

Structural accelerated failure time models for survival analysis in studies with time-varying treatments[†]

Miguel A. Hernán^{1*}, Stephen R. Cole², Joseph Margolick^{2,3}, Mardge Cohen⁴
and James M. Robins^{1,5}

¹*Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA*

²*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

³*Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA*

⁴*Department of Medicine, Cook County Hospital, Chicago, IL, USA*

⁵*Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA*

SUMMARY

Background In the absence of unmeasured confounding and model misspecification, standard methods for estimating the causal effect of time-varying treatments on survival are biased when (i) there exists a time-dependent risk factor for survival that also predicts subsequent treatment and (ii) past treatment history predicts subsequent risk factor level. In contrast, structural models provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent. The parameters of nested structural models are estimated by g-estimation and those of marginal structural models by inverse probability weighting.

Methods We describe a nested structural accelerated failure time model and use it to estimate the total causal effect of highly active antiretroviral therapy (HAART) on the time to AIDS or death among human immunodeficiency virus (HIV)-infected participants of the Multicenter AIDS Cohort and Women's Interagency HIV Studies. The Appendix describes g-estimation and methods to deal with censoring.

Results Comparing the regime 'always treated' to 'never treated,' the AIDS-free survival time ratio was 2.5 (95% confidence interval [CI]: 1.7, 3.3).

Conclusions Our finding of a strongly beneficial effect is consistent with results from randomized trials and from a previous analysis of the same data using a marginal structural Cox model. In contrast, a previous analysis using a standard (non-structural) model did not find an effect of treatment on survival. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS — causal inference; confounding; cohort studies

INTRODUCTION

The goal of many epidemiologic studies is to estimate the causal effect of a time-varying treatment on survival. Two useful models for survival analysis are the Cox proportional hazards model and the accelerated

failure time (AFT) model. The widely used Cox model measures causal effect on the hazard (rate) ratio scale, whereas the less used AFT model^{1,2} measures causal effect on the survival time ratio scale. Both the Cox model and semiparametric versions of the AFT model^{3,4} are models that leave the baseline hazard (or, equivalently, the baseline survival distribution) unspecified.

However, even in the absence of unmeasured confounding and model misspecification, these standard models for survival analysis will provide estimates that fail to have a causal interpretation when

*Correspondence to: Dr Miguel A. Hernán, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA.

E-mail: miguel_hernan@post.harvard.edu

[†]No conflict of interest was declared.

Received 28 January 2004

Revised 1 July 2004

Accepted 27 July 2004

(i) there exists a measured time-dependent risk factor for survival that also predicts subsequent treatment, and (ii) past treatment history predicts subsequent risk factor level.^{5,6} Factors that meet condition (i) are known as time-dependent confounders. For example, when estimating the causal effect of highly active antiretroviral therapy (HAART) on the survival of individuals infected with the human immunodeficiency virus (HIV), condition (i) is met by the variable CD4 cell count because a low CD4 cell count is both a risk factor for survival and used by physicians to decide whether to initiate HAART. Also, condition (ii) is met because prior HAART use increases CD4 cell count. Therefore, including the time-dependent confounder CD4 cell count in a standard Cox or AFT model may not appropriately adjust for confounding.

In contrast to standard Cox and AFT models, structural Cox and AFT models can be used to estimate causal effects when conditions (i) and (ii) hold. We have previously used a marginal structural Cox model to estimate the causal effect of HAART on the hazard of AIDS or death of HIV-infected individuals.⁷ The causal hazard ratio from the marginal structural model was 0.54 (95% confidence interval [CI]: 0.38, 0.78) when comparing continuous treatment with HAART versus no treatment with HAART. This hazard ratio was estimated by inverse probability weighting.⁸

The simultaneous presence of conditions (i) and (ii), and thus the problem of time-dependent confounding by factors affected by prior treatment, is ubiquitous in pharmacoepidemiology. Other examples of time-dependent confounders that are affected by prior treatment are upper gastrointestinal bleeding when studying the effect of NSAIDs on gastric cancer, measures of disease severity when studying the effect of methotrexate on the mortality of patients with rheumatoid arthritis and hematocrit when studying the effect of erythropoietin on the mortality of dialyzed patients. These examples share a common structure that can be represented using causal diagrams.⁶

In this paper, we review the differences between structural models and standard regression models for survival analysis, describe a structural AFT model, and illustrate the application of this model for estimating the effect of HAART on AIDS-free survival in two prospective cohort studies of HIV-infected individuals. The Appendix describes g-estimation and methods to deal with censoring. We first describe the data for the analysis.

DATA AND NOTATION

Our analysis pooled information from two ongoing prospective studies of the natural history of HIV infection:

the Multicenter AIDS Cohort Study (MACS),⁹ which beginning in 1984 enrolled 5622 homosexual men in Baltimore, Chicago, Pittsburgh and Los Angeles; and the Women's Interagency HIV Study (WIHS),¹⁰ which beginning in 1994 enrolled 2628 women in New York, Chicago, Los Angeles, San Francisco and Washington, DC. Every 6 months, participants in both studies completed an extensive interviewer-administered questionnaire with information on antiretroviral therapy use and HIV-related symptoms, and provided a blood sample for the determination of CD4 cell count and plasma HIV-1 RNA concentration. Procedures for determination of HAART use and outcome ascertainment have been described in detail elsewhere.⁷ Once a participant reported initiation of HAART, he or she was considered to remain on HAART for the duration of follow up.

Analyses presented here are limited to the 1498 participants who were HIV positive, AIDS-free and had not initiated HAART before the first eligible study visit. Each participant contributed a maximum of 13 person-visits of follow-up from the baseline visit (first eligible visit after October 1995) to the last visit he or she was seen free of clinically-defined AIDS¹¹ and alive, or the visit before April 2002, whichever came first. The follow-up of participants missing any time-varying characteristic at baseline started at the first subsequent visit when values were observed.

Study visit is indexed by k and takes values from $k = 0$ (October 1995 to April 1996) to 12 (October 2001 to April 2002). We use capital letters to represent random variables and lower case letters to represent possible realizations (values) of random variables. $A_i(k)$ is a time-varying dichotomous variable (0, 1) indicating whether subject i was on HAART at visit k . $L_i(k)$ is a vector of relevant prognostic factors measured at the visit $k - 1$. We use overbars to denote the history of a time-dependent variable. For example, $\bar{A}_i(k)$ represents all treatment indicators for subject i from visit 0 to k , i.e. $\bar{A}_i(k) = [A_i(0), A_i(1), A_i(2), \dots, A_i(k)]$. For modeling purposes, we assumed that $\bar{L}_i(k)$ was appropriately summarized by gender, race, age and calendar year at baseline; non-HAART antiretroviral therapy use, PCP prophylaxis use, CD4 cell count and HIV-1 RNA concentration at baseline and at k ; and number of days since the prior visit. T_i is subject i 's time of death or diagnosis of AIDS, whichever comes first, and is measured in months since baseline.

We often suppress the i subscript denoting individual because we assume that the random vector for each subject is drawn independently from a distribution common to all subjects. Also, we often suppress the time index k when referring to the subject's entire treatment history \bar{A} . We use the symbol \bar{A} to indicate

statistical independence, for example $X \perp\!\!\!\perp Y|Z$ means X is conditionally independent of Y given Z .¹²

STRUCTURAL VERSUS ASSOCIATIONAL MODELS

One common approach to estimate the causal effect of the time-varying treatment A on the survival time T is modeling the hazard (rate) of death at time t , $\lambda_T[t|\bar{A}(t)]$ as a function of past treatment history through t , $\bar{A}(t)$. For example, consider the Cox model

$$\lambda_T[t|\bar{A}(t)] = \lambda_0[t]\exp[\alpha \times \text{dur } \bar{A}(t)]$$

where $\lambda_0[t]$ is the baseline hazard at t , α is the log hazard ratio and treatment history $\bar{A}(t)$ is summarized by the duration of treatment $\bar{A}(t)$ through time t .

The parameter α measures the association between treatment and mortality on the (log) hazard ratio scale. However, α does not generally measure the causal effect of treatment on mortality on any scale. This is so because α may differ from zero even if treatment $\bar{A}(t)$ has no causal effect (causative or preventive) on the hazard of T , whenever there exist time-dependent confounders $\bar{L}(t)$ (e.g. patients with a low CD4 cell count are more likely to receive HAART and they also have a greater risk of death). The parameter α is not appropriate for causal inference because it fails to distinguish causal effects from other sources of association such as confounding or selection bias. The parameters of standard regression models quantify associations. We therefore refer to standard models as associational models.

In contrast, the parameters of structural models quantify causal effects. To describe a structural model, we need to introduce counterfactual (also known as potential) outcomes. Let $T_{\bar{a}}$ be a subject's time of death had she received treatment regime \bar{a} rather than her actual treatment history \bar{A} . Some examples of treatment regimes are 'always treat' $\bar{a} = (1, 1, 1, 1, \dots)$, 'never treat' $\bar{a} = (0, 0, 0, 0, \dots)$ and 'treat at every other visit' $\bar{a} = (1, 0, 1, 0, \dots)$. The subject's actual survival time T is, by definition, her counterfactual survival time under the treatment history \bar{A} that the subject received, i.e. $T = T_{\bar{A}}$.

Let $\lambda_{T_{\bar{a}}}[t]$ be the mortality hazard at time t in a hypothetical study in which all subjects received treatment regime \bar{a} . We can simultaneously model the mortality hazard $\lambda_{T_{\bar{a}}}[t]$ for all treatment regimes \bar{a} by using the Cox model

$$\lambda_{T_{\bar{a}}}[t] = \lambda_{T_0}[t]\exp[\beta \times \text{dur } \bar{a}(t)]$$

where $\lambda_{T_0}[t]$ is the mortality hazard in a hypothetical study in which no subject received treatment by time t , β is the log hazard ratio and treatment $\bar{a}(t)$ is summarized by the duration of treatment $\text{dur } \bar{a}(t)$ through time t . Note that $\beta = 0$ implies that, no matter the treatment regime \bar{a} , the hazards $\lambda_{T_{\bar{a}}}[t]$ and $\lambda_{T_0}[t]$ would be equal. That is, $\beta = 0$ implies that the treatment does not have a causal effect on the outcome on the (log) hazard ratio scale. We refer to models for the distribution of counterfactual outcomes, such as $T_{\bar{a}}$, as structural (or causal) models. The parameter β in the structural Cox model measures the causal effect of treatment on mortality. We refer to $\exp(\beta)$ as the causal rate ratio.

Epidemiologists conducting etiologic studies are much more interested in the causal parameter β than in the associational parameter α because β has direct scientific and policy implications. For example, the value of β for HAART, a treatment proven to be beneficial, will be (strongly) negative. On the other hand, the value of α for HAART may be close to zero or even positive because of time-dependent confounding (i.e. patients with a more severe disease are more likely to receive the treatment).

One attempt to give a causal interpretation to associational parameters is to include the time-dependent confounders L (e.g. CD4 cell count) as covariates in the associational Cox model. Unfortunately, when some time-dependent covariates in L are affected by prior treatment, this approach does not allow one to give a causal interpretation to the treatment parameter α as either the overall effect or the direct effect of treatment on mortality.^{5,6,13}

THE IDENTIFYING ASSUMPTION

Causal parameters can be identified in ideal randomized experiments (that is, causal parameters can be validly estimated). Consider an experiment in which all subjects are untreated at baseline, and they are randomly assigned to either continuous treatment after some random visit ($A(k) = 1$ for all visits k after visit R) or never treatment (i.e. $A(k) = 0$ for all k). Under this design, the survival time distribution in those continuously treated (after certain visit) had they remained always untreated equals the survival time distribution in the continuously untreated. The hazard of failure $\lambda_{T_{\bar{a}}}[t|\bar{A}(t) = \bar{a}(t)] = \lambda_T[t|\bar{A}(t) = \bar{a}(t)]$ at t among those with observed treatment history $\bar{a}(t)$ is equal to the hazard $\lambda_{T_{\bar{a}}}[t]$ at t had, contrary to fact, all subjects received treatment $\bar{a}(t)$. That is, the subjects observed to be alive at t with any particular treatment history $\bar{a}(t)$ are comparable to the subset of the entire study population who would be alive at t had all

subjects been forced to take regime $\bar{a}(t)$. Such comparability or exchangeability is mathematically expressed as $T_{\bar{a}} \perp\!\!\!\perp \bar{A}$, which is read as ‘any counterfactual survival time $T_{\bar{a}}$ and observed treatment history \bar{A} are independent’. An equivalent definition of exchangeability is that the survival time $T_{\bar{a}}$ is independent of the treatment status $A(k)$ given past treatment history $\bar{A}(k-1)$ among those alive at the time of visit k , $u(k)$. That is, $T_{\bar{a}} \perp\!\!\!\perp A(k) | \bar{A}(k-1) = \bar{a}(k-1), T > u(k)$. Exchangeability implies lack of confounding, and therefore in ideal randomized experiments the associational parameter α is a consistent estimator of the causal parameter β .

In observational studies, data suffice to identify associational parameters (because different associational parameters imply different distributions for the observed data), but the data do not identify causal parameters (because of potential confounding by unmeasured factors, a given distribution for the observed data set is consistent with many values of the causal parameter). Hence data from observational studies must be supplemented with an assumption to identify causal parameters. One such identifying assumption is the above exchangeability criterion $T_{\bar{a}} \perp\!\!\!\perp A(k) | \bar{A}(k-1) = \bar{a}(k-1), T > u(k)$, but this is usually too strong an assumption. For example, if untreated patients with low CD4 cell count are more likely to start treatment at k than untreated patients with high CD4 cell count, then those who start treatment and those who remain untreated at k are clearly not exchangeable because the former have, on average, a more severe disease than the latter. The survival time distribution in the treatment initiators, had they remained untreated, would have been worse than the survival time distribution in those who remained untreated. In other words, given past treatment, treatment $A(k)$ predicts the counterfactual survival time under no treatment $T_{\bar{a}=0}$ and exchangeability does not hold.

A weaker and more plausible assumption would be that, among those with low CD4 cell count at visit k , those who receive treatment and those who do not receive treatment at k are exchangeable (i.e. they have the same counterfactual survival time distributions). An even more plausible assumption is that the treated and the untreated are (conditionally) exchangeable at time k when they have the same treatment and covariate (CD4 cell count, viral load etc.) history. This is mathematically expressed as

$$T_{\bar{a}} \perp\!\!\!\perp A(k) | \bar{A}(k-1) = \bar{a}(k-1), \bar{L}(k) = \bar{l}(k), \\ T > u(k) \quad \text{for all } \bar{a}(k-1) \quad \text{and } \bar{l}(k) \quad (\text{A})$$

which is read as ‘ $T_{\bar{a}}$ and $A(k)$ are conditionally independent treatment history $\bar{A}(k-1)$ and covariate history $\bar{L}(k)$, for all histories, among those alive at visit k ’.

Assumption (A) says that there is no confounding given the measured covariates in $\bar{L}(k)$ and is therefore known as the assumption of no unmeasured confounders. Assumption (A) would automatically hold if treatment had been randomly assigned at each visit k even if the probability of receiving treatment were not equal for all subjects, as long as the probability of receiving treatment was the same for subjects with identical treatment and covariate histories till k . For this reason, assumption (A) is also known as the sequential randomization assumption.

Methods for causal inference from observational data provide causal estimates only if assumption (A)—or variations of it—holds. Unfortunately, the assumption is uncheckable and therefore all causal inferences from observational data are risky. Because we have measured, and included in $\bar{L}(k)$, the most important clinical and laboratory indicators that are currently used by physicians to determine the severity of HIV disease and decide whether to prescribe HAART, we will assume throughout this paper that, in our data, assumption (A) is approximately true for treatment initiation at each visit k .

STRUCTURAL AFT MODEL

For simplicity, we first consider a dichotomous, non-time-varying treatment A . Let T_a be the counterfactual survival time and $\lambda_{T_a}[t]$ the counterfactual hazard at time t under treatment value a (i.e. either 0 or 1). We have seen that a structural Cox model $\lambda_{T_a}[t] = \lambda_{T_0}[t] \exp[\beta a]$ uses the causal rate ratio $\exp[\beta] = \lambda_{T_{a=1}}[t] / \lambda_{T_0}[t]$ to measure the effect of treatment A on mortality. Consider the model $T_a = T_0 \exp[-\psi^* a]$ that replaces the counterfactual hazards by the counterfactual survival times, and thus uses the survival time ratio $\exp[-\psi^*] = T_{a=1} / T_{a=0}$ to measure the causal effect of treatment A on each subject’s mortality with ψ^* constant across subjects. This is a structural AFT model.

The survival time ratio $\exp[-\psi^*]$ is the expansion (or contraction) in survival time attributable to treatment. Positive values of ψ^* indicate that a subject’s survival time when treated is shorter than the survival time when untreated (i.e. treatment is harmful) whereas negative values of ψ^* indicate that the survival time when treated is longer than the survival time when untreated (i.e. treatment is beneficial). When $\psi^* = 0$, there is no causal effect of treatment on survival time. For reasons

that will become clear later, we prefer to present this AFT model as

$$T_0 = T_a \exp[\psi^* a]$$

The above AFT model is deterministic because it assumes that, for each subject, the counterfactual survival time under no treatment $T_0 = T_{a=0}$ can be computed without error as a function of $T_{a=1}$ and ψ^* . Under this model, if subject i would die before subject j had they both been untreated, i.e. $T_{i,a=0} < T_{j,a=0}$, then subject i would also die before subject j had they been treated, i.e. $T_{i,a=1} < T_{j,a=1}$. We therefore say that this AFT model assumes rank preservation of the subjects' death times across treatment regimes. (Although non-rank-preserving AFT models¹⁴ are not covered in this paper, we note the method of g-estimation can be used unchanged if the data follow an AFT model regardless of whether or not it is rank preserving.)

We need to generalize the AFT model to accommodate time-varying treatments. To do so, let us first take a closer look at how T_0 is computed for non-time-varying treatments. Consider a subject with survival time $T_{a=1} = 2.3$ months. Under the AFT model, his survival time T_0 when untreated equals $T_{a=1} \exp[\psi^*] = 2.3 \times \exp[\psi^*] = \exp[\psi^*] + \exp[\psi^*] + 0.3 \times \exp[\psi^*]$. The value of T_0 can be viewed as the sum of the time-specific contributions.

Now consider time-varying treatment regimes $\bar{a} = [a(0), a(1), a(2) \dots]$ where some $a(t)$ equal 1 and others equal 0, and a subject with survival time $T_{\bar{a}} = 3.7$ months under the regime 'treat every other visit' $\bar{a} = [1, 0, 1, 0 \dots]$. (For simplicity, we assume that visits take place the first day of each month.) If we again compute T_0 , the subject's survival time under no treatment, as the sum of time specific contributions then T_0 equals $\exp[\psi^*] + 1 + \exp[\psi^*] + 0.7$. At each time, the contribution to the value of T_0 is either $\exp[\psi^*]$ if regime \bar{a} includes treatment at time t , $a(t) = 1$, or 1 if regime \bar{a} does not include treatment at time t , $a(t) = 0$. The value of T_0 is the area under a step function like the one shown in Figure 1 for a positive value of ψ^* . In general, the step function becomes a curve because treatment may change at any time, and therefore T_0 is the area under the curve, i.e. the integral of $\exp[\psi^* a(t)]$ from 0 to $T_{\bar{a}}$. Thus a more general representation of a structural AFT model for time-varying treatments is

$$T_0 = \int_0^{T_{\bar{a}}} \exp[\psi^* a(t)] dt$$

where $\exp[-\psi^*]$ is the expansion (or contraction) factor in survival time when comparing continuous

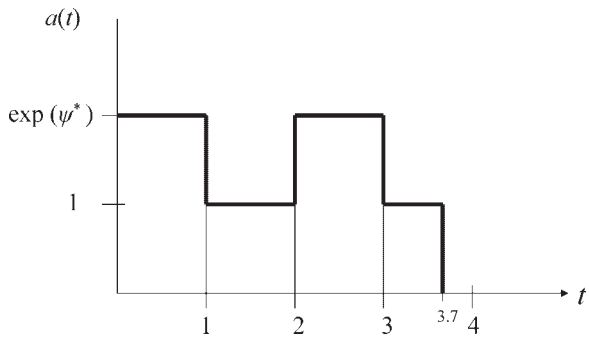


Figure 1. Counterfactual T_0 is the area under the curve

treatment (from time zero until death occurs) versus no treatment, i.e.

$$\exp[-\psi^*] = \frac{T_{\bar{a}}}{T_0} \quad \text{for } \bar{a} = [1, 1, 1, \dots]$$

As presented, the structural AFT model includes only the generally unobserved variables T_0 and $T_{\bar{a}}$. To have any hope of consistently estimating ψ^* , we need to link the model to observed variables. Recall that a subject's observed survival time T is, by definition, her counterfactual survival time under the treatment regime \bar{A} that she received, i.e. $T = T_{\bar{A}}$. Thus, the structural AFT model implies that

$$T_0 = \int_0^T \exp[\psi^* A(t)] dt$$

holds for all subjects. We have now linked the structural model and the counterfactual variable T_0 to the observed data. But, in general, we do not know the value of either T_0 or ψ^* . Therefore, the usual representation of our structural AFT model is

$$H(\psi) = \int_0^T \exp[\psi A(t)] dt \quad (1)$$

where ψ can take any value and $H(\psi)$ is the counterfactual survival time under no treatment only when the true value ψ^* of the parameter ψ is used, i.e. $H(\psi^*) = T_0$. The notation $H(\psi)$ reminds us that the value of H is a function of the value of the parameter ψ .

Conventional statistical procedures for the estimation of model parameters are of little help to estimate the true value ψ^* of the parameter ψ of our structural AFT model. Rather, ψ^* must be estimated by either inverse probability weighting or g-estimation. Both approaches require modeling the probability of receiving treatment at each time k as a function of

previous treatment $\bar{A}(k-1)$ and covariate $\bar{L}(k)$ histories. The method of g-estimation is described in the Appendix. We suggest that readers unfamiliar with g-estimation read the Appendix now. For the analyses presented in this article, we used g-estimation to estimate ψ^* by fitting the pooled logistic model

$$\begin{aligned} \text{logit Pr}[A(k) = 1 | \bar{A}(k-1), \bar{L}(k), T > u(k), H(\psi)] \\ = \theta_0(k) + \theta_1 A(k-1) + \theta_2 L(k) + \theta_3 H(\psi) \end{aligned} \quad (2)$$

where $\theta_0(k)$ is a time-varying intercept (e.g. a cubic spline for time of follow-up), and past treatment and confounder history are summarized by their baseline value and last value prior to $A(k)$ (other specifications could be used, e.g. last two values, cumulative exposure, cumulative average exposure etc.).

NESTED STRUCTURAL AFT MODEL

Structural AFT models are *nested* models. This section explains why. To do so, we first need to re-express model (1).

The counterfactual survival time $H(\psi^*)$ is the interval from the beginning of the follow-up, $t=0$, to the subject's death under no treatment. Consider someone alive at visit k , $t=u(k)$ and let $H(k, \psi)$ be the interval from visit k to her death time had she not received any treatment from visit k on. Following the reasoning of the previous section, this counterfactual survival time is

$$H(k, \psi) = \int_{u(k)}^T \exp[\psi A(t)] dt \quad \text{for } k = 0, 1, 2, \dots \quad (3)$$

for $\psi = \psi^*$. Models (1) and (3) are logically equivalent.

Model (1) can be expressed as the sum $H(\psi) = \int_0^{u(k)} \exp[\psi A(t)] dt + \int_{u(k)}^T \exp[\psi A(t)] dt$ for any visit k . We have partitioned the integral into two integrals: one from $t=0$ to (the instant before) visit k , and another from visit k to T . The second integral is precisely $H(k, \psi)$ from model (3). The first integral $\int_0^{u(k)} \exp[\psi A(t)] dt$ is a function of $\bar{A}(k-1)$ but not of $A(k)$. That is, once we condition on past treatment history $\bar{A}(k-1)$ in model (2), $\int_0^{u(k)} \exp[\psi A(t)] dt$ becomes fixed (non-random) or constant. A constant is always conditionally independent of $A(k)$ (and of any other variable) no matter what value of ψ is being considered. Hence the first integral is irrelevant for g-estimation purposes because it contains no information for the estimation of ψ .

In summary, $H(\psi)$ from model (1) is equal to $H(k, \psi)$ from model (3) plus a non-informative constant. We can then use g-estimation based on model (3) rather than on model (1) (i.e. by including $H(k, \psi)$ rather than $H(\psi)$ in the logistic model (2)) with no loss of information. Our analyses are based on model (3).

We have not yet explained the meaning of the word 'nested' but, by using model (3), we are closer to such explanation. Let us review the main parts of the analysis for each value ψ : we compute as many different counterfactual survival times $H(k, \psi)$ for each subject as the number of study visits that she attended, and we then pool all these subject-visits to fit logistic model (2). The contributions to this pooled analysis can be conceptualized in two ways: (1) each subject contributes a number of visits, or (2) each visit contributes a number of subjects. Both conceptualizations are mathematically equivalent and lead to an identical pooled analysis. We have so far used conceptualization (1) because it is an easier way to introduce nested structural AFT models for longitudinal studies with time varying-exposures. Under conceptualization (1) and assumption (A), we said that subjects in the observational study are like participants in a sequentially randomized trial. At each visit k , treatment is randomly assigned to each subject given her past history (e.g. using different treatment probabilities for each history).

Under conceptualization (2) and assumption (A), a 'randomized' trial starts at each visit k , and subjects in the observational study may simultaneously participate in several of these trials. The first trial starts at visit 0: all subjects in the study are 'randomized' to either the treatment group or the no treatment group, and then followed until their death time. $H(0, \psi)$ is computed for all of them so each subject can contribute an observation to the pooled logistic model. The second trial starts at visit 1 and includes all subjects who are still alive by that time. $H(1, \psi)$ is computed so each subject can contribute another observation to the pooled logistic model. At this time alive subjects are simultaneously participating in the first and second trials (although possibly in different treatment groups). That is, the second trial is *nested* within the first trial. Similarly, the third trial (starting at visit 2) is nested within the second trial and so on.

In each trial k , we assume in the analysis described below that treated and untreated subjects are conditionally exchangeable or comparable only at visit k . We make no assumptions about the comparability of the treated and the untreated after time k . In this sense, our approach resembles the intention-to-treat principle often applied to the analysis of randomized trials.

However, under assumption (A) and no model misspecification, our approach yields an unbiased estimate of the causal effect of received treatment, whereas a standard intention-to-treat analysis of a randomized trial yields a biased estimate of the causal effect of received treatment. This is so because in our approach $H(k, \psi)$ is computed using the treatment actually received by the subjects at k and later times; in contrast, an intention-to-treat analysis of the trial initiated at k compares the survival distributions in the group of subjects assigned to (and receiving) treatment at k and the group assigned to (and receiving) no treatment at k , ignoring all future treatment. The intention-to-treat estimate of effect for the trial initiated at k will be an unbiased estimate of the effect of received treatment only if no subject changed treatment subsequent to k , i.e. there was complete and continuing adherence to the treatment originally 'assigned' at k .

Interestingly, we can easily modify our g-estimation analysis to obtain valid intention-to-treat estimates by simply computing $H(k, \psi)$ with all $A(t)$ for $t > k$ artificially set to the value $A(k)$. It follows that in a true randomized trial (where assumption (A) holds by design for assigned treatment) in which adherence to the randomly assigned treatment regime is poor, one can correct for this non-compliance and estimate the effect that would have been observed had, contrary to fact, compliance been perfect by fitting a nested structural AFT model that uses a subject's actual observed treatment history to compute $H(k, \psi)$.¹⁷ This approach, in contrast to the usual approaches to estimate the effect of received treatment, guarantees the absence of bias under the null hypothesis of no treatment effect in any subject.

APPLICATION TO THE MACS/WIHS

We used the nested structural AFT model (3) to estimate the causal effect of HAART on time to AIDS or death in the MACS/WIHS. The 1498 eligible participants were followed for up to 6.5 years (median 5.4 years). During 6763 person-years of follow-up, 323 incident cases of AIDS and 59 deaths occurred. The g-estimate of ψ^* was $\hat{\psi} = -0.92$ (95%CI: $-1.19, -0.54$), which means that continuous HAART increases the subjects' AIDS-free survival time by 2.5-fold, i.e. $\exp(-\psi)$, compared with no HAART. We adjusted for censoring as described in the Appendix.

As explained in the previous section, $\hat{\psi}$ is a g-estimate of the causal effect of *received* treatment. We also described how one can use a modified g-estimation procedure to obtain an intention-to-treat g-estimate of

the causal effect of *assigned* treatment (see previous section and Appendix). Our analysis, however, has a built-in intention-to-treat approach even though we did not use the modified g-estimation procedure. This is so because we assumed that once a participant reported initiation of HAART, he or she remained on HAART for the duration of follow-up. We adopted this approach for consistency with our previous analysis using a marginal structural Cox model.⁷ From a practical standpoint, our approach has little impact on the estimates because the treatment status of 94% of the observed person-time was correctly classified (i.e. few individuals stopped HAART).

An alternative to ψ^* for measuring effects in survival analysis is the causal hazard ratio $\exp(\beta)$, i.e. the hazard had all subjects been continuously treated divided by the hazard had all subjects remained untreated. When $H(k, \psi)$ follows a Weibull distribution with parameters λ and ϕ and survival probabilities $\Pr[H(k, \psi)t] = \exp[-(\lambda t^\phi)]$, there is a simple relation between ψ^* and β : $\beta = \phi\psi^*$. We estimated $\hat{\phi} = 0.95$ and therefore the causal rate ratio estimate was 0.42, which is comparable to the 0.54 causal rate ratio estimated by using a marginal structural Cox model.⁷

Our analyses were conducted using a SAS macro specifically designed for g-estimation of the parameters of structural AFT models. The program can be downloaded from <http://www.hsph.harvard.edu/causal/downloads.htm>.

CONCLUSION

Nested structural AFT models and marginal structural Cox models can be used to consistently estimate the effect of a time-dependent exposure on survival in the presence of time-dependent confounders affected by prior exposure. On the other hand, standard models for survival analysis may yield biased estimates of causal effect because they adjust for time-dependent confounding by including the confounders as covariates in the model. To avoid this problem, structural models adjust for time-dependent confounding by g-estimation or inverse probability weighting.

Using a nested structural AFT model, we estimated that continuous HAART increased survival time by 2.5-fold in the MACS/WIHS. Our causal effect estimates from a structural AFT model are consistent with those from a marginal structural Cox model.⁷ It is reassuring that these two very different methods for estimating causal effects yield similar results, and that both arrive at the same qualitative conclusion as a previously conducted randomized trial.¹⁸ In contrast, a standard associational Cox model did not find a

substantially lower mortality rate among those treated compared with those untreated with HAART.⁷

The methods described in this article can also be applied to pharmacoepidemiologic data arising from claims databases. A major difference between classic cohort studies (e.g. the MACS/WIHS) and databases is that in the former data on treatments and confounders are recorded only at scheduled study visits, whereas in the latter these data are continuously recorded. In our MACS/WIHS analyses, we used two time scales: study visit k for times of treatment and confounder changes, and month t for times of outcome occurrence. When using databases, the analyst may use the same time scale (say, months) for changes in treatments and confounders as for the occurrence of the outcome.

Nested structural AFT models have been applied to estimate the effect of various time-varying exposures on time to death or other outcomes.^{19–24} Survival analysis with nested structural AFT models has two main advantages over survival analysis with marginal structural Cox models. First, nested structural AFT models can be naturally extended to estimate how the effect of treatment received at time t is modified by a time-varying covariate (such as CD4 cell count) at time t , and to compare dynamic treatment regimes. For example, consider the two-parameter nested structural AFT model

$$H(k, \psi) = \int_{u(k)}^T \exp[\psi_1 A(t) + \psi_2 A(t)B(t)] dt$$

where $B(t)$ is one when the subject's CD4 cell count has ever been under 350 cells/L before time t , and zero otherwise. The parameter ψ_2 measures how much the treatment effect is modified by the time-varying CD4 cell count. The parameters of structural nested AFT models with interaction terms between time-dependent treatment and covariates can only be estimated by g-estimation (not by inverse probability weighting).

Second, biological hypotheses are easier to translate into parameters of structural AFT models than into those of structural Cox models. This is especially true when using non-deterministic and non-rank preserving structural AFT models.¹⁴

Causal inference from observational data requires assumptions. Specifically, the g-estimates of the parameters of nested structural AFT models can only be endowed with a causal interpretation if assumption (A) holds, and both the AFT model (3) and the model for treatment (2) are correctly specified. (The assumption of no misspecification of the model for treatment can be relaxed by using doubly-robust estimators.²⁵)

KEY POINTS

Standard methods for survival analysis may yield biased estimates in the presence of time-dependent confounders affected by previous treatment.

Nested structural models and marginal structural models appropriately adjust for time-dependent confounding

In this example, estimates of the effect of HAART using structural models were similar to those from randomized trials.

On the other hand, endowing the parameters of standard models with a causal interpretation requires not only assumption (A) and no model misspecification but also the additional assumption that the time-dependent confounders are not caused (or do not share common causes) with prior treatment.⁶ This assumption is unlikely to hold in applications like ours, and thus structural models are often a less biased approach to survival analysis than standard associational models.

ACKNOWLEDGEMENTS

The Multicenter AIDS Cohort Study is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute (grants U01-AI-35042, 5-MO1-RR-00722 (General Clinical Research Center), U01-AI-35043, U01-AI-37984, U01-AI-35039, U01-AI-35040, U01-AI-37613 and U01-AI-35041). The Women's Interagency HIV Study is funded by the National Institute of Allergy and Infectious Diseases, with supplemental funding from the National Cancer Institute, the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, the National Institute of Dental and Craniofacial Research, the Agency for Health Care Policy and Research and the Centers for Disease Control and Prevention (grants U01-AI-35004, U01-AI-31834, U01-AI-34994, U01-AI-34989, U01-HD-32632 (National Institute of Child Health and Human Development), U01-AI-34993, U01-AI-42590 and M01-RR00083). (Study websites are located at <http://www.statepi.jhsph.edu>.) Dr Miguel Hernán was supported by National Institutes of Health grant K08-AI-49392, and Dr James Robins was supported by National Institutes of Health grant R01-AI-32475. Data were collected by the Multicenter AIDS Cohort Study Investigators and the Women's Interagency HIV Study Collaborative Study Group. Study

centers/groups (and Principal Investigators) are as follows: Multicenter AIDS Cohort Study—Johns Hopkins Bloomberg School of Public Health (Dr Joseph B. Margolick and Dr Alvaro Muñoz), Baltimore, MD, U.S.A.; Howard Brown Health Center and Northwestern University Medical School (Dr John Phair), Chicago, IL, U.S.A.; University of California, Los Angeles, U.S.A. (Dr Roger Detels and Dr Beth Jamieson), Los Angeles, CA, U.S.A.; and University of Pittsburgh (Dr Charles Rinaldo), Pittsburgh, PA, U.S.A.; Women's Interagency HIV Study—New York City/Bronx Consortium (Dr Kathryn Anastos); Brooklyn, NY, U.S.A. (Dr Howard Minkoff); Washington, DC, U.S.A., Metropolitan Consortium (Dr Mary Young); Connie Wofsy Study Consortium of Northern California (Dr Ruth Greenblatt and Dr Phyllis Tien); Los Angeles County/Southern California Consortium (Dr Alexandra Levine); Chicago Consortium (Dr Mardge Cohen); and Data Coordinating Center (Dr Alvaro Muñoz).

REFERENCES

- Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: Wiley, 1980.
- Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman and Hall, 1984.
- Robins JM, Tsiatis AA. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* 1992; **79**: 311–319.
- Robins JM. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika* 1992; **79**: 321–334.
- Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period application to the healthy worker survivor effect [errata appear in *Math Modelling* 1987; **14**: 917–921]. *Math Model* 1986; **7**: 1393–1512.
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004; **15**: 615–625.
- Cole SR, Hernán MA, Robins JM, Anastos K, Chmiel J, Detels R, Ervin C, Feldman J, Greenblatt R, Kingsley L, Lai S, Young M, Cohen M, Muñoz A. Marginal structural models to evaluate the effect of highly active antiretroviral therapies on time to AIDS or death. *Am J Epidemiol* 2003; **158**: 687–694.
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint effect of non-randomized treatments. *J Am Stat Assoc* 2001; **96**: 440–448.
- Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR. The Multicenter AIDS Cohort Study: rationale, organization and selected characteristics of the participants. *Am J Epidemiol* 1987; **126**: 310–318.
- Barkan SE, Melnick SL, Preston-Martin S, et al. The Womens Interagency HIV Study: WIHS Collaborative Study Group. *Epidemiology* 1998; **9**: 117–125.
- 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Morbidity and Mortality Weekly Report* 1992; **41**: 1–19.
- Dawid AP. Conditional independence in statistical theory (with discussion). *J R Stat Soc B* 1979; **41**: 1–31.
- Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002; **31**: 163–165.
- Robins JM. Structural nested failure time models. In: *Survival Analysis*, Andersen PK, Keiding N, Section Editors. *The Encyclopedia of Biostatistics*, Armitage P, Colton T (eds). Chichester, UK: John Wiley & Sons, 1997; 4372–4389.
- Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Services Research Methodology: A Focus on AIDS*, Sechrest L, Freeman H, Mulley A (eds). Washington, DC: U.S. Public Health Service, National Center for Health Services Research, 1989; 113–159.
- Robins JM. Analytic methods for estimating HIV treatment and cofactor effects. In *Methodological Issues of AIDS Mental Health Research*, Ostrow DG, Kessler R (eds). New York: Plenum Publishing, 1993.
- Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial. *Control Clin Trials* 1993; **14**: 79–97.
- 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 cells counts of 200 per cubic millimeter or less: AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* 1997; **337**: 725–733.
- Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients [errata appear in *Epidemiology* 1993; **4**: 189]. *Epidemiology* 1992; **3**: 319–336.
- Mark SD, Robins JM. Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. *Stat Med* 1993; **12**: 1605–1628.
- Robins JM, Greenland S. Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *J Am Stat Assoc* 1994; **89**: 737–749.
- Witteaman JC, d'Agostino RB, Stijnen T, et al. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Study. *Am J Epidemiol* 1998; **148**: 390–401.
- Keiding N, Filiberti M, Esbjerg S, Robins JM, Jacobsen N. The graft versus leukemia effect after bone marrow transplantation: a case study using structural nested failure time models. *Biometrics* 1999; **55**: 23–28.
- Tilling K, Sterne JA, Szklo M. Estimating the effect of cardiovascular risk factors on all-cause mortality and incidence of coronary heart disease using G-estimation: the atherosclerosis risk in communities study. *Am J Epidemiol* 2002; **155**: 710–718.
- Robins JM. Commentary on 'Using inverse weighting and predictive inference to estimate the effects of time-varying treatments on the discrete-time hazard' by Dawson and Lavori. *Stat Med* 2002; **21**: 1663–1680.
- Thompson WA, Jr. On the treatment of grouped observations in life studies. *Biometrics* 1977; **33**: 463–470.
- Robins JM, Finkelstein D. Correcting for non-compliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**: 779–788.

Appendix

G-ESTIMATION

To introduce g-estimation, we will assume that the death time of all subjects is observed (see section ‘Censoring’ for a discussion of censoring), and that the model is correctly specified for all subjects.

Under assumption (A), all counterfactuals are independent of actual treatment status at any time (conditionally on previous treatment and covariate history). Because $H(\psi^*)$ is a counterfactual outcome, it must be conditionally independent of treatment $A(k)$ for all k , given past history $\bar{A}(k-1)$, $\bar{L}(k)$. That is, the conditional probability of receiving treatment is the same whether we have or do not have information on $H(\psi^*)$, i.e. $\Pr[A(k) = 1 | \bar{L}(k), \bar{A}(k-1), T > u(k)] = \Pr[A(k) = 1 | \bar{L}(k), \bar{A}(k-1), T > u(k), H(\psi^*)]$. In other words, if we regress treatment $A(k)$ on past history and $H(\psi^*)$, then the parameter for $H(\psi^*)$ would be zero. For example, consider the pooled logistic model (2) with $H(\psi)$ replaced by $H(\psi^*)$. If assumption (A) is correct, then θ_3 is equal to zero.

We still do not know the value of $H(\psi^*)$ for all subjects, but we now have a method to rule out potential candidates. For example, suppose someone suggests that the true value ψ^* is 0.5. We can use model (1) to compute $H(0.5)$ for all subjects. Once we have this new variable in our data set, we can replace $H(\psi^*)$ by $H(0.5)$ in model (2). Suppose that the parameter estimate $\hat{\theta}_3 = 1.3$ (p value of test for $\theta_3 = 0$: 0.001). We then conclude that most likely $H(0.5)$ is not a counterfactual outcome and that ψ^* is not 0.5. Let us try another candidate value for ψ^* , say $\psi = -0.7$. Again, we use model (1) to compute $H(-0.7)$ for all subjects, and then replace $H(\psi^*)$ by $H(-0.7)$ in model (2). Suppose that the parameter estimate $\hat{\theta}_3 = 0.001$ (p value of test for $\theta_3 = 0$: 0.91). We cannot reject the null hypothesis (using the conventional $p < 0.05$) that $\theta_3 = 0$ and therefore $H(-0.7)$ may be the counterfactual outcome under no treatment. This in turn implies that ψ^* may be equal to -0.7 . We are now ready to try another value of ψ , until we find all values of ψ that may be equal to ψ^* . This is g-estimation.^{14–16}

Ideally, the search for values of ψ should continue until all possible values have been tested. The point estimate of ψ^* , $\hat{\psi}$, is the value of ψ that makes $\hat{\theta}_3 = 0$. The 95% CI value for ψ^* includes all the values of ψ that produce values of θ_3 with p values over 0.05 (that is, the confidence interval of ψ^* is obtained by inverting the test for $\theta_3 = 0$). In practice not all possible values can be tested because there is an infinite number of values of ψ

in any given interval. One option is to conduct a very fine grid search over many prespecified values of ψ (e.g. from -3 to 3 by increments of 0.01). The finer the search, the more precise the estimates. Section ‘Estimation Details’ describes less computationally intensive approaches to find the point estimate and the 95% confidence limits, and how to modify the procedure to account for censoring.

CENSORING

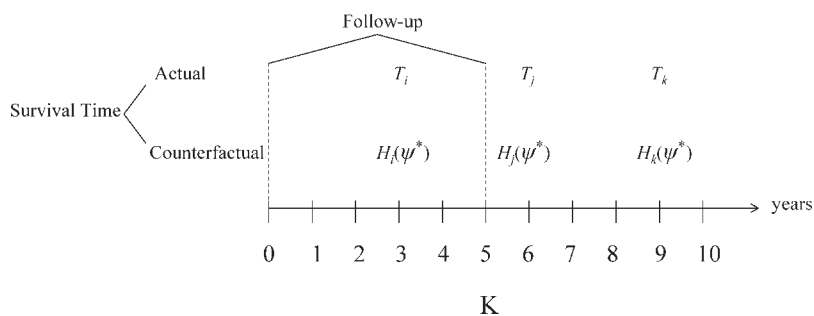
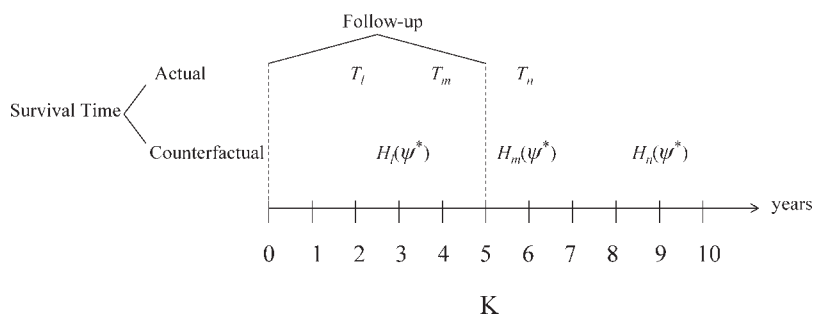
G-estimation requires computing $H(\psi)$ using model (1) or, equivalently, $H(k, \psi)$ using model (3). To compute $H(\psi)$, we need two pieces of information: the time of death T and the treatment history $\bar{A}(T)$ up to that time. However, this information is unknown for individuals with (right-)censored death times, i.e. those whose death time T is not observed. A subject’s death time may be censored because either (i) he remained alive at the planned end of the study (857 subjects in our study), or (ii) because he was lost to follow-up (259 subjects in our study), or died from a different cause than the one under study (when studying outcomes other than total mortality). We refer to censoring (i) as administrative censoring. For now, let us assume that all censoring is administrative (i.e. no subject drops out of the study or is lost to follow-up).

One naive strategy to deal with administrative censoring is ignoring all subjects who did not die before the administrative end of follow-up and restricting the analysis to uncensored subjects. Below we explain why this strategy is biased and describe a non-biased approach to deal with administrative censoring. For simplicity, we first consider a non-time-varying treatment.

A biased approach

Consider an oversimplified study with three untreated subjects (i, j and k) and three treated subjects (l, m and n). All subjects are followed until their death time or the administrative end of the study, whichever comes first. The administrative censoring time K is equal to 5 years for all subjects.

The actual survival times of the untreated subjects i, j and k are 3, 6 and 9 years respectively as shown in Figure 2. The investigators only observe the survival time of subject i because subjects j and k die after the end of follow-up, i.e. their survival times are administratively censored. Because subjects i, j and k are untreated, their actual survival times T are, by definition, equal to their counterfactual survival times under no treatment $H(\psi^*)$.

Figure 2. Untreated subjects, $\psi^* > 0$ Figure 3. Treated subjects, $\psi^* > 0$

The actual survival times of the treated subjects l , m and n are 2, 4 and 6 years, respectively, as shown in Figure 3. The survival time of subject n is administratively censored. The counterfactual survival times under no treatment of the subjects l , m and n are 3, 6 and 9 years, respectively. These counterfactual times under no treatment $H(\psi^*)$ are unknown to the investigators. In this study, treatment is harmful because the survival times when treated are shorter than those when untreated.

Note that the three untreated subjects are comparable or exchangeable with the three treated subjects because, had the treated subjects been untreated, they would have had the same survival time distribution as the untreated subjects. In other words, there is no confounding. Assumption (A) holds because data on $H(\psi^*)$, were it available, would not improve the prediction of whether a subject is treated or untreated. For example, knowing that a subject's $H(\psi^*)$ is greater than five does not make it more likely that he is treated or untreated because two thirds of the treated and two thirds of the untreated have $H(\psi^*) > 5$.

Now suppose that the analysis includes only subjects with known death times T , i.e. treated subject i and untreated subjects l and m . In this selected group, exchangeability is lost because, had the treated subjects l and m been untreated, they would have had

a greater average survival time than the untreated subject i . In other words, the restriction to subjects with known death times creates selection bias because the analysis is conditional on having an observed survival time less than 5 years, and survival time is a variable affected by treatment. Assumption (A) does not hold because $H(\psi^*)$ helps predict a subject's treatment status. For example, a subject with $H(\psi^*)$ greater than 5 is more likely to be treated because half of the treated and none of the untreated have $H(\psi^*) > 5$.

In summary, investigators cannot include the six subjects in the analysis because the actual survival time T of three of them (j , k and n) was not observed, and they cannot restrict the analysis to the three subjects with observed survival time T because that would create selection bias. This problem is exacerbated in complex longitudinal data with time-varying treatments and high-dimensional covariates.

An unbiased approach

The solution to the above problem is restricting the analysis to subjects whose survival time would have been observed whether they had been treated or untreated, i.e. subjects i and l . This strategy preserves exchangeability because the distribution of $H(\psi^*)$ in

the treated equals that of the untreated. Under this approach, assumption (A) holds because $H(\psi^*)$ does not predict a subject's treatment status. Note that, to keep exchangeability, the data of some subjects with known survival time T need to be thrown away. This approach is unbiased regardless of the value of ψ^* , i.e. whether treatment is harmful or beneficial.

In more complex situations with time-varying treatments, this approach needs to be generalized because subjects cannot be classified into 'treated' and 'untreated' (they can be treated at some time points and untreated at others). The generalized unbiased approach is then restricting the analysis to subjects whose survival time would have been observed under any possible treatment regime. Under model (1), this approach boils down to identifying the subjects whose death time would have been observed under the regimes 'always treated' and 'never treated', and including them in the analysis.

We now define this approach more formally. Let $\Delta(\psi^*)$ be a dichotomous variable that takes the value 1 if the subject is included in the analysis when using the unbiased approach of the previous paragraph, and 0 otherwise. Let $K(\psi^*)$ be the minimum survival time under no treatment that could possibly correspond to a subject who actually died at time K (the administrative end of follow-up), i.e. $K(\psi^*) = \inf\{\int_0^K \exp[\psi a(t)] dt\}$ over all possible treatment histories $\bar{a}(t)$. Then $\Delta(\psi^*) = 1$ if $H(\psi^*) < K(\psi^*)$ and 0 otherwise. For a dichotomous treatment, $K(\psi^*) = \int_0^K \exp[\psi \times 0] dt = K$ if $\psi^* > 0$ (i.e. treatment is harmful), $K(\psi^*) = \int_0^K \exp[\psi \times 1] dt = K \exp[\psi]$ if $\psi^* < 0$ (i.e. treatment is beneficial) and $K(\psi^*) = \int_0^K \exp[0] dt = K$ if $\psi^* = 0$ (i.e. treatment has no effect).

All subjects who are censored by the administrative end of follow-up (i.e. $T \geq K$) have $\Delta(\psi^*) = 0$ because there is at least one treatment regime (the one they actually received) under which their survival time is greater than the length of follow-up, i.e. $H(\psi^*) \geq K(\psi^*)$. Some of the subjects who are not censored by administrative end of follow-up (i.e. $T < K$) also have $\Delta(\psi^*) = 0$ and are excluded from the analysis to avoid selection bias. The exclusion of uncensored subjects is sometimes referred to as artificial censoring.

A key point is that, for a given ψ^* , the variable $H(\psi^*)$ cannot be computed for all subjects whereas the variable $\Delta(\psi^*)$ can be computed for all subjects, regardless of whether their actual survival time T is observed. Another key point is that $\Delta(\psi^*)$ is a function of the counterfactual $H(\psi^*)$ (and of K) but not a function of current treatment $A(t)$. Under assumption (A), all such functions of a counterfactual survival time are conditionally indepen-

dent of current treatment given past history and K . Therefore, assumption (A) implies

$$\begin{aligned} \Delta(\psi^*) \prod A(k) \bar{A}(k-1) &= \bar{a}(k-1), \bar{L}(k) = \bar{l}(k), \\ T > u(k), K &\text{ for all } \bar{a}(k-1) \text{ and } \bar{l}(k) \end{aligned} \quad (\text{B})$$

Note that the administrative censoring time K may differ across subjects (when there is staggered entry into the study, for example) but, for each subject, the value of K is known at the start of the follow-up, whether or not the subject was actually followed for all that time. For example, all subjects in our study whose follow-up started in July 1996 had the same potential follow-up time K , even though those who either died or dropped out of the study before 2002 had a shorter actual follow-up time.

Condition (B) implies that $\theta_3 = 0$ in the logistic model

$$\begin{aligned} \text{logit Pr}[A(k) = 1 | \bar{A}(k-1), \bar{L}(k), T > u(k), K, \Delta(\psi^*)] \\ = \theta_0(k) + \theta_1 A(k-1) + \theta_2 L(k) + \theta_3 \Delta(\psi^*) + \theta_4 K \end{aligned} \quad (4)$$

We can now use model (4) to estimate ψ^* and its 95%CI by g-estimation, even in the presence of administrative censoring.

Any function of $(H(\psi^*), K)$ may be substituted for $\Delta(\psi^*)$ in the right-hand side of model (4). The choice of the function of $(H(\psi^*), K)$ does not affect the consistency of the point estimate for ψ^* , but it determines the width of its confidence interval. The most efficient function of $H(\psi^*)$ (i.e. the one that produces the narrowest confidence interval) is not $\Delta(\psi^*)$ but a rather complex one.¹⁶ We elected to use the function $\Delta(\psi^*)$ because it combines good efficiency with computational ease.

When the analysis is based upon $H(k, \psi)$ rather than $H(\psi) \equiv H(0, \psi)$, we replace $\Delta(\psi^*)$ above by $\Delta(k, \psi^*)$, where $\Delta(k, \psi^*) = 1$ if $H(k, \psi^*) < K(k, \psi^*)$ and 0 otherwise, and $K(k, \psi^*) = \inf\{\int_{u(k)}^K \exp[\psi a(t)] dt\}$. We used this method to account for administrative censoring in our study.

We sometimes may want to use a censoring time shorter than the administrative censoring time K . For example suppose that the goal of the analysis is to estimate the intention-to-treat effect of treatment. If the AFT model for received treatment is correct and the correlation of treatment received at k and that received at $k + \delta$ decreases with increasing δ , then one is essentially modeling a series of nested trials with increasing effective compliance for increasing k , and

one would expect the magnitude of the intention-to-treat effect to increase with time. Therefore the parameter of the AFT model for the intention-to-treat effect will be hard to interpret. A solution to this problem is the following. At each visit k , restrict the analysis to subjects who are either treatment initiators or non-users of treatment at that time. Then choose a fixed length of follow-up (say, $h = 36$ months) for every nested trial, and censor participants in every nested trial at time $u(k) + h$. Then define $\Delta(k, \psi^*)$ as above except that K is replaced by $u(k) + h$. The analysis now mimics a series of randomized trials for estimating the intention-to-treat effect of treatment initiation with a fixed length of follow-up.

Non-administrative censoring

In addition to administrative censoring, there may be censoring due to loss to follow-up or competing risks. This non-administrative censoring may be informative and hence a source of selection bias.⁶ To adjust for selection bias due to non-administrative censoring, we use inverse probability weighting (IPW). The idea behind IPW is to assign a weight to each selected subject so that she accounts in the analysis not only for herself but also for those with similar characteristics that were censored. The weight is the inverse of the probability of being uncensored. For example, if there are four untreated women, aged 40–45, with CD4 count >500 in our study, and three of them were lost to follow-up, then these three women do not contribute to the analysis (i.e. they receive a zero weight) while the remaining woman receives a weight of four. In other words, the (estimated) conditional probability of remaining uncensored until the end of the study is $1/4 = 0.25$, and therefore the (estimated) weight for the uncensored subject is $1/0.25 = 4$. IPW creates a hypothetical population where the four subjects of the original population are replaced by four copies of the uncensored subject.¹⁴

We now describe how to use IPW to adjust for non-administrative censoring in model (3). Let us define a dichotomous variable $C(k)$ that takes value 1 when a subject was (non-administratively) censored by visit k and 0 otherwise. At each visit k , each uncensored subject receives the weight

$$W(k) = \left\{ \prod_{v=k+1}^V \Pr[C(v) = 0 | \bar{L}(v-1), \bar{A}(v-1), C(v-1) = 0, T > u(v)] \right\}^{-1}$$

where V is the subject's last visit before death time or administrative end of follow-up, whichever comes first, and $\text{logit } \Pr[C(k) = 0 | \bar{L}(k-1), \bar{A}(k-1), C(k-1) = 0, T > u(k)]$ is the subject's probability of remaining uncensored at k given that she was uncensored at $k-1$ and given past treatment and covariate history. The weight $W(k)$ equals the inverse of the probability that the subject remained uncensored until his death time or the administrative end of follow-up, whichever came first. The probability of being uncensored at each visit k can be estimated by a pooled logistic model such as

$$\text{logit } \Pr[C(k) = 0 | \bar{L}(k-1), \bar{A}(k-1), C(k-1) = 0, T > u(k)] = \gamma_0(k) + \gamma_1 A(k-1) + \gamma_2^T L(k-1) \tag{5}$$

where $\gamma_0(k)$ is a time-varying intercept (e.g. a cubic spline for time of follow-up), and past treatment and confounder history are summarized by their baseline value and their last value prior to $C(k)$. Note that the probability of remaining uncensored at visit k is estimated only among subjects that are uncensored at visit $k-1$ and alive at visit k , i.e. when the probability of censoring is small at all visits k , the parameters of this pooled logistic model are approximately equal to those of a Cox model in which the outcome is time to censoring.²⁶ Subjects who are non-administratively censored during the follow-up receive weights $W(k) = 0$ at all k .

G-estimation is then applied to the pseudo-population created by weighting. In each nested trial k , the pseudo-population consists of the eligible subjects that remained (non-administratively) uncensored, weighted by their estimates of $W(k)$. The causal effect of treatment A in the pseudo-population is equal to the effect of treatment A in the original population had nobody been (non-administratively) censored, but only under the assumption that censoring occurs at random given past treatment and covariate history,²⁷ that is,

$$\begin{aligned} T_{\bar{a}} \prod C(k) | C(k-1) = 0, \bar{L}(k-1) = \bar{l}(k-1), \\ \bar{A}(k-1) = \bar{a}(k-1), \\ T > u(k) \text{ for all } \bar{l}(k-1) \text{ and } \bar{a}(k-1) \end{aligned} \tag{C}$$

Unfortunately, assumption (C) cannot be tested, and thus the adjustment for selection bias by IPW depends on this untestable assumption. When condition (C) holds, censoring is ignorable given past history. We used IPW with weights $W(t)$ to adjust for potential

selection bias due to non-administrative censoring in our study. More complex variations of the weights $W(t)$ can be used to obtain a narrower confidence interval for ψ^* . Again, we elected to use the weights $W(t)$ because they combine good efficiency with computational ease.

Unlike IPW for non-administrative censoring, the method that we described to adjust for administrative censoring was assumption-free. However, it is often reasonable to assume that administrative censoring is also ignorable given past history. In principle, the additional information provided by this assumption would lead to greater efficiency (although this is true only when using fully efficient estimators). To incorporate this assumption, we redefine $C(k)$ to be 1 if the subject was either administratively or non-administratively censored at visit k (0 otherwise) and replace the administrative censoring time K by a constant time c^* which is slightly less than the longest follow-up time of any of the n subjects in the study, i.e. $c^* < \max(K_i = 1, \dots, n)$ so that the probability of being uncensored at time c^* is not zero.

ESTIMATION DETAILS

The g-estimate of ψ^* is the value of ψ that yields $\hat{\theta}_3 = 0$ in model (4) or, equivalently, the value of ψ that minimizes the score test statistic (χ^2 with one degree of freedom) for $\theta_3 = 0$. Minimizing the score test statistic for θ_3 is equivalent to finding the solution to the estimating equation $U(\psi) = 0$ where

$$U(\psi) = \sum_{i=1}^N \sum_{k=0}^V W_i(k) \Delta_i(k, \psi) \{A_i(k) - E[A(k) | \bar{L}(t-1), \bar{A}(t-1), T > u(k), K]\}$$

The value of ψ that solves the estimating equation $U(\psi) = 0$ is a consistent g-estimate $\hat{\psi}$ of ψ^* . The two conceptualizations of the nested model described in the main text (i.e. subject-based or trial-based) can be derived by simply reversing the order of the two sums of the estimating function $U(\psi)$. Note that $E[A(k) | \bar{L}(t), \bar{A}(t-1), T > u(k), K]$ is equal to $\Pr[A(k) = 1 | \bar{L}(k), \bar{A}(k-1), T > u(k), K]$ when $A(k)$ is dichotomous.

The estimating function $U(\psi)$ cannot be computed for any value of ψ when, as it is usually the case, $E[A(k) | \bar{L}(t), \bar{A}(t-1), T > u(k), K]$ and $W_i(k)$ are unknown. However we can replace $E[A(k) | \bar{L}(t), \bar{A}(t-1), T > u(k), K]$ in $U(\psi)$ by the estimates of $\Pr[A(k) = 1 | \bar{L}(k), \bar{A}(k-1), T > u(k), K]$ from model (4) with no covariate $\Delta_i(k, \psi)$. We fit this

model restricted to person-visits with $A(k-1) = 0$, because $\Pr[A(k) = 1 | \bar{L}(k), \bar{A}(k-2), A(k-1) = 1, T > u(k), K] = 1$ under our assumption that subjects never discontinue treatment. We can replace $W_i(k)$ in $U(\psi)$ by the inverse of the product of the estimates of the probabilities $\Pr[C(t) = 0 | \bar{L}(t-1), \bar{A}(t-1), C(t-1) = 0, T > u(k)]$ from model (5). The new estimating function $\hat{U}(\psi)$ that contains the estimates of $E[A(k) | \bar{L}(t), \bar{A}(t-1), T > u(k), K]$ and $W_i(k)$ can also be used to consistently estimate ψ^* . Our g-estimate $\hat{\psi}$ of ψ^* was the value that solved the estimating equation $\hat{U}(\psi) = 0$.

Standard optimization methods (e.g. Newton–Raphson) to solve estimating equations are based on either the first or second derivatives of a quadratic form of the estimating function. These standard methods are not very useful to find the value $\hat{\psi}$ because the estimating function $\hat{U}(\psi)$ is not monotone or differentiable on ψ . Therefore, we need to find the value $\hat{\psi}$ by using either a grid search or non-gradient-based optimizers. Examples of these optimizers are direct search methods (e.g. Nelder–Mead simplex), simulated annealing, and genetic algorithms.

The 95% confidence limits of $\hat{\psi}$ can be obtained by computing the score test statistic $\hat{U}(\psi)^T [\Sigma(\psi)]^{-1} \hat{U}(\psi)$ where $\Sigma(\psi)$ is the covariance matrix of $\hat{U}(\psi)$, and finding the values of ψ that make the test statistic equal to 3.84 (χ^2 value for one degree of freedom). To correctly estimate the covariance matrix $\Sigma(\psi)$, we need to take into account that the contributions to the estimating equation are no longer uncorrelated when weighting is used to adjust for censoring. This is so because of two distinct reasons: the weight $W(k)$ depends on each subject's treatment and covariate history beyond k , and the weight $W(k)$ is replaced by an estimate that depends on all the data.¹⁴

In our study, we replaced $\Sigma(\psi)$ by the empirical covariance matrix of the subject-specific contributions. This approach ignores how the censoring weights $W(k)$ were estimated and therefore yields a conservative 95%CI for ψ^* (i.e. it includes ψ^* more than 95% of the time). The confidence interval can also be estimated by the bootstrap method. In our study, the bootstrap 95%CI based on 200 samples was -1.20 to -0.64 (mean bootstrap estimate -0.91) for our main result.

G-estimation based on the estimating function $\hat{U}(\psi)$ is inefficient but easy to carry out. On the other hand, the efficient g-estimator involves functions of the data that are hard to compute.¹⁶ A counterintuitive consequence of using an inefficient estimator is that throwing away a substantial part of the data may have little impact on, or even decrease, the width of the

confidence interval for ψ^* . We now illustrate this property of the inefficient g-estimator with an example from our study. Most subjects included in our analyses were followed from 1996 and thus could have potentially been followed for about 6 years (because the end of the follow-up for our analyses was 2002). A few subjects became eligible to participate in our study after 1996 and therefore their potential follow-up time was less than 6 years. The median potential follow-up time was 6.0 years (minimum: 2.0, maximum: 6.6). Suppose that we disregard any data on our subjects after 5 years of follow-up, and then deal with censoring as described in the previous section. (That is, we mimic a study with a fixed length of follow-up of 5 years by replacing K_i by 5 in those subjects for whom K_i exceeded 5.) Because this forced censoring strategy implies throwing away part of the data, we would generally expect the confidence interval from the forced censoring analysis to be wider than that of the original analysis. Figure 4 shows the g-estimates and their 95%CI when the potential follow-up time K was set at a maximum of 2, 3, 4, 5 and 6 years of follow-up in our study. (In the absence of model misspecification, all these g-estimates of $\hat{\psi}$ converge to ψ^* regardless of the length of follow-up.) The width of the confidence interval is similar for any value of $K \geq 4$.

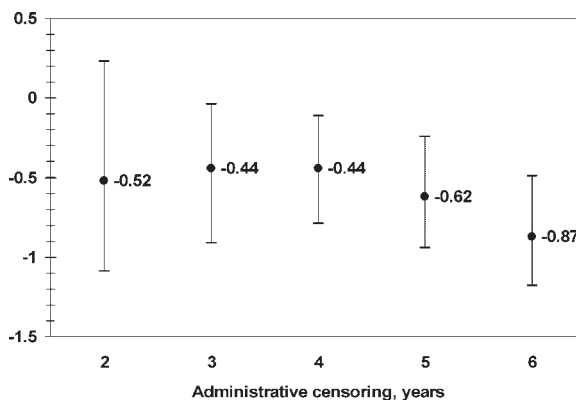


Figure 4. G-estimates for several maximum administrative censoring times

Similar counterintuitive findings may appear when administrative censoring is assumed to be ignorable and adjusted for by IPW as described above. That is, the variance of a (non fully efficient) IPW estimator that assumes ignorable administrative censoring can be less than the variance of the forced censoring estimator, because the efficiency loss attributable to the large size of some of the inverse probability weights may more than offset the efficiency gain due to a larger number of events.