Lifetime Data Anal (2018) 24:201–223
https://doi.org/10.1007/s10985-017-9393-4

Estimation of the cumulative incidence function under multiple dependent and independent censoring mechanisms

Judith J. Lok[1] · **Shu Yang**[2] · **Brian Sharkey**[3] · **Michael D. Hughes**[1]

Received: 16 October 2014 / Accepted: 14 February 2017 / Published online: 25 February 2017

Abstract Competing risks occur in a time-to-event analysis in which a patient can experience one of several types of events. Traditional methods for handling competing risks data presuppose one censoring process, which is assumed to be independent. In a controlled clinical trial, censoring can occur for several reasons: some independent, others dependent. We propose an estimator of the cumulative incidence function in the presence of both independent and dependent censoring mechanisms. We rely on semi-parametric theory to derive an augmented inverse probability of censoring weighted (AIPCW) estimator. We demonstrate the efficiency gained when using the AIPCW estimator compared to a non-augmented estimator via simulations. We then apply our method to evaluate the safety and efficacy of three anti-HIV regimens in a randomized trial conducted by the AIDS Clinical Trial Group, ACTG A5095.

Keywords Competing risks · Cumulative incidence function · Dependent censoring · Inverse probability weighting

✉ Judith J. Lok
jlok@hsph.harvard.edu

Shu Yang
syang24@ncsu.edu

Brian Sharkey
bri.sharkey@gmail.com

Michael D. Hughes
mhughes@sdac.harvard.edu

[1] Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA

[2] Department of Statistics, North Carolina State University, Raleigh, NC, USA

[3] Incyte, Wilmington, DE, USA

1 Introduction

Competing risks often arise in medical studies. In the competing risks setting, as opposed to the standard survival analysis setting, the failure event is classified into one of several mutually exclusive types, and occurrence of one type of event precludes the occurrence of an event of another type. For example, if interest is in death due to cardiovascular disease, a patient experiencing death due to cancer would be precluded from experiencing the event of interest. Standard statistical methods for the analysis of competing risks data are described in, for example, Andersen et al. (1993), Kalbfleisch and Prentice (1980) and Pintilie (2006).

We focus our attention on the cumulative incidence function (CIF), defined as the probability of a particular type of failure by time t, in an environment where other causes of failure may occur. There have been significant developments in statistical inference based on the CIF. Gray (1988) developed a class of tests for comparing CIFs of a particular type of failure among different groups. Lin (1997) constructed confidence bands for the CIF. Fine and Gray (1999) proposed a semi-parametric proportional hazards model for the subdistribution of a competing risk. Other work has focused on modeling the CIF directly, see for example Fine (2001), Bryant and Dignam (2004) and Jeong and Fine (2006).

Previous work has assumed that follow-up of patients is subject to only one censoring process, which is assumed to be independent. However, a patient's follow-up time may be censored for one of many reasons, some of which may be independent and some may be dependent. For example, so-called administrative censoring occurs when patients reach the end of a study, often inducing independent censoring (although, as noted in e.g. Lok and Hughes (2016), this censoring may be dependent if, for example, patients with characteristics which suggest that they might be harder to follow, are under-represented in the patient population enrolling early, or if sites with distinctly different patient populations start enrollment at different times). On the other hand, patients may prematurely drop out of a study prior to the study's planned end of follow-up, which may induce dependent censoring if the patients who dropped out are not representative of the entire sample (e.g. sicker patients drop out of the study with higher probability than healthier patients). Thus, dependent censoring may more accurately reflect situations that arise in clinical studies. If dependent censoring is present, use of methods which assume independent censoring can lead to biased estimates of parameters or functions of interest. Rotnitzky et al. (2007, 2009) estimated survival curves in the presence of dependent censoring. The purpose of this article is to adapt these methods to estimate the CIF for competing risk data in the presence of multiple censoring mechanisms, some of which may be dependent.

This paper is organized as follows. In Sect. 2 we introduce the AIDS Clinical Trial Group (ACTG) A5095 randomized trial, which motivated the methodological developments. In Sect. 3, we introduce our notation and data structure and in Sect. 4, our assumptions. We introduce our estimator of the cumulative incidence function in Sect. 5. In Sect. 6, a simulation study is conducted to evaluate the performance of our estimator in finite samples. In Sect. 7, we illustrate the application of our methods to an analysis of the A5095 study. We end with a discussion in Sect. 8.

2 The ACTG A5095 study: a motivating example

ACTG A5095 was a multicenter, randomized, double-blind, placebo-controlled clinical trial designed to compare the safety and efficacy of two 3-drug regimens versus a 4-drug regimen for initial treatment of HIV-1 infection (Gulick et al. 2004, 2006). One 3-drug regimen was discontinued early on the recommendation of the data and safety monitoring board. Our focus is therefore on the comparison of the remaining 3-drug regimen (zidovudine, lamivudine, and efavirenz) and the 4-drug regimen (zidovudine, lamivudine, abacavir, and efavirenz).

The primary efficacy outcome measure in A5095 was the time to virologic failure (VF), defined as the time to the first of two successive HIV-1 RNA levels of 200 copies/ml of plasma or greater at or 16 weeks of follow-up. This was analyzed using an intention-to-treat analysis ignoring the changes from the randomized regimens which occurred in a reasonable proportion of study participants, often due to treatment limiting adverse events (TLAEs), sometimes due to treatment limiting other events (TLOEs) such as pregnancy and death. Clinically, there is therefore also considerable interest in comparing regimens with respect to regimen failure, with the competing outcome types of VF, TLAE, and TLOE. These are competing risks in that discontinuation of treatment due to a TLAE or TLOE precludes follow-up for VF while on that randomized treatment. However, some participants discontinue randomized treatment prior to the planned administrative end of follow-up of the study for reasons other than VF, TLAE, or TLOE, and there is often concern that this censoring of follow-up on study treatment may be dependent (Dudley et al. 1995; Ioannidis et al. 1997; Arici et al. 2002; Lanoy et al. 2006; Andersen et al. 2007; Krishnan et al. 2010; Fleishman et al. 2012). For example, if patients who feel bad on treatment discontinue treatment and therefore leave the trial, censoring due to dropout might be dependent. Developing statistical methods that allow for dependent censoring is therefore important, particularly for checking the sensitivity of study conclusions to the handling of such discontinuations.

3 Notation and goal

We consider a study that has staggered entry and maximum follow-up time ν^*. Let T^* and C^* be non-negative time to event and time to censoring random variables, respectively. Let $J \in \{1, 2, \ldots j^*\}$ denote the type of failure and $R \in \{1, 2, \ldots r^*\}$ denote the reason for censoring. In order for our estimator to converge properly (specifically, to ensure that regularity condition (2), defined below, holds), we will need to discard data that were recorded after time $\nu = \nu^* - \epsilon$, where ϵ is a small positive number. We then define the event time as $T = \min(T^*, \nu)$ and the censoring time as $C = \min(C^*, \nu)$. We assume that we observe n independent and identically distributed copies of $O = (X, \Delta, \Delta J, (1 - \Delta)R, \bar{V}_X)$ where $X = \min(T, C)$, $\Delta = \mathbf{1}(T \leq C)$, and $\bar{V}_t = (V_s : s \leq t)$ where $\mathbf{1}(\cdot)$ is the indicator function taking value 1 if its argument is true and 0 otherwise, and V_s is the vector of covariates measured at time s. Note that when $X = T$ we observe a patient's full covariate history, $\bar{V}_T$. We assume that either the type of failure, J, or the reason for censoring, R, is observed, but not both.

Going forward, we include the type of failure or reason for censoring in $\bar{V}_X$. Our goal is to estimate the cumulative incidence function on the interval $[0, \nu)$, defined as

$$F_j(t) = P(T \leq t, J = j),$$

in the presence of multiple reasons for censoring, some (or all) of which may be dependent.

4 Assumptions

We shall assume that for $r = 1, ..., r^*$:

$$\lambda_{C,r}(t|\bar{V}_T, T, J, T > t) = \lambda_{C,r}(t|\bar{V}_t, T > t) \tag{1}$$

where $\lambda_{C,r}(t|\bar{V}_T, T, J, T > t) = \lim_{h\to 0} \frac{P(t\leq C<t+h, R=r|C\geq t, \bar{V}_T, T, J, T>t)}{h}$. In words, we assume that the hazard of censoring at time t for reason r, depends only on the measured variables up to time t and not on any future observed or unobserved variables, failure time, or failure type. When assuming (1), we make the non-identifiable assumption that data on all time-dependent and time-independent covariates that are predictive of both failure and censoring are available and included in $\bar{V}_t$. Equation (1) is equivalent to assuming that the data are coarsened at random (CAR) (Heitjan and Rubin 1991) or missing at random (MAR) (Rubin 1976). For ease of notation we will denote $\lambda_{C,r}(t|\bar{V}_t, T > t)$ by $\lambda_{C,r}(t|\bar{V}_t)$.

We impose the regularity condition that for some constant ξ,

$$\lambda_{C,r}(t|\bar{V}_T, T, J, T > t) < \xi \tag{2}$$

with probability 1, for t in the interval $[0, \nu)$. Condition (2) would be false if we took $\nu = \nu^*$ since, with probability 1, all patients who are uncensored just before the administrative end of study ν^* will be censored when the study ends (Rotnitzky et al. (2007)).

When $\bar{V}_t$ is high dimensional, we cannot estimate $\lambda_{C,r}(t|\bar{V}_t)$ non-parametrically due to the curse of dimensionality. Thus, we specify a model for $\lambda_{C,r}(t|\bar{V}_t)$. In this paper, we use Cox's proportional hazards model, of the form:

$$\lambda_{C,r}(t|\bar{V}_t) = \lambda_{0,r}(t) \exp[\gamma_r' w_r(t, \bar{V}_t)], \tag{3}$$

where $\lambda_{0,r}(t)$ is an unknown, non-negative function of t, $w_r(t, \bar{V}_t)$ is a specified function of t and $\bar{V}_t$, and γ_r is an unknown parameter vector.

5 Estimation

5.1 Inverse probability of censoring weighted (IPCW) estimator

In the absence of censoring, $F_j(t)$ could be estimated non-parametrically by solving

$$\sum_{i=1}^{n} \left\{ \mathbf{1}(T_i \leq t, J_i = j) - F_j(t) \right\} = 0.$$

Due to censoring we must modify this expression. The main idea underlying our estimator of $F_j(t)$ is that of "non-uniform-pseudo-redistribution" to the right via inverse probability weighting of uncensored patients (Robins et al. 1995; Rotnitzky et al. 2009). That is, when a patient is censored, our estimator redistributes his or her weight among "similar" remaining uncensored patients. Following Rotnitzky et al. (2007), we define the inverse weights, $\pi(t|\bar{V}_t; \Lambda_0)$, as follows:

$$\begin{aligned}
\pi(t|\bar{V}_t; \Lambda_0) &= \exp\left(- \int_0^t \lambda_C(u|\bar{V}_T, T, J, T > u) du \right) \\
&= \exp\left(- \int_0^t \sum_{r=1}^{r^*} \lambda_{0,r}(u) \exp[\gamma_r' w_r(u, \bar{V}_u)] du \right) \\
&= \prod_{r=1}^{r^*} \prod_{0 \leq u \leq t} [1 - \exp[\gamma_r' w_r(u, \bar{V}_u)] d\Lambda_{0,r}(u)],
\end{aligned}$$

with the cumulative baseline hazard, $\Lambda_{0,r}(t)$, defined as $\Lambda_{0,r}(t) = \int_0^t \lambda_{0,r}(s) ds$.

We can find an estimate of $F_j(t)$ as the solution to the following equation:

$$\sum_{i=1}^{n} \frac{\tilde{\Delta}_i}{\pi_i(\tilde{T}_i|\bar{V}_{i,\tilde{T}_i}; \Lambda_0)} \left\{ \mathbf{1}(T_i \leq t, J_i = j) - F_j(t) \right\} = 0 \tag{4}$$

where $\tilde{T}$ is the minimum time such that $\mathbf{1}(T \leq t, J = j)$ is observed, i.e. $\tilde{T} = \min(T, t)$, and $\tilde{\Delta} = \mathbf{1}(\tilde{T} < C)$. As shown in the "Appendix", Eq. (4) is an unbiased estimating equation for $F_j(t)$ since under CAR, $\Pr(\tilde{\Delta} = 1|V_{\tilde{T}}) = \pi(\tilde{T}|\bar{V}_{\tilde{T}}; \Lambda_0)$. Note that if regularity condition (2) were false, we would be dividing by 0; this is called a positivity violation, where some patients have probability 0 of remaining uncensored, and IPCW fails (Robins et al. 1995).

Note that using Δ and T instead of $\tilde{\Delta}$ and $\tilde{T}$ will result in a less efficient estimator of $F_j(t)$. Intuitively this makes sense: by using Δ and T, censored patients would contribute nothing to Eq. (4). However, for those patients who were censored after time t, we know the value of $\mathbf{1}(T \leq t, J = j)$. As a result, we can use this information to construct a more efficient estimator.

Estimation of the inverse weights $\pi(t|\bar{V}_t; \Lambda_0)$ first requires estimation of $\gamma = (\gamma_1, \gamma_2, ..., \gamma_{r^*})$ and $\Lambda_0(t) = (\Lambda_{0,1}(t), \Lambda_{0,2}(t), ..., \Lambda_{0,r^*}(t))$. Because of the missing

at random assumption we can estimate γ_r, which is the unknown parameter vector in Eq. (3), using standard software via a Cox proportional hazards model with time dependent covariates. To estimate γ_r, treat censoring due to reason r as a "failure" in the time dependent Cox proportional hazards model. All events and censoring due to causes other than r are treated as "censored" observations. This process is repeated to estimate all γ_r's.

Once we have an estimate of γ, $\hat{\gamma}$, we can estimate the cumulative baseline hazard, $\Lambda_0(t) = (\Lambda_{0,1}(t), ..., \Lambda_{0,r^*}(t))$, using Breslow's estimator (Andersen et al. 1993),

$$\hat{\Lambda}_{0,r}(t) = \int_0^t \frac{\sum_{i=1}^n dN_{C_i,r}(u)}{\sum_{i=1}^n \exp[\hat{\gamma}_r' w_r(u, \bar{V}_{i,u})] I(X_i \geq u)} \tag{5}$$

where $N_{C,r}(u) \equiv \mathbf{1}(C \leq u, R = r, C \leq T)$ is the counting process of observing censoring of type r. Next, $\pi(t|\bar{V}_t; \Lambda_0)$ can be estimated by

$$\hat{\pi}(t|\bar{V}_t; \hat{\Lambda}_0) = \prod_{r=1}^{r^*} \prod_{0 \leq s \leq t} [1 - \exp[\hat{\gamma}_r' w_r(s, \bar{V}_s)] d\hat{\Lambda}_{0,r}(s)]. \tag{6}$$

We can now find an estimate of $F_j(t)$ as the solution to

$$\sum_{i=1}^n \frac{\tilde{\Delta}_i}{\hat{\pi}_i(\tilde{T}_i|\bar{V}_{i,\tilde{T}_i}; \hat{\Lambda}_0)} \left\{\mathbf{1}(T_i \leq t, J_i = j) - F_j(t)\right\} = 0. \tag{7}$$

Denote the estimator solving Eq. (7) as $\hat{F}_j^{NA}(t)$. $\hat{F}_j^{NA}(t)$ is known as a (non-augmented) inverse probability of censoring weighted (IPCW) estimator. We can use the non-parametric bootstrap to estimate the variance of $\hat{F}_j^{NA}(t)$.

5.2 Augmented inverse probability of censoring weighted (AIPCW) estimator

We can improve the efficiency of $\hat{F}_j^{NA}(t)$ by introducing an augmentation term (Tsiatis 2006; Rotnitzky and Robins 2005). Consider the solution to the following equation

$$\sum_{i=1}^n \left\{ \frac{\tilde{\Delta}_i}{\pi_i(\tilde{T}_i|\bar{V}_{\tilde{T}_i}; \Lambda_0)} \left\{\mathbf{1}(T_i \leq t, J_i = j) - F_j(t)\right\} - A_i\{F_j(t), \gamma, b(\cdot)\} \right\} = 0 \tag{8}$$

where $A_i\{F_j(t), \gamma, b(\cdot)\}$ is the augmentation term and is defined as

$$A\{F_j(t), \gamma, b(\cdot)\} \equiv \sum_{r=1}^{r^*} \int \frac{b(u, \bar{V}_u)}{\pi(u\text{-}|\bar{V}_u; \Lambda_0(u))} dM_{C,r}(u) \tag{9}$$

where $b(u, \bar{V}_u)$ is a user specified, left-continuous function of u and $\bar{V}_u$, where $\pi\left(u\text{-}|\bar{V}_u; \Lambda_0(u)\right)$ indicates the left-continuous version of π, and where

$$M_{C,r}(u) = N_{C,r}(u) - \int_0^u 1(X \geq s) \exp\left\{\gamma_r' w_r(s, \bar{V}_s)\right\} d\Lambda_{0,r}(s). \tag{10}$$

The process $M_{C,r}(u)$ is a mean zero martingale with respect to the filtration $\mathcal{F}(u)$, where we define $\mathcal{F}(u)$ as the increasing sequence of sigma algebras generated by $\sigma\left\{\mathbf{1}(C \leq x), \bar{V}_x, 0 \leq x \leq u\right\}$. In the "Appendix", we show that Eq. (8) is an unbiased estimating equation for $F_j(t)$.

For efficiency reasons (to be discussed below), we choose $b(u, \bar{V}_u)$ as follows:

$$b(u, \bar{V}_u) = -\mathbf{E}\left[\left\{\mathbf{1}(T \leq t, J_i = j) - F_j(t)\right\} \middle| \bar{V}_{u\text{-}}, T \geq u\right]. \tag{11}$$

If we can consistently estimate $b(u, \bar{V}_u)$ as defined in Eq. (11), we can find an estimate of $F_j(t)$ as the solution to

$$\sum_{i=1}^{n} \left\{ \frac{\tilde{\Delta}_i}{\hat{\pi}_i(\tilde{T}_i|\bar{V}_{\tilde{T}_i}; \hat{\Lambda}_0)} \left\{\mathbf{1}(T_i \leq t, J_i = j) - F_j(t)\right\} - \hat{A}_i(F_j(t), \hat{b}(u, \bar{V}_u), \hat{\gamma}) \right\} = 0, \tag{12}$$

with

$$\hat{A}(F_j(t), \hat{b}(u, \bar{V}_u), \hat{\gamma}) = \sum_{r=1}^{r^*} \int_0^{\tilde{T}} \frac{-\hat{\mathbf{P}}[(T \leq t, J = j)|\bar{V}_{u\text{-}}, T \geq u] + F_j(t)}{\hat{\pi}\left(u\text{-}|\bar{V}_u; \hat{\Lambda}_0(u)\right)} d\hat{M}_{C,r}(u)$$

and

$$\hat{M}_{C,r}(u) = N_{C,r}(u) - \int_0^u \exp\left\{\hat{\gamma}_r' w_r(s, \bar{V}_s)\right\} d\hat{\Lambda}_{0,r}(s).$$

The estimator that solves Eq. (12) is denoted by $\hat{F}_j^A(t)$ and is an augmented inverse probability of censoring weighting (AIPCW) estimator of $F_j(t)$. Again, we can use the non-parametric bootstrap to estimate the variance of $\hat{F}_j^A(t)$.

If we can consistently estimate $b(u, \bar{V}_u)$ as defined in (11), then $\hat{F}_j^A(t)$ would be doubly robust (Rotnitzky and Robins 2005). That is, $\hat{F}_j^A(t)$ is consistent and asymptotically normal if the model for the censoring process, $\pi(t|\bar{V}_t; \Lambda_0)$, is correctly specified or the conditional model, $\mathbf{E}[\mathbf{1}(T \leq t, J = j)|\bar{V}_{u\text{-}}, T \geq u]$ is correctly specified. Also, if both the model for the censoring process and the conditional model are correctly specified then $\hat{F}_j^A(t)$ is locally semi-parametric efficient (Robins and Rotnitzky 1992; Tsiatis 2006). The function $b(u, \bar{V}_u)$ is not arbitrary and is chosen to equal $-\mathbf{E}[\{\mathbf{1}(T \leq t, J = j) - F_j(t)\}|\bar{V}_{u\text{-}}, T \geq u]$ in order to gain the greatest efficiency among estimators that solve an equation such as Eq. (8) (Rotnitzky et al. 2007).

In practice, estimating the conditional expectation $\mathbf{E}[\{\mathbf{1}(T \leq t, J = j) - F_j(t)\}|\bar{V}_{u\text{-}}, T \geq u]$ can be difficult, because the information considered in $\bar{V}_{u\text{-}}$ in Eq. (11) is time-dependent. Thus, in order to make the problem more tractable, one can instead consider estimating

$$\mathbf{E}[\{\mathbf{1}(T \leq t, J = j) - F_j(t)\}|\bar{V}_0, T \geq u], \tag{13}$$

where $\bar{V}_0$ are the baseline covariates. One way to estimate (13) is as follows:

$$\hat{\mathbf{P}}[(T \leq t, J = j)|\bar{V}_0, T \geq u] = \frac{\int_u^t \hat{S}(a\text{-}|\bar{V}_0) d\hat{\Lambda}_j(a|\bar{V}_0)}{\hat{S}(u\text{-}|\bar{V}_0)}, \tag{14}$$

where $\hat{\Lambda}_j(a|\bar{V}_0) = \hat{\Lambda}_{0,j}(a) \exp\left\{\hat{\beta}_j' f(\bar{V}_0)\right\}$ and $\hat{\beta}_j'$ is the estimated parameter vector obtained from fitting a Cox proportional hazards model to the overall time T in each treatment group:

$$\lambda_{T,j}(t|\bar{V}_0) = \lambda_{0,j}(t) \exp\{\beta_j' f(\bar{V}_0)\}, \tag{15}$$

where $\lambda_{T,j}(t|\bar{V}_0) = \lim_{h \to 0} \frac{P(t \leq T < t+h, J=j|\bar{V}_0, T>t)}{h}$, and $\lambda_{0,j}(t)$ is an unknown non-negative function of t. The cumulative baseline hazard, $\Lambda_{0,j}(a)$, can be estimated using Breslow's estimator, and $\hat{S}(u|\bar{V}_0) = \exp\{-\sum_j \hat{\Lambda}_j(u|\bar{V}_0)\}$.

The resulting estimator of $F_j(t)$ will not be doubly robust or locally semi-parametric efficient. However, as shown in the "Appendix", it is still a consistent and asymptotically normal semi-parametric estimator for $F_j(t)$, if the censoring models (3) are correctly specified.

We compared our augmented and non-augmented IPCW estimators with the standard estimator of the cumulative incidence function which assumes independent censoring (Andersen et al. 1993), the Aalen–Johansen estimator $\hat{F}_j^0(t)$:

$$\hat{F}_j^0(t) = \sum_{i|t_{ji}<t} d_{ji} n_{ji}^{-1} \hat{S}(t_{ji}),$$

where d_{ji} is the number of people with failure type j at time $t_{ji}, t_{j1} < t_{j2} < \cdots < t_{jk_j}$ are the failure times for failures of type j, n_{ji} is the number of people at risk at time t_{ji}, and $\hat{S}(t)$ is the Kaplan–Meier estimator of the overall survival function (i.e. for all failure types combined). We calculated the standard error for $\hat{F}_j^0(t)$ using the delta method (Pintilie 2006).

6 Simulation study

We conducted a simulation study, which is a modification of the simulation study in Rotnitzky et al. (2007), in order to evaluate the performance of our estimators in finite samples.

We generated T^* according to an exponential distribution with mean equal to 1.25. We assumed there were 2 event types, with probabilities equal to 0.35 for event type

1, and 0.65 for event type 2. Here, the type of failure is independent of T^*. We then generated two covariates, one time-independent (V_{TI}) and one time-dependent (V_{TD}). The time-independent covariate was generated from a Bernoulli distribution with mean equal to 0.55. The time dependent covariate was a 1×3 row vector generated from a multivariate normal distribution with mean equal to $(T^*, T^*, T^*)'$ (so as to create a dependence between V_{TD} and T^*) and covariance equal to $0.7^{|i-j|}$, where $i, j = 1, 2, 3$. This vector represents the values of $V_{TD}(t)$ at times $t_1 = 0$, $t_2 = 0.5$, and $t_3 = 1$. $V_{TD}(t)$ at $t = 0$ represents a baseline measurement. We assumed that the time-dependent variable remains constant between measurements.

We assumed the maximum follow up time was $\nu^* = 1.35$. We generated the censoring times for the independent censoring process, C_2^*, according to a uniform(.55,1.35) distribution, to represent an administrative censoring process. We chose a uniform(.55,1.35) distribution as opposed to a uniform(0,1.35) distribution because many HIV clinical trials are designed to follow all patients for a pre-specified duration after the last patient is enrolled.

Next, we generated the censoring times for the dependent censoring process, C_1^*, according to the following hazard rate: $\lambda_c(t|\bar{V}_t) = \lambda_0(t)\exp[\gamma' w(t, \bar{V}_t)]$, where $\gamma' w(t, \bar{V}_t) = \gamma_1 V_{TI} + \gamma_2 V_{TD}(t)$ with $\lambda_0(t) = 1.5$, $\gamma_1 = 0.15$ and $\gamma_2 = 0.8$. Generating the censoring times according to the time-dependent model was done sequentially. This was done because the hazard of censoring in the time interval $t_1 = 0$ to $t_2 = 0.5$ differs from the hazard of censoring in the next interval. The algorithm to construct C_1^* for each simulated patient is as follows:

- Generate a censoring time, C_1, compatible with the hazard function for the first time interval, $[t_1, t_2)$, (where $t_1 = 0$) using the method of Bender et al. (2005). Note that this step is simply generating a censoring time that is compatible with a Cox proportional hazards model with time-independent covariates and constant baseline hazard.
- If C_1 is contained within the first time interval $[t_1, t_2)$, then set $C_1^* = C_1$.
- If C_1 is not contained within the interval $[t_1, t_2)$, generate a censoring time, C_2, compatible with the model for the second time interval, $[t_2, t_3)$.
- If C_2 is contained within the interval $[0, t_3 - t_2)$, then set $C_1^* = C_2 + t_2$.
- If C_2 is not contained within the time interval $[0, t_3 - t_2)$, then repeat the previous two steps for the last time interval $[t_3, \infty)$.

Finally, C^* was defined as $\min(C_1^*, C_2^*)$.

We repeated the simulations with the same setting, but with $\lambda_0(t) = 2.5$ and $\lambda_0(t) = 0.04$ instead of $\lambda_0(t) = 1.5$, so as to introduce varying levels of dependent censoring. Furthermore, in order to evaluate the performance of the augmented and non-augmented IPCW estimators when the Aalen–Johansen estimator is consistent, that is, when censoring is independent, we also simulated a scenario with only independent censoring. In this scenario, the distribution of the outcomes was the same as before, but C_1^* was uniform(0,1.35) and C_2^* was uniform(0.55,1.35).

Practically, in order to ensure the regularity condition that $\lambda_{C,r}(t|\bar{V}_T, T, J, T > t) < \xi$, we can treat the last observation (or last x observations) in each dataset as a failure (Robins and Rotnitzky 1992) and set $T = T^*$ and $C = C^*$. Alternatively, we could have chosen an arbitrary ϵ, set $\nu = \nu^* - \epsilon$, $T = \min(T^*, \nu)$ and $C =$

$\min(C^*, v)$. Both methods would ensure that $\lambda_{C,r}(t|\bar{V}_T, T, J, T > t) < \xi$ with probability 1. Here, we treated the last 5 observations as failures; this was chosen ad hoc. We also examined treating only the last observation as a failure, and then taking the last 10 observations as failures; our results were not sensitive to this condition. We generated 1000 datasets with 250 patients each. We estimated $F_j(t)$ at 8 time points: 0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95, 1.1.

In the first simulation scenario, the average censoring rate for the 1000 simulations was 55%, and the average dependent censoring rate was 33%. The results are presented in Table 1A. In the second simulation scenario, the average censoring rate for the 1000 simulations was 58%, and the average dependent censoring rate was 43%. The results are presented in Table 1B. In the third simulation scenario, the average censoring rate for the 1000 simulations was 50%, and the average dependent censoring rate was 15%. The results are presented in Table 2A. In these three simulation scenarios, the augmented IPCW estimator had bias very close to zero for each failure type. As expected in these scenarios, the Aalen–Johansen estimator had substantially larger bias and mean squared error. The non-augmented IPCW estimator had reduced bias compared with the Aalen–Johansen estimator but still appeared to show some bias, particularly at later follow-up times and the highest percentage of dependent censoring. At earlier follow-up times and lower percentage of dependent censoring, the bias of both IPCW estimators did not substantially contribute to the mean squared error. Also, for most time points the augmented IPCW estimator had smaller mean squared error and hence was more efficient than the IPCW estimator, even though (11) was misspecified here. The efficiency gains increased over time. The gain obtained by augmenting the IPCW estimator was larger in the scenario with more dependent censoring.

Table 2B displays the results for the scenario where censoring was independent. In this simulation scenario, the average censoring rate for the 1000 simulations was 48%. In this scenario, all three estimators are consistent. As can be seen in Table 2B, IPCW and augmented IPCW hardly inflated the mean squared errors. This indicates that adjusting for dependent censoring can be done without paying a price in the form of a substantial increase in precision.

7 Analysis of competing risks in ACTG A5095

Our event of interest is failure of the initial treatment regimen and can be classified as one of three types: (1) virologic failure (VF), (2) discontinuation of initial treatment due to treatment limiting adverse event (TLAE), or (3) discontinuation of initial treatment due to treatment limiting other event (TLOE). TLOEs included required discontinuation of study treatment because of the need for medications which could not be taken with study treatment, clinical events, pregnancy, and death. In addition to administrative censoring, arising when the study closes to follow-up, patients may discontinue randomized treatment for reasons other than VF, TLAE, or TLOE (for example, loss of follow-up). Supposing that discontinuing treatment for other reasons could in principle be avoided, our aim is to describe what might happen in the setting where treatment is only discontinued because of VF, TLAE, or TLOE. Therefore,

Table 1 Simulation results under dependent censoring: scenarios 1 and 2, with more dependent censoring

Estimator	Time (t)	0.05	0.20	0.35	0.50	0.65	0.80	0.95	1.10
	$F_1(t)$ (truth)	0.014	0.052	0.085	0.114	0.142	0.166	0.186	0.204
	$F_2(t)$ (truth)	0.025	0.096	0.159	0.215	0.264	0.306	0.346	0.380
A: *33% dependent censoring*									
$\hat{F}_1^A(t)$	Bias	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.002
	SD	0.007	0.015	0.018	0.022	0.024	0.027	0.031	0.037
	rMSE	0.007	0.015	0.018	0.022	0.024	0.027	0.031	0.037
	% Dec. in rMSE	0.48	2.31	5.44	8.52	14.54	17.55	22.51	26.60
$\hat{F}_1^{NA}(t)$	Bias	0.001	0.002	0.003	0.004	0.005	0.008	0.012	0.016
	SD	0.007	0.015	0.019	0.024	0.027	0.031	0.036	0.044
	rMSE	0.007	0.015	0.019	0.024	0.027	0.032	0.038	0.047
$\hat{F}_1^{AJ}(t)$	Bias	0.001	0.004	0.009	0.015	0.020	0.026	0.031	0.034
	SD	0.008	0.015	0.020	0.024	0.026	0.029	0.033	0.037
	rMSE	0.008	0.016	0.022	0.028	0.033	0.039	0.045	0.050
$\hat{F}_2^A(t)$	Bias	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.000
	SD	0.010	0.018	0.023	0.027	0.030	0.035	0.042	0.049
	rMSE	0.010	0.018	0.023	0.027	0.030	0.035	0.042	0.049
	% Dec. in rMSE	0.68	2.77	7.04	12.71	23.58	37.06	36.82	47.35
$\hat{F}_2^{NA}(t)$	Bias	0.000	0.002	0.003	0.006	0.009	0.014	0.019	0.027
	SD	0.010	0.018	0.025	0.029	0.036	0.045	0.054	0.067
	rMSE	0.010	0.018	0.025	0.030	0.037	0.047	0.057	0.072
$\hat{F}_2^{AJ}(t)$	Bias	0.001	0.007	0.015	0.026	0.036	0.047	0.055	0.061
	SD	0.010	0.019	0.025	0.029	0.033	0.037	0.042	0.047
	rMSE	0.010	0.020	0.029	0.039	0.049	0.060	0.069	0.077

Table 1 continued

Estimator	Time (t)	0.05	0.20	0.35	0.50	0.65	0.80	0.95	1.10
	$F_1(t)$ (truth)	0.014	0.052	0.085	0.114	0.142	0.166	0.186	0.204
	$F_2(t)$ (truth)	0.025	0.096	0.159	0.215	0.264	0.306	0.346	0.380
B: *43% dependent censoring*									
$\hat{F}_1^A(t)$	Bias	0.000	0.001	0.001	0.001	0.001	0.001	0.002	0.003
	SD	0.007	0.015	0.019	0.022	0.025	0.029	0.034	0.045
	rMSE	0.007	0.015	0.019	0.022	0.025	0.029	0.034	0.045
	% Dec. in rMSE	0.64	3.50	8.52	12.67	16.91	22.22	26.21	27.80
$\hat{F}_1^{NA}(t)$	Bias	0.001	0.002	0.003	0.006	0.008	0.012	0.017	0.025
	SD	0.008	0.015	0.020	0.025	0.029	0.036	0.044	0.057
	rMSE	0.008	0.015	0.020	0.026	0.030	0.038	0.047	0.062
$\hat{F}_1^{AJ}(t)$	Bias	0.001	0.006	0.012	0.020	0.027	0.034	0.040	0.044
	SD	0.008	0.016	0.022	0.026	0.028	0.032	0.036	0.041
	rMSE	0.008	0.017	0.025	0.033	0.039	0.047	0.054	0.060
$\hat{F}_2^A(t)$	Bias	0.000	0.001	0.000	0.001	0.001	0.001	0.002	0.003
	SD	0.010	0.018	0.024	0.028	0.032	0.038	0.036	0.045
	rMSE	0.010	0.018	0.024	0.028	0.032	0.038	0.036	0.045
	% Dec. in rMSE	0.68	5.78	12.11	20.42	38.02	51.02	52.63	53.60
$\hat{F}_2^{NA}(t)$	Bias	0.000	0.003	0.004	0.010	0.014	0.021	0.031	0.044
	SD	0.010	0.019	0.027	0.032	0.043	0.055	0.069	0.086
	rMSE	0.010	0.019	0.027	0.034	0.045	0.059	0.076	0.097
$\hat{F}_2^{AJ}(t)$	Bias	0.001	0.010	0.021	0.036	0.050	0.063	0.072	0.078
	SD	0.010	0.020	0.027	0.032	0.035	0.041	0.047	0.051
	rMSE	0.010	0.022	0.034	0.048	0.061	0.075	0.086	0.093

rMSE is the square root mean squared error, and % Dec. in rMSE is the percentage decrease in rMSE of the augmented estimator (superscript A) compared to the non-augmented estimator (superscript NA). Superscript AJ refers to the Aalen–Johansen estimator

Table 2 Simulation results under less or no dependent censoring

Estimator	Time (t)	0.05	0.20	0.35	0.50	0.65	0.80	0.95	1.10
	$F_1(t)$ (truth)	0.014	0.052	0.085	0.114	0.142	0.166	0.186	0.204
	$F_2(t)$ (truth)	0.025	0.096	0.159	0.215	0.264	0.306	0.346	0.380
A: *15% dependent censoring*									
$\hat{F}_1^A(t)$	Bias	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.002
	SD	0.008	0.014	0.018	0.021	0.022	0.025	0.027	0.030
	rMSE	0.008	0.014	0.018	0.021	0.022	0.025	0.027	0.030
	% Dec. in rMSE	0.12	0.56	1.43	2.00	2.61	4.38	8.17	10.02
$\hat{F}_1^{NA}(t)$	Bias	0.001	0.001	0.002	0.002	0.002	0.003	0.004	0.006
	SD	0.008	0.015	0.018	0.021	0.023	0.026	0.029	0.032
	rMSE	0.008	0.015	0.018	0.021	0.023	0.026	0.029	0.033
$\hat{F}_1^{AJ}(t)$	Bias	0.001	0.002	0.004	0.006	0.008	0.011	0.014	0.016
	SD	0.008	0.015	0.019	0.022	0.024	0.026	0.029	0.031
	rMSE	0.008	0.015	0.019	0.023	0.025	0.028	0.032	0.035
$\hat{F}_2^A(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	−0.001
	SD	0.010	0.018	0.023	0.026	0.029	0.031	0.034	0.037
	rMSE	0.010	0.018	0.023	0.026	0.029	0.031	0.034	0.037
	% Dec. in rMSE	0.13	0.79	1.80	3.18	5.98	7.03	12.81	18.52
$\hat{F}_2^{NA}(t)$	Bias	0.000	0.001	0.000	0.002	0.002	0.004	0.005	0.007
	SD	0.010	0.018	0.023	0.027	0.030	0.034	0.039	0.043
	rMSE	0.010	0.018	0.023	0.027	0.030	0.034	0.039	0.044
$\hat{F}_2^{AJ}(t)$	Bias	0.000	0.002	0.005	0.009	0.014	0.019	0.022	0.025
	SD	0.010	0.019	0.023	0.028	0.029	0.033	0.036	0.040
	rMSE	0.010	0.019	0.024	0.029	0.033	0.038	0.042	0.047

Table 2 continued

Estimator	Time (t)	0.05	0.20	0.35	0.50	0.65	0.80	0.95	1.10
	$F_1(t)$ (truth)	0.014	0.052	0.085	0.114	0.142	0.166	0.186	0.204
	$F_2(t)$ (truth)	0.025	0.096	0.159	0.215	0.264	0.306	0.346	0.380
B: *independent censoring*									
$\hat{F}_1^A(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SD	0.007	0.014	0.018	0.022	0.025	0.028	0.035	0.044
	rMSE	0.007	0.014	0.018	0.022	0.025	0.028	0.035	0.044
	% Dec. in rMSE	0.07	−0.13	0.40	0.66	0.48	0.40	−0.91	6.23
$\hat{F}_1^{NA}(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SD	0.007	0.014	0.018	0.022	0.025	0.029	0.035	0.047
	rMSE	0.007	0.014	0.018	0.022	0.025	0.029	0.035	0.047
$\hat{F}_1^{AJ}(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SD	0.006	0.014	0.018	0.022	0.024	0.028	0.034	0.043
	rMSE	0.006	0.014	0.018	0.022	0.024	0.028	0.034	0.043
$\hat{F}_2^A(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SD	0.010	0.019	0.024	0.028	0.033	0.038	0.046	0.060
	rMSE	0.010	0.019	0.024	0.028	0.033	0.038	0.046	0.060
	% Dec. in rMSE	−0.06	0.50	1.32	0.99	0.75	−0.12	0.45	9.23
$\hat{F}_2^{NA}(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SD	0.010	0.019	0.024	0.029	0.033	0.038	0.046	0.065
	rMSE	0.010	0.019	0.024	0.029	0.033	0.038	0.046	0.065
$\hat{F}_2^{AJ}(t)$	Bias	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SD	0.010	0.019	0.024	0.028	0.032	0.038	0.045	0.058
	rMSE	0.010	0.019	0.024	0.028	0.032	0.038	0.045	0.058

SD is the standard deviation, rMSE is the square root mean squared error, and % Dec. in rMSE is the percentage decrease in rMSE of the augmented estimator (superscript A) compared to the non-augmented estimator (superscript NA). Superscript AJ refers to the Aalen–Johansen estimator

Table 3 Types of failure and censoring

Regimen	VF	TLAE	Admin. censoring	Non-admin censoring	TLOE
4-drug ($N = 382$)	65	35	213	56	13
3-drug ($N = 376$)	81	23	219	40	13

we censor patients if they discontinue treatment for reasons other than VF, TLAE, or TLOE. This may lead to dependent censoring.

A total of 758 patients were randomized, including 382 patients who received the 4-drug regimen and 376 patients who received the 3-drug regimen. Of the 758 patients, 146 had failure of their initial randomized regimen due to virologic failure (VFs), 58 discontinued their initial regimen due to treatment-limiting adverse events (TLAEs), and 26 discontinued their initial treatment due to treatment limiting other events (TLOEs, including 5 deaths). Of the remaining 528 patients who were censored, 432 patients were still on their initial randomized regimen at completion of the study, and so were administratively censored. The remaining 96 patients were non-administratively censored, mainly due to loss of follow-up while on their initial randomized regimen. The types of failure among the two regimens as well as the types of censoring are presented in Table 3.

We based the model for non-administrative censoring on a literature review of variables that might predict losses to follow-up in HIV-infected patients (Dudley et al. 1995; Ioannidis et al. 1997; Arici et al. 2002; Lanoy et al. 2006; Andersen et al. 2007; Krishnan et al. 2010; Fleishman et al. 2012), and used the same set of variables for administrative censoring. Many of these variables have also been associated with the competing outcomes of interest, and so dependent censoring is a reasonable concern. We therefore used the following variables in Eq. (3), for the hazard of censoring, $w_r(t, \bar{V}_t)$:

$$\left\{\text{CD4 Count}_t, \text{Log Viral Load}_0, \text{Sex}, \text{Age}, \text{IV drug use}, \text{Black}, \text{Hispanic}\right\},$$

for $r = 1, 2$ (administrative and non-administrative censoring), where CD4 count is a time dependent variable coded as 1 for counts $\leq$200 and 0 otherwise; Log Viral Load is the $\log_{10}$ HIV viral load in the blood; Sex is coded as 1 for males, 0 for females; IV drug use is coded as 1 for patients who reported ever using illicit intravenous (IV) drugs, and 0 otherwise; Black and Hispanic are the indicator variables for patients of black non-Hispanic and Hispanic race/ethnicity, respectively, with reference category white non-Hispanic. Table 4 presents the parameter estimates for the two censoring models, one for administrative and one for non-administrative censoring. We found no significant predictors of administrative censoring though there was some evidence of an increased odds of administrative censoring among men ($p = 0.07$) and Hispanic patients ($p = 0.06$) for the 3-drug regimen. Given that, due to randomization, treatment and administrative censoring are unrelated, this could well be a statistical artifact. Reported use of IV drugs was highly predictive of non-administrative censoring in both treatment arms ($p = 0.006$) for the 3-drug regimen, $p = 0.02$ for the

Table 4 Parameter estimates (95%-confidence intervals) for the models for administrative and non-administrative censoring

Covariate	3-drug regimen		4-drug regimen	
	Admin. censoring	Non-admin. censoring	Admin. censoring	Non-admin. censoring
Sex (male vs. female)	0.33 (−0.03, 0.69) $p = 0.07$	0.31 (−0.58, 1.19) $p = 0.50$	−0.03 (−0.41, 0.35) $p = 0.87$	−0.57 (−1.16, 0.03) $p = 0.06$
IV Drug use (ever vs. never)	−0.47 (−1.18, 0.24) $p = 0.20$	1.18 (0.34, 2.02) $p = 0.006$	−0.05 (−0.50, 0.41) $p = 0.85$	0.83 (0.14, 1.51) $p = 0.02$
Age ≤ 30 (vs. >30 years)	0.18 (−0.16, 0.52) $p = 0.30$	0.54 (−0.14, 1.21) $p = 0.12$	−0.07 (−0.40, 0.25) $p = 0.65$	0.28 (−0.31, 0.87) $p = 0.36$
Hispanic (vs. white, non-Hispanic)	0.36 (−0.003, 0.72) $p = 0.06$	0.76 (−0.08, 1.59) $p = 0.08$	0.13 (−0.23, 0.48) $p = 0.49$	0.34 (−0.34, 1.02) $p = 0.33$
Black, non-Hispanic (vs. white, non-Hispanic)	−0.13 (−0.45, 0.19) $p = 0.42$	0.56 (−0.20, 1.32) $p = 0.15$	0.22 (−0.10, 0.55) $p = 0.18$	0.10 (−0.54, 0.74) $p = 0.76$
Log Viral Load (per 1 $\log_{10}$ copies/ml)	−0.05 (−0.25, 0.15) $p = 0.63$	−0.12 (−0.56, 0.32) $p = 0.59$	−0.03 (−0.21, 0.16) $p = 0.79$	−0.24 (−0.61, 0.13) $p = 0.21$
Time-dependent CD4 count ≤200	−0.01 (−0.64, 0.61) $p = 0.97$	−0.45 (−1.37, 0.47) $p = 0.34$	0.19 (−0.43, 0.82) $p = 0.55$	0.31 (−0.37, 0.99) $p = 0.37$

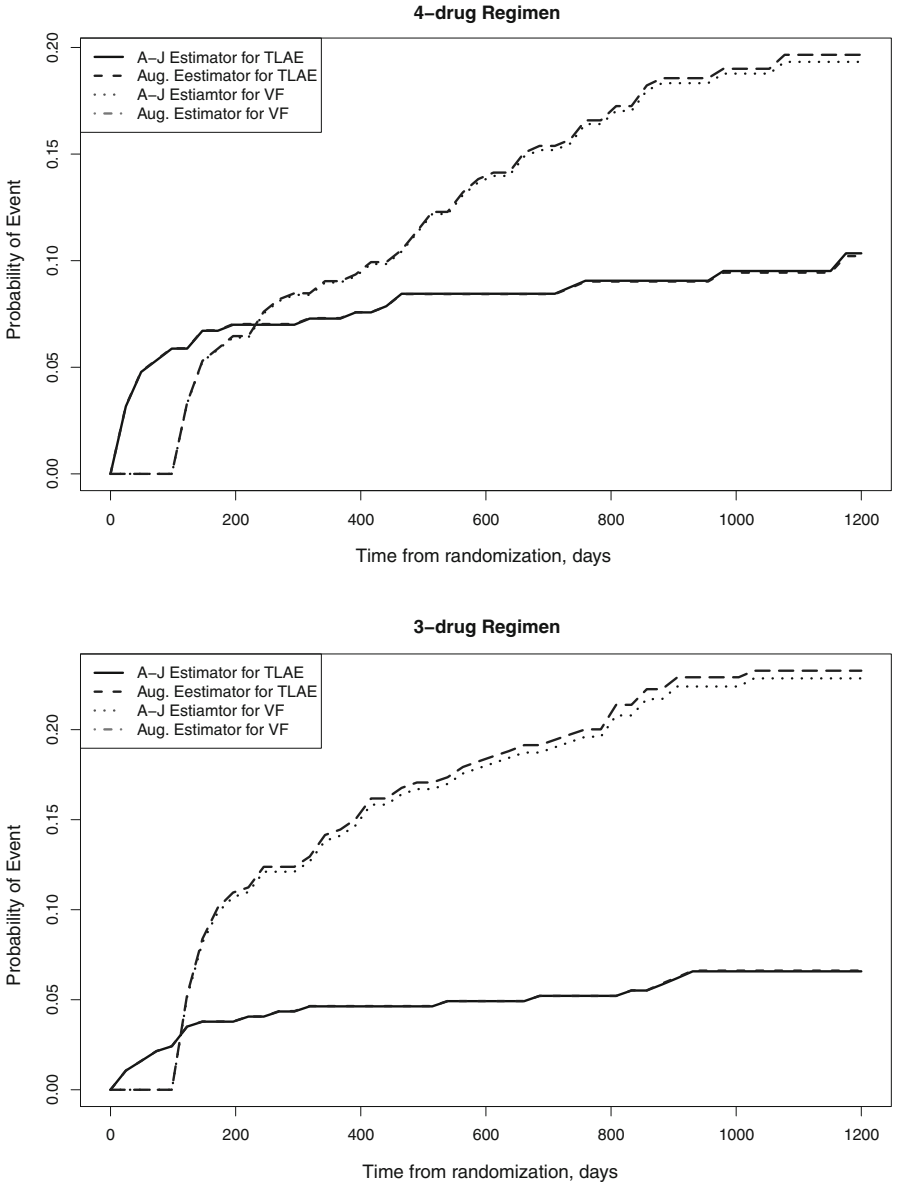

Fig. 1 Cumulative incidence curves, by regimen

4-drug regimen). Hispanic race/ethnicity was marginally significantly associated with an increased odds of non-administrative censoring in those on the 3-drug regimen ($p = 0.08$), and male sex was marginally significantly associated with a reduced odds of non-administrative censoring in those on the 4-drug regimen ($p = 0.06$). Thus, the assumption of independent censoring is violated if the time or type of event depends on, for example, use of IV drugs.

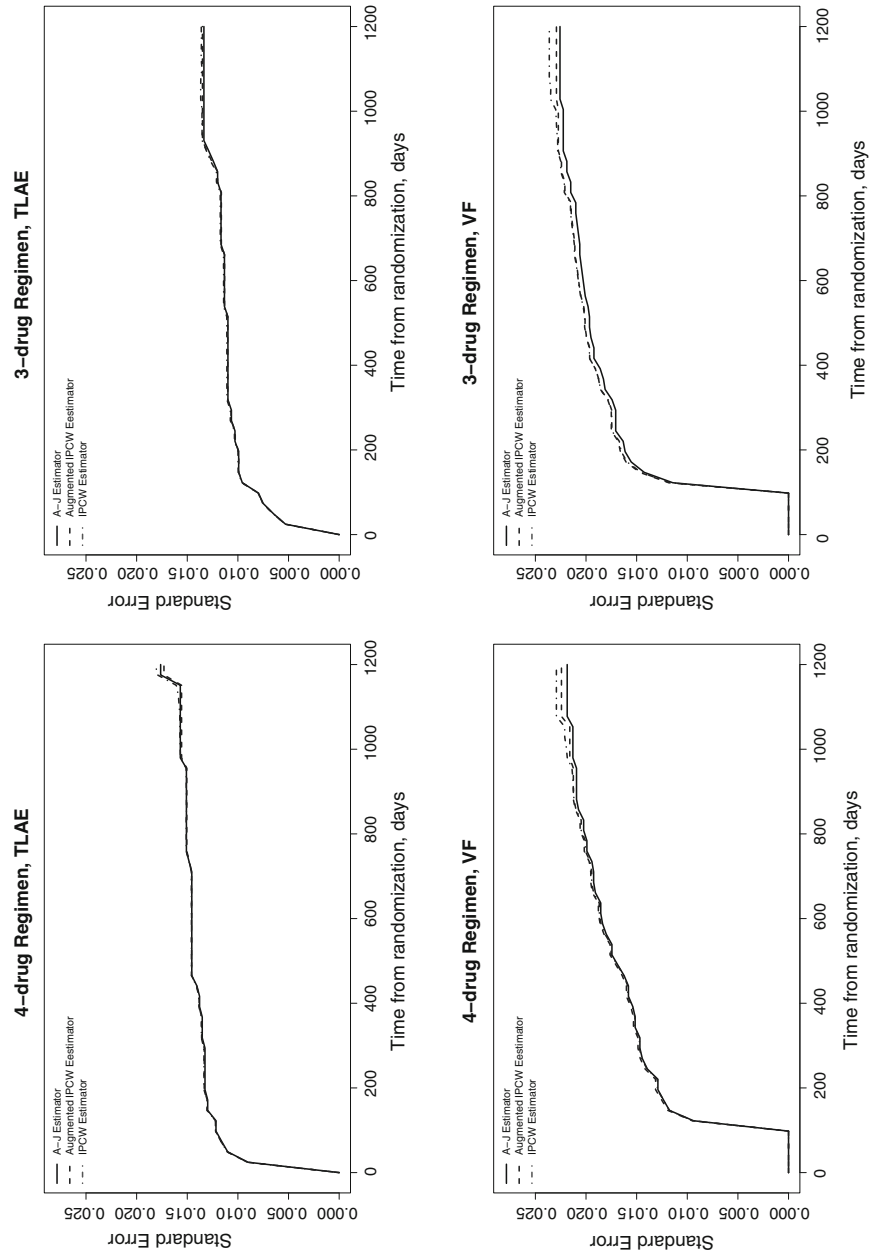

Fig. 2 Standard errors of $\hat{F}_j(t)$, by regimen and type of failure, using bootstrap variance estimation

The estimated cumulative incidence curves for VF and TLAE are shown in Fig. 1. Since there were only 26 TLOEs, we do not present the cumulative incidence curve for TLOE. The non-augmented IPCW estimator was essentially identical to the augmented IPCW estimator and is not shown here. Despite the fact that there were strong predictors of censoring, and the concern that these might also be predictors of the competing outcomes of interest, there was little difference between the standard estimate, $\hat{F}_j^0(t)$, and the augmented IPCW estimate. For this application, of considerable importance, the conclusions that might be drawn from the study have been shown to be not sensitive to potentially dependent censoring, a concern that was well motivated by the fact that some predictors of loss to follow-up are also predictors of the outcome of interest. Comparing treatments, there were general trends for higher rates of VF but lower rates of TLAE for the 3-drug versus the 4-drug regimen.

The standard errors of our non-augmented and augmented IPCW estimators as well as the standard error for $\hat{F}_j^0(t)$ are shown in Fig. 2. The standard errors for $\hat{F}_j^{NA}(t)$ and $\hat{F}_j^A(t)$ were obtained using the non-parametric bootstrap with 500 bootstrap samples. The difference in standard errors between the augmented and non-augmented IPCW estimators is generally small, suggesting that for this application, there is only a slight efficiency advantage to using the more complicated augmented estimator. Furthermore, for this application, the standard error of the augmented IPCW estimator is very similar to that of the Aalen–Johansen estimator. This is particularly important, because our estimator remains valid in the presence of dependent censoring, which is not so for the Aalen–Johansen estimator. Thus, we can rely on the conclusions even though the assumption of independent censoring may be violated.

8 Discussion

In this paper we have developed a method to estimate the cumulative incidence function with multiple types of censoring. The use of methods of analysis which more appropriately address the challenges of competing risk data and potentially dependent censoring may be very valuable in understanding the relative balance of safety outcomes (e.g. TLAE) and efficacy outcomes (e.g. VF), and how this balance evolves with time on treatment and compares among treatments. Such analyses will likely be important complements to analyses of composite outcome measures (e.g. time to first of TLAE or VF) which can be difficult to interpret because they are often complex mixes of efficacy and safety outcomes. In addition, being able to handle dependent censoring in statistical analyses is important, where assessment of the sensitivity of the conclusions to the handling of different reasons for censoring should be part of standard analyses.

We investigated four simulation scenarios. When censoring was dependent, the augmented IPCW estimator had substantially reduced bias as compared to the Aalen–Johansen estimator, which assumes independent censoring; the IPCW estimator was in between, with small bias for earlier time points and larger bias later, especially where there was more dependent censoring. The decrease in root mean squared error obtained from augmenting the IPCW estimator (as opposed to using the non-augmented IPCW estimator) increased with time and with percentage of dependent censoring, and was

substantial for later time points and larger percentages of dependent censoring. When there was less dependent censoring, the IPCW estimator had comparable standard errors as the Aalen–Johansen estimator, and less bias; and the augmented IPCW estimator outperformed the IPCW estimator in terms of root MSE by a smaller percentage. In the scenario with independent censoring, where all three estimators are consistent, IPCW and augmented IPCW did not substantially inflate standard errors, as compared to the Aalen–Johansen estimator.

In our application, the results were not appreciably changed by allowing for the possibility of dependent censoring, but there may be other applications where doing so is important. Furthermore, this analysis provides more confidence in the resulting estimates since it incorporates the possibility of dependent censoring. As shown, this does not need to be at the expense of larger standard errors.

Even if there are only baseline covariates which predict censoring, IPCW is sometimes preferable over basing the analysis on a Cox proportional hazards model for the cause specific hazards. The reason is that we don't need to assume a semi-parametric Cox model for the cause specific hazards of the competing outcomes; we only need to specify (3), which is automatically correctly specified if independent censoring does turn out to hold.

One direction of possible future research is to relax the assumption that data on all time-dependent and independent covariates that are prognostic for both failure and censoring are recorded and available. This future research would rely on sensitivity analyses in order to handle the non-ignorable missingness.

Acknowledgements The authors would like to thank Andrea Rotnitzky for her review and valuable comments on this paper. We are grateful to the ACTG for providing data used in the motivating application. This work was partially supported by Grants AI024643, AI068634, and AI007358 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Compliance with ethical standards

Ethical approval Analysis of data from ACTG A5095 was approved by the Institutional Review Board of the Harvard School of Public Health. Informed consent was obtained from all individual participants included in the study.

Appendix

Equation (3) is an unbiased estimating equation for $F_j(t)$ since

$$\begin{aligned}
&\mathbf{E}\left(\frac{\tilde{\Delta}}{\pi(\tilde{T}|\bar{V}_{\tilde{T}};\Lambda_0)}\left\{\mathbf{1}(T\le t, J=j)-F_j(t)\right\}\right)\\
&=\mathbf{E}\left(\mathbf{E}\left\{\frac{\tilde{\Delta}}{\pi(\tilde{T}|\bar{V}_{\tilde{T}};\Lambda_0)}\left\{\mathbf{1}(T\le t, J=j)-F_j(t)\right\}|\bar{V}_T,T\right\}\right)\\
&=\mathbf{E}\left(\frac{\mathbf{E}[\tilde{\Delta}|\bar{V}_T,T]}{\pi(\tilde{T}|\bar{V}_{\tilde{T}};\Lambda_0)}\left\{\mathbf{1}(T\le t, J=j)-F_j(t)\right\}\right)
\end{aligned}$$

$$= \mathbf{E}\left(1 \cdot \left\{\mathbf{1}(T \leq t, J = j) - F_j(t)\right\}\right)$$
$$= 0.$$

The third equality follows from the fact that $\mathbf{E}\left(\tilde{\Delta}|\bar{V}_T, T\right) = Pr[\tilde{\Delta} = 1|\bar{V}_T, T]$ which under our coarsening at random assumption equals $\pi(\tilde{T}|\bar{V}_{\tilde{T}}; \Lambda_0)$. From this it is clear that Eq. (3) is an unbiased estimating equation for $F_j(t)$.

One can use the results in Robins and Rotnitzky (1992) to prove that with one type of censoring the solution to

$$\sum_{i=1}^{n}\left\{\frac{\tilde{\Delta}_i}{\pi_i(\tilde{T}_i|\bar{V}_{\tilde{T}_i}; \Lambda_0)}\left\{\mathbf{1}(T_i \leq t, J_i = j) - F_j(t)\right\} - A_i\{F_j(t), \gamma, b(\cdot)\}\right\} = 0 \quad (16)$$

with

$$A_i\{F_j(t), \gamma, b(\cdot)\} \equiv \int \frac{b(u, \bar{V}_u)}{\pi(u\text{-}|\bar{V}_u; \Lambda_0(u))} dM_C(u) \quad (17)$$

and $b(u, \bar{V}_u)$ the same as in Eq. (9), is a doubly robust, locally efficient estimator for $F_j(t)$. Here, $dM_C(u) = dN_C(u) - d\Lambda(u)$. In our situation, with multiple types of censoring, it is easy to show that $dM_C(u) = \sum_{r=1}^{r^*} dM_{C,r}(u)$, which leads to our augmentation term in (8). Note that with one type of failure and one type of censoring, our method reduces to that of (Rotnitzky and Robins 2005).

Now, it can be shown that for each r, since

$$\frac{b(u, \bar{V}_u)}{\pi(u\text{-}|\bar{V}_u; \Lambda_0(u))}$$

is a bounded and predictable process, defined on the same filtration as $M_{C,r}(u)$,

$$\int \frac{b(u, \bar{V}_u)}{\pi(u\text{-}|\bar{V}_u; \Lambda_0(u))} dM_{C,r}(u)$$

is a mean zero martingale Fleming and Harrington (1991, Thm 1.5.1). Note that the left-continuous versions of $b(u, \bar{V}_u)$ and $\pi(u|\bar{V}_u; \Lambda_0(u))$ are needed here. As a result

$$\sum_{r=1}^{r^*} \int \frac{b(u, \bar{V}_u)}{\pi(u\text{-}|\bar{V}_u; \Lambda_0(u))} dM_{C,r}(u)$$

is also a mean zero martingale. Thus, $A_i\{F_j(t), \gamma, b(\cdot)\}$ has mean zero. Since $A_i\{F_j(t), \gamma, b(\cdot)\}$ has mean zero, it also follows trivially that Eq. (7) is an unbiased estimating equation for $F_j(t)$.

References

Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer series in statistics. Springer, New York

Andersen JW, Fass R, van der Horst C (2007) Factors associated with early study discontinuation in AACTG studies, DACS 200. Contemp Clin Trials 28:583–592

Arici C, Ripamonti D, Maggiolo F, Rizzi M, Finazzi MG, Pezzotti P, Suter F (2002) Factors associated with the failure of HIV-positive persons to return for scheduled medical visits. HIV Clin Trials 3(1):52–57

Bender R, Augustin T, Blettner M (2005) Generating survival times to simulate cox proportional hazards models. Stat Med 24(11):1713–1723

Bryant J, Dignam JJ (2004) Semiparametric models for cumulative incidence functions. Biometrics 60(1):182–190

Dudley J, Jin S, Hoover D, Metz S, Thackeray R, Chmiel J (1995) The multicenter AIDS cohort study: retention after 9 1/2 years. Am J Epidemiol 142(3):323–330

Fine JP (2001) Regression modeling of competing crude failure probabilities. Biostatistics 2(1):85–97

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94(446):496–497

Fleishman JA, Yehia BR, Moore RD, Korthuis PT, Gebo KA, HIV Reseach Network (2012) Establishment, retention, and loss to follow-up in outpatient HIV care. J Acquir Immune Defic Syndr 60(3):249–259

Fleming T, Harrington D (1991) Counting processes and survival analysis. Wiley, New York

Gray RJ (1988) A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat 16(3):1141–1154

Gulick R, Ribaudo H, Shikuma C, Lustgarten S, Squires K, Meyer W III, Acosta E, Schackman B, Pilcher C, Murphy R et al (2004) Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. N Engl J Med 350(18):1850–1861

Gulick R, Ribaudo H, Shikuma C, Lalama C, Schackman B, Meyer W III, Acosta E, Schouten J, Squires K, Pilcher C et al (2006) Three-vs four-drug antiretroviral regimens for the initial treatment of HIV-1 infection. JAMA 296(7):769–781

Heitjan DF, Rubin DB (1991) Ignorability and coarse data. Ann Stat 19(4):2244–2253

Ioannidis JPA, Bassett R, Hughes MD, Volberding PA, Sacks HS, Lau J (1997) Predictors and impact of patients lost to follow-up in a long-term randomized trial of immediate versus deferred antiretroviral treatment. J Acquir Immune Defic Syndr Hum Retrovirol 16(1):22–30

Jeong JH, Fine J (2006) Direct parametric inference for the cumulative incidence function. J R Stat Soc Ser C (Appl Stat) 55(2):187–200

Kalbfleisch JD, Prentice RL (1980) The statistical analysis of failure time data. Wiley, New York

Krishnan S, Wu K, Smurzynski M, Bosch RJ, Benson CA, Collier AC, Klebert MK, Feinberg J, Koletar SL, ALLRT, A5001 team (2010) 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 Clin Trials 12(4):190–200

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 (2006) Predicators identified for losses to follow-up among HIV-seropositive patients. J Clin Epidemiol 59:829–835

Lin DY (1997) Non-parametric inference for cumulative incidence functions in competing risks studies. Stat Med 16(8):901–910

Lok JJ, Hughes MD (2016) Evaluating predictors of competing risk outcomes when censoring depends on time-dependent covariates, with application to safety and efficacy of HIV treatment. Stat Med 35(13):2183–2194

Pintilie M (2006) Competing risks: a practical perspective. Wiley, New York

Robins JM, Rotnitzky A, Zhao LP (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J Am Stat Assoc 90(429):106–121

Robins J, Rotnitzky A (1992) Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell NP, Dietz K, Farewell VT (eds) AIDS Epidemiology (Methodological issues), Birkhäuser Boston, pp. 297–331

Rotnitzky A, Farall A, Bergesio A, Scharfstein D (2007) Analysis of failure time data under competing censoring mechanisms. J R Stat Soc Ser B (Stat Methodol) 69(3):307–327

Rotnitzky A, Bergesio A, Farall A (2009) Analysis of quality-of-life adjusted failure time data in the presence of competing, possibly informative, censoring mechanisms. Lifetime Data Anal 15(1):1–23

Rotnitzky A, Robins JM (2005) Inverse probability weighted estimation in survival analysis. In: Armitage P, Coulton T (eds) Encyclopedia of biostatistics, 2nd edn. Wiley, New York
Rubin DB (1976) Inference and missing data. Biometrika 63(3):581–592
Tsiatis AA (2006) Semiparametric theory and missing data. Springer, Berlin