Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments

Miguel A Hernán, Babette Brumback & James M Robins

To cite this article: Miguel A Hernán, Babette Brumback & James M Robins (2001) Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments, Journal of the American Statistical Association, 96:454, 440-448, DOI: 10.1198/016214501753168154

To link to this article: http://dx.doi.org/10.1198/016214501753168154

Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments

Miguel A. HERNÁN, Babette BRUMBACK, and James M. ROBINS

Even in the absence of unmeasured confounding factors or model misspecification, standard methods for estimating the causal effect of time-varying treatments on survival are biased when (a) there exists a time-dependent risk factor for survival that also predicts subsequent treatment, and (b) past treatment history predicts subsequent risk factor level. In contrast, methods based on marginal structural models (MSMs) can provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent. MSMs are a new class of causal models whose parameters are estimated using a new class of estimators—inverse-probability-of-treatment weighted estimators. We use a marginal structural Cox proportional hazards model to estimate the joint effect of zidovudine (AZT) and prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of HIV-positive men in the Multicenter AIDS Cohort Study, an observational study of homosexual men. We obtained an estimated causal mortality rate (hazard) ratio of .67 (conservative 95% confidence interval .46–.98) for AZT and of 1.14 (.79, 1.64) for prophylaxis therapy. These estimates will be consistent for the true causal rate ratios when the functional forms chosen for our models are correct and data have been obtained on all time-independent and time-dependent covariates that predict both subsequent treatment and mortality.

KEY WORDS: Causal inference; Confounding; Counterfactual variables; Dependent censoring; Intermediate variables; Semiparametric models, Survival analysis.

1. INTRODUCTION

This article describes the application of marginal structural models (MSMs), a new class of causal models (Robins 1999), to estimate the joint effect of time-dependent nonrandomized treatments, zidovudine (AZT) therapy and prophylaxis therapy, for *Pneumocystis carinii* pneumonia (PCP) on survival among HIV-positive subjects participating in the Multicenter AIDS Cohort Study (MACS), an observational study of homosexual men. The parameters of a MSM can be consistently estimated using a new class of estimators—the inverse-probability-of-treatment weighted estimators. The use of MSMs is an alternative to the semiparametric g-computation algorithm estimator (Robins 1986) and to g-estimation of structural nested models (SNMs) (Robins 1998a).

It is well understood that causal effects can generally be estimated from observational studies only when data on all relevant time-independent and time-dependent confounding factors have been obtained. What is less well known is that standard approaches to confounder control can be biased, even when the causal null hypothesis of no treatment effect is true and there are no unmeasured confounding factors. Specifically, the standard approach to the estimation of the causal effect of a time-varying treatment on survival has been to model the hazard of failure at t as a function of treatment history with a time-dependent proportional hazards model. Robins and colleagues have shown that even in the absence of unmeasured confounding factors or model misspecification, the usual approach may be biased even under the causal null hypothesis, whether or not one adjusts further for the past history of measured covariates in the analysis, when (a) there exists a time-dependent risk factor (say CD4 count and/or PCP history) for survival that also predicts subsequent treatment, and (b) past treatment history predicts subsequent risk factor level (Robins 1986, 1998a; Robins and Greenland 1994). Specifically condition (a) implies that the analysis that does not adjust for covariates is biased because of confounding by CD4 count and/or PCP. Condition (b) implies that the analysis that includes current CD4 count and/or PCP history as a regressor is biased because it adjusts for a variable (CD4 count and/or PCP history) affected by past treatment (see Robins, Greenland, and Hu 1999 for additional details). We show that both conditions (a) and (b) are true in the MACS data. In contrast to standard methods, estimation methods based on MSMs provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent.

2. THE MULTICENTER AIDS COHORT STUDY

The MACS is an ongoing cohort study of more than 5,000 homosexual men from Baltimore, Chicago, Los Angeles, Pittsburgh, and Washington, DC. Study enrollment took place between 1984 and 1991. Follow-up visits are scheduled for every 6 months. During each clinic visit, a structured interview (including questions on demographic variables, therapeutic drugs, and AIDS-related symptoms) is administered, and a physical examination is performed. Blood is collected for a complete blood count, T-cell phenotyping, and assays for HIV-1 antibody. The design and methods of this study have been described previously (Graham et al. 1992). The MACS dataset is available through the National Technical Information Service.

Our analysis concerns two therapies commonly used by HIV-infected patients in the MACS and elsewhere: AZT and prophylaxis for PCP. (In the MACS, aerosolized pentamidine, trimethoprim-sulfamethoxazole, and dapsone were all used as prophylaxis therapy.) AZT temporarily prevents the decline of CD4 lymphocyte count, slows the progression of HIV/AIDS, and, in clinical trials has increased the survival of HIV-infected individuals. PCP is an opportunistic infection that afflicts AIDS patients. Patients may suffer repeated bouts

Miguel Hernán is Instructor, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115 (E-mail: *mhernan@hsph.harvard.edu*). Babette Brumback is Assistant Professor, Department of Biostatistics, University of Washington, Seattle, WA 98195. James Robins is Professor of Epidemiology and Biostatistics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115. This research was supported by NIH grant R01 A-132475.

of PCP, which can be fatal. Prophylaxis therapy might prolong survival by preventing further episodes of PCP.

We restricted the cohort to HIV-positive men alive in the period during which AZT was available for use; that is, after study visit 5 (March 1986–March 1987). Follow-up began at the later of visit 5 and first visit at which the subject was known to be infected. Follow-up ended at study visit 21 October 1994, death, or 24 months after the last visit, whichever came first. Our analysis includes the 2168 HIV-positive men who had at least one visit between visits 5 and 21 and who did not have AIDS and were not on AZT or prophylaxis for PCP on the visit at which follow-up commenced. By the end of the follow-up (median duration 67 months), 1,286 men had initiated AZT treatment, 912 had initiated prophylaxis therapy for PCP, and 738 had died.

3. NOTATION

Time is denoted by t and is measured in months since the beginning of a subject's follow-up. Here we use capital letters to represent random variables and lower-case letters to represent possible realizations (values) of random variables. $A_{1i}(t)$ and $A_{2i}(t)$ are time-varying dichotomous variables indicating whether patient i is on AZT or prophylaxis at time t. Then $A_i(t) = [A_{1i}(t), A_{2i}(t)]$ is the treatment vector at t. $L_i(t)$ is a vector of relevant prognostic factors for (i.e., predictors of) survival, such as CD4 count and number of PCP bouts. For the purposes of this analysis, we assumed that if a treatment or covariate process jumped in month t, then it did so on the first day of the month. We also assumed that once an individual started treatment with AZT, he remained on it thereafter, and similarly for prophylaxis, and that $L(t)$ is temporally earlier than $A(t)$ because physicians commonly obtained data recorded in $L(t)$, such as CD4 count, before deciding on a treatment $A(t)$ to be given in month t. For any time-dependent variable, we use overbars to denote the history of that variable up to and including t; for example, $\overline{L}_i(t) = [L_i(0), L_i(1), L_i(2), \ldots, L_i(t)]$ is the covariate process through t. 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. We use the symbol $\coprod$ to indicate statistical independence; for example, $A \coprod B \mid C$ means that A is conditionally independent of B given C.

4. MARGINAL STRUCTURAL MODELS

Before describing the marginal structural proportional hazard models for a survival outcome, we describe the simpler marginal structural logistic model for a dichotomous outcome measured at the end of follow-up.

4.1 Inverse-Probability-of-Treatment Weighted Estimator

Let Y be a dichotomous outcome (e.g., $Y = 1$ if the Karnofsky score is greater than 70, which indicates an acceptable functional status, and 0 otherwise) measured at the end of follow-up at time $K+1$. Our goal is to estimate the causal effect of the time-dependent treatment $A(t)$ on the mean of Y. In this section we assume that there is no loss to follow-up, so Y is observed on each study subject. To motivate the need for MSMs, we first study the association (i.e., regression) model that states that the mean of Y, given treatment history $\overline{A} \equiv \overline{A}(K)$, is a linear logistic function of a subject's duration of AZT and prophylaxis therapies. That is,

$$E[Y \mid \overline{A}] = g(\overline{A}; \gamma),$$

where

$$g(\overline{A}; \gamma) = \frac{\exp[\gamma_0 + \gamma_1 dur(\overline{A}_1) + \gamma_2 dur(\overline{A}_2)]}{1 + \exp[\gamma_0 + \gamma_1 dur(\overline{A}_1) + \gamma_2 dur(\overline{A}_2)]} \quad (1)$$

Here $dur(\overline{A}_j) = \sum_{k=0}^{K} A_j(k)$ is the subject's duration of AZT ($j = 1$) or prophylaxis ($j = 2$) treatment in months, and $A_j(k)$ equals 1 if the subject is on treatment j in month k and 0 otherwise. The maximum likelihood estimator of $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ maximizes $\prod_{i=1}^{n} lik_i(\gamma)$, with $lik(\gamma) = g(\overline{A}; \gamma)^Y [1 - g(\overline{A}; \gamma)]^{1-Y}$ being the likelihood contribution for a single subject.

Assuming that the association model (1) is correct, when do γ_1 and γ_2 have a causal interpretation? Suppose that treatment at each time t was assigned completely at random and noncompliance was absent. Then, treatment at t will be independent of the history up to t of both measured and unmeasured prognostic factors (i.e., there is no confounding). In the absence of confounding, association implies causation, and we would expect γ_1 to represent the effect of AZT treatment and γ_2 to represent the effect of prophylaxis therapy on the mean of Y. More generally, we define a treatment process to be "causally exogenous or ancillary" if the conditional probability of receiving a treatment $A(t)$ at month t given past treatment and prognostic factor history (measured and unmeasured) depends only on past history of treatment $\overline{A}(t-1)$. It is well recognized in the social sciences, econometrics, epidemiologic, and biostatistical literature that the treatment parameters of a correctly specified association model will have a causal interpretation if treatment is causally exogenous. Our defintion of causal exogeneity is essentially equivalent to that of Leamer (1985), but differs somewhat from that given by Engle, Hendry, and Richard (1983) and Manski (1995). It is also closely related to Kalbfleisch and Prentice's (1980) concept of an "external covariate process of the ancillary type."

To help assess whether AZT and prophylaxis therapy may be causally exogenous, we introduce the concept of "statistical exogeneity." We say that treatment $A(t)$ is a "statistically exogenous or ancillary" process if the probability of receiving treatment at time t does not depend on the history of measured time-dependent prognostic factors up to t conditional on treatment history before t; that is,

$$\overline{L}(t) \coprod A(t) \mid \overline{A}(t-1).$$

Note that a nearly necessary condition for $A(t)$ to be "causally exogenous" is for it to be "statistically exogenous." But the fact that a process is "statistically exogenous" does not imply that it is "causally exogenous," because unmeasured prognostic factors for the outcome (i.e., confounders) may predict the probability of treatment $A(t)$ at time t given past treatment history. We can empirically test whether $A(t)$ is statistically exogenous, but not whether it is causally exogenous.

Suppose that we can correctly model the probability of receiving prophylaxis therapy in each month t as a function of past treatment and measured prognostic factor history. We could then quantify the degree to which the treatment process is statistically nonexogenous through month t by the random quantity

$$SW(t) = \prod_{k=0}^{t} \frac{f[A(k)|\overline{A}(k-1)]}{f[A(k)|\overline{A}(k-1), \overline{L}(k)]}, \tag{2}$$

where $f[A(k)|\overline{A}(k-1), \overline{L}(k)]$ is by definition the conditional probability mass function $f_{A(k)|\overline{A}(k-1),\overline{L}(k)}[a(k) \mid \bar{a}(k-1), \bar{\ell}(k)]$, with $(a(k), \bar{a}(k-1), \bar{\ell}(k))$ evaluated at the random argument $(A(k), \overline{A}(k-1), \overline{L}(k))$. Informally, the denominator in each term in $SW(t)$ is the probability that a subject received his own observed treatment, $A(k)$, at time k given his past AZT, prophylaxis, and prognostic factor history. Informally, the numerator is the probability that a subject received his observed treatment conditional only on his past AZT and prophylaxis history. Note that the numerator and denominator of $SW(t)$ are equal for all t with probability 1 if and only if the treatment process is statistically exogenous; that is, $\overline{L}(t) \coprod A(t) \mid \overline{A}(t-1)$. In practice $SW(t)$ will have to be estimated from the data, but for pedagogical purposes, assume for now that it is known.

When $A(t)$ is statistically nonexogenous, we consider estimating γ by a weighted logistic regression in which a subject is given the weight $SW \equiv SW(K)$. Standard software packages for logistic regression will allow the user to specify the subject-specific weight SW (e.g., option "weight" in SAS proc genmod). The weighted logistic regression estimator, which we call an inverse-probability-of-treatment weighted (IPTW) estimator, is the maximizer of $\prod_{i=1}^{n}[lik_i(\gamma)]^{SW_i}$. This weighted logistic regression would agree with the usual unweighted analysis described earlier only in the case in which $A(t)$ were statistically exogenous. The IPTW estimator is an extension to longitudinal causal inference models of estimators proposed by Horvitz and Thompson (1952), Kalbfleisch and Lawless (1988), Flanders and Greenland (1991), Rosenbaum (1987), and Robins and Rotnitzky (1992) for missing-data models.

The somewhat surprising result is that if the vector of prognostic factors recorded in $\overline{L}(t)$ constitutes all relevant time-dependent prognostic factors (i.e., confounders), then, whether or not the treatment process is statistically exogenous, the weighted logistic regression estimators of γ_1 and γ_2 will converge to quantities β_1 and β_2 that can be interpreted as the causal effects of AZT and prophylaxis therapy on the mean of Y (on the log odds ratio scale). In contrast, when $A(t)$ is statistically nonexogenous, the usual unweighted logistic regression estimators will still converge to the association parameters γ_1 and γ_2 , but now γ_1 and γ_2 will have no causal interpretation. We now give a formal mathematical meaning to the informal concepts of the causal effects of AZT and prophylaxis history on the mean of Y.

4.2 Counterfactuals and Marginal Structural Models

To formalize our results, we use counterfactual or potential outcomes. Neyman (1923) introduced counterfactual outcomes to analyze the causal effect of time-independent treatments in randomized experiments. Rubin (1978) and Holland (1986) championed Neyman's idea and emphasized the usefulness of counterfactuals in the analysis of the causal effects of time-independent treatments from observational data. Robins (1986, 1987) proposed a formal counterfactual theory of causal inference that extended Neyman's time-independent treatment theory to longitudinal studies with both direct and indirect effects and sequential time-varying treatments and confounders. In this theory, for any fixed AZT and prophylaxis history $\bar{a} = (\bar{a}_1, \bar{a}_2)$, $Y_{\bar{a}}$ is defined to be the random variable representing a subject's outcome had (possibly contrary to fact) the subject been treated with $\bar{a}$ rather than his observed treatment $\overline{A} = (\overline{A}_1, \overline{A}_2)$. Note that $\bar{a}$ is a possible realization of the random variable $\overline{A}$. For each possible history $\bar{a}$, we are assuming that a subject's response $Y_{\bar{a}}$ is well defined, although generally unobserved. Indeed, we observe $Y_{\bar{a}}$ only for that treatment history $\bar{a}$ equal to a subject's actual treatment history $\overline{A}$; that is, $Y = Y_{\overline{A}}$. This identity is the fundamental "consistency" assumption that links the counterfactual data $Y_{\bar{a}}$ to the observed data $(Y, \overline{A})$.

Note that if at each month t, $A_j(t)$ can take but one of two values (0 for untreated and 1 for treated) with $j \in \{1, 2\}$ and the study duration is K months, then 4^K different $Y_{\bar{a}}$ values are associated with each subject. Then, formally our statement that the effect of treatment history on the mean of Y is a linear logistic function of duration of AZT and prophylaxis therapy is the statement that for each $\bar{a} = (\bar{a}_1, \bar{a}_2)$,

$$E[Y_{\bar{a}}] = g(\bar{a}; \beta),$$

where

$$g(\bar{a}; \beta) = \frac{\exp(\beta_0 + \beta_1 dur(\bar{a}_1) + \beta_2 dur(\bar{a}_2))}{1 + \exp(\beta_0 + \beta_1 dur(\bar{a}_1) + \beta_2 dur(\bar{a}_2))}, \tag{3}$$

$\beta = (\beta_0, \beta_1, \beta_2)$, and $dur(\bar{a}_j) = \sum_{k=0}^{K} a_j(k)$ is the duration of treatment j under the treatment history $\bar{a}$. We call this model a MSM for the effect of AZT and prophylaxis therapy on the mean of Y, because it is a model for the marginal distribution of counterfactual variables and, in the econometric and social science literature, causal models (i.e., models for counterfactual variables) are often referred to as structural. Let $\{Y_{\bar{a}}\}$ denote the set of all 4^K counterfactuals. Then our MSM is a semiparametric model for $(\{Y_{\bar{a}}\}, \overline{L}, \overline{A})$ in which the marginal distribution of each $Y_{\bar{a}}$ is known up to the unknown parameter β but the joint distribution of $\{Y_{\bar{a}}\}$ is otherwise completely unrestricted. For each subject, only 1 of these 4^K counterfactuals are observed, which can create substantial problems for naive users of Bayesian and other likelihood-based methods of inference. For example, in the E step of the EM algorithm one would in principle have to impute for each subject $4^K - 1$ unobserved counterfactuals from the conditional distribution of the unobserved counterfactuals given the single observed counterfactual. If, as in the MACS study, K is approximately 100, this would be computationally infeasible unless we assume a nearly degenerate joint distribution. [See Robins et al. (1999) for an alternative but still computationally demanding approach to estimation of an MSM by maximum likelihood.] Furthermore, the fact that our model

does not specify a parametric form for this conditional distribution creates difficulties in implementing the Monte Carlo *EM* algorithm even when K is moderate. In contast, the non–likelihood-based semiparametric IPTW estimator is easily computed with standard off-the-shelf software.

The parameters β of our MSM encode the magnitude of the average causal effects of the treatments on the outcome. By definition, the causal effect of treatment regimen $\bar{a}$ on the outcome Y for a given study subject is the difference $Y_{\bar{a}} - Y_{\bar{0}}$ between his outcome $Y_{\bar{a}}$ when treated with regimen $\bar{a}$ and his outcome $Y_{\bar{0}}$ when never treated with either AZT or prophylaxis therapy. Thus the average causal effect of regimen $\bar{a}$ is $E[Y_{\bar{a}} - Y_{\bar{0}}] = E[Y_{\bar{a}}] - E[Y_{\bar{0}}] = g(\bar{a};\beta) - g(\bar{0};\beta)$, which depends on β. For example if β_1 is 0, then we say that there is no effect of AZT treatment $\bar{a}_1$ on the outcome, because $E[Y_{\bar{a}}] - E[Y_{\bar{0}}]$ is the same for all $\bar{a}_1$. In contrast, the association parameter γ_1 lacks a causal interpretation.

4.3 Formal Definitions of Causal Exogeneity and No Unmeasured Confounders

We are now in a position to offer more mathematically precise definitions of causal exogeneity and of no unmeasured confounders. Formally, we say that the treatment process $A(t)$ is causally exogenous if for all treatment histories $\bar{a}$,

$$Y_{\bar{a}} \coprod A(t) \mid \overline{A}(t-1), \tag{4}$$

which is mathematically equivalent to the statement that $Y_{\bar{a}}$ is independent of $\overline{A}$. Note that even when $A(t)$ is "causally exogenous," if the treatment has an effect on the outcome, then the observed outcome $Y = Y_{\overline{A}}$ will not be independent of $\overline{A}$, because $Y_{\overline{A}}$ is a function of a subject's observed treatment history $\overline{A}$ itself. Given the covariates recorded in $L(t)$, following Robins et al. (1992), we say that there are no unmeasured confounders for the effect of $A(t)$ on Y if, for each $\bar{a}$,

$$Y_{\bar{a}} \coprod A(t) \mid \overline{A}(t-1), \overline{L}(t). \tag{5}$$

We also refer to the assumption of no unmeasured confounders (5) as the assumption that treatment $A(t)$ is sequentially randomized given the past. This assumption generalizes Rosenbaum and Rubin's (1983) assumption of ignorable treatment assignment to longitudinal studies with time-varying treatments and confounders. The assumption states that, conditional on treatment history and the history of all recorded covariates up to t, treatment at t is independent of the counterfactual random variables $Y_{\bar{a}}$. This will be true if all prognostic factors for (i.e., predictors of) Y used by physicians to determine whether treatment is given at t are recorded in $\overline{L}(t)$ and $\overline{A}(t-1)$. For example, because physicians tend to administer prophylaxis to subjects with previous bouts of PCP, and in untreated subjects PCP predicts Y, the assumption of no unmeasured confounders would be false if $\overline{L}(t)$ does not contain PCP history. The primary goal of the epidemiologists conducting an observational study is to collect data on a sufficient number of covariates to ensure that the assumption of no unmeasured confounders will be at least approximately true.

In an observational study, the assumption of no unmeasured confounders cannot be guaranteed to hold even approximately, and it is not subject to empirical test. Thus investigating the sensitivity to violations of the assumption through a formal sensitivity analysis may be useful. Robins, Greenland, and Hu (1999) and Robins, Rotnitzky, and Scharfstein (1999) have provided details.

Robins (1999) proved that when there are no unmeasured confounders, (a) statistical exogeneity implies causal exogeneity (4), (b) the weighted logistic regression estimator using the weights SW converges to the parameter β of the MSM (3) for $E[Y_{\bar{a}}]$, and (c) the probability limit γ of the usual unweighted logistic estimator generally differs from the causal parameter β of the MSM unless the treatment process is statistically exogenous. Here we provide an informal heuristic argument for (b). View each person as a member of a pseudopopulation comprising SW copies of themselves. In the Appendix, we show that in this new pseudopopulation, $\overline{L}(t)$ does not predict treatment at t given past treatment history, and thus we have created a pseudopopulation in which treatment is statistically and thus causally exogenous. Furthermore, Robins (1999) proved that the causal effect of treatment on Y is the same in the pseudopopulation as in the original population. That is, if $E[Y_{\bar{a}}] = g(\bar{a};\beta)$ in the true population, then the same will be true of the pseudopopulation. Hence we would like to do ordinary logistic regression in the pseudopopulation. But that is what our weighted logistic regression estimator is doing, because the weights create, as required, SW copies of each subject.

We can generalize our MSM (3) slightly and model the marginal distribution of $Y_{\bar{a}}$ within levels of a subset V of the pretreatment (baseline) covariates $L(0)$. Then our marginal structural logistic model (3) could be modified to

$$E[Y_{\bar{a}} \mid V] = \frac{\exp(\beta_0 + \beta_1 dur(\bar{a}_1) + \beta_2 dur(\bar{a}_2) + \beta_3' V)}{1 + \exp(\beta_0 + \beta_1 dur(\bar{a}_1) + \beta_2 dur(\bar{a}_2) + \beta_3' V)}.$$

An IPTW estimator of the parameter β can be obtained by weighted logistic regression with weights SW except now the logistic model includes $dur(\overline{A}_1)$, $dur(\overline{A}_2)$, and V as regressors, and $SW(t)$ is redefined to be

$$SW(t) = \prod_{k=0}^{t} \frac{f[A(k) \mid \overline{A}(k-1), V]}{f[A(k) \mid \overline{A}(k-1), \overline{L}(k)]}. \tag{6}$$

Note that V is already included in the denominator, because V is a subset of the variables in $L(0)$.

4.4 Marginal Structural Cox Proportional Hazards Model

MSMs can easily be extended to failure time outcomes by specifying a marginal structural Cox proportional hazards model. In our analysis of the MACS, we specify the MSM

$$\lambda_{T_{\bar{a}}}(t \mid V) = \lambda_0(t) \exp(\beta_1 a_1(t) + \beta_2 a_2(t) + \beta_3' V), \tag{7}$$

where $T_{\bar{a}} = T_{\bar{a}_1, \bar{a}_2}$ is the subject's time to death if he had followed AZT history $\bar{a}_1$ and prophylaxis history $\bar{a}_2$, $\lambda_{T_{\bar{a}}}(t \mid V)$ is the hazard of $T_{\bar{a}}$ at t conditional on having pretreatment variables V, $\lambda_0(t)$ is an unspecified baseline hazard function, $\exp(\beta_1)$ and $\exp(\beta_2)$ are the causal rate ratios for the effects of AZT and prophylaxis therapy, and V is the vector of baseline regressors comprised of age, calendar year, CD4 count,

CD8 count, white blood cell (WBC) count, red blood cell (RBC) count, platelets, and the following symptoms: fever, diarrhea, herpes zoster, thrush, oral leukoplakia, and weight loss. For variety, we have chosen a model that specifies that the hazard of death at time t depends on current treatment status only. We considered other dose–response models and obtained qualitatively similar results to those reported herein. We make the assumption of no unmeasured confounders with $L(t)$ comprised of CD4, CD8, WBC, RBC, platelets, AIDS-defining illness, number of episodes of PCP, and symptoms at month t. The baseline covariates V are equal to $L(0)$.

Let T be a subject's observed failure (i.e., death) time, so that $T = T_{\bar{A}}$. Arguing as earlier, Robins (1999) showed that in the absence of censoring, a consistent estimator of the unknown parameter $\beta = (\beta_1, \beta_2, \beta_3')'$ is obtained by fitting the ordinary time-dependent Cox model $\lambda_T(t \mid \overline{A}_1, \overline{A}_2(t), V) = \lambda_0(t)\exp(\gamma_1 A_1(t) + \gamma_2 A_2(t) + \gamma_3' V)$, except that the contribution of subject i to a risk set calculation at time t is weighted by $SW_i(t)$, as defined in (6), with $T > k$ added to the conditioning event. Note that the subject-specific weights change with time. Unfortunately, few standard Cox proportional hazard software programs allow for time-varying weights. To avoid this software problem, we fit a weighted pooled logistic regression, treating each person-month as an observation and allowing for a time-dependent intercept. That is, we fit, by weighted logistic regression using weights $SW(t)$, the model

$$\begin{aligned}\operatorname{logit}\Pr[D(t) = 1 \mid D(t-1) = 0, \overline{A}(t-1), V] \\ = \gamma_0(t) + \gamma_1 A_1(t-1) + \gamma_2 A_2(t-1) + \gamma_3' V, \quad (8)\end{aligned}$$

where $D(t) = 0$ if a subject was alive at month t and 1 if the subject died at month t, and $\gamma_0(t)$ is a time (i.e., month)-specific intercept. This method offers the advantage of being easy to program in any standard statistical package. Under our assumptions, we thereby obtain a consistent estimator of the parameter vector β of the MSM,

$$\begin{aligned}\operatorname{logit}\Pr[D_{\bar{a}}(t) = 1 \mid D_{\bar{a}}(t-1) = 0, V] \\ = \beta_0(t) + \beta_1 a_1(t-1) + \beta_2 a_2(t-1) + \beta_3' V. \quad (9)\end{aligned}$$

Because the death rate in any given month t is small, the parameters of (9) and (7) closely approximate one another.

Because of the weights, the standard error estimates outputted by a standard logistic program are invalid and may be either too large or too small. To overcome this difficulty, (8) should be fit using a generalized estimating equations (GEE) (Liang and Zeger 1986) program (e.g., option "repeated" in SAS proc genmod), which outputs robust variance estimators. The robust variance GEE estimator provide a conservative confidence interval for the β (Robins 1999); that is, the 95% Wald confidence interval calculated as $\beta \pm 1.96\times$ (robust) standard error is guaranteed to cover the true β at least 95% of the time in large samples. GEE intervals are conservative because they do not account for the fact that the weights are estimated (as described in Sec. 4.5), and estimating the weights shrinks the variance of our IPTW estimator of β. The observation that an IPTW estimator that uses estimated weights has smaller variance than one that uses the true weights has been discussed by Robins (1999). Robins provided a nonconservative variance estimator; however, this estimator cannot be easily computed using widely available software.

We now describe how to accommodate censoring in the analysis. We defined a subject as right censored at time t [i.e., $C(t) = 1$] if by time t he either dropped out of the study or reached administrative end of follow-up alive. No new idea is required to account and adjust for right censoring. To want to adjust for censoring is only to say that our interest is in estimating the effect of $\bar{a}$ when $\bar{c} \equiv 0$; that is, when censoring is abolished. Thus the marginal structural Cox model that we are interested in is now

$$\lambda_{T_{\bar{a},\bar{c}\equiv 0}}(t \mid V) = \lambda_0(t)\exp[\beta_1 a_1(t) + \beta_2 a_2(t) + \beta_3' V],$$

where $T_{\bar{a},\bar{c}\equiv 0}$ is a subject's failure time when treated with regimen $\bar{a}$ in the absence of censoring.

We say that censoring is ignorable or noninformative if the conditional cause-specific hazard of being censored at k among subjects alive and uncensored up to k does not depend on the failure times $T_{\bar{a},\bar{c}\equiv 0}$ given AZT $\overline{A}_1(k-1)$, prophylaxis $\overline{A}_2(k-1)$, and the time-dependent covariate $\overline{L}(k-1)$ history before k. Under the assumptions of ignorable censoring and no unmeasured confounding, Robins (1999) showed that we still obtain from (8) consistent estimators of β if we weight a subject alive and uncensored at month t by $SW(t) \times SW^{\dagger}(t)$, where

$$SW^{\dagger}(t) = \prod_{k=0}^{t} \frac{\Pr[C(k) = 0 \mid \overline{C}(k-1) = 0, \overline{A}(k-1), V, T > k]}{\Pr[C(k) = 0 \mid \overline{C}(k-1) = 0, \overline{A}(k-1), \overline{L}(k), T > k]}$$

is informally the inverse of the ratio of a subject's probability of remaining uncensored up to month t divided by that probability calculated as if there had been no time-dependent determinants of censoring except past treatment history and V, and we modify our definition (6) of $SW(t)$ to add $C(k) = 0$ to the conditioning events in both the numerator and the denominator. The denominator of $SW(t) \times SW^{\dagger}(t)$ is informally the probability that a subject would have his own observed treatment and censoring history through month t.

4.5 Estimation of the Weights

We now describe how to estimate the unknown weights $SW(t)$ and $SW^{\dagger}(t)$. First, consider estimation of the $SW(t)$. Let $P_1(k) = \Pr[A_1(k) = 0 | \overline{A}_1(k-1) = 0, \overline{A}_2(k-1), \overline{L}(k), \overline{C}(k) = 0, T \geq k]$ and $P_2(k) = \Pr[A_2(k) = 0 | A_1(k), \overline{A}_2(k-1) = 0, \overline{L}(k), C(k) = 0, T > k]$ denote the conditional probability of remaining off AZT and prophylaxis in month k given past treatment and time-dependent covariate history $\overline{L}(k)$. Let $P_1^*(k)$ and $P_2^*(k)$ be defined analogously, except that $\overline{L}(k)$ is replaced by V.

We obtained estimates $\widehat{P}_j(k)$ of the $P_j(k)$, $j = 1, 2$, by fitting to subjects at risk to initiate therapy with treatment j at time k the pooled logistic model for the binary response A_j,

$$P_j(k) = \operatorname{expit}\{\alpha_{j0}(k) + \alpha_{j1}' Q_j(k)\},$$

where the $\alpha_{j0}(k)$ are month-specific intercepts, $Q_1(k) = (A_2(k-1), L(k)', V')'$, and $Q_2(k) = (A_1(k), A_1(k-1), L(k)', V')'$. We obtained analogous estimates $\widehat{P}_j^*(k)$ of the $P_j^*(k)$ by

removing $L(k)$ from $Q_j(k)$. Our estimate of the denominator of $SW_i(t)$ is $\widehat{\overline{P}}_{1i}(t)\widehat{\overline{P}}_{2i}(t)$, where

$$\widehat{\overline{P}}_{ji}(t) = \prod_{u=0}^{t} \widehat{P}_{ji}(u)$$

if subject i did not start treatment A_j up to month t, and

$$\widehat{\overline{P}}_{ji}(t) = \left[1 - \widehat{P}_{ji}(k)\right] \prod_{u=0}^{k-1} \widehat{P}_{ji}(u)$$

if subject i started treatment A_j in month k for $k \leq t$. Note that in calculating $\widehat{\overline{P}}_{ji}(k)$, we have used the assumption that no subject stops either treatment once he begins it. We obtained estimates $\widehat{\overline{P}}^{*}_{1i}(t)\widehat{\overline{P}}^{*}_{2i}(t)$ of the numerator of $SW_i(t)$ by using the $\widehat{P}^{*}_{ji}(k)$ in place of the $\widehat{P}_{ji}(k)$. To estimate $SW^{\dagger}(k)$, we fit logistic models for the binary responses $C(k)$ in exact analogy to the logistic models for $P_1(k)$ and $P^{*}_1(k)$.

Robins (1999) proved that our IPTW estimator will be consistent if the models for treatment initiation and censoring used in estimating the denominators of $SW(t)$ and $SW^{\dagger}(t)$ are correctly specified, regardless of whether or not the numerator models are misspecified.

There is one further detail we have yet to discuss. Under the asymptotic theory that motivated our estimation procedure, the IPTW estimator of β will be $n^{1/2}$ consistent only if our estimate of $SW(t)$ converges at a rate of $n^{1/4}$ or better (Robins 1999). The practical implication of this result is that for our IPTW estimator of β to perform well in moderate-sized samples, our estimate $\widehat{SW}(t)$ of $SW(t)$ cannot be exceedingly variable. To ensure this, we reduced the number of free parameters in the logistic model for $P_j(t)$ by not fitting a separate intercept $\alpha_{j0}(k)$ for each month k. Rather, we modeled $\alpha_{j0}(k)$ using natural cubic splines with five knots (at months 23, 44, 71, 94, and 100).

5. DATA ANALYSIS

We first demonstrate that conditions (a) and (b) stated in Section 1 are true in the MACS data. To determine whether current CD4 count and PCP were independent risk factors for death, we fit an (unweighted) time-dependent Cox model for the hazard of death at time t that included the baseline covariates V and time-dependent covariates PCP, CD4 count, AZT, and prophylaxis. The mortality hazard ratio for a PCP episode before t (1 = yes, 0 = no) was 3.77 ($p < .001$). The hazard ratios for low CD4 count (<200) and moderate CD4 count (200–500) relative to a normal CD4 count (≥500) were 16.5 ($p < .001$) and 3.48 ($p < 001$). To determine whether current CD4 and PCP predicted subsequent treatment with AZT and prophylaxis, we fit pooled linear logistic models for treatment initiation with AZT and with prophylaxis that included the baseline covariates V and the time-dependent covariates PCP and CD4 count. In the model for initiation of prophylaxis, the (discrete) hazard ratios corresponding to PCP history, low CD4 count, and moderate CD4 count were 1.87 ($p < .001$), 4.94 ($p < .001$), and 2.82 ($p < .001$). In the model for the initiation of AZT, the corresponding hazard ratios were 2.18 ($p < .001$), 3.38 ($p < .001$), and 2.56 ($p < .001$). Taken together, these results imply that condition (a) of Section 1 holds.

To determine whether condition (b) holds, we first modeled the logit of the probability of developing a first episode of PCP in month t given the baseline covariates V, AZT, and prophylaxis at time $t-1$, and the time-dependent covariates $L(t-2)$. The discrete hazard ratios were 1.03 ($p = .64$) for AZT and 0.77 ($p < .001$) for prophylaxis. This implies that condition (b) holds in the MACS data, because prophylaxis is a protective risk factor of the development of subsequent PCP. Finally, we fit a linear model for the mean of CD4 count in month t given the baseline covariates V, AZT and prophylaxis at time $t-1$, and the remaining time-dependent covariates $L(t-2)$. In this model neither the estimated AZT nor prophylaxis coefficient was a significant predictor of higher CD4 count.

Having demonstrated that the standard analytic approaches cannot be used to validly estimate the joint effects of AZT and prophylaxis on mortality in the MACS dataset, we now proceed with our MSM analysis. Table 1 summarizes the empirical distribution at two time points (24 and 84 months of follow-up) of the estimated weight $\widehat{SW}(t) = \widehat{\overline{P}}^{*}_{1i}(t)\widehat{\overline{P}}^{*}_{2i}(t)/\widehat{\overline{P}}_{1i}(t)\widehat{\overline{P}}_{2i}(t)$ and the treatment-specific contributions to the numerator [$\widehat{\overline{P}}^{*}_{1i}(t)$ for AZT and $\widehat{\overline{P}}^{*}_{2i}(t)$ for prophylaxis] and denominator [$\widehat{\overline{P}}_{1i}(t)$ and $\widehat{\overline{P}}_{2i}(t)$] of $\widehat{SW}(t)$. We do not include the estimated weights $\widehat{SW}^{\dagger}(t)$ for censoring in the table, because they were all close to 1, indicating little selection bias due to censoring.

Table 2 compares our IPTW estimates and conservative standard errors with the usual unweighted estimates and standard errors obtained by fitting (8). The difference between the weighted and unweighted analyses is especially striking for AZT. By exponentiating, we can obtain 95% conservative confidence intervals (CIs) for the treatment hazard (i.e., mortality) ratios. The estimated mortality rate ratios were .67 (conservative 95% CI of .46, .98) for AZT users versus nonusers, and 1.14 (conservative 95% CI of .79, 1.64) for prophylaxis users versus nonusers. The statistically significant beneficial effect of AZT agrees with the results of randomized clinical trials. The hazard ratios estimated by the unweighted model were substantially greater: 1.85 (95% CI of 1.49, 2.30) for AZT and 1.58 (95% CI of 1.31, 1.89) for prophylaxis. This difference indicates a large amount of residual confounding by the time-varying prognostic factors recorded in $L(t)$, even after adjusting for the baseline confounders V. Indeed, both treatments appear significantly harmful when we fail to adjust for time-varying confounding. The large degree of confounding occurs because there exist time-dependent covariates such as CD4 count and PCP that are strong predictors of both subsequent mortality and treatment.

6. DISCUSSION

Before introducing MSMs, Robins and colleagues introduced three methods for estimating the causal effect of a time-varying treatment: the semiparametric g-computation algorithm formula estimator (Robins 1986, 1998b), g-estimation of structural nested models (Robins 1998b; Robins, Blevins, Ritter, and Wulfsohn 1992), and the iterative conditional expectations

Table 1. Probability of Having One's Own Observed Treatment History at 24 and 84 Months of Follow-Up, MACS

		Mean (SD)[a]	*Median (IQR)*[a]	*Percentile 1*	*Percentile 99*
24 months ($n = 2,048$)					
Probability of having observed AZT					
• Given baseline covariates[b]	$\widehat{\overline{P}}^{*}_{1i}(24)$	.521(.337)	.689(.771)	.006	.864
• Given time-varying covariates[c]	$\widehat{\overline{P}}_{1i}(24)$	.538(.354)	.676(.821)	.002	.916
Probability of having observed prophylaxis history					
• Given baseline covariates[b]	$\widehat{\overline{P}}^{*}_{2i}(24)$	.747(.323)	.936(.342)	.002	.964
• Given time-varying covariates[c]	$\widehat{\overline{P}}_{2i}(24)$	.751(.329)	.933(.318)	.001	.972
Estimated weight $\widehat{SW}$		1.07(1.01)	.93(.23)	.29	3.92
84 months ($n = 801$)					
Probability of having observed AZT history					
• Given baseline convariates[b]	$\widehat{\overline{P}}^{*}_{1i}(84)$	.116(.170)	.013(.235)	.003	.520
• Given time-varying covariates[c]	$\widehat{\overline{P}}_{1i}(84)$	.155(.237)	.016(.254)	.002	.697
Probability of having observed AZT history					
• Given baseline covariates[b]	$\widehat{\overline{P}}^{*}_{2i}(84)$	.332(.345)	.181(.784)	.002	.855
• Given time-varying covariates[c]	$\widehat{\overline{P}}_{2i}(84)$	.356(.360)	.244(.773)	.001	.903
Estimated weight $\widehat{SW}$		1.10(1.24)	.75(.74)	.15	5.39

[a] *SD* = standard deviation; *IQR* = interquartile range.
[b] Age (years), calendar year (1985, 1986, 1987–89, ≥1990), CD4 (<200, 200–499, 500/μL, CD8 (<500, 500–999, 1000/0 × b5L), WBC (<3000, 3000–4999, 5000/μL), RBC (<35, 35–44, 45 × 10^5/μL), platelets (<150, 150–249, 250 × 10^3/μL), presence of symptoms, AZT use, and prophylaxis use.
[c] Baseline covariates plus most recent CD4, CD8, WBC, RBC (<35, 35–44, 45 × 10^5/μL), platelets, presence of symptoms, presence of an AIDS-defining illness, and number of previous PCP bouts (0, 1, 2 or more).

Table 2. IPTW Estimates of the Parameters of a Marginal Structural Model for the Causal Effects of AZT and Prophylaxis on Mortality, MACS

		Weighted model			*Unweighted model*[b]		
Variable[a]		*Parameter estimate*	*Robust standard error*	*p value*	*Parameter estimate*	*Standard error*	*p value*
Zidovudine		−.396	.192	.040	.615	.110	<.001
Prophylaxis		.131	.185	.477	.455	.094	<.001
Age		.041	.009	<.001	.028	.005	<.001
Year	1985	.358	.527	.497	.652	.509	.200
	1986	.325	.540	.547	.579	.515	.261
	1987–1989	.429	.595	.472	.247	.527	.639
	≥1990	0			0		
CD4 (/μL)	<200	1.847	.258	<.001	1.658	.136	<.001
	[200,500)	.599	.150	<.001	.519	.089	<.001
	≥500	0			0		
CD8 (/μL)	<500	−.599	.188	.001	−.488	.123	<.001
	[500,1,000]	−.486	.151	.001	−.338	.089	<.001
	≥1,000	0			0		
WBC (/μL)	>3000	.815	.266	.002	.356	.221	.108
	[3,000,5,000)	.305	.149	.040	.109	.086	.208
	≥5000	0			0		
RBC (×10^5/μL)	<35	−.093	.623	.882	.536	.349	.180
	[35,45)	.166	.198	.403	.345	.105	.001
	≥45	0			0		
Platelets (×10^3/μL)	<150	.521	.239	.030	.305	.133	.022
	[150,250)	.010	.136	.939	.013	.086	.880
	≥250	0			0		
Thrush		.454	.198	.022	.480	.112	<.001
Oral leukoplakia		.318	.383	.407	.516	.202	.011
Weight loss		.641	.220	.004	.538	.136	.004
Herpes zoster		.089	.294	.761	.073	.184	.691
Diarrhea		.469	.305	.124	.203	.186	.275
Fever		.354	.277	.200	.452	.278	.105

[a] All variables were measured at baseline except AZT and prophylaxis (1 = ever user, 0 = never user). Weights $\widehat{SW}(t) \times \widehat{SW}^{\dagger}(t)$ are as defined in the text.
[b] Unweighted estimates do not have a causal interpretation and are shown for comparison purposes only.

(ICE) estimator (Robins 1998b). Arjas and Eerola (1993) proposed an approach that is a particular example of the semiparametric g-computation algorithm estimator. Lavori and Dawson (1994) described a different approach that is applicable only under rather restrictive assumptions.

ICE estimators can be used only rarely, because they often lead to incompatible models (Robins 1998b). Inference based on SNMs and MSMs is preferable to that based on the semiparametric g-computation algorithm estimator, because models based on the g-computation algorithm are often unable to represent the null hypothesis of no treatment effect (Robins 1986, 1998b).

A major advantage of MSMs is that they resemble standard models, whereas SNMs often do not. For example, the logistic MSM and the Cox proportional hazards MSM described herein are natural ways to extend the ordinary logistic and time-dependent Cox models to allow for estimation of causal effects. SNMs cannot be conveniently used to estimate the effect of treatment on dichotomous $(0, 1)$ outcomes unless the outcome is rare (Robins 1998a,b, 1999), because logistic SNMs cannot be fit by g-estimation.

MSMs do have limitations, however. For example, Robins (1999) showed that in contrast to SNMs, MSMs cannot be used to estimate effects in studies in which all subjects with a particular covariate value, say $l^*(k)$, are certain to receive the same treatment at k , as would be true in a study of the effect of exposure to an industrial chemical on the mortality of a cohort of chemical workers; all workers with the covariate value "off work at time k" receive no exposure to the chemical at k. Another advantage of SNMs over MSMs is that although MSMs are useful for estimating both interactions between treatment and baseline covariates V and the effect of prespecified treatment regimens $\bar{a}$, they are much less useful than SNMs for estimating either interactions between treatment and time-dependent covariates or the effect of dynamic treatment regimens in which treatment in a given month is decided in part based on a subject's evolving covariate history (Robins 1998a, 1999). Finally, if data on an instrumental variable can be obtained, SNMs, in contrast to MSMs, can still be consistently estimated even in the presence of confounding by unmeasured factors (Robins, 1999).

In summary, we have used IPTW estimation of a marginal structural Cox proportional hazards model to estimate the joint effect of AZT and prophylaxis therapy on the mortality of HIV-positive patients in the MACS. We found a significant beneficial effect of AZT on mortality. IPTW estimators appropriately adjust for time-dependent confounding. An unweighted analysis that adjusted only for baseline confounding found an artifactual statistically significant adverse effect on mortality for both AZT and prophylaxis therapy.

The correctness of the causal inferences from our weighted analysis depends on three key assumptions. First, we must assume that the covariates in $L(t)$ are sufficient to adjust for both confounding and selection bias due to loss to follow-up. In an observational study, this assumption is not testable. But, as described earlier, methods have recently been developed to quantify the sensitivity of effect estimates to increasing violation of this assumption. Second, we need to assume that our MSM for the effect of AZT on mortality is correctly specified. Finally, we must assume that our models for initiation of treatment and censoring, given past covariate and treatment history, are correctly specified. This last assumption can be weakened if we use the doubly-robust MSM estimators described in Robins (2000) and in Section 3.3 of Robins (1999) instead of IPTW estimators. Given the first two assumptions, doubly-robust estimators are consistent for the parameters of an MSM when either (but not necessarily both) our model for treatment and censoring is correct or a model for the remaining, as yet unmodeled, part of the joint distribution of the observables is correct.

The foregoing assumptions may appear heroic. Nonetheless, even when estimating the effect of a time-independent treatment using standard statistical models, the same assumptions (no unmeasured confounders, noninformative censoring, and no model misspecification) are needed to endow the parameters with a causal interpretation. Furthermore, when estimating the effect of a time-varying treatment, our assumptions are less restrictive than those required by standard unweighted analyses. Our approach does not require for validity the absence of confounding by time-dependent covariates affected by previous treatment.

APPENDIX: LEMMA

Lemma A.1. In the pseudopopulation created by weighting by SW, treatment is statistically exogenous.

Proof. By definition, the density of $(\overline{L}_K, \overline{A}_K)$ in the pseudopopulation is the weighted density $f^*(\overline{L}_K, \overline{A}_K) = f(\overline{L}_K, \overline{A}_K)sw(\overline{L}_K, \overline{A}_K)/f(\overline{L}_K, \overline{A}_K)sw(\overline{L}_K, \overline{A}_K)d\mu(\overline{L}_K, \overline{A}_K)$, where $sw(\overline{L}_K, \overline{A}_K) = SW$. Now, using the defintion of SW, a direct calculation shows that the denominator is 1 and the numerator is $\prod_{k=0}^{K} f(L_k \mid \overline{L}_{k-1}, \overline{A}_{k-1}) \prod_{k=0}^{K} f(A_k \mid \overline{A}_{k-1})$, which implies that $f^*(A_k \mid \overline{L}_k, \overline{A}_{k-1}) = f^*(A_k \mid \overline{A}_{k-1}) = f(A_k \mid \overline{A}_{k-1})$, thus proving the lemma.

[Received August 1998. Revised December 2000.]

REFERENCES

Arjas, E., and Eerola, M. (1993), "On Predictive Causality in Longitudinal Studies," *Journal of Statistical Planning and Inference*, 34, 361–384.

Engel, R. F., Hendry, D. F., and Richard, J. F. (1983), "Exogeneity," *Econometrica*, 51, 277–304.

Flanders, W. D., and Greenland, S. (1991), "Analytic Methods for Two-Stage Case-Control Studies and Other Stratified Designs," *Statistics in Medicine*, 10, 739–747.

Graham N. M. H., Zeger S. L., Park L. P., Vermund, S. H., Detels, R., Rinaldo, C. R., and Phair, J. P. (1992), "The Effects on Survival of Early Treatment of Human Immunodeficiency Virus Infection," *New England Journal of Medicine*, 326, 1037–1042.

Holland, P. (1986), "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–961.

Horvitz, D. G., and Thompson, D. J. (1952), "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Medical Association*, 47, 663–685.

Kalbfleisch, J. D., and Lawless, J. F. (1988), "Likelihood Analysis of Multistage Models for Disease Incidence and Mortality," *Statistics in Medicine*, 7, 149–160.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Lavori, P. W., Dawson, R., and Mueller, T. B. (1994), "Causal Estimation of Time-Varying Treatments in Observational Studies: Application to Depressive Disorder," *Statistics in Medicine*, 13, 1089–1100.

Leamer, E. E. (1985), "Vector Autoregressions for Causal Inference?," *Carnegie-Rochester Conference Series on Public Policy*, 22, 255–304.

Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9," translated in *Statistical Science*, 1990, 5, 465–480.

Robins, J. M. (1986), "A New Approach to Causal Inference in Mortality Studies With Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7, 1393–1512.

——— (1987), Addendum to "A New Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods—Application to Control of the Healthy Worker Survivor Effect," *Computers and Mathematics with Applications*, 14, 923–945.

——— (1998a), "Structural Nested Failure Time Models," in *The Encyclopedia of Biostatistics*, eds. P. Armitage and T. Colton, Chichester, U.K.: Wiley, pp. 4372–4389.

——— (1998b), "Correction for Non-Compliance in Equivalence Trials," *Statistics in Medicine*, 17, 269–302.

——— (1999), "Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference," in *Statistical Models in Epidemiology: The Environment and Clinical Trials*, eds. M. E. Halloran and D. Berry, New York: Springer-Verlag, pp. 95–134.

——— (2000). "Robust Estimation in Sequentially Ignorable Missing Data and Causal Inference Models," in *Proceedings of the Section on Bayesian Statistical Science, American Statistical Association*, pp. 6–11.

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; errata 4, 189.

Robins, J. M., and Greenland, S. (1994), "Adjusting for Differential Rates of PCP Prophylaxis in High- Versus Low-Dose AZT Treatment Arms in an AIDS Randomized Trial," *Journal of the American Statistical Association*, 89, 737–749.

Robins, J. M., Greenland, S., and Hu, F.-C. (1999), "Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome" (with discussion), *Journal of the American Statistical Association*, 94, 687–712.

Robins, J. M., and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," in *AIDS Epidemiology—Methodological Issues*, eds. N. Jewell, K. Dietz, and V. Farewell, Boston: Birkhäuser, pp. 297–331.

Robins, J. M., Rotnitzky A., and Scharfstein, D. (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*, eds. M. E. Halloran and D. Berry, New York: Springer-Verlag, pp. 1–94.

Rosenbaum, P. R. (1987), "Model-Based Direct Adjustment," *Journal of the American Statistical Association*, 82, 387–394.

Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34–58.