. Predicators identified for losses to follow-up among HIV-seropositive patients. Journal of Clinical Epidemiology. 2006; 59:829–835. [PubMed: 16828676]

23. Andersen JW, Fass R, van der Horst C. Factors associated with early study discontinuation in AACTG studies, DACS 200. Contemporary Clinical Trials. 2007; 28:583–592. [PubMed: 17395549]

24. Krishnan S, Wu K, Smurzynski M, Bosch RJ, Benson CA, Collier AC, Klebert MK, Feinberg J, Koletar SL. for the ALLRT/A5001 team. Incidence Rate of and Factors Associated with Loss to Follow-up in a Longitudinal Cohort of Antiretroviral-Treated HIV-Infected Persons: An AIDS Clinical Trials Group (ACTG) Longitudinal Linked Randomized Trials (ALLRT) Analysis. HIV Clinical Trials. 2010; 12(4):190–200. [PubMed: 22044855]

25. Fleishman JA, Yehia BR, Moore RD, Korthuis PT, Gebo KA. for the HIV Reseach Network. Establishment, retention, and loss to follow-up in outpatient HIV care. Journal of Acquired Immune Deficiency Syndrom. 2012; 60(3):249–259.

26. Cole SR, Herna´n MA. Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology. 2008; 186(6):656–664. [PubMed: 18682488]

Table 1

Censoring models: Pooled (over the periods) logistic regression models for the log odds of remaining in follow-up (i.e. uncensored): model 1 is for censoring due to loss to follow-up; and model 2 is for censoring due to end of study follow-up (administrative censoring). Estimated odds ratios (95%-confidence intervals) for remaining in follow-up.

Covariate	Model for remaining in follow-up when censoring due to loss to follow-up	Model for remaining in follow-up when censoring due to end of study
ART: 4 drugs (vs 3 drugs)	0.75 (0.49,1.15), p=0.19	0.90 (0.63,1.29), p=0.56
sex: male (vs female)	1.16 (0.69,1.96), p=0.58	0.97 (0.61,1.55), p=0.90
age ≤30 (vs > 30years)	0.69 (0.43,1.10), p=0.12	0.99 (0.64,1.51), p=0.95
injection drug use: ever (vs never)	0.41 (0.23,0.72), p=0.002	0.91 (0.47,1.77), p=0.78
black non-Hispanic[1]	0.73 (0.44,1.21), p=0.22	1.04 (0.67,1.59), p=0.88
Hispanic[1]	0.60 (0.35,1.04), p=0.07	0.66 (0.42,1.03), p=0.07
time-dependent CD4 count ≤200	0.79 (0.46,1.36), p=0.39	1.07 (0.47,2.44), p=0.86
viral load (per $\log_{10}$ copies/ml higher)	1.23 (0.91,1.64), p=0.18	0.91 (0.71,1.17), p=0.46
period 3[2]	3.38 (1.46,7.86), p=0.005	–
period 4[2]	1.44 (0.75,2.75), p=0.27	–
period 5-8[2]	5.69 (2.93,11.03), p<0.0001	–
period 9-13[2]	5.18 (2.75,9.76), p<0.0001	–
period 14-18[2]	7.53 (3.46,16.43), p<0.0001	–
period 17[3]	–	1.03 (0.66,1.60), p=0.90
period 18[3]	–	0.75 (0.49.1.15), p=0.19

[1] versus white non-Hispanic/other.

[2] versus periods 1-2.

[3] versus period 16.

Table 2

Univariate analyses. Estimated odds ratios for events within 144 weeks of starting treatment (95%-confidence intervals).

Baseline covariate	VF[4]	TLAE[5]	Composite outcome (VF[4]/TLAE[5]/TLOE[6])
ART: 4 drugs (vs 3 drugs)	0.76 (0.52,1.08) p=0.12	1.41 (0.82,2.56), p=0.20	0.89 (0.64,1.22), p=0.48
sex: male (vs female)	1.02 (0.64,1.76) p=0.90	0.70 (0.38,1.55), p=0.33	0.81 (0.53,1.26), p=0.34
age (per 10 years older)	0.97 (0.79,1.18) p=0.81	1.37 (0.99,1.88), p=0.06	1.16 (0.97,1.42), p=0.09
age ≤30 (vs > 30 years)	1.04 (0.65,1.57) p=0.89	0.73 (0.31,1.37), p=0.34	0.85 (0.57,1.24), p=0.39
injection drug use: ever (vs never)	2.21 (1.23,3.75) p=0.01	0.90 (0.18,2.03), p=0.80	1.92 (1.10,3.24), p=0.02
black non-Hispanic[1]	2.17 (1.44,3.39) p<0.001	0.79 (0.39,1.47), p=0.46	1.53 (1.06,2.21), p=0.03
Hispanic[1]	1.13 (0.66,1.91) p=0.66	0.96 (0.43,1.87), p=0.89	1.01 (0.64,1.57), p=0.96
sqrt[2] CD4 count (per 3 units higher)[3]	0.96 (0.88,1.05) p=0.38	1.01 (0.89,1.16), p=0.83	0.99 (0.91,1.07), p=0.75
viral load (per $\log_{10}$ copies/ml higher)	1.25 (0.96,1.63) p=0.10	1.06 (0.74,1.54), p=0.76	1.17 (0.93,1.46), p=0.17

[1] versus white non-Hispanic/other, model with all race/ethnicities included.

[2] square root.

[3] For example, a CD4 count of 100 versus 49, 294 versus 200, or 643 versus 500 cells/mm^3.

[4] Virologic Failure.

[5] Treatment Limiting Adverse Event.

[6] Treatment Limiting Other Event. Number of bootstrap samples: 5000.

Stat Med. Author manuscript; available in PMC 2017 June 15.

Table 3

Multivariate analysis. Estimated odds ratios for events within 144 weeks of starting treatment (95%-confidence intervals).

Baseline covariate	VF[2]	TLAE[3]	events (VF[2]/TLAE[3]/TLOE[4])
ART: 4 drugs (vs 3 drugs)	0.94 (0.62,1.43), p=0.78	1.59 (0.88,3.05), p=0.12	1.07 (0.75,1.54), p=0.70
age (per 10 years older)	0.93 (0.74,1.16), p=0.53	1.36 (0.99,1.87), p=0.05	1.14 (0.95,1.39), p=0.16
black non-Hispanic[1]	2.11 (1.43,3.19), p<0.001	0.85 (0.43,1.49), p=0.56	1.55 (1.10,2.20), p=0.01
viral load (per $\log_{10}$ copies/ml higher)	1.29 (0.99,1.73), p=0.06	1.03 (0.70,1.52), p=0.89	1.17 (0.93,1.48), p=0.18
injection drug use: ever (vs never)	3.75 (1.64,9.44), p=0.003	1.64 (0.000,5.05)[5], p=0.55	3.53 (1.60,9.99), p=0.003
ART*injection drug use	0.32 (0.07,1.09), p=0.07	0.25 (0.000,7×10^{-5})[5], p=0.21	0.27 (0.06,0.87), p=0.03

[1] versus white non-Hispanic/other/Hispanic.

[2] Virologic Failure.

[3] Treatment Limiting Adverse Event.

[4] Treatment Limiting Other Event.

[5] Confidence interval may be unreliable because of small number of events (see text). Number of bootstrap samples: 20000.