Practice of Epidemiology

Time Scale and Adjusted Survival Curves for Marginal Structural Cox Models

Daniel Westreich, Stephen R. Cole*, Phyllis C. Tien, Joan S. Chmiel, Lawrence Kingsley, Michele Jonsson Funk, Kathryn Anastos, and Lisa P. Jacobson

* Correspondence to Dr. Stephen R. Cole, McGavran-Greenberg Hall, CB7435, Chapel Hill, NC 27599-7435 (e-mail: cole@unc.edu).

Typical applications of marginal structural time-to-event (e.g., Cox) models have used time on study as the time scale. Here, the authors illustrate use of time on treatment as an alternative time scale. In addition, a method is provided for estimating Kaplan-Meier–type survival curves for marginal structural models. For illustration, the authors estimate the total effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome (AIDS) or death in 1,498 US men and women infected with human immunodeficiency virus and followed for 6,556 person-years between 1995 and 2002; 323 incident cases of clinical AIDS and 59 deaths occurred. Of the remaining 1,116 participants, 77% were still under observation at the end of follow-up. By using time on study, the hazard ratio for AIDS or death comparing always with never using highly active antiretroviral therapy from the marginal structural model was 0.52 (95% confidence interval: 0.35, 0.76). By using time on treatment, the analogous hazard ratio was 0.44 (95% confidence interval: 0.32, 0.60). In time-to-event analyses, the choice of time scale may have a meaningful impact on estimates of association and precision. In the present example, use of time on treatment yielded a hazard ratio further from the null and more precise than use of time on study as the time scale.

acquired immunodeficiency syndrome; antiretroviral therapy, highly active; bias (epidemiology); causal inference; confounding factors (epidemiology); proportional hazards model; survival curve; survival time

In time-to-event data analysis, the epidemiologist must articulate the event of interest as well as the time origin and time scale. The "time origin" is the point at which each individual starts being at risk for the event of interest, and the "time scale" is the metric on which the risk sets are constructed and times to event are measured (1). For example, in a study where the event of interest is sexually acquired human immunodeficiency virus (HIV), the origin might be the age of sexual debut, with the associated time scale being duration of sexual activity.

Choice of origin and time scale (hereafter, collectively time scale) will not affect estimates of the rate ratio from Poisson regression models (unless those estimates are stratified by time); in such models, person-time is considered exchangeable, and the sum of person-years and number of events are invariant to the choice of time scale. However, the choice of time scale may affect estimates of the hazard ratio from Cox proportional hazards models (2, 3). Although the sum of person-years and number of events both again remain invariant, the composition of the risk sets, and therefore who serves as a comparator to whom, may differ with choice of time scale.

In randomized clinical trials where treatment is assigned at the beginning of the trial, a natural time origin is randomization date, and the associated time scale is time since randomization. Disregarding any run-in period, the time since randomization coincides with the time on study. Similarly, in this setting, the time on study typically coincides with time on treatment (for at least 1 study arm). In contrast, in observational cohort studies, the time on study does not

typically coincide with the time on treatment, though some have called for such new-user designs (4).

To date, implementations of marginal structural (5) Cox proportional hazards (6) models in observational cohort studies have typically used time on study as the time scale (7–17); these studies have not considered the role of time scale on estimates of effect (18, 19). Here, we compare the use of time on study and time on treatment as competing time scales in marginal structural Cox proportional hazards models. We also extend time-fixed inverse probability of treatment weighted adjusted survival curves (20, 21) to allow for 1) time-varying inverse probability of treatment and censoring (IPTC) weights and 2) time on treatment as the time scale.

MATERIALS AND METHODS

Study population

This analysis used information from the Multicenter AIDS Cohort Study (MACS) (22) that, beginning in 1984, enrolled 6,972 homosexual and bisexual men in Baltimore, Maryland; Chicago, Illinois; Pittsburgh, Pennsylvania; and Los Angeles, California, as well as from the Women's Interagency HIV Study (WIHS) (23) that, beginning in 1994, enrolled 3,772 women in New York, New York; Chicago, Illinois; Los Angeles and San Francisco, California; and Washington, DC. Every 6 months, participants in both studies completed an extensive interviewer-administered questionnaire regarding antiretroviral therapy use and provided a blood sample for the determination of CD4 cell count and HIV viral load. Exposures and covariates were updated at these 6-month study visits, while outcomes were under continuous monitoring. Positive enzyme-linked immunosorbent assays with confirmatory Western blots were used to determine HIV seropositivity. Institutional review boards approved all protocols and informed consent forms, which were completed by study participants in both cohorts.

This report focuses on reanalysis (9) of 1,498 participants who were HIV positive, were free of clinical acquired immunodeficiency syndrome (AIDS), and had not initiated highly active antiretroviral therapy (HAART) at a study visit by September 1995, just prior to the availability of HAART in the United States; follow-up began between September 1995 and March 2000. Participants were followed through April 2002 for primary diagnosis of clinical AIDS or death. Clinical AIDS was defined by 1993 Centers for Disease Control and Prevention criteria and required diagnosis of a clinical AIDS-defining illness; that is, a CD4 count of <200 cells/mm^3 or a CD4 percent of <14 by itself was not considered clinical AIDS (24). Deaths were ascertained by using death certificate abstractions upon notification and National Death Index searches, as described previously (9). Participants who missed 2 consecutive study visits were classified as dropouts and censored at the date of the last visit. Participants alive and free of clinical AIDS in April 2002 were administratively censored.

The primary exposure was initiation of treatment with HAART. The definition of HAART was based on recommendations from the Department of Health and Human Services and Kaiser Panel guidelines (25) and has been described previously (9). Once participants reported HAART initiation, they were assumed to have remained on HAART for the duration of follow-up, which correctly classified 94% of the observed person-time (9).

Statistical methods

We estimated the parameters of a marginal structural Cox proportional hazards model of the form

$$\lambda_{T^{\bar{x}}}(t) = \lambda_0(t) \times \exp(\alpha_1 X_{it} + \alpha_p' L_{i0}), \quad (1)$$

where X_{it} is a time-varying indicator of HAART initiation prior to week t for individual $i = 1$ to N and week $t = 1$ to T_i from study entry; T_i is the time (in weeks) from study entry to AIDS or death for participant i; $\bar{x} = \bar{x}_{iT} = \{x_{i1}, x_{i2}, \ldots, x_{it}, \ldots x_{iT}\}$ is one of $T + 1$ *possible* treatment histories under the intent-to-treat assumption, "once initiated, always treated," namely, never treated $\{0, 0, \ldots, 0\}$ to always treated $\{1, 1, \ldots, 1\}$; $T^{\bar{x}}$ is the *potential* time from study entry to AIDS or death under treatment history $\bar{x}$, and $\lambda_{T^{\bar{x}}}(t)$ is the *potential* hazard of AIDS or death at time t from study entry under treatment history $\bar{x}$, $\lambda_0(t)$ is an unspecified hazard of AIDS or death at time t from study entry in the reference (i.e., untreated) group, α_1 is the log hazard ratio comparing always treated with never treated, and α_p' is the transpose of the column vector of log hazard ratios for the p components of baseline covariate vector L_{i0}. This marginal structural model can be estimated by using IPTC weights W_{it} (defined in Appendix 1) under the assumptions of consistency (26, 27), positivity (28, 29), no unmeasured confounding or selection bias, and correct model specification as

$$\lambda_{T^{\bar{X}}}^{W_{it}}(t) = \lambda_0(t) \times \exp(\beta_1 X_{it} + \beta_p' L_{i0}), \quad (2)$$

where the superscript W_{it} indicates that the hazard of AIDS at time t from study entry for *observed* treatment history $\bar{X}$, that is, $\lambda_{T^{\bar{X}}}(t)$, is weighted by W_{it}, β_1 is the log hazard ratio, and $\lambda_{T^{\bar{X}}}^{W_{it}}(t) \to \lambda_{T^{\bar{x}}}(t)$, and therefore $\beta_1 \to \alpha_1$, if the above stated assumptions are met, as previously described (5, 8).

To accommodate time-varying IPTC weights, we approximate the marginal structural Cox model by a pooled logistic regression model (30) of the form

$$\log\{\Pr(Y_{it} = 1)^{W_{it}} / [1 - \Pr(Y_{it} = 1)^{W_{it}}]\}$$
$$= \delta_{0t} + \delta_1 X_{it} + \delta_p' L_{i0}, \quad (3)$$

where Y_{it} is an indicator of AIDS or death in week t for participant i; δ_1 is a discrete-time approximation to the continuous-time hazard ratio; and $\delta_1 \cong \beta_1$ if the risk of AIDS or death in any week is less than 10% (31). This condition is met in our example. Note that if one were to use the cumulative log-log link function in equation 3, rather than the logistic link function, the resulting δ_1 would be equivalent to the continuous-time hazard ratio, but the difference due to choice of link function is trivial in our case. Fitting equation 3 with δ_{0t} parameterized by a series of
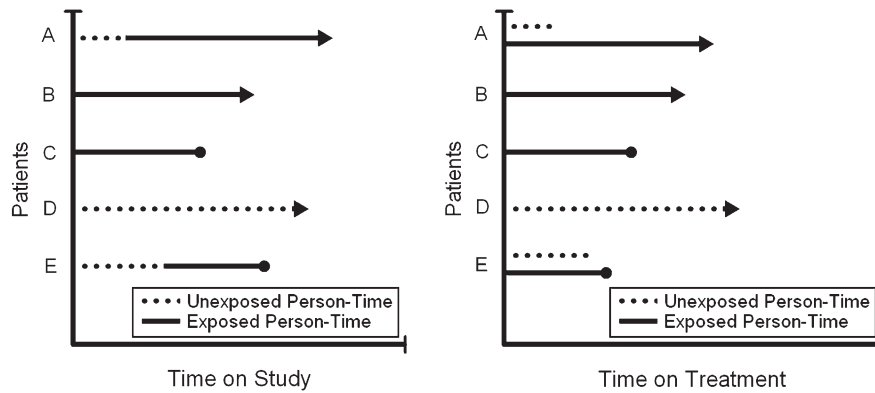
Figure 1.   Schematic diagram comparing time-on-study time scale and time-on-treatment time scale.

indicators for weeks allows the reference group hazard to be largely unrestricted (only assumed constant within weeks); we fit equation 3 by parameterzing δ_{0t} with a series of indicator variables for 6-month increments in follow-up. Note that, if one used the cumulative log-log link and parameterized δ_{0t} by using an intercept and either the single variable t or $\log(t)$, the resulting model is a discrete-time version of the Gompertz or Weibull survival model for t, respectively (32).

Under the assumptions of no unmeasured confounding or selection bias, the IPTC-weighted study population provides an unbiased representation of the association between exposure and outcome in that population (5). Therefore, 2-dimensional graphical depictions (e.g., histograms, survival curves, and so on) contrasting the weighted treatment groups provide an unbiased illustration of the causal contrast (20).

Define Q_i as time of first exposure (i.e., $Q_i = t$ if $x_{it-1} = 0$ and $x_{it} = 1$; otherwise $Q_i = T_i$). Formally, $Q_i = \left[\sum_{t=1}^{T_i} (1 - x_{it}) \right] + x_{iT_i}$. Q_i is therefore the time of HAART initiation for those who are observed to initiate HAART and is the time to AIDS, death, or censoring for those that do not initiate HAART. The person-time contributions using time on study as the time scale are then $[0, Q_i]$ among the untreated and $[Q_i, T_i]$ among the treated.

The IPTC-weighted Kaplan-Meier survival curves for the treated ($x = 1$) and untreated ($x = 0$) are then defined as

$$S_x(t) = \prod_t 1 - \frac{d_{tx}}{r_{tx}}, \quad (4)$$

where $d_{tx} = \sum_{i=1}^{N} W_{it} \times Y_{it} \times 1(X_{it} = x)$ is the weighted number of events for group x at week t, $1(\cdot)$ is the indicator function for the argument $\cdot$; similarly, $r_{tx} = \sum_{i=1}^{N} W_{it} \times 1(a < t \leq b) \times 1(X_{it} = x)$ is the weighted risk set at week t (20, 21), where $\{a, b\}$ are defined as $\{0, Q_i\}$ if $x = 0$ and $\{Q_i, T_i\}$ if $x = 1$. This is simply an IPTC-weighted version of the extended (to allow for late entries) Kaplan-Meier estimator (3) and can be calculated simply as the extended Kaplan-Meier in the weighted study population. The risk sets defined above allow a single individual to

contribute to both untreated and treated risk sets at disjoint times from study entry.

Now, we redefine the parameter we wish to estimate by considering time on treatment (i.e., time since initiation of HAART) V as an alternative to time on study T. Note that V does not represent baseline covariates. The person-time contributions are then $[0, Q_i]$ among the untreated and $[Q_i - Q_i, T_i - Q_i]$ among the treated. If participant i does not initiate HAART during follow-up (i.e., $X_{iT_i} = 0$), then $Q_i = T_i$ by definition, and participant i contributes only untreated person-time; otherwise, rewriting the treated time as $[0, T_i - Q_i]$ shows that individual i may contribute to both untreated and treated risk sets at the same time, in terms of V. For example, participant i may contribute in the risk set for the untreated at time 6 months from study entry and in the risk set for treated at time 6 months from HAART initiation. Under this alternative time scale, each measured unit of person-time appears only once in the analysis, although 2 units of person-time for the same individual may appear in the same risk set. Furthermore, because each person can contribute only a single event to the analysis, variance correction for repeated events is unnecessary. A schematic diagram comparing the 2 time scales is shown in Figure 1. As shown in Figure 1, this change in time scale is achieved in practice by "resetting the clock" for participants when they initiate HAART; here, contributions to untreated person-time are invariant by time scale.

Formally, a time v under time scale V relates to a time t under time scale T as a function of t, x_{it}, and Q_i as $v = [(1 - x_{it})(t)] + [(x_{it})(t - Q_i + 1)]$. A marginal structural Cox model analogous to equation 1 but using this alternative time scale V in place of T is

$$\lambda_{V^{\bar{x}}}(v) = \lambda_0(v) \times \exp(\phi_1 X_{iv} + \phi_p' L_{i0}), \quad (5)$$

where $\phi_1 \neq \alpha_1$, except in settings where the choice of time scale is irrelevant (refer to Discussion). Again, as in equation 2, we estimate ϕ under the assumptions of no unmeasured confounding or selection bias using IPTC weights, and we approximate ϕ by using a weighted pooled logistic model analogous to equation 3. We compared the precision of estimated hazard ratios using confidence interval ratios (i.e., the ratio of the upper to the lower confidence limit).

Table 1. Characteristics of 1,498 HIV-positive US Men and Women at Study Entry, 1995–2000

	Men (n = 506)		Women (n = 992)	
	No.	%	No.	%
Age[a]	42 (37, 46)		36 (31, 42)	
Caucasian race	398	78.7	163	16.4
Antiretroviral therapy				
None	330	65.2	568	57.3
Monotherapy	61	12.1	260	26.2
Combination therapy	115	22.7	164	16.5
HAART	0	0	0	0
CD4 count, cells/mm^3				
<50	9	1.8	38	3.8
50–199	61	12.1	150	15.1
200–350	129	25.5	244	24.6
>350	307	60.7	560	56.5
CD4 count, cells/mm^{3,a}	410 (275, 568)		389 (247, 553)	
HIV RNA, copies/mL				
≤400	73	14.4	349	35.2
401–10,000	124	24.5	148	14.9
>10,000	309	61.1	495	49.9
Log$_{10}$ HIV RNA, copies/mL^{a,b}	4.4 (3.9, 4.9)		4.6 (4.1, 5.5)	
≥1 HIV-related symptom[c]	105	20.8	291	29.3

Abbreviations: HAART, highly active antiretroviral therapy; HIV, human immunodeficiency virus.

[a] Median (quartiles).

[b] Among persons with detectable levels (i.e., >400 copies/mL).

[c] Persistent fever, diarrhea, night sweats, and weight loss.

IPTC-weighted Kaplan-Meier curves for this alternate time scale are created analogously to equation 4 but, among the treated, $\{a, b\}$ are defined as $\{0, T_i - Q_i\}$. SAS, version 9.1, software (SAS Institute, Inc., Cary, North Carolina) was used to implement the above-described methods, and code to implement the IPTC-weighted Kaplan-Meier curves is provided in Appendix 2. We plot the complement of the IPTC-weighted Kaplan-Meier survival curves as an estimate of the cumulative incidence.

In addition to the time scales discussed here, we controlled for calendar time and age in adjusted as well as weight and structural models using restricted cubic splines (with knots at the 5th, 35th, 65th, and 95th percentiles). Control for additional confounders and functional forms of weight models are described in Appendix 1.

RESULTS

Table 1 shows characteristics at study entry by gender/cohort for the 1,498 participants. The median age at study entry was 42 years among men and 36 years among women; 79% of men and 16% of women were Caucasian. About 14% of men and 19% of women entered this study with CD4 cell counts less than 200 cells/mm^3, while 61% and 57%, respectively, had CD4 counts greater than 350
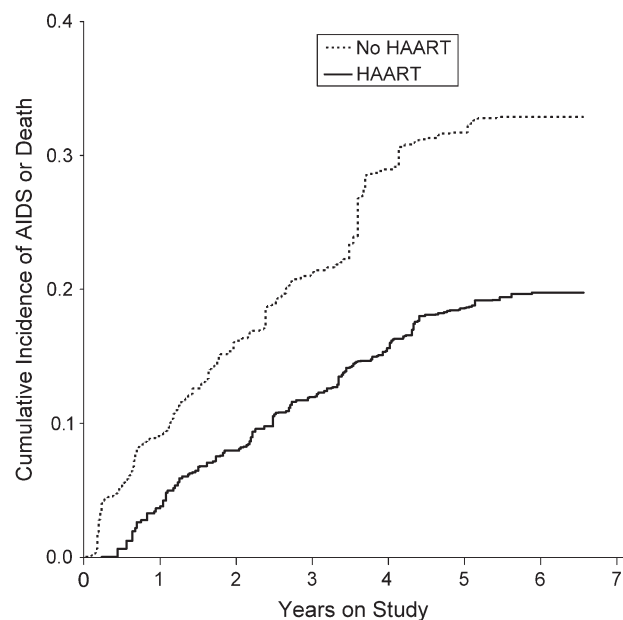
Figure 2. Inverse probability of treatment and censoring weighted cumulative incidence by treatment with highly active antiretroviral therapy (HAART) on time to acquired immunodeficiency syndrome (AIDS) or death for 1,498 human immunodeficiency virus-positive US men and women, 1995–2002. Time scale is time on study.

cells/mm^3. More women than men (35% vs. 14%) entered this study with plasma RNA levels less than or equal to 400 copies/mL. During follow-up, 323 incident cases of clinical AIDS and 59 deaths occurred, a rate of approximately 6 per 100 person-years for the combined endpoint. The remaining 1,116 participants were censored; 77% of those 1,116 were still under observation at study completion in April 2002.

Figure 2 shows the cumulative incidence by treatment group with the *x*-axis being time on study, while Figure 3 shows the cumulative incidence by treatment group with the *x*-axis being time on treatment. As noted in Materials and Methods, the untreated curve is identical in Figures 2 and 3, and both curves are adjusted for confounding and censoring by using IPTC weights.

Results from unadjusted, adjusted, and weighted (i.e., marginal structural) models are provided in Table 2 for both time scales. Using time on study, we observed no apparent benefit of treatment for unadjusted and adjusted models and a strong beneficial effect for the weighted model. Specifically, the hazard ratio from the marginal structural model was 0.52 (95% confidence interval (CI): 0.35, 0.76). These results replicated previously reported results (9). Previously published results were as follows: unadjusted hazard ratio = 0.98 (95% CI: 0.76, 1.26); adjusted hazard ratio = 0.81 (95% CI: 0.61, 1.07); and weighted hazard ratio = 0.54 (95% CI: 0.38, 0.78). The slight differences are due to differences in the coarseness of time categorization.

Using time on treatment, we observed essentially equivalent results in unadjusted and adjusted models and a notably stronger hazard ratio in the marginal structural model. However, in contrast to using time on study, with time on
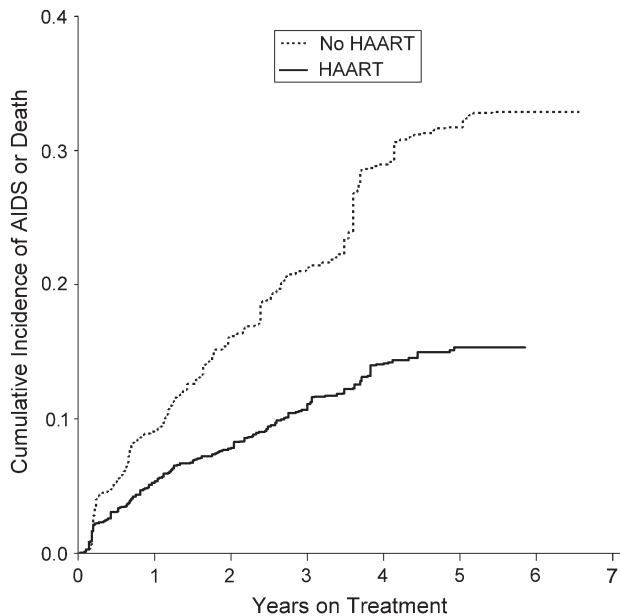
Figure 3. Inverse probability of treatment and censoring weighted Kaplan-Meier curves for the effect of highly active antiretroviral therapy (HAART) on time to acquired immunodeficiency syndrome (AIDS) or death for 1,498 human immunodeficiency virus-positive US men and women, 1995–2002. Time scale is time on treatment (i.e., time since HAART initiation).

treatment as the time scale there is a benefit of treatment apparent in the unadjusted and adjusted models. Specifically, we observed an unadjusted hazard ratio of 0.70 (95% CI: 0.57, 0.86) and an adjusted hazard ratio of 0.67 (95% CI: 0.53, 0.85). The hazard ratio from the marginal structural model was 0.44 (95% CI: 0.32, 0.60). All 3 results using the time-on-treatment time scale were more precise

Table 2. Estimated Effect of Highly Active Antiretroviral Therapy on Time to Acquired Immunodeficiency Syndrome or Death for 1,498 HIV-positive US Men and Women, 1995–2002

	Hazard Ratio	95% Confidence Interval
Unadjusted		
Time on study	0.97	0.76, 1.24
Time on treatment	0.70	0.57, 0.86
Adjusted[a]		
Time on study	0.81	0.61, 1.06
Time on treatment	0.67	0.53, 0.85
Weighted[a]		
Time on study	0.52	0.35, 0.76
Time on treatment	0.44	0.32, 0.60

Abbreviations: HAART, highly active antiretroviral therapy; HIV, human immunodeficiency virus.

[a] Accounting for age, race, sex, baseline and current CD4 count, baseline and current HIV RNA, follow-up time, calendar time, prophylaxis for *Pneumocystis carinii*, and non-HAART antiretroviral therapy.

(i.e., had smaller confidence interval ratios) than the analogous results using time on study as the time scale.

The hazard ratios from the marginal structural models were calculated by using stabilized weights (refer to Appendix 1), which means that they are conditional, rather than purely marginal estimates of hazard ratio; these results are therefore not strictly interpretable as what we would expect to obtain from a randomized controlled trial. In contrast, Figures 2 and 3 represent unconditional estimates of the hazard ratio, and they were drawn by using unstabilized weights; the marginal hazard ratios corresponding directly to these figures are hazard ratio = 0.59 (95% CI: 0.38, 0.92) for time on study and hazard ratio = 0.45 (95% CI: 0.33, 0.62) for time on treatment. As expected, these unstabilized (purely marginal) estimates have wider confidence intervals than the analogous estimates calculated by using stabilized IPTC weights.

Last, we investigated the effect of change of time scale on constancy of the hazard ratios before and after the median event time in each time scale (87 weeks under time on study and 58 weeks under time on treatment). The effect of HAART on the hazard or death of AIDS increased with time under both time scales, but the coefficient from the product term and associated *P* value were both stronger under the time-on-treatment time scale (Table 3).

DISCUSSION

The choice of time scale may affect both visual and numerical descriptions of effects in settings in which a time-varying treatment is subject to time-varying confounding affected by prior treatment. In the example presented, using time on treatment as the time scale yielded a 1.18-fold stronger estimate of the hazard ratio (0.44 compared with 0.52) and a 1.16-fold tighter confidence interval than using time on study as the time scale. Time-on-treatment results are closer to what we would expect from a randomized trial of HAART versus non-HAART. For example, Hammer et al. (33) reported a hazard ratio of 0.50 (95% CI: 0.33, 0.76) for the effect of HAART compared with dual therapy for the outcome of AIDS or death. We would expect to see a stronger effect because we are comparing HAART with a combination of dual, mono, and nontherapy (34) and because of the noncompliance reported by Hammer et al. (33). Thus, in the present example, time on treatment appears to be a more appropriate time scale for addressing the central research question.

To date, marginal structural models have typically used time on study as the time scale. However, there are advantages of using time on treatment as the time scale. Given the stated assumptions, a marginal structural model estimates the same parameter one would recover from a fully compliant randomized trial free from dropout. However, using time on study as the time scale, the marginal structural model recovers the *average* hazard ratio across a series of randomized trials, with randomization at each sequential timepoint. In contrast, using time on treatment as the time scale, the average hazard ratio more closely mimics what one would obtain in a standard randomized trial with the average taken over time since treatment initiation, albeit with the

Table 3. Estimated Time-dependent Effect of Highly Active Antiretroviral Therapy on Time to Acquired Immunodeficiency Syndrome or Death for 1,498 HIV-positive US Men and Women, 1995–2002[a,b]

	Hazard Ratio	95% Confidence Interval	Categorical Interaction at Median Event Time				$P_{interaction}$
			Hazard Ratio (≤Median)	95% Confidence Interval	Hazard Ratio (>Median)	95% Confidence Interval	
Time on study	0.52	0.35, 0.76	0.68	0.40, 1.16	0.46	0.30, 0.73	0.268
Time on treatment	0.44	0.32, 0.60	0.61	0.42, 0.90	0.32	0.21, 0.48	0.017

Abbreviations: HAART, highly active antiretroviral therapy; HIV, human immunodeficiency virus.

[a] Models account for age, race, sex, baseline and current CD4 count, baseline and current HIV RNA, follow-up time, calendar time, prophylaxis for *Pneumocystis carinii*, and non-HAART antiretroviral therapy.

[b] Results are from weighted analysis only. The table gives separate estimates of hazard ratio up to and after the median event time (87 weeks under time on study and 58 weeks under time on treatment).

above-stated caveat that the inclusion of baseline covariates in the final structural model renders the estimate conditional rather than fully marginal. Moreover, departures from the proportional hazards assumption may be more scientifically compelling, or apparent, using time on treatment rather than time on study as the time scale, as illustrated above.

However, there are many cases when time on study is biologically meaningful, such as when study time begins with an acute event of interest, such as a diagnosis of HIV or an acute myocardial infarction. There are also settings in which results will be (nearly) invariant to choice of time scale. Outside the context of marginal structural models, prior work suggests that, in time-to-event analyses, one chooses the origin and time scale that are most strongly associated with the event (2) or that one adequately (i.e., flexibly) models such important time scales when using an alternate time scale (19). In the setting in which the time to event follows an exponential distribution (i.e., a constant hazard), the choice of time scale will matter little. When the origin of the chosen time scale is a strong risk factor for the event of interest, but unrelated to treatment, differences in the obtained hazard ratio may be largely due to the non-collapsibility of the hazard ratio (35, 36) rather than due to confounding by time.

The present work is subject to limitations. First, defining time on treatment is complicated if participants move on and off treatment during follow-up. This complication is similar to the choice of time scale with repeated events within the same participant (37). Here, we made an "intent-to-treat" assumption, namely, that participants remained on treatment after initiation regardless of time scale. This intent-to-treat assumption renders the complication of participants moving on and off treatment moot, at the cost of misclassifying 6% of person-time. Second, typical implementations of marginal structural models stabilize the IPTC weights using a subset of measured baseline covariates. Stabilization improves the variance of the final estimates of association, as was demonstrated here. However, if baseline factors are used to stabilize the weights, these baseline factors must then be included in the final structural model to ensure control of confounding; this inclusion means that the final model estimates are both conditional on these baseline covariates (rather than fully marginal) and imposes the additional assumption that the inclusion of such covariates is

properly specified in the structural model. Because IPTC-weighted Kaplan-Meier survival curves are calculated from the simple weighted study population, one must stratify the weighted study population curves to take advantage of stabilization by baseline factors. To obtain unstratified Kaplan-Meier curves that control for confounding due to baseline and time-varying factors, the epidemiologist may exclude baseline covariates from stabilization of the IPTC weights. We took this latter approach to constructing the Kaplan-Meier curves. Last, there are subtly different specifications for the effect of calendar time in the structural models under the 2 time scales; care must be taken when (as in this case) calendar time is a strong confounder. When fitting a marginal structural model using pooled logistic regression (where the time scale is an explicit component of the model), if one models all time scales simultaneously, then one can build survival curves based on any time metric present in the model (38). However, inclusion of multiple time scales when unnecessary may adversely affect precision.

This report makes 2 contributions. First, we illustrate differences in causal effects that may be observed when using competing time scales in the context of marginal structural models. Second, we provide a method for displaying IPTC-weighted Kaplan-Meier survival curves adjusted for both time-fixed and time-varying confounding and emigrative selection bias. Both contributions have the potential to increase understanding of marginal structural models—the first because time on treatment may be a more appropriate time scale in many scientific applications, and the second because prior applications of marginal structural models have rarely included confounding and censoring-adjusted survival curves, which help to visually depict associations.

REFERENCES

1. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. New York, NY: Springer; 2003.

2. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am J Epidemiol*. 1997;145(1):72–80.

3. Lamarca R, Alonso J, Gómez G, et al. Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. *J Gerontol A Biol Sci Med Sci*. 1998;53(5): M337–M343.

4. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158(9): 915–920.

5. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.

6. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc (B)*. 1972;34(2):187–220.

7. Barrón Y, Cole SR, Greenblatt RM, et al. Effect of discontinuing antiretroviral therapy on survival of women initiated on highly active antiretroviral therapy. *AIDS*. 2004;18(11):1579–1584.

8. Choi HK, Hernán MA, Seeger JD, et al. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *Lancet*. 2002;359(9313):1173–1177.

9. Cole SR, Hernán MA, Robins JM, 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):687–694.

10. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. *Am J Epidemiol*. 2002;155(11):1045–1053.

11. Fox MP, Brooks DR, Kuhn L, et al. Role of breastfeeding cessation in mediating the relationship between maternal HIV disease stage and increased child mortality among HIV-exposed uninfected children. *Int J Epidemiol*. 2009;38(2): 569–576.

12. 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.

13. Hernán MA, Brumback BA, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc*. 2001;96(454):440–448.

14. López-Gatell H, Cole SR, Hessol NA, et al. Effect of tuberculosis on the survival of women infected with human immunodeficiency virus. *Am J Epidemiol*. 2007;165(10):1134–1142.

15. López-Gatell H, Cole SR, Margolick JB, et al. Effect of tuberculosis on the survival of HIV-infected men in a country with low tuberculosis incidence. *AIDS*. 2008;22(14):1869–1873.

16. Sterne JA, Hernán MA, Ledergerber B, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet*. 2005;366(9483): 378–384.

17. Wiesbauer F, Heinze G, Mitterbauer C, et al. Statin use is associated with prolonged survival of renal transplant recipients. *J Am Soc Nephrol*. 2008;19(11):2211–2218.

18. Gail MH, Graubard B, Williamson DF, et al. Comments on 'Choice of time scale and its effect on significance of predictors in longitudinal studies' by Michael J. Pencina, Martin G. Larson, and Ralph B. D'Agostino, Statistics in Medicine 2007;26:1343–1359. *Stat Med*. 2009;28(8):1315–1317.

19. Pencina MJ, Larson MG, D'Agostino RB. Choice of time scale and its effect on significance of predictors in longitudinal studies. *Stat Med*. 2007;26(6):1343–1359.

20. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004;75(1):45–49.

21. 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.

22. Kaslow RA, Ostrow DG, Detels R, et al. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol*. 1987; 126(2):310–318.

23. Barkan SE, Melnick SL, Preston-Martin S, et al. The Women's Interagency HIV Study. WIHS Collaborative Study Group. *Epidemiology*. 1998;9(2):117–125.

24. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Recomm Rep*. 1992;41(RR-17):1–19.

25. Panel on Clinical Practices for Treatment of HIV Infection. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Bethesda, MD: AIDSinfo (formerly HIV/AIDS Treatment Information Service), National Institutes of Health, 2000. (http://www.aidsinfo.nih.gov).

26. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009; 20(1):3–5.

27. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes (Lond)*. 2008;32(suppl 3):S8–S14.

28. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008; 168(6):656–664.

29. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006; 60(7):578–586.

30. D'Agostino RB, Lee ML, Belanger AJ, et al. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med*. 1990;9(12): 1501–1515.

31. Abbott RD. Logistic regression in survival analysis. *Am J Epidemiol*. 1985;121(3):465–471.

32. Allison PD. Discrete-time methods for the analysis of event histories. In: *Sociological Methodology* Vol. 13. Hoboken, NJ: John Wiley & Sons; 1982. 61–98.

33. Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med*. 1997;337(11): 725–733.

34. Detels R, Muñoz A, McFarlane G, et al. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study Investigators. *JAMA*. 1998;280(17):1497–1503.

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

36. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615–625.

37 Cain LE, Cole SR, Chmiel JS, et al. Effect of highly active antiretroviral therapy on multiple AIDS-defining illnesses among male HIV seroconverters. *Am J Epidemiol*. 2006; 163(4):310–315.

38. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–779.

39. Cole SR, Hernán MA, Anastos K, et al. Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using a marginal structural left-censored mean model. *Am J Epidemiol*. 2007;166(2):219–227.

40. Cole SR, Hernán MA, Margolick JB, et al. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *Am J Epidemiol*. 2005; 162(5):471–478.

APPENDIX 1

Inverse Probability of Treatment and Censoring Weights

We account for measured confounding and emigrative selection bias in the marginal structural model using stabilized inverse probability of treatment and censoring (IPTC) weights (5). In general, IPTC weights are defined as $W_{it} = W_{it}^X \times W_{it}^C$, where

$$W_{it}^X = \prod_{k=0}^{t} f[X_{ik}|\ \bar{X}_{ik-1}, \bar{C}_{ik-1} = 0, L_{i0}]/$$
$$f[X_{ik}|\ \bar{X}_{ik-1}, \bar{C}_{ik-1} = 0, \bar{L}_{ik-1}]$$

and

$$W_{it}^C = \prod_{k=1}^{t+1} \Pr[C_{ik} = 0|\ \bar{C}_{ik-1} = 0, \bar{X}_{ik}, L_{i0}]/$$
$$\Pr[C_{ik} = 0|\ \bar{C}_{ik-1} = 0, \bar{X}_{ik}, \bar{L}_{ik-1}] \quad,$$

where $f[\ \cdot\ |\ \cdot\ ]$ is the conditional density function evaluated at the observed covariate values for a given participant, which was assumed to be logistic (9), and $\bar{L}_{ik-1}$ is the vector of time-varying covariate histories measured up to week $k-1$ including baseline covariates L_{i0}, and C_{ik} is 1 if participant i is censored because of dropout by week k and 0 otherwise. In our case, the intent-to-treat assumption leads to a simplification of the treatment history, and the treatment weights are as follows: if $X_{it-1} = 1$, then $W_{it}^X = W_{it-1}^X$; else

$$W_{it}^X = \prod_{k=0}^{t} f[X_{ik}|\ \bar{X}_{ik-1} = 0, \bar{C}_{ik-1} = 0, L_{i0}]/$$
$$f[X_{ik}|\ \bar{X}_{ik-1} = 0, \bar{C}_{ik-1} = 0, \bar{L}_{ik-1}] \quad.$$

The covariates included in the weights were as follows: age and calendar time (both modeled as restricted cubic splines with knots at the 5th, 35th, 65th, and 95th percentiles), race, sex, CD4 count (baseline and updated), viral RNA (baseline and updated), use of non-HAART antiretroviral therapy, use of cotrimoxazole prophylaxis, and presence of HIV symptoms (reported persistent fever, diarrhea, night sweats, or weight loss). All time-varying variables were lagged 1 visit. For more detail on covariate selection, refer to reference 9, and for further details on the estimation of IPTC weights in general, refer to reference 28. The distribution of the weights was as follows: mean, 1.03; median, 0.91; first percentile, 0.21; 99th percentile, 4.41; minimum, 0.12; and maximum, 28.5. The mean (minimum/maximum) weights for treated and

untreated person-time were 1.03 (0.15/28.5) and 1.03 (0.12/10.3), respectively. These data exhibit adequate positivity, and when weights were truncated at first and 99th percentiles, the hazard ratio changed by 0.01 under both time scales.

Note that IPTC-weighted Kaplan-Meier curves were estimated by using alternate weights in which L_{i0} was removed from the numerator (but not the denominator) of both W_{it}^X and W_{it}^C. As a result, and in contrast to prior analyses (9, 39, 40), models that use these alternate weights need not include L_{i0} in the final structural model. This allows the weights to control for time-fixed and time-varying confounding, and therefore the Kaplan-Meier curves drawn from such weighted data account for both time-fixed and time-varying confounding. A cost of using such weights is a loss in efficiency because the weights are less stable than those defined above.

APPENDIX 2

SAS Code for Adjusted Survival Curves

```
/******************************************************************/
/* Code written by Daniel Westreich and Stephen R. Cole */
/* Released under GNU Public License */
/******************************************************************/
/******************************************************************/
/* Notes: */
/* A. Outcomes occur AFTER time point listed, before next time */
/*    point */
/* B. timeunit is the interval by which subsequent DTHM data */
/*    elements are organized. Thus, if the timeunit is month, */
/*    we assume that for a given pid, the data structure goes */
/*    month=1, month=2, etc. */
/* C. timeunit count starts at 1, not 0 */
/******************************************************************/

%macro
WeightedCurvesData (data, pid, timeunit, outcome, treatment, weight, type);

* if this is an 'treatment' KM curve, first reset visit numbers;
%if &type = "treatment" %then %do;
    proc sort data=&data; by &pid;
    data __foo; set &data; by &pid;
        __lagexp = lag(&treatment);
        if first.&pid then __lagexp = &treatment;
    data __foo2; set __foo;
        __resetflag=0;if &treatment ne __lagexp then
__resetflag=1;
    data __foo3; set __foo2;by &pid;
        retain __reset; if first.&pid then __reset=0;
        if __resetflag=1 then __reset=&timeunit-1;
    data __foo4; set __foo3;by &pid; &timeunit = &timeunit-__reset;
        drop __lagexp __resetflag __reset;

* for calculation of Treatment HRs;
data FinalTreatmentData; set __foo4;run;
    * Remove nonessential data from dataset to improve performance;
    data __main0; set __foo4;keep &timeunit &outcome &weight &treatment;
        where &weight ne .;run;
    %end;

* otherwise, keep visit numbers as they are;
%if &type = "study" %then %do;
```

```
      * Remove nonessential data from dataset to improve performance;
      data __main0; set &data; keep &timeunit &outcome &weight &treatment;
           where &weight ne .;run;
      %end;
* weight the outcome;
data __main1; set __main0;__wgtoutcome =&weight*&outcome;
proc sort data=__main1; by &treatment &timeunit;
proc univariate data=__main1 noprint; var &weight &outcome __wgtoutcome;
      by &treatment &timeunit;
      output out=__main2 n=atRisk n=nTemp1 n=nTemp2
           sum=sumWgts sum=totalDead sum=wgtDead;
data __main3; set __main2;drop nTemp1 nTemp2;

* calculate final KM parameters;
proc sort data=__main3; by &treatment &timeunit;
data FinalKMData; set __main3; by &treatment &timeunit;
      retain km0 kmw;
      if first.&treatment and first.&timeunit then
           do; km0=1; kmw=1; end;
      else do;
           if km0 ne 0 then km0 = km0*(1-totalDead/atRisk);
           if kmw ne 0 then kmw = kmw*(1-wgtDead/sumWgts);
      end;
run;
%mend;

/******************************************************************/
/* Additional sample code to invoke macro and display curves */
/* Note that ''treatment'' can be replaced with ''study'' in the */
/* invocation below to draw study time scale output curves */
/******************************************************************/

%WeightedCurvesData (temp,pid,visit,event,treatment,weight,"treatment");
goptions reset=all;
axis1 label=(angle=90) order=(0 to 1 by .2) offset=(0,0);
symbol1 v=none i=steplj c=black line=1 width=2;
symbol2 v=none i=steplj c=black line=2 width=2;
proc gplot data=FinalKMData;
      label kmw='Weighted AIDS-free survival';
      label visit='Years of treatment';
      plot kmw*visit=treatment / vaxis=axis1;
      run; quit;

/******************************************************************/
/* Additional sample code to calculate treatment hazard ratio */
/* Note that FinalTreatmentData will only be defined if the macro */
/* has been invoked with the ''treatment'' option */
/******************************************************************/

proc genmod descending data=FinalTreatmentData;
      class pid visit;
      title "Treatment time scale";
      model event = treatment visit / dist=bin link=logit;
      weight weight;
      repeated subject=pid / type=ind;
      estimate "Effect" treatment 1 / exp;
run;
```