

Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men

Miguel A. Hernán,¹ Babette Brumback,² and James M. Robins^{1,2}

From the Departments of ¹Epidemiology and ²Biostatistics,
Harvard School of Public Health, Boston, MA

Running title: MSMs FOR SURVIVAL ANALYSIS

Word count: 196 (abstract); 4416 (text)

Address correspondence and reprint requests to Dr. Miguel Hernán, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115

Phone: (617) 432 0101; Fax: (617) 566 7805

e-mail: miguel_hernan@post.harvard.edu

ABSTRACT

Standard methods for survival analysis, such as the time-dependent Cox model, may produce biased effect estimates when there exist time-dependent confounders that are themselves affected by previous treatment or exposure. Marginal structural models (MSMs) are a new class of causal models whose parameters are estimated through inverse-probability-of-treatment weighting, and that allow for appropriate adjustment for confounding. We describe the marginal structural Cox proportional hazards model and use it to estimate the causal effect of zidovudine on the survival of HIV-positive men participating in the Multicenter AIDS Cohort Study. In this study, CD4 lymphocyte count is both a time-dependent confounder of the causal effect of zidovudine on survival, and is affected by past zidovudine treatment. The crude mortality rate ratio (95% confidence interval) for zidovudine was 2.3 (2.0, 2.7), which reflects the presence of confounding. After controlling for baseline CD4 count and other baseline covariates using standard methods, the mortality rate ratio decreased to 1.7 (1.4, 1.9). Using a marginal structural Cox model to further control for time-dependent confounding due to CD4 count and other time-dependent covariates, the mortality rate ratio was 0.7 (95% conservative confidence interval: 0.6, 0.9). We compare MSMs to previously proposed causal methods.

Keywords: *counterfactuals; causality; epidemiologic methods; longitudinal data; survival analysis; structural models; confounding; intermediate variables; AIDS.*

Introduction

Marginal structural models (MSMs) can be used to estimate the causal effect of a time-dependent exposure in the presence of time-dependent confounders that are themselves affected by previous treatment.^{1,2} The use of MSMs is an alternative to g-estimation of structural nested models (SNMs).³

In our companion paper we describe inverse-probability-of-treatment weighted (IPTW) estimation of a marginal structural logistic model.⁴ In this paper we introduce the marginal structural Cox proportional hazards model, show how to estimate its parameters by inverse-probability-of-treatment weighting, provide practical advice on how to use standard statistical software to obtain the IPTW estimates, and include, as an appendix, the SAS code necessary for the analysis. We use this Cox proportional hazards MSM to estimate the effect of zidovudine on the survival of HIV-positive men enrolled in an observational

cohort study: the Multicenter AIDS Cohort Study (MACS). We conclude by comparing methods based on MSMs to previously proposed methods based on g-estimation of SNMs and on the g-computation algorithm.

We now begin by describing the MACS and then summarize why standard methods for survival analysis are not appropriate for estimating the effect of zidovudine on mortality in this cohort.

The Multicenter AIDS Cohort Study and Bias of standard methods

Between 1984-91, the MACS enrolled 5622 homosexual and bisexual men, with no prior AIDS-defining illness, from the metropolitan areas of Los Angeles, Baltimore-Washington, Pittsburgh, and Chicago. Study participants were asked to return every 6 months to complete a questionnaire, undergo physical examination, and provide blood samples. The design and methods of the MACS have been described in detail elsewhere.^{5,6}

We restricted our cohort to HIV-positive men alive in the period during which zidovudine was available for use (i.e., after the study visit 5; March 1986 to March 1987). Follow-up ended at study visit 21 (March to November 1994), death, or last contact, whichever came first. Our analysis include the 2132 men that attended at least one visit between visits 5 and 21 while HIV-positive, and that did not have AIDS and were not on zidovudine at the first eligible visit. By the end of the follow-up (median duration: 71 months), 1275 had initiated zidovudine treatment and 926 subjects had died.

The usual approach to the estimation of the effect of a time-varying exposure, such as zidovudine, on survival is to model the hazard of failure at a given time as a function of past exposure history using a time-dependent Cox proportional hazards model. However, Robins has shown this approach may be biased, whether or not one further adjusts for past covariate history, whenever (1) there exists a time-dependent covariate that is both a risk factor for mortality and also predicts subsequent exposure; and (2) past exposure history predicts the risk factor.⁷ Covariates satisfying (1) are called time-dependent confounders. Past CD4 count is a time-dependent confounder for the effect of zidovudine on survival since it is a risk factor for mortality and a predictor of subsequent initiation of zidovudine therapy,⁶ and past zidovudine history is an independent predictor of subsequent CD4 count.⁸ In fact, all standard methods (e.g., Cox or Poisson regression) that predict the mortality rate at each time using a summary of zidovudine history up to that time may produce biased estimates of the causal effect of zidovudine whether or not one adjusts for past CD4 count in the analysis.

Marginal structural Cox proportional hazards model

In the absence of time-dependent confounding, a time-dependent Cox proportional hazards model is typically used. We treat visit 5, or the earliest subsequent visit at which a man was HIV-positive, as a start of follow-up time for our analysis. Let T be a subject's time of death with time measured in months since start of follow-up. Let $A(t)$ be one if a subject was on zidovudine at the last visit to t . We use overbars to represent a covariate history so, for example, $\bar{A}(t) = \{A(u); 0 \leq u < t\}$ is a subject's treatment history up to t . Finally, let V be a vector of time-independent baseline covariates measured prior to start of follow-up. Then, the conditional hazard of death (i.e., mortality rate) $\lambda_T(t | \bar{A}(t), V)$ given treatment history $\bar{A}(t)$ and baseline covariates V is modelled as

$$\lambda_T(t | \bar{A}(t), V) = \lambda_0(t) \exp(\gamma_1 A(t) + \gamma_2 V) .$$

The subscript T in $\lambda_T(t | \bar{A}(t), V)$ merely identifies this hazard function as being that corresponding to the variable T . In our analysis, the covariates in V are age, CD4 count, white blood cell count (WBC),

hematocrit, and presence of symptoms. Symptomatic status was defined by the presence during the previous six months of one or more of the following clinical symptoms or signs: fever (temperature $> 37.9^\circ\text{C}$) for ≥ 2 weeks, oral candidiasis, diarrhea for ≥ 2 weeks, weight loss of ≥ 4.5 kg, oral hairy leukoplakia, or herpes zoster. We assume, for simplicity, that patients remain on therapy once they start it and that the hazard of death at time t depends on a subject’s zidovudine history only through its current value, but alternative specifications are possible. Suppose, for the moment, no censoring occurs, i.e., death times T are observed for all subjects.

In the presence of time-dependent covariates $L(t)$ satisfying the conditions (1) and (2), the estimate of γ_1 by maximized Cox partial likelihood is a biased estimate of the effect of zidovudine on mortality, whether or not we additionally include the time-dependent covariates $L(t)$ as regressors.

Arguing as in our companion paper,⁴ we can eliminate or reduce this bias by fitting the above time-dependent Cox model with the contribution of a subject to a risk-set calculation performed at time t weighted by the “stabilized” weights

$$sw_i(t) = \prod_{k=0}^{int(t)} \frac{pr(A(k) = a_i(k) \mid \bar{A}(k-1) = \bar{a}_i(k-1), V = v_i)}{pr(A(k) = a_i(k) \mid \bar{A}(k-1) = \bar{a}_i(k-1), \bar{L}(k) = \bar{l}_i(k))},$$

to obtain an inverse probability of treatment weighted (IPTW) partial likelihood estimate. In the above $\bar{A}(-1)$ is defined to be zero. Here, $int(t)$ is the largest integer less than or equal to t and k is an integer-valued variable denoting whole months since start of follow-up. Since a subject’s recorded treatment changes at most once per month, each factor in the denominator of $sw_i(t)$ is, informally, the probability that the subject received his own observed treatment at time k , given his past treatment and prognostic factor history (V is included in $L(0)$). Each factor in the numerator is, informally, the probability that the subject received his observed treatment conditional on his past treatment history and baseline covariates, but not further adjusting for his past time-dependent prognostic factor history. “Non stabilized” weights $w_i(t)$, in which the numerator of $sw_i(t)$ is replaced by 1, can be used in lieu of $sw_i(t)$. Although the choice of weights will not influence the consistency of our causal estimates, the stabilized weights $sw_i(t)$ are preferred because they lead to more efficient estimates (i.e., narrower confidence intervals). In a latter section we described how these stabilized inverse-probability-of-treatment weights $sw_i(t)$ can be estimated from the data.

Suppose all relevant time-dependent confounders are measured and included in $L(t)$. Then, weighting by $sw_i(t)$ effectively creates, for a risk-set at time t , a pseudo-population in which (1) no longer predicts initiation of zidovudine at t (i.e., $\bar{L}(t)$ is not a confounder), and (2) the causal association between zidovudine and mortality is the same as in the original study population.¹ As argued in Ref. 4, this implies that an IPTW estimator, say $\hat{\beta}_1$ of the parameter γ_1 of our time-dependent Cox model will converge to a quantity β_1 that can be appropriately interpreted as the causal effect, on the log rate ratio scale, of zidovudine on mortality.

To formalize the above, we introduce counterfactual outcomes.⁴ For each possible treatment history $\bar{a} = \{a(t); 0 \leq t\}$, let $T_{\bar{a}}$ be the random variable representing the subject’s time to death had he followed, possibly contrary to fact, the zidovudine history \bar{a} from the start of follow-up, rather than his observed history $\bar{A}(T)$. For example, $T_{\bar{a}}$ with \bar{a} such that $a(t) = 0$ for $t < 2.5$ and $a(t) = 1$ for $t \geq 2.5$ is the subject’s survival time when he started zidovudine therapy 2.5 months after the start of follow-up. We only observe $T_{\bar{a}}$ for those treatment histories \bar{a} that agree with the subject’s observed treatment history $\bar{A}(T)$ until the subject’s observed death time T . For these histories $T_{\bar{a}}$ equals T . For each \bar{a} we specify

the marginal structural Cox proportional hazards model

$$\lambda_{T_{\bar{a}}}(t | V) = \lambda_0(t) \exp(\beta_1 a(t) + \beta_2 V)$$

where $\lambda_{T_{\bar{a}}}(t | V)$ is the hazard of death at t among subjects with baseline covariates V in the source population had, contrary to fact, all subjects followed zidovudine history \bar{a} through time t , the constant β_1 and the row vector β_2 are unknown parameters, and $\lambda_0(t)$ is an unspecified baseline hazard. We refer to this model as a MSM because, within levels of V , it is a structural (i.e., causal) model for the marginal distribution of the counterfactual variables $T_{\bar{a}}$.

The parameter β_1 of our MSM is the causal log rate ratio for zidovudine. Hence, $\exp(\beta_1)$ has a causal interpretation as the ratio of the mortality (hazard) rate at any time t had all subjects been continuously exposed to zidovudine compared to the hazard rate at time t had all subjects remained unexposed. β_1 is consistently estimated by our IPTW estimator $\hat{\beta}_1$, under the untestable assumption of no unmeasured confounders given the measured risk factors in $L(t)$.¹ We shall make this assumption with $L(t)$ being the covariate vector with the following elements: most recently recorded CD4, WBC, hematocrit, AIDS diagnosis, and symptomatic status at t .

It is difficult to get standard Cox model software to compute our IPTW estimator $\hat{\beta}_1$ because our subject-specific weights $sw_i(t)$ vary over time, and most standard Cox model software programs, even those that allow for subject-specific weights, do not allow for subject-specific time-varying weights. The approach we shall adopt to overcome this software problem is to fit a weighted pooled logistic regression treating each person-month as an observation. (In the MACS, our 2132 men contribute 137,647 person-months of observation.) That is, we will fit, by weighted logistic regression using weights $sw_i(t)$, the model

$$\text{logit } pr [D(t) = 1 | D(t-1) = 0, \bar{A}(t-1), V] = \beta_0(t) + \beta_1 A(t-1) + \beta_2 V$$

where, henceforth t , like k , is integer valued denoting whole months since start of follow-up, $D(t) = 0$ if a subject was alive in month t and 1 if the subject died in month t , and $\beta_0(t)$ is a time- (i.e., month-) specific intercept. This method has the advantage of being easily programmed in many standard statistical packages. In the unweighted case, it is essentially equivalent to fitting an unweighted time-dependent Cox model, because the hazard on any single day is extremely small.⁹ However, the use of weights induces within-subject correlation, which invalidates the standard error estimates outputted by a standard logistic program (they can be either too large or too small). To overcome this difficulty, the above weighted logistic model should be fit using a generalized estimating equations¹⁰ program (e.g., option ‘repeated’ in SAS Proc Genmod) which outputs ‘robust’ variance estimators that allow for correlated observations. The robust variance estimator provides a conservative confidence interval for the β . That is, the 95% confidence interval calculated as $\hat{\beta} \pm 1.96 \times \text{robust standard error}$ is guaranteed to cover the true β at least 95% of the time in large samples.

Censoring

The analysis just described assumes that there is no drop-out or censoring by end of follow-up. We define the censoring indicator $C(t)$ to be 1 if a subject is right-censored by time t and $C(t) = 0$ otherwise, where a subject is right-censored if he either dropped out of the study or reached the administrative end of follow-up alive. To estimate β_1 in the presence of censoring, we fit a weighted Cox model in which, for

a subject at risk at month t , we use the weight $sw_i(t) \times sw_i^\dagger(t)$ where

$$sw_i^\dagger(t) = \prod_{k=0}^t \frac{pr [C(k) = 0 \mid \overline{C}(k-1) = 0, \overline{A}(k-1) = \overline{a}_i(k-1), V = v_i]}{pr [C(k) = 0 \mid \overline{C}(k-1) = 0, \overline{A}(k-1) = \overline{a}_i(k-1), \overline{L}(k-1) = \overline{\ell}_i(k-1)]},$$

where $\overline{C}(-1)$ is defined to be zero. $sw_i^\dagger(t)$ is, informally, the ratio of a subject's probability of remaining uncensored up to month t , calculated as if there had been no time-dependent determinants of censoring except past zidovudine history, divided by the subject's probability of remaining uncensored up to month t . The denominator of the product $sw_i(t) \times sw_i^\dagger(t)$ is, informally, the probability that a subject had had his observed zidovudine and censoring history through month t . Since $sw_i(t)$ and $sw_i^\dagger(t)$ are unknown, they must be estimated from the data as described below. Weighting by $sw_i(t) \times sw_i^\dagger(t)$ produces a consistent estimate of the causal parameter β_1 under the assumption that the measured covariates are sufficient to adjust for both confounding and for selection bias due to loss to follow-up.⁴

Estimation of the weights

The practical problem faced by the investigator is how to obtain the quantities $sw_i(t) \times sw_i^\dagger(t)$ necessary to run the pooled weighted logistic regression model. Consider first the estimation of $sw_i(t)$. We need to consistently estimate the denominator and numerator of $sw_i(t)$ for each subject and time point. Since any subject starting zidovudine was assumed to remain on it thereafter, we can regard the time to starting zidovudine as a failure time variable and model the probability of starting zidovudine through a pooled logistic model that treats each person-month as an observation and allows for a time-dependent intercept. Specifically, we can fit the model and obtain estimates $\hat{\alpha} = (\hat{\alpha}_0(k), \hat{\alpha}_1, \hat{\alpha}_2)$ for the unknown parameters. It is only necessary to fit the model for subjects who had yet to begin zidovudine (i.e., the 87,840 person-months in the MACS with $\overline{A}(k-1) = 0$).

The estimated predicted values $\hat{P}_i(k) = \text{expit}(\hat{\alpha}_0(k) + \hat{\alpha}_1 L_i(k) + \alpha_2 V_i)$ from this model are the estimated probabilities of subject i not starting zidovudine in month k given that zidovudine had not been started by month $k-1$, where $\text{expit}(x) = e^x / (1 + e^x)$. Our estimate of the denominator of $sw_i(k)$ for person i on month k is the product $\overline{\overline{P}}_i(k) = \prod_{u=0}^k \hat{P}_i(u)$ if subject i did not start zidovudine up to month k , and is $\overline{\overline{P}}_i(k) = \left[1 - \hat{P}_i(t)\right] \prod_{u=0}^{t-1} \hat{P}_i(u)$ if subject i started zidovudine on month t for $t \leq k$. (Note that, in calculating $\overline{\overline{P}}_i(k)$, we have used the fact that no subject stops zidovudine once begun.) Similarly, we estimate the numerator of $sw_i(k)$ by fitting the above logistic model except with the covariates $L(k)$ removed from the model.

There is a small but important technical detail we have yet to discuss. For our IPTW estimates of β to be consistent, it is necessary that the denominator of $sw_i(t)$ be consistently estimated. To do so, we cannot estimate a separate intercept $\alpha_0(k)$ for each month k . Rather, we need to “borrow strength” from subjects starting zidovudine on months other than k to estimate $\alpha_0(k)$. This can be accomplished by assuming that $\alpha_0(k)$ is constant in windows of, say, 3 months. An alternative approach is to assume the $\alpha_0(k)$ are a smooth function of $\alpha_0(k)$ and thus can be estimated by smoothing techniques (such as regression splines, smoothing splines, or kernel regression).¹¹

In order to correct for censoring, we estimate $sw_i^\dagger(k)$ in a manner analogous to the estimation of $sw_i(k)$ except with $A(k)$ replaced by $C(k)$ as the outcome variable, with $A(k-1)$ added as an additional regressor, and not conditioning on $\overline{A}(k-1) = 0$ but rather on $\overline{C}(k-1) = 0$.

Causal effect of zidovudine in the MACS

Using a standard Cox proportional hazards model — or the equivalent pooled logistic regression model — with no covariates, the crude mortality rate ratio for zidovudine was 2.3 (95% confidence interval: 2.0, 2.7). The addition of the baseline covariates V to the model decreased this rate ratio to 1.7 (1.4, 1.9).

To further adjust for confounding due to the time-dependent factors $L(t)$, we calculated a stabilized weight $s\hat{w}_i(t) \times s\hat{w}_i^\dagger(t)$ for each person-month and fit a weighted pooled logistic model, whose parameters estimate those of a marginal structural Cox model. The estimated causal mortality rate ratio, $\exp(\beta_1)$, was 0.7 (95% conservative confidence interval: 0.6, 0.9), indicating that, under our assumptions, zidovudine therapy appears to decrease the risk of death. When non stabilized weights $\hat{w}_i(t) \times \hat{w}_i^\dagger(t)$ were used, the rate ratio was virtually identical but the 95% conservative confidence interval was 10% wider, compared to the stabilized results (table 1). We also report invalid model-based intervals obtained using an ordinary weighted logistic regression program that does not account for within-subject correlations. The point estimates and 95% conservative confidence intervals for each of the parameters of our marginal structural Cox model are displayed in table 2.

The stabilized weights were calculated by means of four pooled logistic regression models, as described in the previous section. In two of the models the outcome was ‘initiation of zidovudine’. Using the estimated predicted values from each of these models, we calculated two quantities for each observation: the probability of each person having his own observed zidovudine history up to month t given baseline covariates V , and, then, given also time-varying covariates $L(t)$. Similar models were fit for the outcome ‘censoring,’ after adding zidovudine history as a time-varying dichotomous variable indicating whether the subject was on zidovudine on month $t - 1$.

Table 3 shows the center and dispersion parameters of the distribution of the four probabilities estimated for all patients in the study at two arbitrary time points: 24 and 84 months of follow-up. The estimated probabilities of having one’s own observed zidovudine history at 24 months of follow-up, given time-varying covariates, range from 0.939 to 0.002. This would be translated into (non stabilized) inverse-probability-of-treatment weights $\hat{w}_i(t)$ ranging from 1.06 ($1/0.939$) to 500 ($1/0.002$). Thus, in the pseudo-population, some observations would be represented by 1.06 copies of themselves, whereas others should be represented by 500 copies. The use of stabilized inverse-probability-of-treatment weights $s\hat{w}_i(t)$ “normalizes” or stabilizes the range of these inverse probabilities and increases the efficiency of the analysis by preventing just a few people from contributing most of the observations in the pseudo-population. Thus, the values $s\hat{w}_i(t)$ for $t = 24$ are centered around 1.00 and show a narrower range (0.10 to 6.72).

The estimated probabilities of being uncensored at 24 months follow a more peaked distribution centered around values close to one (0.946), and range from 0.987 to 0.883. This is expected, as 90% of the men were uncensored at 24 months of follow-up. Inverse probabilities $\hat{w}_i^\dagger(t)$ range from 1.01 to 1.13. The stabilized weights $s\hat{w}_i^\dagger(t)$ for $t = 24$, are centered around one and range from 0.99 to 1.01. The estimated probabilities of being uncensored at 84 months are lower as expected.

The distribution of the final weights, which combine information on zidovudine and censoring history, is presented in figures 1 and 2 for several follow-up times (a logarithmic transformation was applied for display purposes only). Two sets of weights were estimated: the stabilized weights $s\hat{w}_i(t) \times s\hat{w}_i^\dagger(t)$ and the non stabilized weights $\hat{w}_i(t) \times \hat{w}_i^\dagger(t)$. The distribution of stabilized weights is symmetric and centered around 1 at all times, while their variability increases over time. The distribution of the non stabilized weights is skewed and shows a much greater variability than that of the stabilized weights.

The weight estimates were robust with respect to the choice of method to estimate the time-dependent

baseline hazard $\alpha_0(k)$ in the logistic models for zidovudine and censoring, provided that sufficient flexibility was allowed. The weights in figures 1 and 2 were obtained by modelling the time-dependent intercept $\alpha_0(k)$ with natural cubic splines with five knots (on months 14, 46, 71, 90, and 96, which correspond to the percentiles 5, 27.5, 50, 72.5, and 95, respectively).

Comparison of MSMs with previously proposed methods

Prior to introducing MSMs, Robins and co-workers introduced three methods for estimation of the causal effect of a time-varying treatment in the presence of time-varying confounders: the parametric g-computation algorithm formula,^{12,13} g-estimation of structural nested models,^{12,14,15} and the iterative conditional expectations (ICE) estimator.^{3,12} Inverse-probability-of-treatment weighted estimation of MSMs constitutes a fourth method. When (i) both treatment and the confounders are discrete variables, (ii) they are measured at only a few time points, and (iii) the sample size is large, then estimation can be carried out using fully saturated models (i.e., non-parametrically) and all four methods are precisely equivalent. They differ when, due to sparse multivariate data, one must introduce modelling assumptions.

ICE estimators can only rarely be used because they often lead to incompatible models and will not be discussed further.¹² Of the remaining three methods, inference based on SNMs and MSMs is preferable to that based on the parametric g-computation algorithm. The reason is that MSM and SNM models, in contrast to models based on the g-computation algorithm formula, include parameters that represent the null hypothesis of no treatment effect.^{12,15} As a consequence, when using the parametric g-computation algorithm estimator, it is quite difficult to determine whether one's confidence interval for the treatment effect includes the null hypothesis of no effect.

MSMs have two major advantages over SNMs. Although useful for survival time outcomes, continuous measured outcomes (e.g., blood pressure), and Poisson count outcomes, logistic SNMs cannot be conveniently used to estimate the effect of treatment on dichotomous (0, 1) outcomes unless the outcome is rare.^{1,2,12} This is because logistic SNMs cannot be fit by g-estimation. In contrast, as we have seen,⁴ IPTW estimation of logistic MSMs can be used to estimate the effect of a time-dependent treatment on a binary outcome.

The second major advantage of MSMs is that they resemble standard models, whereas SNMs do not. For example, the logistic MSM described in our companion paper⁴ and the Cox proportional hazards MSM described here are the natural way to extend the ordinary logistic and time-dependent Cox models to allow for estimation of causal effects. The close resemblance of MSMs to standard statistical models makes their application more intuitive for researchers and easier for programmers.

However, SNMs have a number of advantages over MSMs. For example, as discussed in Ref. 4, MSMs cannot be used to estimate the causal effect of a time-dependent exposure on the mortality of an occupational cohort in the presence of the healthy worker survivor effect.^{1,2}

A second major drawback of MSMs is that one must be able to specify a correct model for the probability of exposure, $pr(A(k) = a(k) | \bar{L}(k) = \bar{\ell}(k), \bar{A}(k-1) = \bar{a}(k-1))$, for each time k up to end of follow-up. This is unfortunate since, if the $L(k)$ and $A(k)$ are discrete, we could use non-parametric saturated models for small values of k , (say $k = 0, 1, 2$), but for large k we require strong modelling assumptions because there are so many variables in $\bar{\ell}(k) = (\ell(0), \ell(1), \dots, \ell(k))$ and in $\bar{a}(k-1)$. It is unlikely that these assumptions would be precisely correct.

The use of g-estimation of SNMs overcomes the above difficulties. For example, one can use SNMs to estimate the effect of an exposure on mortality in occupational cohort studies.^{14,16} Similarly, one can unbiasedly estimate the causal parameter of a SNM without having to model the probability of treatment

given the past through end of follow-up. Instead, in the setting of a discrete $A(k)$ and $L(k)$ described above, one can unbiasedly estimate the parameters of SNMs by using a saturated model for the probability of exposure $A(k)$ given the past restricted to the first several periods, thus preventing bias due to model misspecification. Of course, as always in statistical analysis, there will be a loss of efficiency of estimation associated with this protection against bias.

Another advantage of SNMs over MSMs is that, although MSMs are useful for estimating the causal effect of the pre-specified treatment regime \bar{a} (e.g., always treat, treat on alternate months, etc.), they are less useful than SNMs for estimating the effect of dynamic treatment plans in which treatment on a given month is decided in part based on a subject's evolving covariate history.^{1,2} It is important to recognize that actual medical treatment regimes are usually dynamic, since if a patient develops a toxic reaction to a drug, the drug must be stopped. Nonetheless, causal questions concerning pre-specified treatment plans, such as estimating the effect of a continuous exposure at a certain level versus no exposure, are of great interest in many areas of epidemiology, such as nutritional, and environmental.

Discussion

We have used a marginal structural Cox proportional hazards model to estimate the causal effect of zidovudine on mortality of HIV-positive patients in the MACS. This method was utilized because standard statistical methods are not appropriate when there exists time-dependent confounding by variables, such as CD4 count, that are affected by previous exposure.

Because of the presence of confounding, the crude mortality rate ratio for zidovudine was 2.3 (95% confidence interval: 2.0, 2.7), erroneously suggesting an increased risk of death among zidovudine users. The rate ratio estimated by the (unweighted) standard model that included only baseline covariates, and that therefore does not adjust for time-dependent confounding, was 1.7 (95% confidence interval: 1.4, 1.9), which still suggests a detrimental effect of zidovudine, probably due to the presence of a substantial amount of time-dependent confounding.

In fact, the mortality rate ratio for zidovudine was 0.7 (95% conservative confidence interval: 0.6, 0.9) in the weighted model that provides, under our assumptions, an unbiased estimate of the causal rate ratio, $\exp(\beta_1)$, of the marginal structural Cox model, even in the presence of time-dependent confounding by intermediate variables.

The difference between the unweighted and weighted estimates is an indication of the amount of confounding due to the time-dependent prognostic factors. The weights can be interpreted as the number of copies of each observation that are created to form a pseudo-population in which censoring does not exist and in which the time-dependent prognostic factors do not predict initiation of zidovudine history (i.e., treatment is unconfounded).

The correctness of the causal inferences from our analyses is dependent on several assumptions. First, we assume that the covariates in $L(t)$ are sufficient to adjust for both confounding and for selection bias due to loss to follow-up. This implies that we have available, on each month, data recorded in $L(t)$ on the history of all time-dependent covariates that (i) are independent predictors of death and (ii) independently predict the probability of starting zidovudine and/or of being censored on that month. Unfortunately, as in all observational studies, these two assumptions cannot be tested by the data. In our analysis we assume this goal has been realized, while recognizing that, in practice, this would never be precisely or sometimes even approximately true. Recently, Robins and co-workers have developed extensions of IPTW estimation of MSMs that allow one to evaluate the sensitivity of one's estimates to increasing violation of these fundamental assumptions.¹⁶

Second, we assume that the models for initiation of zidovudine and censoring, given the past, are

correctly specified. Last, we assume that our MSM for the effect of zidovudine on mortality, within levels of baseline covariates V , is correctly specified.

Although the stated assumptions of MSMs may seem heroic, note that, in point-exposure studies, the same assumptions (no unmeasured confounders, non informative censoring, and no misspecification of the model) are required to give a causal interpretation to the parameters of standard statistical models. Furthermore, when studying the effect of a time-dependent treatment like zidovudine, the assumptions of MSMs are less restrictive than those of standard methods: MSMs do not require the absence of time-dependent confounding by variables affected by previous exposure.

REFERENCES

1. Robins, J.M. (1998). Marginal structural models. *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*, pp. 1-10.
2. Robins J.M. (1999). Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. In: **Statistical Models in Epidemiology: The Environment and Clinical Trials**. Halloran, M.E. and Berry, D., eds. NY: Springer-Verlag, pp. 95-134 (to appear).
3. Robins, J.M. (1998). Structural nested failure time models. **Survival Analysis**, P.K. Andersen and N. Keiding, Section Editors. **The Encyclopedia of Biostatistics**, P. Armitage and T. Colton, Editors. Chichester, UK: John Wiley & Sons, pp. 4372-4389.
4. Robins, J.M., Hernán, M., and Brumback, B. (1998). Marginal structural models and causal inference in epidemiology (submitted to *Epidemiology*).
5. Kaslow, R.A., Ostrow, D.G., Detels, R., Phair, J.P., Polk, B.F., Rinaldo, C.R. (1987). The Multi-center AIDS Cohort Study: Rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, 126, 310-318.
6. Graham, N.M.H., Zeger, S.L., Park, L.P., et al. (1992). The effects on survival of early treatment of human immunodeficiency virus infection. *New England Journal of Medicine*, 326, 1037-1042.
7. Robins, J.M. (1997). Causal inference from complex longitudinal data. In: **Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)**, M. Berkane, Editor. NY: Springer Verlag, 69-117.
8. Kinloch-de Loes, S., Hirschel, B.J., Hoen, B., et al. (1995). A controlled trial of zidovudine in primary human immunodeficiency virus infection. *New England Journal of Medicine*, 333, 408-413.
9. d'Agostino, R.B., Lee, M-L., Belanger, A.J. (1990). Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. *Statistics in Medicine*, 9, 1501-1515.
10. Liang, K-Y., and Zeger, S.L. (1986), Longitudinal Data Analysis Using Generalized Linear Model. *Biometrika*, 73, 13-22.
11. Hastie, T.J., Tibshirani, R.J. (1995). **Generalized additive models**. New York: Chapman & Hall, 335.
12. Robins, J.M. (1998). Correction for non-compliance in equivalence trials. *Statistics in Medicine*, 17:269-302.

13. Robins, J.M. (1987). A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *Journal of Chronic Disease* (40, Supplement), 2:139s-161s.
14. Robins, J.M., Blevins D, Ritter G, and Wulfsohn M. (1992). G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, 3:319-336. [published errata appear in *Epidemiology*, 4:189.]
15. Robins, J.M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. **Health Service Research Methodology: A Focus on AIDS**. Eds: Sechrest L., Freeman H., Mulley A. Washington, D.C.: U.S. Public Health Service, National Center for Health Services Research. pp. 113-159.
16. Robins JM, Scharfstein D, Rotnitzky A. (1999). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: **Statistical Models in Epidemiology: The Environment and Clinical Trials**. Halloran, M.E. and Berry, D., eds. NY: Springer-Verlag, pp. 1-94 (to appear).

APPENDIX 1: SAS code for the marginal structural Cox proportional hazards model

In this appendix we provide SAS code to fit the Cox proportional hazards MSM described in the text. The original MACS data file contains one record per man, but here we use a transformed, or pooled, file (MAIN) with each person-month as a separate record. This file format is necessary to fit pooled logistic models. The code used to generate the pooled data set from the original one is available from the first author upon request. The records in the file MAIN must be sorted by patient identification number (variable ID) and, within each ID level, by month of follow-up (MONTH). The SAS code shown below is organized as follows. First, we use Proc Logistic to fit four pooled logistic models (two for the probability of remaining off zidovudine, and two for the probability of remaining uncensored) and obtain their predicted values. Second, we use a SAS data step to calculate the weights for each person-month based on the predicted values of the previous four models. Last, we use Proc Genmod to fit the final weighted pooled logistic model that estimates the causal parameter of interest and its robust standard error.

The outcome variable in Models 1 and 2 is a dichotomous variable A indicating whether the patient had started ($A = 1$) or remained off ($A = 0$) zidovudine on that month. When the option descending is not specified, Proc Logistic models the probability that the outcome variable is 0. Hence Models 1 and 2 model the probability of remaining off zidovudine. The **where** statement restricts the analysis to patients not previously on zidovudine by specifying that either month of follow-up (Month) is less than or equal to month of onset of zidovudine (ZDV_M), or zidovudine was never initiated during the follow-up period (ZDV_M is coded as missing if this is the case). Model 1 includes as regressors a time-dependent intercept and the baseline covariates V : age, baseline CD4, WBC, and hematocrit; and presence of symptoms. Model 2 includes, in addition, the time-dependent covariates $L(t)$: most recently available CD4, WBC, hematocrit, symptoms, and AIDS diagnosis. We estimate the time-dependent intercept by a smooth function of the time since beginning of follow-up (Month) using natural cubic splines with five knots. To do so, we need to include, as regressors, the variables Month1, Month2, and Month3, that are specific polynomial functions of Month, calculated with the cubic splines SAS macro RCSPLINE in survrisk.pak, by Frank Harrel, publicly available on ??).

The outcome variable in Models 3 and 4 is a dichotomous variable C indicating whether the patient was censored ($C = 1$) or uncensored ($C = 0$) on that month. Thus models 3 and 4 model the probability of remaining uncensored for each person-month. All available person-months are used. Model 3 includes the baseline covariates and the time-dependent intercept, while Model 4 includes the time-dependent covariates (to which we add A) as well.

For each model, we output a new data file (option **out=** in Proc Logistic) that contains, for each person-month, the original variables plus the predicted values from the model (option **p=**). As an example, the first Proc Logistic creates the data set MODEL1 with its predicted values as the variable PZDV_0.

In the following data step we merge the four output files in the file MAIN_W that contains the predicted values from the four logistic models. We then compute the numerator $K2_0$ and the denominator $K2_W$ of the weights $sw_i^\dagger(t)$ by forming the product up to month t of the subject-specific predicted values from Models 3 and 4.

Similarly, we calculate the numerator $K1_0$ and the denominator $K1_W$ of the weights $sw_i(t)$ for months on which the subject has not yet started zidovudine by forming the product up to that month of the subject-specific predicted values from Models 1 and 2. For a month on which a subject did begin zidovudine, we multiply by 1 minus the predicted value. For months subsequent to starting zidovudine, we no longer update $K1_0$ and $K1_W$. Then we use the numerators and denominators to calculate the

“stabilized” weights $sw_i(t) \times sw_i^\dagger(t)$ (STABW), and use the denominators alone to calculate the “non stabilized” weights $w_i(t) \times w_i^\dagger(t)$ (NSTABW).

Finally, we call Proc Genmod to fit a weighted pooled logistic model for survival to obtain consistent estimates of the parameters of our Cox MSM. The outcome variable of this model, D , is a dichotomous variable indicating whether the patient died ($D = 1$) or remained alive ($D = 0$) on that month. The program will provide robust standard errors for the model parameters when the option `repeated` is included. The patient identification variable and the independent working correlation matrix (`subject=ID/type=ind`) must be specified. We fit the model using the stabilized weights by specifying the variable STABW in the `scwgt` statement. Specifying the variable NSTABW fits the model with non stabilized weights.

The SAS code given below can also be used to fit the logistic MSM of our companion paper. The only difference is that the final weighted logistic model in Proc Genmod includes a single observation per person using as outcome variable the logistic variable Y of our companion paper, rather than the survival variable D considered in this paper.

```

/* Model 1 */
proc logistic data=MAIN;
  where MONTH<=ZDV_M or ZDV_M=.;
  model A = AGE CD4_01 CD4_02 WBC_01 WBC_02 HCT_01 HCT_02 SYMPT_0
           MONTH MONTH1-MONTH3;
  output out=model1 p=pzdv_0;
run;

/* Model 2 */
proc logistic data=MAIN;
  where MONTH<=ZDV_m or ZDV_M=.;
  model A = AGE CD4_01 CD4_02 WBC_01 WBC_02 HCT_01 HCT_02 SYMPT_0
           CD4_1 CD4_2 WBC_1 WBC_2 HCT_1 HCT_2 SYMPT AIDS MONTH MONTH1-MONTH3;
  output out=model2 p=pzdv_w;
run;

/* Model 3 */
proc logistic data=MAIN;
  model C = AGE CD4_01 CD4_02 WBC_01 WBC_02 HCT_01 HCT_02 SYMPT_0
           MONTH MONTH1-MONTH3;
  output out=model3 p=punc_0;
run;

/* Model 4 */
proc logistic data=MAIN;
  model C = AGE CD4_01 CD4_02 WBC_01 WBC_02 HCT_01 HCT_02 SYMPT_0 ZDV

```

```
CD4_1 CD4_2 WBC_1 WBC_2 HCT_1 HCT_2 SYMPT AIDS MONTH MONTH1-MONTH3;
```

```
output out=model4 p=punc_w;
```

```
run;
```

```
data main_w;
```

```
merge model1 model2 model3 model4;
```

```
by ID MONTH;
```

```
/* variables ending with _0 refer to the numerator of the weights  
variables ending with _w refer to the denominator of the weights */
```

```
/* reset the variables for a new patient */
```

```
if first.id then do;
```

```
k1_0=1; k2_0=1; k1_w=1; k2_w=1;
```

```
end;
```

```
retain k1_0 k2_0 k1_w k2_w;
```

```
/* Inverse probability of censoring weights */
```

```
k2_0=k2_0*punc_0;
```

```
k2_w=k2_w*punc_w;
```

```
/* Inverse probability of treatment weights */
```

```
/* patients not on zidovudine */
```

```
if zdv_m>day or zdv_m=. then do;
```

```
k1_0=k1_0*pzdv_0;
```

```
k1_w=k1_w*pzdv_w;
```

```
end;
```

```
/* patients that start zidovudine today */
```

```
else if zdv_m=day then do;
```

```
k1_0=k1_0*(1-pzdv_0);
```

```
k1_w=k1_w*(1-pzdv_w);
```

```
end;
```

```
/* patients that have already started zidovudine */
```

```
else do;
```

```
k1_0=k1_0;
```

```
k1_w=k1_w;
```

```
end;
```

```
/* Stabilized and non stabilized weights */
```

```
stabw=(k1_0*k2_0)/(k1_w*k2_w);
nstabw=1/(k1_w*k2_w);
run;

proc genmod data=main_w;
  class id;
  make 'classlevels' noprint;
  make 'parminfo' noprint;
  model D= ZDV AGE CD4_01 CD4_02 WBC_01 WBC_02 HCT_01 HCT_02 SYMPT_0
          MONTH MONTH1-MONTH3/ link=logit dist=bin;
  scwgt stabw;
  repeated subject=ID/ type=ind;
run;
```

TABLE 1. Inverse-probability-of-treatment weighted estimates of the causal effect of zidovudine therapy on mortality, MACS

	RR*	95% CI*	
Unweighted estimates, [†]			
unadjusted	2.33	2.01 - 2.70	
only baseline covariates	1.66	1.43 - 1.94	
		Valid	Invalid
	RR*	95% Conservative CI*	Model-based [§] 95% CI*
Weighted estimates, [‡]			
stabilized weights	0.69	0.56 - 0.85	0.60 - 0.79
non stabilized weights	0.72	0.57 - 0.91	0.69 - 0.76

*RR = mortality rate ratio (zidovudine users versus non users), CI = confidence interval

[†]Non causal models, shown for comparison purposes only. The crude model includes only intercept and zidovudine (yes, no). The model with baseline covariates includes also: age, baseline CD4 count (< 200, 200-499, ≥ 500 cells/ μ L), baseline WBC (< 3000, 3000-4999, ≥ 5000 cells/ μ L), baseline hematocrit (< 35, 35-39, ≥ 40%), presence of symptoms at baseline (yes if fever, oral candidiasis, diarrhea, weight loss, oral hairy leukoplakia, or herpes zoster; no if otherwise).

[‡]Weights calculated as described in the text using data on baseline covariates plus most recent CD4 count, WBC, hematocrit, presence of symptoms, AIDS diagnosis, and previous zidovudine use (only in models for censoring).

[§]The model-based intervals are not valid for weighted models because they fail to account for the within-subject covariances induced by weighting.

TABLE 2. Inverse-probability-of-treatment weighted estimates of the parameters of a marginal structural model for the causal effect of zidovudine on mortality, MACS

Covariates [†]	Parameter estimate	Robust standard error	Conservative 95% CI*
Zidovudine	-0.375	0.109	-0.588, -0.162
Age	0.028	0.009	0.011, 0.045
Baseline CD4, cells/ μL			
< 200	1.874	0.176	1.528, 2.219
200-499	0.581	0.108	0.371, 0.792
≥ 500	0.000		
Baseline WBC, cells/ μL			
< 3000	0.026	0.244	-0.452, 0.503
3000-4999	0.188	0.099	-0.006, -0.382
≥ 5000	0.000		
Baseline hematocrit, %			
< 35	1.034	0.295	-0.467, 1.612
35-39	0.580	0.139	0.307, 0.669
≥ 40	0.000		
Presence of symptoms	0.443	0.115	0.218, 0.669

*CI = confidence interval

[†]Weighted logistic model including the covariates listed in the table plus a time-varying intercept (not shown). Weights were estimated by $s\hat{w}_i(t) \times s\hat{w}_i^\dagger(t)$ as defined in the text.

TABLE 3. Estimated probability of having one’s own observed treatment history (estimated denominator of $sw_i(t)$) and censoring history (estimated denominator of $sw_i(t)$) at 24 and 84 months of follow-up, MACS

24 MONTHS ($n = 1919$)	Mean (SD)*	Median (IQR)*	Minimum	Maximum
Probability of having observed zidovudine history				
- given baseline covariates [†]	0.56 (0.34)	0.72 (0.79)	0.0040	.860
- given time-varying covariates [‡]	.58 (0.36)	0.71 (0.85)	0.002	0.94
Probability of being uncensored				
- given baseline covariates [†]	0.95 (0.01)	0.95 (0.02)	0.89	0.99
- given time-varying covariates [‡]	0.95 (0.01)	0.95 (0.02)	0.88	0.99
84 MONTHS ($n = 771$)	Mean (SD)*	Median (IQR)*	Minimum	Maximum
Probability of having observed zidovudine history				
- given baseline covariates [†]	0.12 (0.17)	0.01 (0.26)	0.001	0.49
- given time-varying covariates [‡]	0.17 (0.25)	0.02 (0.27)	0.002	0.74
Probability of being uncensored				
- given baseline covariates [†]	0.66 (0.05)	0.66 (0.08)	0.54	0.81
- given time-varying covariates [‡]	0.66 (0.05)	0.66 (0.08)	0.53	0.81

*SD = standard deviation, IQR = interquartile range

Age, baseline CD4 count (< 200, 200-499, ≥ 500 cells/ μ L), baseline WBC (< 3000, 3000-4999, ≥ 5000 cells/ μ L), baseline hematocrit (< 35, 35-39, ≥ 40%), presence of symptoms at baseline (yes if fever, oral candidiasis, diarrhea, weight loss, oral hairy leukoplakia, or herpes zoster; no otherwise).

Baseline covariates plus most recent CD4 count, WBC, hematocrit, presence of symptoms, AIDS diagnosis, and previous zidovudine use (only in models for censoring).