2014

Worth the weight: Using Inverse Probability Weighted Cox Models in AIDS Research

Ashley L. Buchanan
University of Rhode Island, buchanan@uri.edu

Michael G. Hudgens

Stephen R. Cole

Bryan Lau

Adaora A. Adimora

Citation/Publisher Attribution

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Lau, B., & Adimora, A. A.. (2014). Worth the Weight: Using Inverse Probability Weighted Cox Models in AIDS Research. *AIDS Research and Human Retroviruses, 30*(12): 1170-1177.doi: 10.1089/aid.2014.0037
Available at: https://doi.org/10.1089/aid.2014.0037

Worth the weight: Using inverse probability weighted Cox models in AIDS research

Short Title: IP-weighted Cox models

Ashley L. BUCHANAN[1]

Michael G. HUDGENS[1]

Stephen R. COLE[2]

Bryan LAU[3]

Adaora A. ADIMORA[4]

For the Women's Interagency HIV Study

Departments of Biostatistics[1], Epidemiology[2], and Department of Internal Medicine[4],

University of North Carolina, Chapel Hill, NC 27599

Department of Epidemiology[3], Johns Hopkins, Baltimore, MD 21287

Funding for this study was provided by NIH grants U01AI103390, R01AI100654,

U01AI042590, U01AI069918, R56AI102622, and P30AI50410.

Text (abstract) word count: 3206/3500 (151/250)

Title: 69/120

Short title: 20/40

Abstract

Survival analysis methods such as Cox regression can be used in infectious disease research to compare the timing of clinical events between treatment or exposure groups. Randomized clinical trials are the gold standard for estimating the effect of a treatment or exposure on a survival time endpoint. However, clinical trials are not always ethical or feasible. In that case, inference about the effect of interest might be attempted using data from observational studies. Unfortunately, observational studies may be riddled with confounding which can cast doubt on the validity of the results. In this tutorial, we demonstrate how inverse probability weighted Cox models can be used to account for multiple measured confounders, while concentrating inferences on the treatment or exposure effects of central interest and providing graphical summaries of these effects. This approach is illustrated using an example that estimates the effect of injection drug use on AIDS-free survival among HIV-infected women.

Key Words: Bias; Censoring; Cohort studies; Confounding; Cox proportional hazards model; Inverse probability weights; Standardization; Survival analysis; Time-to-event

Objective: The primary objective of this paper is to familiarize the reader with inverse probability weighted Cox models.

Survival analysis can be used in infectious disease research to compare the time to occurrence of clinical events between treatment or exposure groups (1). Randomized trials are the gold standard to estimate exposure effects on survival time, but are not always ethical or feasible. Although observational studies may provide estimates of effects when trial data are unavailable, the estimates they yield are often riddled with confounding (2). Informally, confounding occurs when the exposure and outcome share a common cause. The Cox proportional hazards regression model (3), the standard approach in survival analysis, can account for multiple measured confounders. Unfortunately, the Cox model provides only a single summary measure (i.e., hazard ratio), which can be difficult to interpret (4).

As an alternative to the standard Cox model, we present a method in this paper that uses inverse probability (IP) weights to estimate the effect of an exposure that is fixed at study entry. Under certain assumptions, this method can be used to mimic a randomized trial when only observational data is available. In particular, unlike the standard Cox model, this approach allows for estimation of marginal effects which compare the distribution of outcomes when the entire population is exposed versus when the entire population is unexposed (5). This IP-weighted approach naturally leads to Kaplan-Meier (6) type survival curve estimates that account for confounding by multiple covariates (7, 8). Herein, we refer to IP weighting as standardization, where the standardization is to the entire population under two different exposures (7, 9). We illustrate this standardization method through an example that estimates the effect of injection drug use (IDU) on AIDS-free survival among HIV-infected women.

Motivating Example: AIDS-Free Survival Among Injection Drug Users

The Women's Interagency HIV Study (WIHS) is a prospective, observational, multicenter study of women living with HIV and women at-risk of HIV infection in the US (10). A total of 4,129 women (1,065 HIV-uninfected) were enrolled between October 1994 and December 2012 at six US sites. An institutional review board at each site approved study procedures and all study participants provided written informed consent. We were interested in determining if AIDS-free survival among HIV-infected women differed by IDU, accounting for possible confounding by factors measured at baseline and during study follow-up. We estimated the hazard ratio and the absolute risk difference at ten years to quantify this effect.

The study sample consisted of 1,164 women enrolled in WIHS who were alive, HIV-infected, and free of AIDS on 6 December 1995 (11). The endpoint was either death or a diagnosis of AIDS. Women who did not reach this endpoint by 6 December 2005 were censored at that time or at their last visit where they were known to be alive and AIDS-free, whichever came first. A history of IDU at WIHS enrollment is denoted as $X = 1$ ($X = 0$ otherwise). Denoted as the list (or vector) $\boldsymbol{Z}$, the baseline covariates are: African American race, age, and nadir CD4 count (in cells/uL) measured from WIHS enrollment to baseline (i.e., 6 December 1995). Also, let $Z(t)$ denote the time-varying covariate antiretroviral (ART) initiation during study follow-up, where $Z(t) = 1$ if an individual starts ART by time t since baseline and $Z(t) = 0$ otherwise.

Inverse Probability Weighted Cox Models

Researchers are often interested in estimating effects of an exposure fixed at study entry. IP-weighted Cox models are a method to compare the timing of clinical events under two different exposures, mimicking results in randomized trials. An IP-weighted Cox model is fit by maximizing a weighted partial likelihood, where participant *i* who died or was diagnosed with AIDS at time *t* from baseline contributes the term $\{\exp(\beta X_i) / \sum_{j \in R(t)} \widehat{w}_j(t) \exp(\beta X_j)\}^{\widehat{w}_i(t)}$, where $R(t)$ is the risk set at time t and $\exp(\beta)$ is the hazard ratio for a unit difference in exposure X accounting for confounding measured by covariates through the estimated IP weight $\widehat{w}_i(t)$ (discussed below) (12). Slight modification of the likelihood is needed in the presence of tied survival times. The robust variance estimator (13) can be employed to account for the fact that the IP weights are estimated (14). See the Appendix A for a review of inference for the standard (i.e., unweighted) Cox proportional hazards model.

The estimated IP weight $\widehat{w}_i(t)$ is the product of an estimated time-fixed IP exposure weight $\widehat{w}_{1i}$ and an estimated time-varying IP drop out weight $\widehat{w}_{2i}(t)$ for each participant i at each survival time *t*. The time-fixed IP exposure weights are constructed to account for confounding by covariates measured at baseline. Different versions of these weights have been proposed. We recommend the stabilized IP exposure weight w_{1i} defined as the ratio of the marginal probability of having the exposure that participant *i* had, formally $P(X_i = x_i)$, to the covariate-conditional probability of having the exposure that participant *i* had, formally $P(X_i = x_i|\boldsymbol{Z}_i)$, where $\boldsymbol{Z}_i$ are the measured covariates for participant *i* assumed sufficient to adjust for confounding. If we do not appropriately adjust for confounding, the estimated association between the exposure and study outcome may be far from the truth (i.e., biased). Because these IP weights are unknown, the probabilities of

exposure are estimated using the observed data. Estimation details are provided in the following section tailored to the example.

The time-varying IP drop out weights are constructed to account for possible selection bias due to drop out (12). Participants last observed alive and AIDS-free more than one year prior to 6 December 2005 were considered drop outs (i.e., loss to follow-up). Participants receive a time-varying weight that corresponds to their probability of remaining free from drop out. This stabilized IP weight $w_{2i}(t)$ is defined as the ratio of the marginal probability of remaining free of drop out, formally $P(D_i > t|X_i)$, where D_i is the time from baseline to drop out for participant *i*, to the covariate-conditional probability of remaining free of drop out, formally $P(D_i > t|\boldsymbol{Z}_i, Z_i(t), X_i)$, where $\boldsymbol{Z}_i$ and $Z_i(t)$ are the measured common causes of drop out and the study outcome for participant *i* up to time *t*. (Note the covariates in the drop out model can be different than the covariates in the exposure weight model). If we do not appropriately adjust for the common (time-varying) causes of drop out and study outcome, the estimated association between the exposure and outcome may be biased due to drop out. Again, because these IP drop out weights are unknown, the probabilities of remaining free of drop out are estimated using the observed data. Estimation details are provided in the following section tailored to the example.

Standardized survival curve estimates can be obtained by fitting an IP-weighted Cox model stratified by exposure with no covariates and then nonparametrically estimating the baseline survival functions for the two strata (7). In the absence of weighting, these survival curve estimates will be (asymptotically) equivalent to Kaplan-Meier estimates obtained separately for each of the exposure strata (15).

For all Cox models presented below, we employed Efron's method to account for events that occurred on the same date(16). We obtained confidence intervals for the risk difference at 10 years using a nonparametric bootstrap using 200 random samples with replacement (17). The data analysis for this paper was conducted using SAS software version 9.3 (SAS Institute Inc., Cary, NC). SAS code for analyses in the present paper is provided in the Electronic Supplement.

Illustrative Example

The 1,164 women were 58% African American, median age was 36 years, and median nadir CD4 count was 349 cells/uL at baseline (Table 1). At enrollment, 38% of women reported a history of IDU. During follow-up, 667 (57%) of women initiated ARTs. Women were followed for up to 10 years with a total of 7,090 person-years during which 579 (50%) developed AIDS or died, and 117 (10%) dropped out of the study.

In analyses that did not account for covariates, women with a history of IDU had notably worse AIDS-free survival than women without a history of IDU (Figure 1). The estimated hazard ratio from the unadjusted Cox model was 1.72 (95% confidence interval (CI): 1.46, 2.03; Wald P value < 0.001), suggesting that the hazard of AIDS or death for those with a history of IDU was almost twice the hazard of those without a history of IDU (Table 2). We assessed the proportional hazards assumption graphically by examining whether the log cumulative hazard function estimates (Supplemental Figure 1) were approximately parallel. We also assessed this assumption statistically by inclusion of a product term between history of IDU and time in the Cox model, for which the Wald P value

was 0.40. Neither graphical nor statistical assessment suggested a meaningful departure from proportional hazards.

We then obtained a standardized hazard ratio estimate from the IP-weighted Cox model, which involved two steps. In the first step, using separate logistic regression models, weights were estimated for the probability of exposure (i.e., history of IDU) and for the probability of not dropping out. For the exposure weights, we fit logistic regression models for both the numerator and denominator. The exposure model for the numerator had no covariates, while the exposure model for the denominator included age, race, and nadir CD4 count, as well as all pairwise interactions. Age and nadir CD4 were included as continuous variables using restricted quadratic splines with four knots placed at 5^{th}, 35^{th}, 65^{th}, and 95^{th} percentiles (18). For the drop out weights, time was coarsened into months since baseline (19). Then, using pooled logistic regression(20), the drop out model for the numerator included only exposure (i.e., history of IDU) and time (using restricted quadratic splines), while the drop out model for the denominator included exposure, time (spline), age (spline), race, nadir CD4 count (spline), and ART initiation (time-varying), as well as all pairwise interactions. In the pooled logistic regression model, each person contributed up to 120 records and the weights were cumulatively multiplied for each person. The estimated weights $\widehat{w}_i(t)$ had a mean of 1.00 (with a standard deviation of 0.70), and ranged from 0.46 to 10.85 (Supplemental Table 1). In the second step, the IP-weighted Cox model was fit by weighting participants by their estimated weights, with outcome time to AIDS or death, and history of IDU as the sole covariate.

We obtained the estimated survival functions from an IP-weighted Cox model with no covariates stratified by history of IDU. After standardization for confounding and drop

out by IP weighting, survival curves (Figure 2) showed an attenuated difference in AIDS-free survival compared to the survival curves without accounting for any covariates (Figure 1). Under certain assumptions discussed below, the dashed curve can be interpreted as an estimate of the AIDS-free survival if (contrary to fact) everyone had a history of IDU at enrollment, while the solid curve can be interpreted as an estimate of the AIDS-free survival if (contrary to fact) no one had a history of IDU at enrollment (7, 8). The standardized hazard ratio from the IP-weighted Cox model was 1.53 (95% CI: 1.27, 1.85; Wald P value < 0.001) (Table 2). We again assessed the proportional hazards assumption graphically by examining whether the IP-weighted log cumulative hazard function estimates (Supplemental Figure 2) were approximately parallel. We also assessed this assumption statistically by inclusion of a product term between history of IDU and time, for which the Wald P value was 0.11. Neither graphical nor statistical assessment suggested a meaningful departure from proportional hazards. From the standardized survival curves, the ten-year risk of AIDS or death was 0.60 if (contrary to fact) everyone had a history of IDU at enrollment and 0.46 if (contrary to fact) no one had a history of IDU at enrollment. The 10-year risk difference was 0.14 (bootstrap 95% CI: 0.06, 0.22). For comparison, we also estimated a covariate-adjusted hazard ratio by including history of IDU, age (spline), race, and nadir CD4 count (spline) directly in an unweighted Cox model. The covariate-adjusted hazard ratio estimate was 1.62 (95% CI: 1.35, 1.95; Wald P value < 0.001).

Discussion

IP-weighted Cox models and standardized survival curves were presented as a method to compare the timing of clinical events for two different exposure conditions,

mimicking results in randomized trials under certain assumptions. We compare this method to the traditional covariate-adjusted Cox model and discuss assumptions and caveats below.

Although hazard ratio estimates from the two approaches were comparable in the WIHS example above, the standardized (i.e., IP-weighted) method provides several potential benefits over the covariate-adjusted method. First, the standardized approach can be used to mimic a randomized trial when only observational data is available (under certain assumptions discussed below). In particular, the estimated hazard ratio using the standardized approach can be interpreted the same as the (marginal) hazard ratio one would obtain in a randomized experiment such as a clinical trial where there is no confounding. In contrast, a covariate-adjusted Cox model hazard ratio does not necessarily equal the marginal hazard ratio, even in the absence of confounding, because the Cox model is not collapsible for the hazard ratio parameter(21). A regression model is said to be collapsible for a parameter (in this case, the hazard ratio) if the covariate-adjusted parameter is the same as the unadjusted parameter (22).

Second, the IP weighting approach yields standardized survival curve estimates. Although the hazard ratio is a common summary parameter to compare survival distributions between exposure groups, there are drawbacks to focusing inference on hazard ratios. For instance, the hazard ratio can be difficult to interpret, especially when trying to summarize the effect of a treatment or exposure (4). Presenting estimated survival curves is an alternative to reporting hazard ratios that may be more interpretable because survival curves summarize all information from baseline up to any time *t*. The IP-weighted approach leads to Kaplan-Meier type survival curve estimates that are

standardized to the entire population under two different exposures at baseline while accounting for confounding by multiple covariates. A covariate-adjusted Cox model does not afford such survival curve estimates (5, 8).

Third, the IP-weighted approach with drop out weights requires a weaker assumption about censoring than the Cox model. Specifically, if there are measured time-varying covariates predictive of censoring and survival time, the IP-weighted approach will yield consistent estimates of the hazard ratio, while the covariate-adjusted approach will not (15, 23). The use of IP drop out weights also yields estimators that are more efficient (i.e., less variable) than those from the covariate-adjusted Cox model, even when there is no selection bias (23).

Estimation of the hazard ratio and survival curves using standardization by IP weights requires certain assumptions to yield valid inference about the exposure effect. In particular, this approach assumes positivity, well-defined exposures, correctly specified models, and no unmeasured confounding or selection bias. For each level defined by the covariates, positivity means that there is a positive probability of each level of exposure (14). Well-defined exposures imply that there are not multiple versions of exposure, or if there are, that they are unimportant (14). The standardized hazard ratio estimator and survival curves require correctly specified IP weights (i.e., correct covariate functional forms). It is also assumed that sufficient sets of covariates have been measured to effectively address confounding (i.e., no unmeasured confounding) (7, 12) and selection bias due to drop out (23).

Drop out weights can be included in the IP-weighted Cox model to adjust for baseline and time-dependent covariates predictive of both censoring and survival time (24-

26). In the example, the time-varying covariate ART initiation was not included in the covariate-adjusted Cox model. Typically, when assessing the effect of a baseline exposure, one would not adjust for post-baseline covariates because such covariates may be on the causal pathway from the exposure to the outcome. Thus, adjusting for post-baseline covariates may lead to attenuated estimates of the total effect of the exposure (26-28). On the other hand, time-varying ART initiation may be predictive of both drop out and the survival time, so excluding that variable from the Cox model has the potential to introduce selection bias. In contrast, the use of IP drop out weights provides a valid approach to adjusting for a time-varying covariate associated with drop out and survival (19).

We only discussed exposure groups defined at baseline. When interest focuses on exposures that change over time, methods must be adapted accordingly. When a time-varying confounder is a risk factor for the outcome, predicts later exposure, and is affected by prior exposure, standard statistical methods (e.g., Cox models with time-varying covariates) are biased and fail to provide consistent estimators of effects (29-32). IP weighting can be generalized to account for time-varying confounders (12). For example, in HIV-infected individuals, CD4 count is a risk factor for death, predicts subsequent treatment with antiretroviral therapy, and is affected by prior treatment; thus, the IP-weighted Cox model is appropriate for studying the effect of time-varying antiretroviral therapy on overall survival while adjusting for time-varying CD4 count.

We suggest using expert knowledge to determine which covariates to adjust for prior to model fitting. Many epidemiologists would retain a possible confounder if its inclusion changes the estimate of association by more than 10% or 20% and a great deal of precision is not sacrificed (33). More principled approaches for determining which

covariates to adjust for in a model include conditioning on all causes of the exposure or outcome (34) and constructing causal directed acyclic graphs (based on a priori beliefs or knowledge) to posit a sufficient set of covariates to block all back door paths (35). For the weight models, covariates that are unrelated to the exposure but related to the outcome yield effect estimates with smaller variance and no increase in bias, so they should be included in the model; however, covariates that are related to the exposure but not to the outcome lead to effect estimates with larger variance and no reduction in bias, so they should be excluded from the model (36). Machine learning techniques (37, 38) can be used as an alternative approach to logistic regression for estimating weights.

Although the IP-weighted method used to analyze the WIHS data attempts to adjust for confounding and selection bias, the conclusions from the analysis are still subject to the following considerations. Comparisons of groups from observational studies may be subject to unmeasured confounding bias, as the assumption of no unmeasured confounding is untestable. Similarly, the IP-weighted method assumes drop out is independent of the survival time conditional on observed baseline and time-varying covariates. The absence of unmeasured covariates predictive of both the censoring mechanism and survival time is also an untestable assumption. Finally, as with all methods, error in the measurement of exposure, covariates, or the event status or times could bias the results(39).

In conclusion, we have presented an example of survival data pertinent to infectious disease research and illustrated how to compare groups of study participants using the IP-weighted Cox proportional hazards model. The methods presented here, and in the prior companion review paper (1), have broad applicability in infectious disease research. Careful use of this and other methods for survival analysis will continue to enrich the

evidence base in the field of infectious diseases by providing answers to questions that are difficult or impossible to answer well without explicitly accounting for time. Inverse probability weighted Cox models provide a method to estimate covariate-standardized hazard ratios and survival curves in observational studies, and obtain information about effects of treatments or exposures to prevent infectious diseases or their sequela.

Appendix: Review of the Standard (Unweighted) Cox Proportional Hazards Model

Let uppercase letters denote random variables and lowercase letters possible realizations of random variables or constants. Let $i = 1, \ldots, n$ index the study participants. Let T_i be the time from baseline to AIDS diagnosis or death, D_i be the time from baseline to study drop out, and C_i be the time from baseline to administrative censoring. In practice, only the minimum of T_i, D_i, and C_i is observed, denoted by $T_i^* = \min(T_i, D_i, C_i)$. See the first paper in this tutorial series for a review of survival, hazard, and log cumulative hazard functions (1).

The Cox proportional hazards regression model (3) is one of the most widely used statistical methods in biomedical research. The univariate Cox model is defined as $h_i(t) = h_0(t)\exp(\beta X_i)$, where $h_i(t)$ is the hazard function for individuals with covariate X_i, $h_0(t)$ is the reference hazard at time t for those with $X_i = 0$, and β is the log hazard ratio for a one unit change in X_i.

Heuristically, Cox regression may be understood as a series of logistic regression models, where at each ordered survival time, the log odds of the event are regressed on the exposure groups and any covariates (16). The Cox model is a semiparametric model because no assumption is placed on the probability distribution for the reference survival time distribution. Equivalently, the function $h_0(t)$ is left arbitrary. The parameters of a Cox model are estimated using maximum partial likelihood (40). Assuming no tied survival times, participant *i* who had the event at time *t* contributes the term $\exp(\beta X_i)$ / $\sum_{j \in R(t)} \exp(\beta X_j)$ to the partial likelihood function, where $R(t)$ is the set of participants at risk at time *t*. For the case of a single covariate X_i, the partial likelihood is defined as simply a product of these individual contributions for events, or

$L(\beta) = \prod_{i=1}^{n} \left[\frac{\exp(\beta X_i)}{\sum_{j \in R(T_i)} \exp(\beta X_j)} \right]^{Y_i}$, where Y_i is an event indicator (i.e., $T_i^* = T_i$). Only events contribute to the numerator of the likelihood due to the exponent Y_i. There are several ways to handle tied survival times, including methods ascribed to Peto and Breslow (41, 42), Efron (16) and an exact approach (26), which all return the same results if there are no ties. In the presence of moderate ties and if time is truly continuous, Efron's approximation performs well compared to the other approaches (43).

One of the central assumptions of the Cox model is that the ratios of the hazards defined by levels of the covariates are constant over time. This is the proportional hazards assumption. The proportional hazards assumption can be assessed by fitting the model $h(t) = h_0(t)\exp(\beta_1 X_i + \beta_2 X_i t)$ and testing the null hypothesis that $\beta_2 = 0$, where $X_i t$ is a product of the covariate and time *t*, or some monotonic function of time, such as $\log(t)$.

In general, a $1 - \alpha$ Wald confidence interval (CI) for the hazard ratio is defined as $\exp\left(\hat{\beta} \pm z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\beta})}\right)$, where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ percentile of a standard normal distribution and $\hat{V}(\hat{\beta})$ is the estimated variance of $\hat{\beta}$. A Wald test statistic is defined as $\left(\frac{\hat{\beta}}{\sqrt{\hat{V}(\hat{\beta})}}\right)^2$ and is chi-squared distributed with 1 degree of freedom under the null hypothesis $\beta = 0$.

Acknowledgements

These findings are presented on behalf of the Women's Interagency HIV Study (WIHS). We would like to thank all of the WIHS investigators, data management teams, and patients who contributed to this project. Funding for this study was provided by NIH grants U01AI103390, R01AI100654, U01AI042590, U01AI069918, R56AI102622, and P30AI50410. The views and opinions of authors expressed in this manuscript do not necessarily state or reflect those of the NIH.

References

1. Cole SR, Hudgens MG. Survival analysis in infectious disease research: Describing events in time. AIDS. 2010;24(16):2423.

2. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22(1):189-212.

3. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society Series B (Methodological). 1972;34(2):187-220.

4. Hernán MA. The hazards of hazard ratios. Epidemiology. 2010;21(1):13-5.

5. Kaufman JS. Marginalia: Comparing adjusted effect measures. Epidemiology. 2010;21(4):490-3.

6. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958;53(282):457-81.

7. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed. 2004;75(1):45-9.

8. 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):3089-110.

9. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. Epidemiology. 2003;14(6):680-6.

10. Bacon MC, von Wyl V, Alden C, Sharp G, Robison E, Hessol N, et al. The Women's Interagency HIV Study: An observational cohort brings clinical sciences to the bench. Clin Diagn Lab Immunol. 2005;12(9):1013-9.

11. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009;170(2):244-56.

12. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-60.

13. Lin D, Wei L. The robust inference for the Cox proportional hazards model. Journal of the American Statistical Association. 1989;84(408):1074-8.

14. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168(6):656-64.

15. Collett D. Modelling Survival Data in Medical Research. Boca Raton: CRC press; 2003.

16. Efron B. The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association. 1977;72(359):557-65.

17. Efron B, Tibshriani R. An Introduction to the Bootstrap. London: Chapman Hall. 1994.

18. Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, Eron Jr JJ. Splines for trend analysis and continuous confounder control. Epidemiology. 2011;22(6):874-75.

19. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. Journal of the American Statistical Association. 2001;96(454):440-8.

20. D'Agostino RB, Lee M, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. Stat Med. 1990;9(12):1501-15.

21. Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. Epidemiology. 1996;7(5):498-501.

22. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Statistical Science. 1999;14(1):29-46.

23. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted log rank tests. Biometrics. 2000;56(3):779-88.

24. Hernán MÁ, 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):561-70.

25. Howe CJ, Cole SR, Chmiel JS, Muñoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. Am J Epidemiol. 2011;173(5):569-77.

26. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. Hoboken: Wiley-Interscience; 2002.

27. Cole SR, Hernán MA. Fallibility in estimating direct effects. Int J Epidemiol. 2002;31(1):163-5.

28. Pearl J. Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence; Morgan Kaufmann Publishers Inc.; 2001.

29. Cole SR, Hernán MA, Robins JM, Anastos K, Chmiel J, Detels R, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. Am J Epidemiol. 2003;158(7):687-94.

30. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. Journal of the American Statistical Association. 2001;96(454):440-8.

31. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Statistical Models in Epidemiology, the Environment, and Clinical Trials. Springer; 2000. p. 95-133.

32. Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med. 2013;159(8):560-62.

33. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol. 1989;129(1):125-37.

34. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics. 2011;67(4):1406-13.

35. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37-48.

36. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol. 2006;163(12):1149-56.

37. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. PloS One. 2011;6(3):e18174.

38. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Stat Med. 2010;29(3):337-46.

39. Hernán MA, Cole SR. Invited commentary: Causal diagrams and measurement bias. Am J Epidemiol. 2009;170(8):959-62.

40. Cox DR. Partial likelihood. Biometrika. 1975;62(2):269-76.

41. Peto R, Peto J. Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society Series A (General). 1972;135(2):185-207.

42. Breslow N. Covariance analysis of censored survival data. Biometrics. 1974;58(3):643-9.

43. Hertz-Picciotto I, Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. Biometrics. 1997;53(3):1151-6.

Table 1. Characteristics of 1,164 HIV-infected women in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005

Characteristics [a]	History of Injection Drug Use (IDU) $n = 439$	No History of Injection Drug Use (IDU) $n = 725$	Overall $n = 1{,}164$
Age years	40 (35, 44)	33 (29, 39)	36 (31, 41)
African American race	273 (62%)	399 (55%)	672 (58%)
Nadir CD4+ count cells/uL	352 (208, 522)	348 (216, 505)	349 (213, 517)
Initiated antiretrovirals (ARTs)	208 (47%)	459 (63%)	667 (57%)

[a] Median (interquartile range) or number (percent)

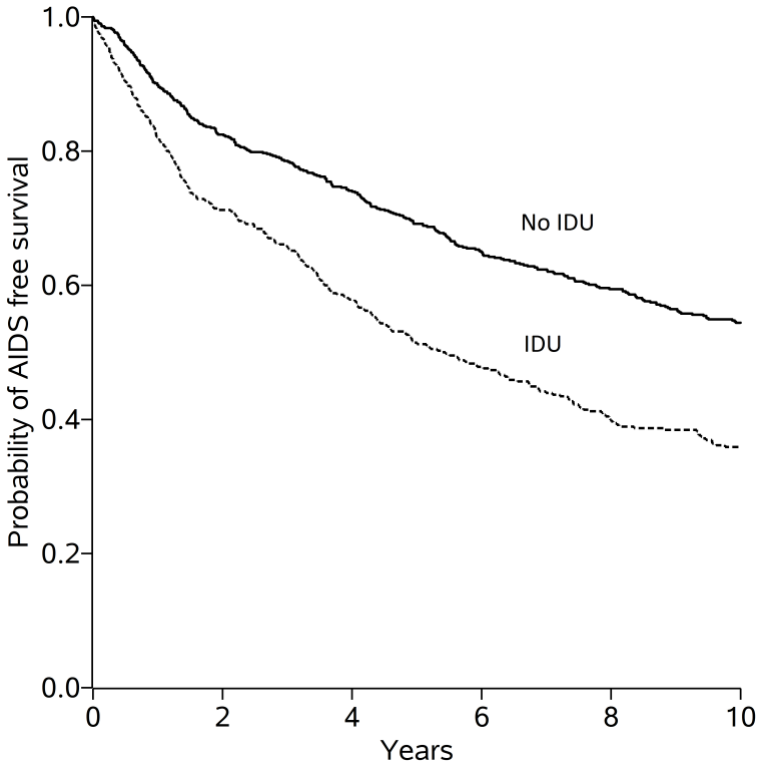
Table 2. Association of history of injection drug use with time to AIDS or death for 1,164 HIV-infected women in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005

	History of Injection Drug Use (IDU) $n = 439$	No History of Injection Drug Use (IDU) $n = 725$	Overall $n = 1{,}164$
Unadjusted			
AIDS cases and deaths	272 (62%)	307 (42%)	579 (50%)
Person-years	2,368	4,721	7,090
Hazard ratio (95% CI)	1.72 (1.46, 2.03)	1	-
10-year risk (95% CI)	0.64 (0.59, 0.68)	0.46 (0.42, 0.49)	0.53 (0.50, 0.56)
10-year risk difference (95% CI)	0.19 (0.13, 0.24)	0	-
Standardized [a]			
AIDS cases and deaths	272 (62%)	307 (42%)	579 (50%)
Person-years	9,804	20,692	30,496
Hazard ratio (95% CI)	1.53 (1.27, 1.85)	1	-
10-year risk (95% CI)	0.60 (0.54, 0.64)	0.46 (0.42, 0.50)	0.51 (0.48, 0.54)
10-year risk difference (95% CI)	0.14 (0.06, 0.22)	0	-

[a] IP weighted to account for confounding and selection bias due to age (spline), race, nadir CD4 (spline), and ART initiation (time-varying)

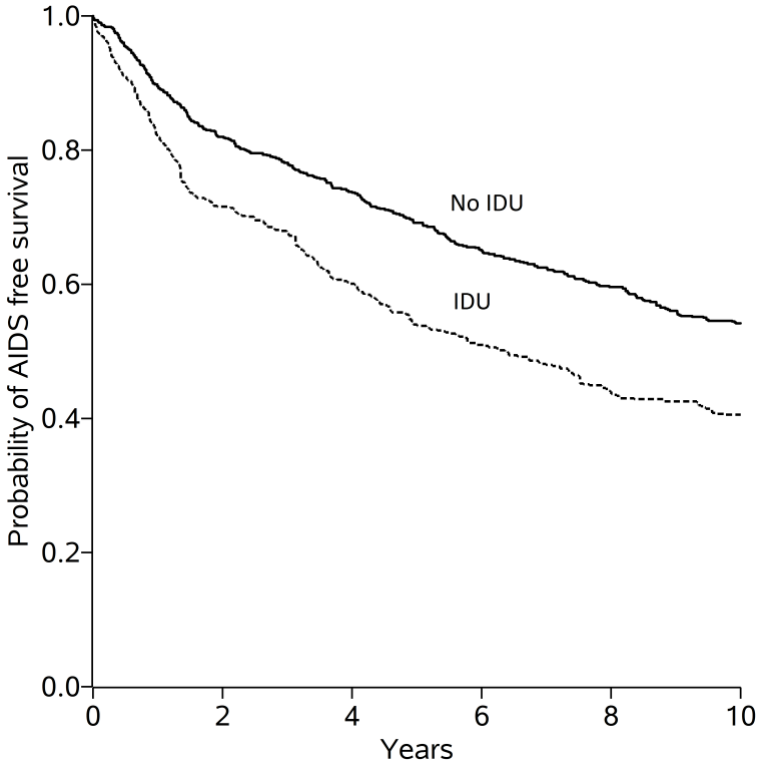
IP-weighted Cox models

Figure 1. Kaplan-Meier estimated AIDS-free survival curves without accounting for any covariates for 1,164 HIV-infected women with and without a history of injection drug use (IDU) in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005



IP-weighted Cox models

Figure 2. Standardized estimated AIDS-free survival curves (accounting for age, race, nadir CD4, and ART initiation) for 1,164 HIV-infected women with and without a history of injection drug use (IDU) in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005

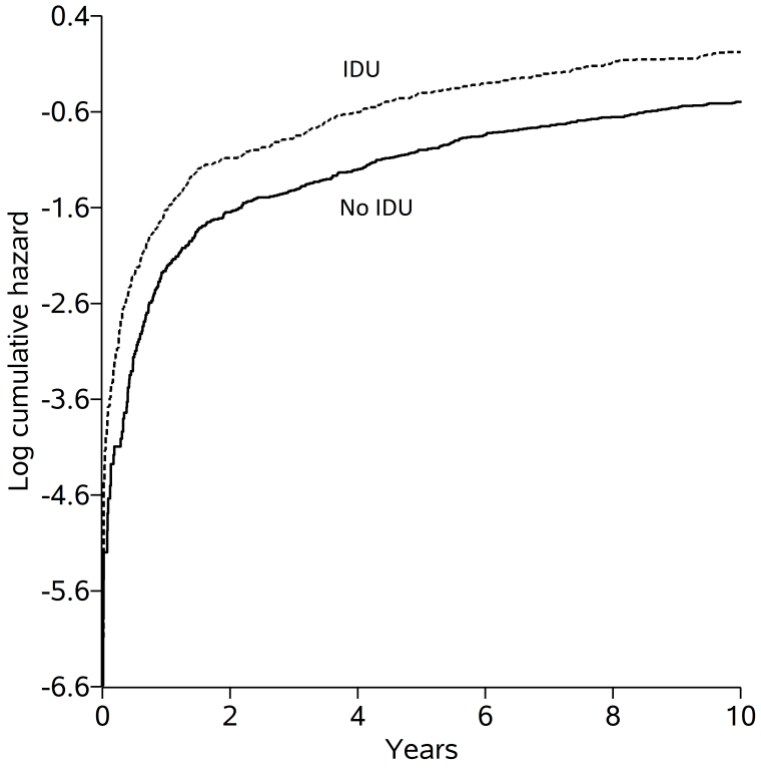


IP-weighted Cox models

Supplemental Table 1. Example individual-level estimated exposure weights, drop out weights, and combined weights for 1,164 HIV-infected women in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005

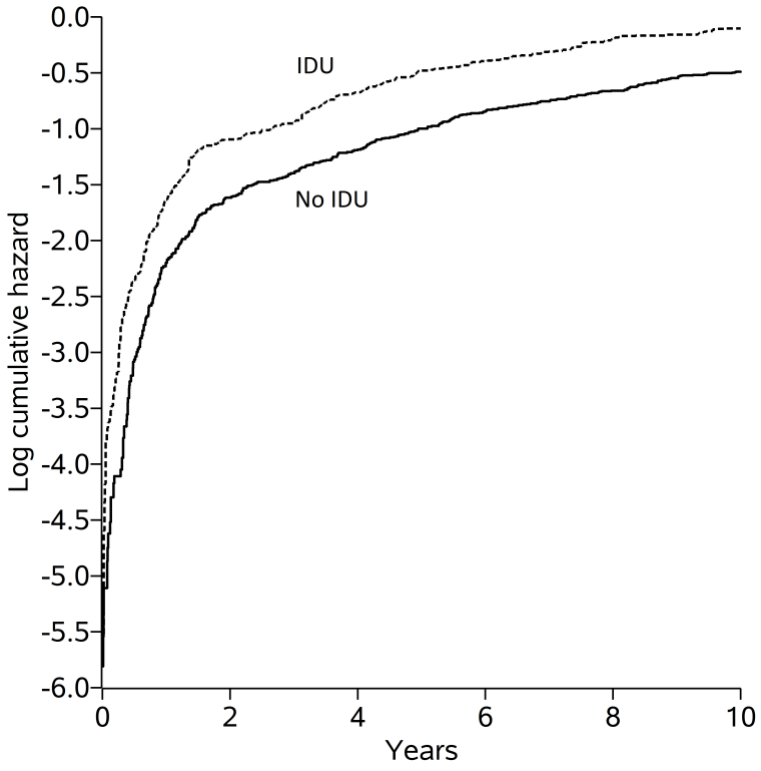
ID	Time in	Time out	History of Injection Drug Use (IDU)	Event	Drop out	Exposure weight	Drop out weight	Combined weight
34	0.00	0.080	Yes	No	No	0.515	1.000	0.515
34	0.08	0.170	Yes	No	No	0.515	1.000	0.515
34	0.17	0.208	Yes	No	Yes	0.515	1.001	0.516
36	0.00	0.080	No	No	No	1.191	1.000	1.191
36	0.08	0.170	No	No	No	1.191	1.000	1.191
36	0.17	0.225	No	No	Yes	1.191	1.001	1.192
37	0.00	0.080	Yes	No	No	1.007	1.000	1.007
37	0.08	0.170	Yes	No	No	1.007	1.000	1.007
37	0.17	0.227	Yes	Yes	No	1.007	1.000	1.007
38	0.00	0.080	No	No	No	0.949	1.000	0.949
38	0.08	0.170	No	No	No	0.949	1.000	0.949
38	0.17	0.250	No	No	No	0.949	0.999	0.948
38	0.25	0.330	No	No	No	0.949	0.999	0.948
38	0.33	0.420	No	No	No	0.949	0.998	0.947
38	0.42	0.427	No	Yes	No	0.949	0.997	0.946
66	0.00	0.080	Yes	No	No	0.545	1.000	0.545
66	0.08	0.170	Yes	No	No	0.545	1.000	0.545
66	0.17	0.250	Yes	No	No	0.545	0.999	0.545
66	0.25	0.330	Yes	No	No	0.545	0.999	0.544
66	0.33	0.378	Yes	No	Yes	0.545	0.998	0.544

Supplemental Figure 1. Estimated log cumulative hazard curves without accounting for any covariates calculated for 1,164 HIV-infected women with and without a history of injection drug use (IDU) in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005



IP-weighted Cox models

Supplemental Figure 2. Standardized estimates of the log cumulative hazard curves (accounting for age, race, nadir CD4, and ART initiation) calculated for 1,164 HIV-infected women with and without a history of injection drug use (IDU) in the Women's Interagency HIV Study December 6, 1995 through December 6, 2005



IP-weighted Cox models