# When to Censor?

**Catherine R. Lesko*, Jessie K. Edwards, Stephen R. Cole, Richard D. Moore, and Bryan Lau**

* Correspondence to Dr. Catherine R. Lesko, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205 (email: clesko2@jhu.ed).

*Initially submitted November 30, 2016; accepted for publication July 14, 2017.*

Loss to follow-up is an endemic feature of time-to-event analyses that precludes observation of the event of interest. To our knowledge, in typical cohort studies with encounters occurring at regular or irregular intervals, there is no consensus on how to handle person-time between participants' last study encounter and the point at which they meet a definition of loss to follow-up. We demonstrate, using simulation and an example, that when the event of interest is captured outside of a study encounter (e.g., in a registry), person-time should be censored when the study-defined criterion for loss to follow-up is met (e.g., 1 year after last encounter), rather than at the last study encounter. Conversely, when the event of interest must be measured within the context of a study encounter (e.g., a biomarker value), person-time should be censored at the last study encounter. An inappropriate censoring scheme has the potential to result in substantial bias that may not be easily corrected.

bias (epidemiology); censoring; epidemiologic methods; loss to follow-up; selection bias; survival analysis

Abbreviations: AIDS, acquired immunodeficiency syndrome; ART, antiretroviral therapy; CI, confidence interval; HIV, human immunodeficiency virus; IDU, injection drug use; IDUs, injection drug users; IPCWs, inverse probability of censoring weights; IQR, interquartile range; LTFU, lost to follow-up; PWID, persons who inject drugs.

Censoring is an endemic feature of time-to-event analysis that precludes observation of the event. Right-censoring occurs when an event may have occurred after the last time a person was under observation, but the specific timing of the event is unknown. Right-censoring may occur at the end of the study period (i.e., administrative censoring) or when a person fails to return for a study visit (i.e., is lost to follow-up (LTFU)). In contrast to administrative censoring, which coincides with the end of the analytical period and can be placed precisely in time, LTFU is a nonevent. As such, there is a lack of consensus on when exactly a study participant becomes "lost" and how to treat person-time analytically for participants who are LTFU. Specifically, if LTFU is defined as 12 months without a clinic encounter, should the censoring date be the date of the last encounter or the date on which the definition of LTFU is met? Despite the near ubiquity of LTFU, to our knowledge, the implications of alternative approaches for dealing with person-time between the last encounter and the time point at which the definition of LTFU is met have not been formally examined.

Herein, we discuss characteristics of the study design and research question that may influence the censoring scheme. Specifically, we make the case that the least biased censoring scheme depends on whether the outcome is detected outside of or within a study encounter. We demonstrate bias that may result from an inappropriate LTFU censoring strategy with a simulation and a simple example.

## CONCEPTUAL FRAME

As we demonstrate in the subsequent simulation, the most appropriate censoring strategy depends on the nature of the outcome of interest. For the purposes of this paper, we dichotomize outcomes into 2 types: those that are *captured* and those that are *measured*. We are not aware of existing, commonly accepted terms that encapsulate this distinction, and we define our usage of these terms as follows. We classify outcomes as *captured* if they are detectable outside of a study visit (e.g., death, hospital admission, cancer diagnosis reported to a registry). We classify outcomes as *measured* if they are only detectable within a study visit (e.g., change in a biomarker requiring a laboratory test, remission of depressive symptoms reportable on a standardized

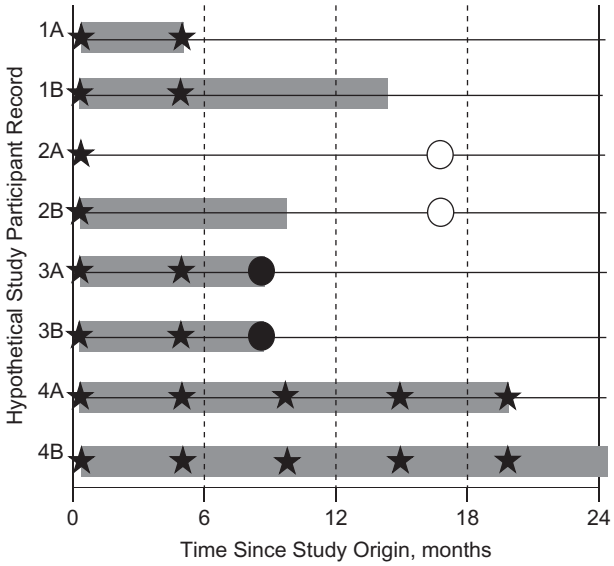


**Figure 1.** Several illustrative study records for hypothetical individuals (numbered) in an interval cohort study under 2 different censoring schemes (last-encounter censoring, denoted A, and lost-to-follow-up (LTFU)-definition censoring, denoted B). Study visits are represented by stars. The outcome is captured (e.g., in a registry) outside of a study visit, and its occurrence is denoted by a circle. Outcomes included in the analysis are denoted by solid circles, while outcomes excluded from analysis (because they occur after the LTFU definition is met) are denoted by hollow circles. LTFU is defined as going 12 months without a study visit *or occurrence of the outcome*. For persons who are administratively censored (e.g., participant 4) or LTFU (e.g., participants 1 and 2), person-time could be included from study entry to the last study visit (last-encounter censoring) or from study entry to the point at which the definition of LTFU is met or to administrative censoring (LTFU-definition censoring). Included person-time under each censoring scheme is denoted by gray shaded rectangles. Note that application of last-encounter censoring results in complete exclusion of persons who never return for a follow-up visit (e.g., participant 2).

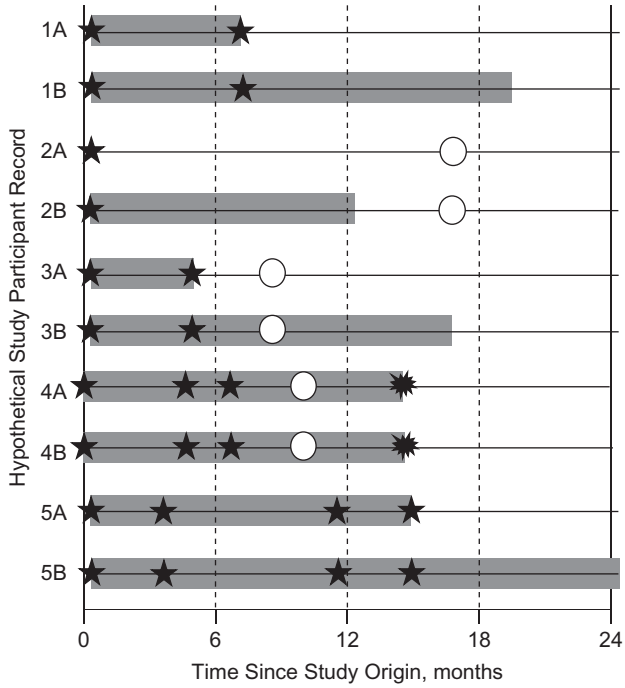


**Figure 2.** Several illustrative study records for hypothetical individuals (numbered) in a clinical cohort study under 2 different censoring schemes (last-encounter censoring, denoted A, and lost-to-follow-up (LTFU)-definition censoring, denoted B). Study visits are represented by stars. A 5-pointed star indicates a visit at which the outcome has not yet occurred; a starburst indicates a visit at which the outcome is detected. The outcome is measured; its timing is unobserved, and its occurrence is denoted by a hollow circle. For persons whose outcome is observed (e.g., participant 4), person-time is included from study entry to the visit at which the outcome occurs. (Although the occurrence of the outcome could be thought of as interval-censored, interval censoring is rarely employed; see Web Appendix 2.) LTFU is defined as going 12 months without a study visit. For persons who are administratively censored (e.g., participant 5) or LTFU (e.g., participants 1–3), person-time could be included from study entry to the last study visit (last-encounter censoring) or from study entry to the point at which the definition of LTFU is met or to administrative censoring (LTFU-definition censoring). Included person-time under each censoring scheme is denoted by gray shaded rectangles. Note that application of last-encounter censoring results in complete exclusion of persons who never return for a follow-up visit (e.g., participant 2). Further note that, because the outcome is not observable outside of the context of a study visit, the outcome for individual 3 is unobserved and is not included in the analysis, despite its occurrence during person-time included under LTFU-definition censoring.

survey). Classification of outcomes as captured or measured will depend on the study design and setting. For example, we cite cancer diagnosis as an example of a captured outcome because in the United States, there are population-based cancer registries in almost all states. However, in other settings, cancer registries may be imperfect or nonexistent and cancer diagnosis might be classified as a measured outcome if it would only be detectable when reported at a study visit.

Herein, we explore 2 possible censoring schemes that represent the most common analytical approaches for addressing LTFU. Using (what we dub) "last-encounter" censoring, participants who are LTFU are censored at their last study encounter. Using "LTFU-definition" censoring, participants who are LTFU are censored when they meet the definition of LTFU.

## METHODS

To illustrate how the date of the last study visit, the point at which the definition of LTFU is met, and the censoring scheme interact to determine the period in which participants are methodologically "at risk" for the outcome, we generated 2 figures with several illustrative person-records and their treatment under each of the 2 censoring schemes. Figure 1 assumes a captured outcome and Figure 2 assumes a measured outcome. In both cases, LTFU was defined as 12 months with no study visits. To increase the applicability of our results, we chose to depict 2 main cohort study designs: an interval cohort study in Figure 1 and a clinical cohort study in Figure 2 (1).

### Simulation

Full details of the simulation are provided in Web Appendix 1 (available at https://academic.oup.com/aje). Briefly, we simulated 1,000 cohorts of 2,000 persons. For each individual,

we simulated a visit schedule such that visits occurred at irregular intervals approximately every $M$ months. Next we assigned binary baseline exposure $A$ and binary time-varying exposure $X_j$ (which changed at most once from 0 to 1, in a month corresponding to a visit), where $j$ indexed month of follow-up. We explored bias for a time-fixed exposure ($A$) versus time-varying exposure ($X_j$) in separate analyses. The outcome $Y$ could occur in any month and was a function of $A$ and $X_j$. The cumulative incidence of $Y$ due to $A$ and the cumulative incidence of $Y$ due to $X$ (simulated truth) are illustrated in Web Figures 1 and 2, respectively.

For captured outcomes, we assigned event time $T$ equal to the month $J$ in which $Y$ occurred. For measured outcomes, we assigned event time $T$ equal to the next clinic visit $j_{k+1}$, where $k$ indexes visits. The measured event is therefore, in actuality, interval-censored, although most analyses do not treat measured events as interval-censored. We explore this issue more in Web Appendix 2; results of analyses treating measured events as if they were interval-censored appear in Web Table 1.

We estimated the bias of last-encounter censoring and LTFU-definition censoring under several scenarios. In each scenario, we simulated random variable $D$ (months of follow-up) and assigned observations a last visit $L = \max(J_K \leq D)$. We simulated $D$ for each scenario as follows:

1. $D$ is random.
2. $D$ is associated with $A$.
3. $D$ is associated with $X_j$.

We defined LTFU as 12 months without a study visit, such that the definition of LTFU was met at $L + 12$. The average probability of LTFU under each scenario above is reported in Web Table 2. Captured outcomes were included (i.e., the LTFU definition was not met because it was precluded by an event) and the end of follow-up was set to $T$ if $T \leq L + 12$; if $T > L + 12$, outcomes were excluded (all person-time and events after $L + 12$ were censored). Measured outcomes were included and the end of follow-up was set to $T$ if $T \leq L$; if $T > L$, outcomes were excluded. We set the end of follow-up to $L$ under last-encounter censoring and to $L + 12$ under LTFU-definition censoring.

We estimated the 10-year risk, risk difference, risk ratio, and hazard ratio for the effect of $A$ on $Y$ and $\bar{X}$ on $Y$ in each simulation. Here $\bar{X}$ denotes history of treatment, set to always ($X_j = 1$ for all $j$) or never ($X_j = 0$ for all $j$). For scenarios in which censoring was associated with $A$ or $X_j$, we estimated inverse probability of censoring weights (IPCWs) (assuming the censoring mechanism was known) and calculated weighted estimates of the 10-year risk, risk difference, risk ratio, and hazard ratio. We estimated average bias as $\hat{\beta} - \beta$ and percent bias as average bias/$\beta \times 100$. We also report the median values and interquartile ranges of biases ($\hat{\beta} - \beta$). We do not report the variance or mean squared error because we used the same estimators with both censoring schemes, and thus variances were similar across simulations and all differences in mean squared error were driven by bias.

### Example

To show the impact of different censoring schemes on parameter estimates in real-world data, we report the association between history of injection drug use (IDU) as a risk factor for acquisition of human immunodeficiency virus (HIV) and time from clinic enrollment to 1) initiation of antiretroviral therapy (ART) (a measured outcome) and 2) death (a captured outcome) in the Johns Hopkins HIV Clinical Cohort. The Johns Hopkins HIV Clinical Cohort includes data on all adults receiving continuity care at the Moore Clinic for HIV Care (Baltimore, Maryland) who have agreed to share their data (>90%). We followed ART-naive patients from enrollment at the clinic between January 1, 1998, and August 30, 2015, to 1) ART initiation, death, LTFU (12 months without a CD4 cell count, viral load measurement, or clinic visit), or administrative censoring (at 5 years or on August 31, 2015) and to 2) death, LTFU, or administrative censoring, whichever came first. ART initiation was defined as initiation of ≥3 antiretroviral medications on 1 day. Deaths were ascertained through matches against the Social Security Death Index.

When analyzing time to ART initiation, we treated patients who died (a competing event) prior to ART initiation as LTFU (censored them). We acknowledge that best practice is to not conflate competing events and LTFU (2), but for the purposes of illustration we will ignore the question of competing risks until we reach the Discussion section.

To account for differences in patient characteristics at baseline among persons who inject drugs (PWID) and non–injection drug users (non-IDUs), we estimated inverse probability of exposure weights (3, 4) conditional on sex, black race, prior acquired immunodeficiency syndrome (AIDS) diagnosis or exposure to mono- or dual ART, and baseline age, CD4 cell count, and $\log_{10}$ viral load (measurement most proximal to clinic enrollment, drawn up to 12 months prior). To account for possibly nondifferential LTFU, we used IPCWs (5) conditional on baseline covariates and time-varying AIDS diagnosis, most recent CD4 cell count, and $\log_{10}$ viral load, months since baseline, and IDU (exposure). We modeled continuous variables using a restricted quadratic spline (6). This example is similar to the analysis of HIV-infected women described by Buchanan et al. (7).

We applied last-encounter censoring and LTFU-definition censoring in separate analyses. We estimated hazard ratios using an inverse-probability-weighted Cox proportional hazards model (8), and we estimated the conditional 5-year risk for PWID and non-IDUs and the 5-year risk difference and risk ratio based on the complement of inverse-probability-weighted survival curves (9, 10). We report 95% confidence intervals based on the 2.5th and 97.5th percentiles of the distribution of estimates from 1,000 nonparametric bootstrap resamples of the data (11).

## RESULTS

Figures 1 and 2 depict several person-records and their treatment under each censoring scheme. The most illustrative person in both figures is participant 3, who has her last study visit in month 5 of follow-up and experiences an event in month 8. When the event is captured (Figure 1), participant 3 is observed to experience the event, precluding her meeting the definition of LTFU; thus, under both censoring schemes, the amounts of person-time she contributed are equivalent.

However, had participant 3 not had an event, she would have looked like participant 1, who met the definition of LTFU at 17 months. Under LTFU-definition censoring, we still include the person-time after participant 1's last visit, but under last-encounter censoring, person-time from month 5 onward is excluded. This differential treatment of person-time for participants who have and do not have an event under last-encounter censoring can lead to bias when studying captured outcomes.

When the outcome is measured (Figure 2), the fact that participant 3 does not return after month 5 means we would not observe her event in month 8. However, under LTFU-definition censoring, we include person-time from month 5 to month 17, despite not being able to observe an event even if it had occurred, making this person-time functionally immune. Thus, in contrast to captured outcomes, when analyzing time to a measured outcome, participants who meet the definition of LTFU should be censored at their last visit.

## Simulation

The true risks for a captured or measured outcome were similar in our simulated data set. For a *captured outcome*, 10-year risk was 52.4% and 31.3%, respectively, for persons assigned $A = 1$ and $A = 0$; the risk difference was 21.1%, the risk ratio was 1.68, and the hazard ratio was 1.99. When censoring was random (and 66.8% of persons were LTFU; Web Table 2), LTFU-definition censoring resulted in percent bias ≤1% for all estimands (Table 1). However, last-encounter censoring resulted in risk functions that were biased upwards: Percent bias in 10-year risk under $A = 1$ and $A = 0$ was 11.9% and 16.1%, respectively. Some bias canceled out, such that percent bias in the risk difference, risk ratio, and hazard ratio was 5.7%, −6.8%, and −1.0%, respectively. Figure 3 shows the estimated cumulative incidence of $Y$ for persons with $A = 1$ under no censoring, LTFU-definition censoring, and last-encounter censoring. A similar pattern (negligible to no bias under LTFU-definition censoring and substantial upward bias of risk under last-encounter censoring) was seen when LTFU was faster for persons with $A = 1$. When LTFU was more likely when $X_j = 1$, both censoring schemes produced bias, although again bias was stronger under last-encounter censoring. IPCWs mitigated the bias under LTFU-definition censoring but failed to remove the bias under last-encounter censoring (Web Table 3).

For a *measured* outcome and random LTFU, last-encounter censoring resulted in minimal upward bias of risk estimates. Percent bias in 10-year risk under $A = 1$ and $A = 0$ was 2.4% and 3.4%, respectively (Table 2). Percent bias in the risk difference, risk ratio, and hazard ratio was 0.9%, −1.3%, and

**Table 1.** Bias in the Estimated Effects of a Baseline Exposure on a Captured Outcome (e.g., Death) in the Presence of Loss to Follow-Up (LTFU) When Person-Time for Those Lost Is Censored at the Time of the Last Encounter or the Point At Which the Definition of LTFU Is Met[a]

| Estimand | Truth | Last-Encounter Censoring | | | | LTFU-Definition Censoring | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | Bias[b] | % Bias | Median Bias (IQR) | Estimate | Bias[b] | % Bias | Median Bias (IQR) |
| *LTFU is Random* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 52.4 | 58.6 | 6.2 | 11.9 | 5.0 (3.2 to 6.8) | 52.3 | −0.2 | −0.3 | 0.0 (−1.6 to 1.6) |
| $A = 0$ | 31.3 | 36.3 | 5.0 | 16.1 | 6.3 (4.4 to 8.0) | 31.3 | 0.1 | 0.2 | −0.2 (−2.0 to 1.5) |
| RD | 21.1 | 22.3 | 1.2 | 5.7 | 1.2 (−1.4 to 3.8) | 20.9 | −0.2 | −1.1 | −0.2 (−2.7 to 2.3) |
| RR | 1.68 | 1.62 | −0.035 | −6.8 | −0.038 (−0.094 to 0.026) | 1.68 | −0.003 | −0.6 | −0.008 (−0.067 to 0.061) |
| HR | 1.99 | 1.97 | −0.007 | −1.0 | −0.006 (−0.070 to 0.058) | 1.98 | −0.002 | −0.4 | −0.003 (−0.068 to 0.063) |
| *LTFU Occurs at a Faster Rate When $A = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 52.4 | 60.2 | 7.8 | 14.8 | 4.2 (2.6 to 5.7) | 52.5 | 0.1 | 0.1 | 0.0 (−1.5 to 1.4) |
| $A = 0$ | 31.3 | 35.4 | 4.2 | 13.3 | 7.8 (5.7 to 9.8) | 31.2 | −0.1 | −0.2 | 0.1 (−2.0 to 2.1) |
| RD | 21.1 | 24.8 | 3.6 | 17.1 | 3.5 (0.9 to 6.3) | 21.3 | 0.1 | 0.6 | 0.0 (−2.4 to 2.7) |
| RR | 1.68 | 1.71 | 0.014 | 2.8 | 0.013 (−0.043 to 0.068) | 1.69 | 0.004 | 0.8 | 0.005 (−0.060 to 0.062) |
| HR | 1.99 | 2.11 | 0.061 | 9.0 | 0.057 (0.000 to 0.126) | 1.99 | 0.004 | 0.6 | 0.002 (−0.057 to 0.068) |
| *LTFU is More Likely When $X_j = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 52.4 | 59.1 | 6.7 | 12.8 | 5.0 (3.4 to 6.6) | 53.5 | 1.1 | 2.1 | 0.6 (−0.7 to 2.1) |
| $A = 0$ | 31.3 | 36.3 | 5.0 | 16.1 | 6.7 (4.8 to 8.5) | 32.0 | 0.7 | 2.2 | 1.1 (−0.7 to 2.9) |
| RD | 21.1 | 22.8 | 1.7 | 7.9 | 1.7 (−0.6 to 4.1) | 21.5 | 0.4 | 1.9 | 0.4 (−1.8 to 2.6) |
| RR | 1.68 | 1.64 | −0.028 | −5.3 | −0.029 (−0.077 to 0.026) | 1.68 | 0.000 | −0.1 | −0.001 (−0.052 to 0.053) |
| HR | 1.99 | 1.98 | 0.001 | 0.2 | 0.002 (−0.059 to 0.062) | 1.99 | 0.004 | 0.6 | 0.007 (−0.055 to 0.065) |

Abbreviations: HR, hazard ratio; IQR, interquartile range; LTFU, lost to follow-up; RD, risk difference; RR, risk ratio.
[a] Full details of the simulations are provided in Web Appendix 1.
[b] Absolute bias is on the log scale for RR and HR.

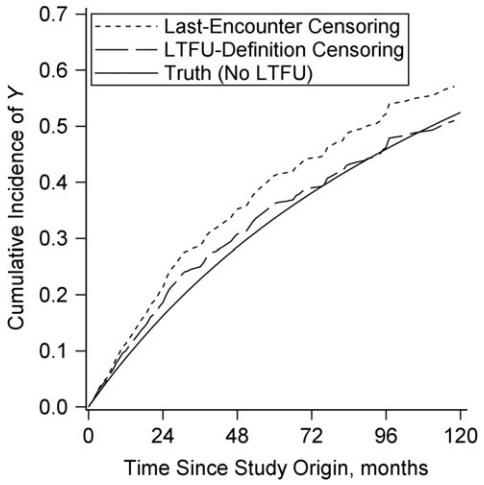


**Figure 3.** Estimated cumulative incidence (simulation results) of a captured event *Y* (event detectable outside a study visit) under $A = 1$ if persons who are lost to follow-up (LTFU) are censored when they met the definition of LTFU (LTFU-definition censoring) or at their last study encounter, compared with truth (absence of LTFU).

0.0%, respectively. In contrast, LTFU-definition censoring resulted in substantial downward bias in the risks. Percent bias in 10-year risk under $A = 1$ and $A = 0$ was −11.5% and −13.3%, respectively. Percent bias in the risk difference, risk ratio, and hazard ratio was −9.0%, 4.4%, and −0.2%, respectively. A similar pattern (minor bias under last-encounter censoring and substantial negative bias under LTFU-definition censoring) was seen when LTFU was faster among persons with $A = 1$. Again, when LTFU was more likely if $X_j = 1$, both censoring schemes produced bias. IPCWs reduced the bias under last-encounter censoring but were ineffective at removing bias under LTFU-definition censoring (Web Table 4).

We hypothesized that the minimal upward bias in the risk estimates for measured outcomes under last-encounter censoring was related to the nonconstant monitoring of individuals (due to the visit structure). Specifically, when a person who is LTFU is censored at her last visit, the Kaplan-Meier estimator begins reallocating that censored person to events immediately following the last visit. However, if she had been observed to have the event, due to the nature of the visit structure, her event would not have been observed until a mean of *M* months later. To explore this theory, we reran the simulation for random LTFU and random censoring, setting the mean visit

**Table 2.** Bias in the Estimated Effects of a Baseline Exposure on a Measured Outcome (e.g., a Biomarker) in the Presence of Loss to Follow-Up (LTFU) When Person-Time for Those Lost Is Censored at the Time of the Last Encounter or the Point at Which the Definition of LTFU Is Met[a]

| Estimand | Truth | Last-Encounter Censoring | | | | LTFU-Definition Censoring | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | Bias[b] | % Bias | Median Bias (IQR) | Estimate | Bias[b] | % Bias | Median Bias (IQR) |
| *LTFU is Random* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 53.7 | 55.0 | 1.3 | 2.4 | 0.7 (−1.3 to 3.2) | 47.5 | −6.2 | −11.5 | −4.5 (−6.2 to −2.5) |
| $A = 0$ | 32.2 | 33.3 | 1.1 | 3.4 | 1.2 (−1.3 to 3.6) | 27.9 | −4.3 | −13.3 | −6.2 (−8.4 to −4.2) |
| RD | 21.5 | 21.7 | 0.2 | 0.9 | 0.1 (−3.3 to 3.6) | 19.6 | −1.9 | −9.0 | −1.9 (−4.9 to 1.0) |
| RR | 1.67 | 1.67 | −0.007 | −1.3 | −0.004 (−0.089 to 0.077) | 1.72 | 0.022 | 4.4 | 0.020 (−0.057 to 0.103) |
| HR | 1.98 | 1.98 | 0.000 | 0.0 | −0.003 (−0.065 to 0.073) | 1.98 | −0.002 | −0.2 | −0.005 (−0.068 to 0.072) |
| *LTFU Occurs at a Faster Rate When A = 1* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 53.7 | 55.6 | 1.9 | 3.5 | 0.7 (−1.3 to 2.7) | 46.8 | −7.0 | −13.0 | −3.9 (−5.6 to −2.2) |
| $A = 0$ | 32.2 | 33.1 | 0.9 | 2.8 | 1.5 (−1.3 to 4.7) | 28.5 | −3.7 | −11.6 | −7.2 (−9.7 to −4.6) |
| RD | 21.5 | 22.5 | 1.0 | 4.6 | 0.8 (−2.9 to 4.5) | 18.3 | −3.3 | −15.0 | −3.4 (−6.5 to −0.2) |
| RR | 1.67 | 1.70 | 0.009 | 1.7 | 0.007 (−0.079 to 0.092) | 1.66 | −0.014 | −2.8 | −0.015 (−0.100 to 0.066) |
| HR | 1.98 | 2.02 | 0.021 | 3.0 | 0.022 (−0.047 to 0.087) | 1.89 | −0.049 | −7.1 | −0.048 (−0.120 to 0.020) |
| *LTFU is More Likely When $X_j = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 53.7 | 56.7 | 3.0 | 5.5 | 1.8 (−0.2 to 4.1) | 49.8 | −3.9 | −7.2 | −3.1 (−4.6 to −1.2) |
| $A = 0$ | 32.2 | 34.3 | 2.1 | 6.6 | 2.8 (0.3 to 5.3) | 29.4 | −2.8 | −8.8 | −4.0 (−6.0 to −1.9) |
| RD | 21.5 | 22.4 | 0.9 | 4.0 | 1.0 (−2.4 to 4.1) | 20.5 | −1.0 | −4.8 | −1.0 (−3.7 to 1.6) |
| RR | 1.67 | 1.67 | −0.007 | −1.4 | −0.004 (−0.084 to 0.072) | 1.71 | 0.019 | 3.8 | 0.019 (−0.057 to 0.093) |
| HR | 1.98 | 1.99 | 0.003 | 0.5 | 0.004 (−0.060 to 0.069) | 1.98 | 0.002 | 0.2 | 0.002 (−0.060 to 0.069) |

Abbreviations: HR, hazard ratio; IQR, interquartile range; LTFU, lost to follow-up; RD, risk difference; RR, risk ratio.
[a] Full details of the simulations are provided in Web Appendix 1.
[b] Absolute bias is on the log scale for RR and HR.

interval to 3 months ($M = 3$) instead of 6 months. Our theory would predict that a shorter visit interval should result in smaller positive bias; and indeed, absolute bias in the risk estimates for $A = 1$ and $A = 0$ was 0.7% and 0.6%, respectively (Web Table 5), compared with 1.3% and 1.1% when the mean visit interval was 6 months. The bias in risk estimates (and contrasts of risk estimates) due to last-encounter censoring for measured outcomes should be minimal when the risk of the outcome is low (indeed, we confirmed this by altering parameters in our simulation; data not shown) or when the visit interval is short; even in our simulation, in the presence of a high risk and long visit intervals, bias associated with last-encounter censoring was less than the bias associated with LTFU-definition censoring.

We next considered effects of time-varying exposure $X$ and found similar results. When the outcome was captured, last-encounter censoring resulted in substantial upward bias of risk estimates and LTFU-definition censoring was unbiased (Table 3). When censoring was associated with covariates, IPCWs minimized bias under LTFU-definition censoring but did not remove the bias under last-encounter censoring (Web Table 6). When the outcome was measured, LTFU-definition censoring produced substantial downward bias of risk estimates, while last-encounter censoring produced minimal upward bias of risk estimates (Table 4). IPCWs minimized bias under last-encounter censoring (but did not eliminate the small positive bias) but did not remove bias under LTFU-definition censoring (Web Table 7).

## Example

Of 2,446 ART-naive patients enrolled in the Johns Hopkins HIV Clinical Cohort between January 1, 1998, and August 30, 2015, we excluded 184 (8%) without a baseline CD4 cell count or viral load measurement. Of the 2,262 patients in the final study sample, the majority were male (64%) and black (77%). The median age was 40 years (interquartile range (IQR), 34–47), the median CD4 cell count was 292 cells/μL (IQR, 120–492), and the median $\log_{10}$ viral load was 4.3 copies/mL (IQR, 3.2–5.0). One-third of the sample (34%) reported IDU as their most probable route of HIV acquisition. Over 5 years of follow-up, 1,541 persons initiated ART and 257 persons died before being LTFU.

*Time to ART initiation.* Based on simulation results, the most appropriate censoring scheme when studying ART initiation would be last-encounter censoring. There were 622 patients LTFU prior to ART initiation, 154 of whom never returned after their baseline visit. Those 154 persons, plus 19 patients who enrolled less than a year before the end of the study period (and so never had the opportunity to meet the definition

**Table 3.** Bias in the Estimated Effects of a Time-Varying Exposure on a Captured Outcome (e.g., Death) in the Presence of Loss to Follow-Up (LTFU) When Person-Time for Those Lost Is Censored at the Time of the Last Encounter or the Point at Which the Definition of LTFU Is Met[a]

| Estimand | Truth | Last-Encounter Censoring | | | | LTFU-Definition Censoring | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | Bias[b] | % Bias | Median Bias (IQR) | Estimate | Bias[b] | % Bias | Median Bias (IQR) |
| *LTFU is Random* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 34.3 | 39.2 | 4.9 | 14.3 | 6.2 (3.7 to 8.8) | 34.2 | −0.1 | −0.3 | −0.5 (−2.8 to 2.0) |
| $A = 0$ | 49.8 | 56.0 | 6.2 | 12.5 | 4.9 (3.0 to 6.8) | 49.5 | −0.3 | −0.6 | −0.1 (−1.9 to 1.6) |
| RD | −15.5 | −16.9 | −1.4 | 8.7 | −1.3 (−4.5 to 1.8) | −15.3 | 0.2 | −1.3 | 0.4 (−2.8 to 3.1) |
| RR | 0.69 | 0.70 | 0.015 | −4.0 | 0.016 (−0.054 to 0.080) | 0.69 | 0.003 | −0.8 | 0.008 (−0.069 to 0.074) |
| HR | 0.61 | 0.60 | −0.016 | 3.2 | −0.016 (−0.091 to 0.059) | 0.61 | −0.001 | 0.1 | 0.001 (−0.072 to 0.072) |
| *LTFU Occurs at a Faster Rate When $A = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 34.3 | 37.7 | 3.4 | 10.1 | 4.3 (2.0 to 6.7) | 32.9 | −1.4 | −4.0 | −2.2 (−4.3 to 0.2) |
| $A = 0$ | 49.8 | 54.1 | 4.3 | 8.7 | 3.5 (1.5 to 5.3) | 47.8 | −2.0 | −4.1 | −1.3 (−3.1 to 0.2) |
| RD | −15.5 | −16.4 | −0.9 | 5.8 | −0.7 (−3.8 to 2.0) | −14.9 | 0.7 | −4.2 | 0.9 (−2.0 to 3.4) |
| RR | 0.69 | 0.70 | 0.012 | −3.2 | 0.017 (−0.051 to 0.077) | 0.69 | 0.000 | −0.1 | 0.007 (−0.070 to 0.071) |
| HR | 0.61 | 0.60 | −0.014 | 2.9 | −0.012 (−0.091 to 0.061) | 0.61 | 0.000 | 0.0 | 0.001 (−0.076 to 0.073) |
| *LTFU is More Likely When $X_j = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 34.3 | 40.2 | 6.0 | 17.4 | 3.8 (1.9 to 5.5) | 34.2 | 0.0 | −0.1 | −0.2 (−1.9 to 1.6) |
| $A = 0$ | 49.8 | 53.5 | 3.7 | 7.5 | 6.0 (4.2 to 7.7) | 49.6 | −0.2 | −0.3 | 0.0 (−1.6 to 1.5) |
| RD | −15.5 | −13.3 | 2.2 | −14.3 | 2.2 (−0.3 to 4.8) | −15.4 | 0.1 | −0.8 | 0.2 (−2.3 to 2.6) |
| RR | 0.69 | 0.75 | 0.087 | −23.3 | 0.087 (0.028 to 0.145) | 0.69 | 0.001 | −0.3 | 0.001 (−0.059 to 0.064) |
| HR | 0.61 | 0.67 | 0.096 | −19.4 | 0.096 (0.028 to 0.164) | 0.61 | 0.000 | −0.1 | 0.000 (−0.066 to 0.068) |

Abbreviations: HR, hazard ratio; IQR, interquartile range; LTFU, lost to follow-up; RD, risk difference; RR, risk ratio.
[a] Full details of the simulations are provided in Web Appendix 1.
[b] Absolute bias is on the log scale for RR and HR.

**Table 4.** Bias in the Estimated Effects of a Time-Varying Exposure on a Measured Outcome (e.g., a Biomarker) in the Presence of Loss to Follow-Up (LTFU) When Person-Time for Those Lost Is Censored at the Time of the Last Encounter or the Point at Which the Definition of LTFU Is Met[a]

| Estimand | Truth | Last-Encounter Censoring | | | | LTFU-Definition Censoring | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | Bias[b] | % Bias | Median Bias (IQR) | Estimate | Bias[b] | % Bias | Median Bias (IQR) |
| *LTFU is Random* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 34.0 | 35.2 | 1.1 | 3.4 | 0.1 (−2.9 to 4.1) | 29.7 | −4.4 | −12.9 | −7.2 (−9.9 to −3.7) |
| $A = 0$ | 52.6 | 53.4 | 0.9 | 1.6 | 1.0 (−1.2 to 3.4) | 45.9 | −6.7 | −12.8 | −4.4 (−6.3 to −2.6) |
| RD | −18.5 | −18.2 | 0.3 | −1.6 | 1.0 (−3.9 to 4.7) | −16.2 | 2.3 | −12.5 | 2.9 (−1.2 to 6.0) |
| RR | 0.65 | 0.67 | 0.017 | −4.0 | 0.029 (−0.078 to 0.114) | 0.65 | −0.001 | 0.2 | 0.009 (−0.091 to 0.092) |
| HR | 0.56 | 0.54 | −0.027 | 4.6 | −0.023 (−0.100 to 0.051) | 0.55 | −0.007 | 1.1 | −0.004 (−0.081 to 0.071) |
| *LTFU Occurs at a Faster Rate When $A = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 34.0 | 33.5 | −0.6 | −1.6 | −1.8 (−4.9 to 1.5) | 28.3 | −5.8 | −16.9 | −8.9 (−11.7 to −5.9) |
| $A = 0$ | 52.6 | 51.3 | −1.3 | −2.4 | −0.7 (−2.9 to 1.7) | 44.0 | −8.5 | −16.2 | −5.8 (−7.6 to −3.9) |
| RD | −18.5 | −17.8 | 0.7 | −3.8 | 1.3 (−2.9 to 5.1) | −15.7 | 2.8 | −15.0 | 3.3 (−0.5 to 6.5) |
| RR | 0.65 | 0.66 | 0.009 | −2.0 | 0.014 (−0.084 to 0.104) | 0.65 | −0.007 | 1.7 | −0.002 (−0.103 to 0.089) |
| HR | 0.56 | 0.54 | −0.024 | 4.1 | −0.021 (−0.105 to 0.054) | 0.55 | −0.004 | 0.8 | −0.003 (−0.086 to 0.079) |
| *LTFU is More Likely When $X_j = 1$* | | | | | | | | | |
| Risk | | | | | | | | | |
| $A = 1$ | 34.0 | 35.2 | 1.1 | 3.3 | 0.2 (−2.3 to 3.1) | 28.7 | −5.4 | −15.7 | −4.5 (−6.6 to −1.8) |
| $A = 0$ | 52.6 | 53.2 | 0.6 | 1.2 | 1.1 (−1.1 to 3.2) | 48.4 | −4.1 | −7.8 | −5.3 (−7.1 to −3.8) |
| RD | −18.5 | −18.0 | 0.5 | −2.8 | 0.8 (−3.2 to 4.2) | −19.8 | −1.3 | 6.8 | −1.2 (−4.3 to 2.0) |
| RR | 0.65 | 0.67 | 0.020 | −4.6 | 0.024 (−0.066 to 0.105) | 0.60 | −0.091 | 20.9 | −0.088 (−0.177 to −0.008) |
| HR | 0.56 | 0.56 | −0.002 | 0.3 | −0.002 (−0.076 to 0.066) | 0.50 | −0.103 | 17.5 | −0.103 (−0.178 to −0.031) |

Abbreviations: HR, hazard ratio; IQR, interquartile range; LTFU, lost to follow-up; RD, risk difference; RR, risk ratio.
[a] Full details of the simulations are provided in Web Appendix 1.
[b] Absolute bias is on the log scale for RR and HR.

of LTFU) but did not return for any visits after enrollment, were completely excluded from the analysis that relied upon last-encounter censoring. When we used last-encounter censoring, we estimated that the 5-year covariate-standardized risk of ART initiation was 85.8 (95% confidence interval (CI): 80.7, 90.1) among PWID and 90.2 (95% CI: 88.1, 92.3) among non-IDUs (Table 5). The risk difference and hazard ratio for ART initiation associated with IDU were −4.5 (95% CI: −10.1, 0.3) and 0.71 (95% CI: 0.63, 0.79), respectively. When we (inappropriately) used LTFU-definition censoring, as in the simulation, risks were biased downward. Risk of ART initiation among PWID was 78.9 (95% CI: 72.6, 85.1), and that among non-IDUs was 86.5 (95% CI: 83.4, 89.2). The risk difference and hazard ratio associated with a history of IDU were −7.6 (95% CI: −14.4, −0.5) and 0.67 (95% CI: 0.59, 0.77), respectively (Table 5).

*Time to death.* Based on simulation results, the most appropriate censoring scheme when studying time to death would be LTFU-definition censoring. There were 1,015 patients LTFU prior to death, 94 of whom never returned after their baseline visit (and did not die within the first year of follow-up). Those 94 persons, along with 77 patients who enrolled less than a year before the administrative censoring date (and so did not meet the definition for LTFU) but also did not return or die between enrollment and the end of the study period, were completely excluded from the analysis that utilized last-encounter censoring. When we used LTFU-definition censoring, we estimated that the 5-year risk of death was 25.9 (95% CI: 20.9, 31.0) among PWID and 12.4 (95% CI: 10.2, 15.0) among non-IDUs (Table 5). IDU was associated with a risk difference and hazard ratio for death of 13.4 (95% CI: 8.2, 19.3) and 2.09 (95% CI: 1.61, 2.79), respectively. As in the simulation, when we (inappropriately) censored patients who met the definition of LTFU the last time they were seen, estimated risks were biased upwards. Mortality risk was 28.7 (95% CI: 23.0, 34.4) among PWID and 13.2 (95% CI: 11.0, 15.8) among non-IDUs. The risk difference and hazard ratio associated with a history of IDU were 15.4 (95% CI: 9.3, 21.2) and 2.19 (95% CI: 1.64, 2.92), respectively (Table 5).

Because deaths were captured for all patients in this cohort, regardless of engagement at the clinic, and because exposure in this example was ascertained at baseline, we did not actually have to censor anyone when estimating time to death associated with IDU; we have only employed various censoring schemes here for the purposes of illustration. The benefit of this approach in this particular instance is that we can compare results using either censoring scheme with the "true" (no

**Table 5.** Associations (Example, Crude, and Standardized) of History of Injection Drug Use With Initiation of Antiretroviral Therapy and Death Among 2,262 Patients Enrolled in the Johns Hopkins HIV Clinical Cohort, Baltimore, Maryland, 1998–2015

| Estimand and Censoring Strategy | 5-Year Risk | | | | Risk Difference | | Risk Ratio | | Hazard Ratio | |
|---|---|---|---|---|---|---|---|---|---|---|
| | PWID | | Non-IDUs | | | | | | | |
| | PE | 95% CI | PE | 95% CI | PE | 95% CI | PE | 95% CI | PE | 95% CI |
| Time from enrollment to ART initiation | | | | | | | | | | |
| Last-encounter censoring | | | | | | | | | | |
| Crude | 86.6 | 82.1, 90.5 | 90.0 | 87.6, 92.1 | −3.5 | −8.1, 0.9 | 0.96 | 0.91, 1.01 | 0.71 | 0.64, 0.79 |
| Standardized[a] | 85.8 | 80.7, 90.1 | 90.2 | 88.1, 92.3 | −4.5 | −10.1, 0.3 | 0.95 | 0.89, 1.00 | 0.71 | 0.63, 0.79 |
| LTFU-definition censoring | | | | | | | | | | |
| Crude | 78.6 | 73.4, 83.3 | 85.4 | 82.7, 88.0 | −6.8 | −12.1, −1.4 | 0.92 | 0.86, 0.98 | 0.66 | 0.59, 0.73 |
| Standardized[a] | 78.9 | 72.6, 85.1 | 86.5 | 83.4, 89.2 | −7.6 | −14.4, −0.5 | 0.91 | 0.84, 0.99 | 0.67 | 0.59, 0.77 |
| Time from enrollment to death | | | | | | | | | | |
| Last-encounter censoring | | | | | | | | | | |
| Crude | 29.5 | 25.3, 33.7 | 12.1 | 10.2, 14.3 | 17.4 | 12.5, 22.3 | 2.43 | 1.91, 3.06 | 2.54 | 1.98, 3.25 |
| Standardized[a] | 28.7 | 23.0, 34.4 | 13.2 | 11.0, 15.8 | 15.4 | 9.3, 21.2 | 2.17 | 1.64, 2.86 | 2.19 | 1.64, 2.92 |
| LTFU-definition censoring | | | | | | | | | | |
| Crude | 25.3 | 21.5, 29.3 | 10.6 | 8.8, 12.6 | 14.7 | 10.4, 19.1 | 2.39 | 1.90, 3.01 | 2.41 | 1.90, 3.10 |
| Standardized[a] | 25.9 | 20.9, 31.0 | 12.4 | 10.2, 15.0 | 13.4 | 8.2, 19.3 | 2.08 | 1.61, 2.75 | 2.09 | 1.61, 2.79 |
| Truth (no censoring) | | | | | | | | | | |
| Crude | 23.7 | 20.6, 27.0 | 11.8 | 10.2, 13.5 | 11.9 | 8.4, 15.7 | 2.01 | 1.68, 2.46 | 2.13 | 1.75, 2.66 |
| Standardized[b] | 23.6 | 20.0, 27.5 | 12.7 | 10.9, 14.6 | 10.9 | 7.1, 15.2 | 1.85 | 1.51, 2.29 | 1.94 | 1.55, 2.45 |

Abbreviations: AIDS, acquired immunodeficiency syndrome; ART, antiretroviral therapy; CI, confidence interval; HIV, human immunodeficiency virus; IDU, injection drug use; IDUs, injection drug users; LTFU, lost to follow-up; PE, point estimate; PWID, persons who inject drugs.

[a] Standardized (with inverse probability of IDU weights) to the distribution of the following covariates in the study sample: male sex, black race, male-male sex as an HIV acquisition risk factor, and baseline age, prior AIDS diagnosis, prior mono- or dual ART exposure, CD4 cell count, and $\log_{10}$ viral load; also weighted for possibly differential loss to follow-up associated with the previously listed baseline covariates and time-varying most recent CD4 cell count, $\log_{10}$ viral load, and AIDS diagnosis.

[b] Standardized with inverse probability of IDU weights as specified above. Because there was no censoring, no censoring weighting was done.

censoring) time to death among PWID and non-IDUs. While LTFU-definition censoring slightly overestimated 5-year mortality risk for PWID and non-IDUs, there was substantially less bias associated with this censoring scheme as compared with last-encounter censoring. Deviations between the "truth" and LTFU-definition censored estimates may have been due to random error or to violations of the assumption of uninformative censoring.

## DISCUSSION

We have demonstrated here that analytical handling of person-time in the presence of censoring should depend on the nature of the study design and the outcome: Persons who are LTFU when analyzing time to a captured outcome should be censored when they meet the LTFU definition; persons who are LTFU when analyzing time to a measured outcome should be censored at the time when they were last seen. Bias due to the choice of a censoring scheme that was incompatible with the outcome under study was not remedied by IPCWs, even though we assumed that the censoring mechanism was known. This bias is perhaps most closely conceptually related to immortal time bias (12, 13), as it arises due to excluding person-time at risk (when using last-encounter censoring for captured outcomes) or to including person-time not at risk (when using LTFU-definition censoring for measured outcomes). The choice of censoring scheme can have a substantial impact on results, as was observed in our application, although the magnitude of the bias will depend upon the risk of the outcome under study and the risk of LTFU.

Herein, we made several simplifying assumptions about the underlying processes for the outcome and LTFU in order to isolate the issue of including versus excluding person-time between the last visit and the point at which the definition of LTFU was met. In reality, there are often additional issues related to analytical handling of censoring that must be considered. For example, we assumed that censoring was ignorable, conditional on covariates (14), and that the censoring mechanism was known (5). When censoring is not ignorable, other approaches may be necessary (15–17). This is a commonly recognized problem in HIV clinics in resource-limited settings, where increasing illness severity may reduce the probability of a patient's returning for a clinic visit and increase the mortality hazard, or where failure to return for a clinic visit may be the direct result of mortality (16, 18). Furthermore, in our conceptual framework, we assumed that ascertainment of captured

outcomes (e.g., death) was complete (e.g., through a population-based registry). In many settings, information on mortality (and other captured outcomes such as emergency room admissions or cancer diagnoses) may not be routinely collected and recorded. We also assumed that visit frequency was the same across exposure groups (19–21) and that the occurrence of an event did not alter visit frequency or the probability of LTFU. This is particularly relevant for measured outcomes in clinical cohorts. In practice, the onset of disease may be associated with a change in health status that would trigger a clinic visit earlier than would be expected in the absence of disease or that might prevent the patient from attending his next visit altogether (e.g., if symptoms limited mobility).

Our conclusions raise additional questions for future research. One question is, Which censoring scheme should be used when the outcome is a composite of captured and measured outcomes (e.g., AIDS or death) or when there are multiple outcomes of interest, as in the case of competing events, that are of different types (e.g., ART initiation, where death is a competing event)? In our example examining time to ART initiation, we censored patients who died prior to ART initiation. The problems with this approach have been well documented (2, 22–24). However, the best censoring strategy to use had we attempted to explicitly model death as a competing event remains unclear. Time between the last study visit and death is functionally immune for the endpoint of ART initiation, so its inclusion should bias the cumulative incidence of ART initiation. Yet that same person-time is at-risk time for the endpoint of death, so its exclusion should bias the cumulative incidence of death.

We determined that when information on the outcome is captured, LTFU-definition censoring is the least biased censoring scheme. Herein, we simulated time-varying treatment $X_j$ such that the value of $X$ could only change in months that included a study visit. Thus, both in our simulations (where changing exposure status was structurally tied to attending a study visit) and in our example (where exposure was time-fixed at baseline), the observed value of treatment was the true value of treatment (assuming no underreporting of history of IDU), and we did not have to adjust for covariates that were affected by prior treatment. However, for other types of study questions, the value of a time-varying exposure may not be fixed (e.g., if treatment is available outside of study visits) or controlling for time-varying covariates may be more crucial. If LTFU-definition censoring is the best analytical option, variable values between the last encounter and the LTFU date are not updated in the analytical data set. Researchers must impute those missing values, either by carrying prior values forward, assuming that they follow some specified decay function, or using a more complicated approach. Although LTFU-definition censoring appropriately accounts for person-time at risk, bias due to measurement error may be introduced because the assumed values and the true values may differ (25). While the impact of covariate measurement error on estimation of inverse probability of treatment weights has been a topic of recent inquiry (26–30), we are unaware of investigations of measurement error for the estimation of IPCWs.

The advent of electronic health records has facilitated the creation of many new clinical cohorts, which are characterized by substantial LTFU, given that the primary purpose of data collection is health-care delivery and billing rather than research. Bias due to use of a censoring strategy that is incompatible with the outcome of interest and the study design will be largest when the rate of the event and the rate of LTFU are high. Thus, this investigation into proper analytical handling of LTFU is particularly timely and clears up some confusion regarding the most appropriate censoring time in time-to-event analyses.

## ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Catherine R. Lesko, Richard D. Moore, Bryan Lau); Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (Jessie K. Edwards, Stephen R. Cole); and Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland (Richard D. Moore, Bryan Lau).

C.R.L., R.D.M., and B.L. were supported in part by National Institutes of Health (NIH) grants U01 DA036935, U01 HL121812, U01 AA020793, and P30 AI094189. J.K.E. and S.R.C. were supported in part by NIH grant R01 AI100654. S.R.C. was also supported in part by NIH grants R24 AI067039, U01 AI103390, and P30 AI50410.

Conflict of interest: none declared.

## REFERENCES

1. Lau B, Gange SJ, Moore RD. Interval and clinical cohort studies: epidemiological issues. *AIDS Res Hum Retroviruses*. 2007;23(6):769–776.
2. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170(2):244–256.
3. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.
4. Hernán MA, Robins J. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC Press. In press.
5. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56(3):779–788.
6. Howe CJ, Cole SR, Westreich DJ, et al. Splines for trend analysis and continuous confounder control. *Epidemiology*. 2011;22(6):874–875.
7. Buchanan AL, Hudgens MG, Cole SR, et al. Worth the weight: using inverse probability weighted Cox models in AIDS research. *AIDS Res Hum Retroviruses*. 2014;30(12):1170–1177.
8. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11(5):561–570.
9. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75(1):45–49.
10. Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24(20):3089–3110.

11. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall, Inc.; 1993.
12. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167(4):492–499.
13. Lash TL, Cole SR. Immortal person-time in studies of cancer outcomes. *J Clin Oncol*. 2009;27(23):e55–e56.
14. Cox DR. Regression models and life-tables [with discussion]. *J R Stat Soc Series B Stat Methodol*. 1972;34(2):187–220.
15. Brinkhof MW, Spycher BD, Yiannoutsos C, et al. Adjusting mortality for loss to follow-up: analysis of five ART programmes in sub-Saharan Africa. *PLoS One*. 2010;5(11): e14149.
16. Schomaker M, Gsponer T, Estill J, et al. Non-ignorable loss to follow-up: correcting mortality estimates based on additional outcome ascertainment. *Stat Med*. 2014;33(1):129–142.
17. Siannis F. Sensitivity analysis for multiple right censoring processes: investigating mortality in psoriatic arthritis. *Stat Med*. 2011;30(4):356–367.
18. Binder N, Schumacher M. Missing information caused by death leads to bias in relative risk estimates. *J Clin Epidemiol*. 2014;67(10):1111–1120.
19. Griffin JT, Fraser C, Gras L, et al. The effect on treatment comparisons of different measurement frequencies in human immunodeficiency virus observational databases. *Am J Epidemiol*. 2006;163(7):676–683.
20. Hernán MA, McAdams M, McGrath N, et al. Observation plans in longitudinal studies with time-varying treatments. *Stat Methods Med Res*. 2009;18(1):27–52.
21. Nevo ON, Lesko CR, Colwell B, et al. Outcomes of pharmacist-assisted management of antiretroviral therapy in patients with HIV infection: a risk-adjusted analysis. *Am J Health Syst Pharm*. 2015;72(17):1463–1470.
22. Allignol A, Schumacher M, Wanner C, et al. Understanding competing risks: a simulation point of view. *BMC Med Res Methodol*. 2011;11:86.
23. Lau B, Cole SR, Moore RD, et al. Evaluating competing adverse and beneficial outcomes using a mixture model. *Stat Med*. 2008;27(21):4313–4327.
24. Lesko CR, Lau B. Bias due to confounders for the exposure-competing risk relationship. *Epidemiology*. 2017;28(1):20–27.
25. Lau B, Gange SJ, Kirk GD, et al. Evaluation of human immunodeficiency virus biomarkers: inferences from interval and clinical cohort studies. *Epidemiology*. 2009;20(5):664–672.
26. Kyle RP, Moodie EE, Klein MB, et al. Correcting for measurement error in time-varying covariates in marginal structural models. *Am J Epidemiol*. 2016;184(3):249–258.
27. McCaffrey DF, Lockwood JR, Setodji CM. Inverse probability weighting with error-prone covariates. *Biometrika*. 2013; 100(3):671–680.
28. Regier MD, Moodie EE, Platt RW. The effect of error-in-confounders on the estimation of the causal parameter when using marginal structural models and inverse probability-of-treatment weights: a simulation study. *Int J Biostat*. 2014;10(1):1–15.
29. Edwards JK, Cole SR, Westreich D, et al. Multiple imputation to account for measurement error in marginal structural models. *Epidemiology*. 2015;26(5):645–652.
30. Cole SR, Jacobson LP, Tien PC, et al. Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death. *Am J Epidemiol*. 2010;171(1):113–122.