STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:1689–1709 (DOI: 10.1002/sim.1144)

Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures

Miguel A. Hernán[1,*,†], Babette A. Brumback[2] and James M. Robins[1,3]

[1]*Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.*
[2]*Department of Biostatistics, University of Washington, 1705 N.E Pacific, Seattle, WA 98195, U.S.A.*
[3]*Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, U.S.A*

SUMMARY

Even in the absence of unmeasured confounding factors or model misspecification, standard methods for estimating the causal effect of a time-varying treatment on the mean of a repeated measures outcome (for example, GEE regression) may be biased when there are time-dependent variables that are simultaneously confounders of the effect of interest and are predicted by previous treatment. In contrast, the recently developed marginal structural models (MSMs) can provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent. We describe an MSM for repeated measures that parameterizes the marginal means of counterfactual outcomes corresponding to prespecified treatment regimes. The parameters of MSMs are estimated using a new class of estimators – inverse-probability of treatment weighted estimators. We used an MSM to estimate the effect of zidovudine therapy on mean CD4 count among HIV-infected men in the Multicenter AIDS Cohort Study. We estimated a potential expected increase of 5.4 (95 per cent confidence interval $-1.8, 12.7$) CD4 lymphocytes/μl per additional study visit while on zidovudine therapy. We also explain the theory and implementation of MSMs for repeated measures data and draw upon a simple example to illustrate the basic ideas. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: repeated measures; time-varying exposure; confounding; causality; HIV disease

1. INTRODUCTION

In this paper we use marginal structural models to estimate the causal effect of zidovudine therapy on mean CD4 count among HIV-infected participants in the Multicenter AIDS Cohort Study (MACS), an observational follow-up study of U.S. men [1]. We then compare our

*Correspondence to: Miguel A. Hernán, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.
†E-mail: miguel_hernan@post.harvard.edu

Contract/grant sponsor: U.S. National Institutes of Healths; contract/grant numbers: NIAID T32AI07358, NIEHS 5T32 ES07142-14 and R01-AI32475

Received November 1998
Accepted September 2001

observational estimate of the effect of zidovudine therapy with those obtained in randomized trials.

Marginal structural models (MSMs) are a new class of causal models [2, 3] whose parameters can be consistently estimated using a new class of estimators: the inverse-probability of treatment weighted (IPTW) estimators. MSMs are an alternative to g-estimation of structural nested models [4–6].

The standard approach to estimation of the effect of a time-varying treatment, such as zidovudine, on the mean of a repeated measures continuous outcome, such as CD4 count, is to model the mean of the outcome at each time as a function of past treatment. However, Robins [7] has shown that the standard approach may be biased, whether or not one further adjusts for past covariate and outcome history in the analysis, when:

A1 – conditional on past treatment history, a time-dependent variable is a predictor of the subsequent mean of the outcome and also a predictor of subsequent treatment;
A2 – past treatment history is an independent predictor of the time-dependent variable.

We refer to covariates satisfying A1 as time-dependent confounders. For example, in the MACS, past CD4 count is a time-dependent confounder for the effect of zidovudine on future CD4 count since it not only predicts future CD4 count but also subsequent initiation of zidovudine therapy [8], and past zidovudine history is an independent predictor of subsequent CD4 count [9]. Thus, standard methods for repeated measures (for example, GEE regression [10, 11]) that model the mean CD4 count at each time using a summary of zidovudine history up to that time may produce biased estimates of the causal effect of zidovudine whether or not one adjusts for past CD4 count in the analysis. Marginal structural models can eliminate the bias of standard methods.

This paper is organized as follows. In Section 2 we give a brief summary of the MACS and describe our notation. Section 3 reviews the standard GEE linear model, and Section 4 reviews the definitions of causal effect and confounding based on counterfactual outcomes. In Section 5, we describe the MSM framework for repeated measures data, including IPTW estimation, an extension to right-censored data, and practical advice on how to estimate the weights required by the method. We present a simple worked example in Section 6 to help clarify the theory presented in Section 5. Section 7 presents our analysis of the MACS data, and Section 8 concludes.

2. THE MACS

Between 1984–1991, the MACS enrolled 5622 homosexual and bisexual men, with no prior AIDS-defining illness, from the metropolitan areas of Los Angeles, Baltimore-Washington, Pittsburgh and Chicago. Study participants were asked to return every 6 months to complete a questionnaire, undergo physical examination and provide blood samples. The design and methods of the MACS have been described in detail elsewhere [1, 8].

We restricted our cohort to HIV-positive men alive in the period during which zidovudine was available for use (that is, after the study visit 5; March 1986 to March 1987). Follow-up ended at study visit 21, October 1994, death, or first missed visit, whichever came first. Our analysis include the 1486 men (contributing 9752 visits) that attended the first two eligible

visits between visits 5 and 21 while HIV-positive, and that did not have an AIDS-defining illness and were not on antiretroviral therapy before the first eligible visit.

Let $Y_i(t)$ be subject's i CD4 count at his tth eligible visit, and $A_i(t)$ be one if subject i was on zidovudine at visit t, zero otherwise, $t = 0, \ldots, 16$. As in previous MACS analyses, we assumed that subjects remained on treatment once they started. Throughout we assume that each subject's data are drawn independently from a distribution common to the N subjects. We can then suppress the subject-specific notation, and we will write, for example, $Y_i(t)$ as $Y(t)$. Let V be a vector of time-independent baseline covariates measured at the first eligible visit. In our analysis, the covariates included in V are age, calendar year, CD4 count, CD8 count, white blood cell count (WBC), red blood cell count (RBC), platelets, and presence of the following symptoms: fever (temperature $> 37.9°$C) for 2 weeks; oral candidiasis; diarrhoea for 2 weeks; weight loss of $\geqslant 4.5$ kg; oral hairy leukoplakia; herpes zoster. Finally, let $L(t)$ be the vector of the time-varying covariates recording CD4, CD8, WBC, RBC, platelets, and presence of symptoms measured at visit t. Note that $L(t)$ includes $Y(t)$. We adopt the convention that variables in $L(t)$ are measured prior to deciding $A(t)$. We use overbars to represent a covariate history so, for example, $\bar{A}(t) = \{A(u); u = 0, \ldots, t\}$ is a subject's treatment history through t. $\bar{A} \equiv \bar{A}(16)$ is the subject's treatment history through the end of the follow-up period. We use the symbol $\amalg$ to indicate statistical independence, for example, $A \amalg B \mid C$ means A is conditionally independent of B given C.

3. A GEE MODEL FOR REPEATED MEASURES

In the absence of confounding, the mean CD4 count given zidovudine treatment history would often be modelled using a GEE model [10, 11]

$$E[Y(t+1)|\bar{A}] = g(\bar{A}(t), \boldsymbol{\gamma}) \tag{1a}$$

where g is a known function and $\boldsymbol{\gamma}$ is a parameter to be estimated. For example, we will use

$$g(\bar{A}(t), \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 \,\mathrm{cum}[\bar{A}(t)] + \gamma_2 t \tag{1b}$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)'$ and $\mathrm{cum}[\bar{A}(t)] = \sum_{k=0}^{t} A(k)$. This association model states that the mean outcome at t depends linearly on time t and on prior cumulative exposure $\mathrm{cum}[\bar{A}(t)]$, but alternative specifications could be entertained. To avoid extraneous complications, until Section 6, we will assume that there is no model misspecification.

A complete specification of the GEE requires a variance function and a correlation structure between repeated outcomes from the same individual. For example, one could assume either a homoscedastic variance $\mathrm{var}[Y(t+1)|\bar{A}] = \sigma^2$, or a heteroscedastic variance, and the 'exchangeable' structure $\mathrm{corr}[Y(t_1), Y(t_2)|\bar{A}] = \rho$, the 'autoregressive' structure $\mathrm{corr}[Y(t_1), Y(t_2)|\bar{A}] = \rho^{|t_2 - t_1|}$ etc., for any pair (t_1, t_2). However, even if these second moment assumptions do not hold, the GEE estimator of $\boldsymbol{\gamma}$ remains asymptotically unbiased [10, 11]. Upon specifying a working covariance matrix, model (1) can be fit by generalized least-squares using standard software (for example, '`repeated`' statement in SAS proc genmod [12]).

Suppose model (1) is correctly specified so that our GEE estimator $\boldsymbol{\gamma}_1$ is consistent and asymptotically normal. This still leaves open the question as to when the parameter γ_1 from model (1) has a causal interpretation. Before we answer this question, we need to give a formal meaning to the causal effect of zidovudine on the mean of CD4.

4. COUNTERFACTUAL OUTCOMES

We begin by characterizing the relevant treatment regimes. Corresponding with the 16 MACS visits, we consider zidovudine treatment regimes $\bar{a} = [a(0), \ldots, a(16)]$ where $a(t)$ is 1 if the regime specifies the subject is to be on treatment at time t and $\bar{a}(t)$ represents treatment history under regime $\bar{a}$ through visit t. Note $\bar{a}(16) = \bar{a}$. Because in the MACS we have assumed subjects remain on zidovudine therapy once begun, we consider 17 possible treatment regimes: regime (0) begin zidovudine in the interval [visit 0, visit 1), i.e., at $t = 0$, regime (1) begin zidovudine in the interval [visit 1, visit 2), i.e., at $t = 1, \ldots$, and regime (16) begin zidovudine at or after visit 16, i.e., $t \geqslant 16$. For our purposes, regime (16) is equivalent to never starting zidovudine, as we do not consider outcomes measured past visit 16. In a small abuse of notation, we denote treatment regime j as $\bar{a} = j$.

Associated with each treatment regime $\bar{a}$ are the potential or counterfactual outcomes $Y_{\bar{a},i}(t)$, which denote subject's i outcome Y at time t had, possibly contrary to the fact, subject i followed treatment plan $\bar{a}$. The subject's observed outcome $Y_i(t)$ is the counterfactual outcome $Y_{\bar{a},i}(t)$ for the treatment regime $\bar{a}$ that the subject did indeed take. That is, $Y_i(t) \equiv Y_{\bar{A},i}(t)$, where $\bar{a} = \bar{A}$ is the subject's observed treatment history. For example, suppose $\bar{a} = 8$ and $\bar{a}' = 11$. Then $Y_{\bar{a},i}(16)$ represents subject's i hypothetical CD4 count at visit 16 when zidovudine is initiated at visit 8, and $Y_{\bar{a}',i}(16)$ represents his hypothetical response when zidovudine is begun three visits later. Had subject i actually initiated treatment at visit 11, then the counterfactual outcome $Y_{\bar{a}',i}(16)$ would have been observed, that is, $Y_{\bar{a}',i}(16) = Y_i(16)$. Otherwise, the counterfactual outcome $Y_{\bar{a},i}(16)$ would be missing.

The random variables $Y_{\bar{a}}$ are counterfactual variables because they represent a subject's outcome had, possibly contrary to fact, the subject been treated with $\bar{a}$ rather than his observed treatment $\bar{A}$. For each possible history $\bar{a}$ we are assuming a subject's response $Y_{\bar{a}}$ is well defined, although generally unobserved [13]. Neyman [14] introduced counterfactual outcomes to analyse the causal effect of time-independent treatments in randomized experiments. Rubin [15, 16] 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 [7, 17] proposed a formal counterfactual theory of causal inference that extended Neyman's [14] time-independent treatment theory to longitudinal studies with both direct and indirect effects and sequential time-varying treatments and confounders.

We are now ready to present formal definitions of a causal effect and confounding. Treatment has a causal effect on subject's i outcome when there is a difference in his counterfactual outcomes under two or more treatment plans. In other words, if zidovudine has no effect on the CD4 count of subject i, then $Y_{\bar{a},i}(t+1) = Y_{\bar{a}',i}(t+1)$ for all $\bar{a}, \bar{a}'$, and $t = 0, \ldots, 15$. Thus, the causal effect of treatment regime $\bar{a}$ on the outcome $Y(t+1)$ for a given subject is the difference $Y_{\bar{a}}(t+1) - Y_{\bar{0}}(t+1)$ between his outcome $Y_{\bar{a}}(t+1)$ when treated with regime $\bar{a}$ and his outcome $Y_{\bar{0}}(t+1)$ when never treated through time t. Similarly, $E[Y_{\bar{a}}(t+1) - Y_{\bar{0}}(t+1)] = E[Y_{\bar{a}}(t+1)] - E[Y_{\bar{0}}(t+1)]$ is the average causal effect of regime $\bar{a}$ in the population.

We say that there is no confounding when for all $\bar{a}$

$$Y_{\bar{a}}(t+1) \amalg \bar{A}(t) \quad t = 0, \ldots, 15$$

that is, when prior treatment history is independent of the counterfactual outcome. We say

that there is no unmeasured confounding when for all $\bar{a}$ and $t \geqslant k$

$$Y_{\bar{a}}(t+1) \amalg A(k) | \bar{A}(k-1), \bar{L}(k) \qquad \text{Assumption (1)}$$

This assumption will be true if all prognostic factors for $Y_{\bar{a}}(t+1)$ that are used by patients and physicians to decide the administration of zidovudine at visit k are recorded in $\bar{L}(k)$ and $\bar{A}(k-1)$. For example, since physicians tend to prescribe zidovudine to subjects experiencing low CD4 count, assumption (1) would be suspect if $\bar{L}(k)$ did not contain history of CD4 count. Robins' MSMs allow one to estimate causal effects when assumption (1) (no unmeasured confounders) holds.

5. A MARGINAL STRUCTURAL MODEL FOR REPEATED MEASURES

Suppose we model the mean of the counterfactual outcomes $Y_{\bar{a}}(t)$ as

$$E[Y_{\bar{a}}(t+1)] = g(\bar{a}(t), \boldsymbol{\beta}) \tag{2a}$$

where g is a known function and $\boldsymbol{\beta}$ is a parameter to be estimated. This is a marginal structural model. It is marginal because it describes the effect of the treatment regimes $\bar{a}$ on the *marginal* distributions of their corresponding counterfactual outcomes $Y_{\bar{a}}$, and structural because models for counterfactual random variables are called *structural* in the social and economic sciences. It is a semi-parametric model because we parameterize the mean but leave the rest of the distribution unspecified. Robins [2, 3] gives an introduction to and a taxonomy for MSMs. We restrict the scope of the present discussion to MSMs for repeated measures data. By assumption, because the future cannot cause the past, $Y_{\bar{a}}(t+1)$ is equal to $Y_{\bar{a}'}(t+1)$ if $\bar{a}$ and $\bar{a}'$ agree prior to time $t+1$. Therefore, $E[Y_{\bar{a}}(t+1)]$ can only depend on $\bar{a}$ through time t.

The MSM we use to analyse the MACS data is the linear mean model

$$g(\bar{a}(t), \boldsymbol{\beta}) = \beta_0 + \beta_1 \operatorname{cum}[\bar{a}(t)] + \beta_2 t \tag{2b}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ and $\operatorname{cum}[\bar{a}(t)] = \sum_{k=0}^{t} a(k)$.

The MSM (2) resembles the standard GEE model (1). The difference between a GEE model and an MSM is that the former models the association between observed treatments and observed outcomes while the latter models the causal relation between treatment regimes and their corresponding potential outcomes. The relationship of MSM (2) to GEE model (1) can be clearly seen by expressing model (1) as

$$E[Y_{\bar{a}}(t+1) | \bar{A} = \bar{a})] = g(\bar{a}(t), \boldsymbol{\gamma}) \tag{3a}$$

where

$$g(\bar{a}(t), \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 \operatorname{cum}[\bar{a}(t)] + \gamma_2 t \tag{3b}$$

(Because we are conditioning on $\bar{A} = \bar{a}$ in equation (3a), it follows that $Y = Y_{\bar{a}}$. Thus, we can substitute Y for $Y_{\bar{a}}$ in (3) and obtain $E[Y(t+1) | \bar{A} = \bar{a})] = g(\bar{a}(t), \boldsymbol{\gamma})$, which is equivalent to (1).) From (3) we see that a regression model is a model for the conditional mean of $Y_{\bar{a}}(t+1)$ given $\bar{A} = \bar{a}$.

As defined above, the average causal effect of regime $\bar{a}$ is $E[Y_{\bar{a}}(t+1)] - E[Y_{\bar{0}}(t+1)]$, which under our MSM (2) is $\beta_1 \operatorname{cum}[\bar{a}(t)]$. Thus, the parameter β_1 from model (2) has a causal

interpretation as the mean change in CD4 count caused by an additional visit of zidovudine therapy.

β_1 is also of important policy interest. To see why, consider a new subject exchangeable with (that is, drawn from the same distribution as) the N study subjects. We must decide which treatment regime $\bar{a}$ to administer to the new subject. We would like to provide the treatment that maximizes the expected CD4 count at each t. That is, we want to find $\bar{a}$ that maximizes $E[Y_{\bar{a}}(t)]$. For example, if the parameter β_1 is negative, we will withhold zidovudine treatment from our subject, since negative β_1 indicates that the expected CD4 count decreases with increasing zidovudine cumulative dose. In contrast to β_1, the parameter γ_1 of our association (GEE regression) model (1) may have no causal interpretation. For example, suppose physicians preferentially started zidovudine treatment on subjects who, as indicated by their prognostic factor history, were doing poorly and that zidovudine has no causal effect on the mean of Y (that is, $\beta_1 = 0$). None the less, the mean of Y will decrease with cumulative zidovudine dose (since patients with poor prognostic factor history, say low previous CD4 count, will have lower CD4 counts and will have received more treatment). Thus γ_1 will be negative. In this setting, we say that the parameter γ_1 of the association model (1) lacks a causal interpretation because it is confounded by the association of the prognostic factors with the subsequent treatment. If we made policy decisions as to the optimal zidovudine dose based on the parameter γ_1 rather than β_1, we may well be doing many of our patients a potentially fatal disservice. For example, γ_1 may be negative even if zidovudine was beneficial and thus β_1 was positive, if the confounding due to physicians preferentially treating subjects with low previous CD4 counts is of greater magnitude than the true beneficial effect of zidovudine on Y as measured by the absolute value of β_1.

The result shown by Robins [2, 3] is that, assuming we have no unmeasured confounders and the positivity assumption

$$f[\bar{a}(k-1), \bar{l}(k)] > 0 \;\Rightarrow\; f[a(k)|\bar{a}(k-1), \bar{l}(k)] > 0 \tag{4}$$

holds, we can obtain an asymptotically unbiased estimate of the causal parameter β_1 of MSM (2) by simply fitting the standard GEE linear model (1) except that each subject is given the time-specific weight

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

where $A(-1) = 0$ for all subjects, and $f[A(k)|\bar{A}(k-1), \bar{L}(k)]$ is, by definition, the conditional probability mass function $f_{A(k)|\bar{A}(k-1),\bar{L}(k)}[a(k) \mid \bar{a}(k-1), \bar{\ell}(k)]$ with $[a(k), \bar{a}(k-1), \bar{\ell}(k)]$ evaluated at the random argument $[A(k), \bar{A}(k-1), \bar{L}(k)]$. Specifically, Robins [2] proved the following theorem:

Theorem 1

Under assumption (1) of no unmeasured confounders and positivity condition (4) $E[Y_{\bar{a}}(t+1)]$ is the unique function $c[\bar{a}(t)]$ of $\bar{a}(t)$ such that $E\{q[\bar{A}(t)]\{Y(t+1) - c[\bar{A}(t)]\}SW(t)\} = 0$ for all functions $q[\bar{A}(t)]$ where the expectation exists.

Informally, the denominator in each term of the weight $SW(t)$ is the probability that a subject received his own observed treatment, $A(k)$, at visit k given his past zidovudine and

covariate history. Informally, the numerator is the probability that a subject received his observed treatment conditional on his past zidovudine history, but not further adjusting for his past covariate history. The true weights SW are unknown but can be estimated from the data, as we explain below. Once the $SW(t)$ are estimated for each subject, standard software packages for GEE linear regression will allow the user to specify the weight $SW(t)$ (for example, option '`weight`' in SAS proc genmod [12]).

Hence, if the vector of covariates recorded in $\bar{L}(t)$ constitutes all relevant time-dependent confounders, then the weighted GEE linear regression estimator of γ_1 will converge to the parameter β_1 that represents the causal effect of zidovudine on the mean of CD4 count [2, 3]. In contrast, when there exists confounding due to $\bar{L}(t)$, the usual unweighted GEE regression estimator will still converge to γ_1, but now γ_1 will have no causal interpretation. The weighted regression estimator, which we will refer to as an inverse-probability of treatment weighted (IPTW) estimator, is an extension to longitudinal causal inference models of estimators proposed by Horvitz and Thompson [18], Kalbfleisch and Lawless [19], Flanders and Greenland [20], Rosenbaum [21] and Robins and Rotnitzky [22] for missing data models.

As discussed above, including the variables $\bar{L}(t)$ as additional regressors in model (1) fails to appropriately adjust for confounding when some variables in $L(t)$ meet conditions A1 and A2. The MSM approach correctly adjusts for the time-dependent covariates $\bar{L}(t)$, not by adding them to the model as regressors, but by using them to calculate subject-specific weights for a weighted GEE analysis.

We can generalize our MSM (2) 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 linear mean model (2) could be modified to

$$E[Y_{\bar{a}}(t+1) \mid V] = g(\bar{a}(t), \boldsymbol{\beta}) = \beta_0 + \beta_1 \,\mathrm{cum}[\bar{a}(t)] + \beta_2 t + \boldsymbol{\beta}_3' V \tag{5}$$

where $\boldsymbol{\beta}_3$ is a parameter vector. An IPTW estimator of the parameter β_1 can be obtained as above except now the GEE linear model includes the variables in V as additional regressors, and $SW(t)$ is redefined to be

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

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

5.1. Details of estimation

The inverse probability of treatment weighted estimator described in the last section is the solution to the weighted GEE estimating equation $Y = (Y_1, \ldots, Y_{16})^{\mathrm{T}}$. The above estimating equation is in fact a member of the general class of inverse probability of treatment weighted estimating equations. We next discuss the estimation of $\boldsymbol{\beta}$ under three different designs: (i) randomization at baseline, (ii) sequential randomization based on time-dependent covariates, (iii) non-randomized assignment, as in an observational study. These settings cover most practical applications.

5.1.1. Randomization at baseline. When individuals are independently randomized at baseline to treatment with $\bar{A}$, with probabilities depending only on the baseline covariates V, the assigned course of treatment $\bar{A}$ is guaranteed to be independent of the potential outcomes $Y_{\bar{a}}(t+1)$,

$t = 0, \ldots, 15$ for all $\bar{a}$, conditional on the baseline covariates V. Under these conditions, $f[A(k)|\bar{A}(k-1), V] = f[A(k)|\bar{A}(k-1), \bar{L}(k)]$ for all k, $SW(t) = 1$ for all person-visits, and the parameter $\boldsymbol{\beta}$ of MSM (5) can be estimated using the usual unweighted GEE regression.

5.1.2. Sequential randomization based on time-dependent covariates. Intermediate between a simple randomized trial and an observational study is a sequentially randomized trial, where treatment at time k is randomly assigned with the randomization probabilities possibly depending on covariate history $\bar{L}(k)$ and past treatment $\bar{A}(k-1)$. Thus, the (randomization) probabilities $f[A(k)|\bar{A}(k-1), \bar{L}(k)]$ are known for all k because they are under the control of the investigator. Assumption (1) automatically holds for sequentially randomized trials.

An IPTW estimator $\boldsymbol{\beta}$ solves

$$\sum_{i=1}^{N} D_i'(\boldsymbol{\beta}) \Sigma_i^{-1} SW_i (Y_i - \mu_i(\boldsymbol{\beta})) = 0 \tag{7}$$

where $Y_i = (Y_i(1), \ldots, Y_i(K+1))'$, $D_i(\boldsymbol{\beta}) = \partial \mu_i(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}$, $\mu_i(\boldsymbol{\beta}) = (g(A_i(0), V_i, \boldsymbol{\beta}), \ldots, g(\bar{A}_i(K), V_i, \boldsymbol{\beta}))'$, $K+1$ is the time when the last measure of Y is obtained, Σ_i is a (possibly estimated) working covariance matrix, and SW_i is the diagonal matrix of weights $SW(t)$, $t = 0, \ldots, K$. To analyse the MACS data, we used the IPTW estimator (7) with $\Sigma_i \equiv \Sigma_i(\sigma) = \sigma^2 I$, where I is the identity matrix of dimension $K+1$. We estimate robust standard errors with a 'sandwich' estimator [10, 11].

The name 'inverse probability of treatment' stems from the denominator of (6). Robins [2, 3] has shown that if we replace SW_i in (7) by the diagonal matrix W_i of terms $W_i(t)$ representing the 'inverse probability of treatment' for subject i through visit t:

$$W_i(t) \equiv \frac{1}{\prod_{k=0}^{t} f[A_i(k)|\bar{A}_i(k-1), \bar{L}_i(k)]}$$

then our estimator of $\boldsymbol{\beta}$ will remain consistent and asymptotically normal. However, when the numerator of (6) assigns uneven probabilities of treatment conditional on V, the weights in W_i will vary much more than those in SW_i. Using such unstable weights can lead to estimators with large variance because they may be dominated by a small number of observations carrying limited information about $\boldsymbol{\beta}$. The weights SW_i are typically more stable than their counterparts in W_i; accordingly, we refer to elements of SW_i as stabilized weights, and to elements of W_i as non-stabilized weights.

The class of IPTW estimating equations for $\boldsymbol{\beta}$, introduced by Robins [2, 3], generalizes to the set of estimating equations of the form

$$\sum_{i=1}^{N} D_i^{+}(\boldsymbol{\beta}) B_i W_i (Y_i - \mu_i(\boldsymbol{\beta})) + \phi(\bar{A}_i(K), \bar{L}_i(K)) = 0$$

where $D_i^{+}(\boldsymbol{\beta})$ is a $\dim(\beta) \times K+1$ matrix whose tth column consists of user-supplied functions of $(\bar{A}_i(t), V_i, \boldsymbol{\beta})$, B_i is a diagonal matrix whose terms $B_i(t)$ are user-supplied functions of $(\bar{A}_i(t), V_i)$, and

$$\phi(\bar{A}(K), \bar{L}(K)) = \sum_{t=0}^{K} \phi_t(\bar{A}(t), \bar{L}(t)) - E[\phi_t(\bar{A}(t), \bar{L}(t))|\bar{A}(t-1), \bar{L}(t)]$$

with the ϕ_t any user-supplied functions. Note in particular that $\phi_t \equiv 0$ is acceptable. Robins [3] shows that $E[D^+(\boldsymbol{\beta})BW(Y-\mu(\boldsymbol{\beta}))|\bar{A}(t),\bar{L}(t)]$ is the most efficient choice for $\phi(\bar{A}(t),\bar{L}(t))$.

5.1.3. Non-randomized assignment as in an observational study. In an observational study, we cannot guarantee assumption (1) no matter how many covariates are represented by $\bar{L}(t)$. The primary goal of epidemiologists conducting an observational study is to collect data on a sufficient number of covariates to ensure that assumption (1) will be at least approximately true. However, in an observational study, whether assumption (1) is true is not subject to empirical test.

Despite these epistemological constraints, causal evidence is often sought from observational data, especially when there are no viable alternatives, as when ethical concerns preclude randomization. We will proceed as if assumption (1) holds and thus treat the observational study as a sequentially randomized study, except that the probabilities $f[a(k)|\bar{a}(k-1),\bar{l}(k)]$ are now unknown and thus must be modelled and estimated from the data.

5.2. Estimation with censored data

Only minor modifications are necessary to accommodate right censoring due to loss-to-follow-up. Let $C(t)=0$ if a subject remains in the study beyond time t, 1 otherwise. Then the outcome $Y(t+1)$ is observed if and only if $C(t)=0$. In the MACS, we censored subjects at their first visit missed. By viewing the censoring process as an additional treatment process, where 'treatment' corresponds to a patient's removal from the study, it is straightforward to show that if censoring is sequentially ignorable, that is, if

$$Y_{\bar{a}}(t+1) \amalg C(t+1)|\bar{C}(t)\equiv 0,\bar{L}(t),\bar{A}(t)$$

and assumption (1) holds with $\bar{C}(t)\equiv 0$ added to the conditioning event, then the discussion in the previous section can be generalized to show that a weighted GEE analysis using the weights $SW(t)\times SW^{\dagger}(t)$ leads to an asymptotically unbiased estimate of the causal parameter $\boldsymbol{\beta}$, where $SW^{\dagger}(t)$ is the random variable corresponding to

$$SW^{\dagger}(t)=\prod_{k=0}^{t}\frac{\Pr[C(k+1)=0|C(k)=0,\bar{A}(k),V]}{\Pr[C(k+1)=0|C(k)=0,\bar{A}(k),\bar{L}(k)]}$$

and we modify our definition of $SW(t)$ to add $C(k)=0$ to the conditioning events both in the numerator and the denominator. In general, the weights $SW(t)\times SW^{\dagger}(t)$ are unknown and have to be estimated from the data.

5.3. Estimation of the weights

We now describe how to estimate the unknown weights $SW(t)$ and $SW^{\dagger}(t)$ in the MACS. We need to estimate the denominator and numerator of $SW(t)$ for each subject and visit. Since any subject starting zidovudine was assumed to remain on it thereafter, we can regard the time to starting zidovudine as a failure time variable and model the probability of starting zidovudine through a pooled logistic model that treats each person-visit as an observation and allows for a time-dependent intercept. As an example, we fit the model

$$\text{logit}\Pr[A(k)=0|\bar{A}(k-1)=0,\bar{L}(k),\bar{C}(k-1)=0]=\alpha_0(k)+\boldsymbol{\alpha}_1'V+\boldsymbol{\alpha}_2'L(k)$$

and obtained estimates for the unknown parameters $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_0(k), \hat{\boldsymbol{\alpha}}_1', \hat{\boldsymbol{\alpha}}_2')$. It is only necessary to fit the model for subjects who had yet to begin zidovudine (that is, the 6775 person-visits in the MACS with $\bar{A}(k-1)=0$). Note this particular model assumes that only the most recent covariate values $L(k)$ are useful for predicting $A(k)$.

The estimated predicted values $\hat{p}_i(k) = \text{expit}[\hat{\alpha}_0(k) + \hat{\boldsymbol{\alpha}}_1' V_i + \hat{\boldsymbol{\alpha}}_2' L_i(k)]$ from this model are the estimated probabilities of subject i not starting zidovudine at visit k given that zidovudine had not been started by visit $k-1$, where $\text{expit}(x) = e^x/(1+e^x)$. Our estimate of the denominator of $SW_i(t)$ for person i at visit t is the product $\bar{\hat{p}}_i(t) = \prod_{u=0}^{t} \hat{p}_i(u)$ if subject i did not start zidovudine up to visit t, and is $\bar{\hat{p}}_i(t) = [1 - \hat{p}_i(k)] \prod_{u=0}^{k-1} \hat{p}_i(u)$ if subject i started zidovudine at visit k for $k \leqslant t$. In calculating $\bar{\hat{p}}_i(t)$ we have used our assumption that no subject stops zidovudine once begun. Similarly, we estimate the numerator of $SW_i(t)$ by fitting the above logistic model except with the covariates $L(k)$ removed from the model.

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

Robins [3] proved our IPTW estimator of $\boldsymbol{\beta}$ will be consistent if the models for treatment initiation and censoring that are 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. Interestingly, when the denominator models are correctly specified, the results of Pierce [23] and Robins *et al.* [24] can be used to show that estimating the weights increases the efficiency of the IPTW estimator. In other words, to increase efficiency one should estimate the weights even when the assignment mechanism is known from the study design, as would be the case in a sequentially randomized trial.

There is one further detail we have yet to discuss. Under the asymptotic theory that motivated our estimation procedure, the IPTW estimator of $\boldsymbol{\beta}$ will be $n^{1/2}$-consistent only if our estimate of $SW(t)$ converges at a rate of $n^{1/4}$ or better [3]. The practical implication of this result is that for our IPTW estimator of $\boldsymbol{\beta}$ 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 $p(k)$ by not fitting a separate intercept $\alpha_0(k)$ for each visit k. Rather, we modelled $\alpha_0(k)$ using a linear term for visit. Intermediate modelling strategies such as using natural cubic splines with five knots yielded similar estimates.

6. A SIMPLE EXAMPLE

It is helpful to work through a simple example with a univariate response Y measured at the end of follow-up. We consider a hypothetical sequentially randomized trial in which $N=200$ patients are randomly assigned at time $t=0$ to zidovudine ($A(0)=1$) with probability $1/2$ and to placebo ($A(0)=0$) otherwise. Patients continue on zidovudine or placebo until their next visit to clinic at time $t=1$, whereupon the attending physician measures CD4 count. Irrespective of previous treatment history, patients with high CD4 count ($L(1)=1$) are again randomized with probability $1/2$ to zidovudine ($A(1)=1$) and to placebo otherwise ($A(1)=0$). Because it is hoped that treatment with zidovudine will improve CD4 count, patients having low CD4 counts at time $t=1$ ($L(1)=0$) are randomized with probabilities $9/10$ and $1/10$ to

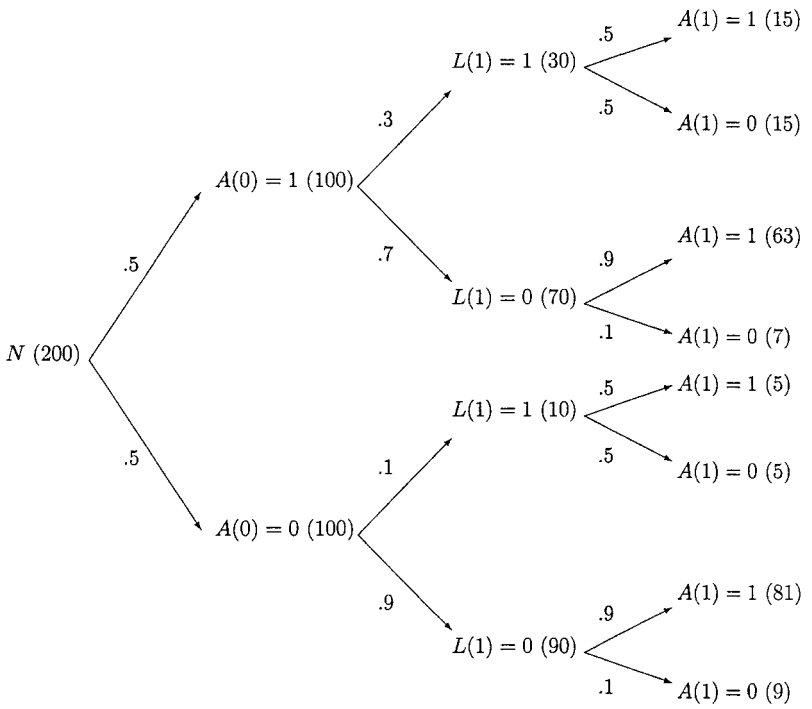

Figure 1. Probability tree diagram for the example of Section 6. Values in parentheses summarize the number of patients at each node. N represents the total number of patients in the study.

Table I. Data corresponding to the example in Section 6*.

$(A(0),L(1),A(1))$	N_O	$E[Y\|\bar{A},\bar{L}]$	W	N_W	$f[A(1)\|A(0)]f[A(0)]$	SW	N_{SW}
(1, 1, 1)	15	100	4	15	0.39	1.56	23.4
(1, 1, 0)	15	100	4	15	0.11	0.44	6.6
(1, 0, 1)	63	90	2.22	35	0.39	0.87	54.6
(1, 0, 0)	7	90	20	35	0.11	2.20	15.4
(0, 1, 1)	5	100	4	5	0.43	1.72	8.6
(0, 1, 0)	5	100	4	5	0.07	0.28	1.4
(0, 0, 1)	81	90	2.22	45	0.43	0.95	77.4
(0, 0, 0)	9	90	20	45	0.07	1.40	12.6

*$W = \dfrac{1}{f[A(1)|L(1),A(0)]f[A(0)]}$, $SW = \dfrac{f[A(1)|A(0)]f[A(0)]}{f[A(1)|L(1),A(0)]f[A(0)]}$

treatment and placebo at time $t = 1$. The probability tree diagram in Figure 1 summarizes the progression of the trial through time $t = 1$.

At time $t = 2$, the patients return to clinic for a final evaluation and their absolute CD4 count Y is measured. Table I summarizes the full data set. The first column defines eight strata of patients according to the observed treatment and covariate histories $(A(0),L(1),A(1))$. Columns 2 and 3 provide, respectively, the number of patients in each stratum and their

average CD4 counts at time 2. In this much simplified data set, the inverse probability of treatment W, given in column 4, can be estimated non-parametrically. As an example, for subjects with $(A(0), L(1), A(1)) = (1, 1, 1)$, we first read from the probability tree that $\Pr[A(1) = 1 | L(1) = 1, A(0) = 1] = 0.5$ and $\Pr[A(0) = 1] = 0.5$. The inverse probability of treatment is then computed as $W = 1/(0.5 \times 0.5) = 4$. Column 5 gives the adjusted population counts when the non-stabilized inverse probabilities W are used as weights. These counts have been renormalized by multiplying each by the same constant so that the total number of patients remains fixed at 200. Columns 6 to 8 present the numerator of the stabilized weights SW, the stabilized weights SW, and the resulting adjusted population counts. Notice that the ratio of the largest to smallest stabilized weight is $2.2/0.28 = 7.86$, whereas that for the non-stabilized weights is $20/2.2 = 9.1$. We will find in the MACS analysis that this difference can be much more pronounced.

Suppose we are interested in estimating the causal effect

$$E[Y_{\bar{a}} - Y_{\bar{a}'}]$$

where $\bar{a}$ represents treatment with $(a(0), a(1)) = (1, 1)$ and $\bar{a}'$ represents treatment with $(a(0), a(1)) = (0, 0)$. The data in the first and third rows of Table I represent subjects treated with zidovudine for the duration of the trial for whom we observe $Y_{\bar{a}}$; the sixth and eighth rows represent observations on $Y_{\bar{a}'}$ in the untreated subjects. To estimate $E[Y_{\bar{a}}]$, we can combine the data from the first and third rows in a variety of ways. A standard approach weights the two averages by the observed N_O, giving

$$(15 \times 100 + 63 \times 90)/(15 + 63) = 91.92$$

The analogous estimate of $E[Y_{\bar{a}'}]$ is

$$(5 \times 100 + 9 \times 90)/(5 + 9) = 93.57$$

and thus the corresponding estimate of our causal effect is $91.92 - 93.57 = -1.65$. However, this result is likely to be biased because $L(1)$ is a confounder for the effect of $A(1)$ on Y.

On the other hand, we can combine rows to obtain a causally valid answer using the MSM approach with either the non-stabilized or the stabilized weights. We see that both methods produce the same result:

$$E[Y_{\bar{a}}] = \frac{15 \times 100 + 35 \times 90}{15 + 35} = \frac{23.4 \times 100 + 54.6 \times 90}{23.4 + 54.6} = 93 \tag{8}$$

$$E[Y_{\bar{a}'}] = \frac{5 \times 100 + 45 \times 90}{5 + 45} = \frac{1.4 \times 100 + 12.6 \times 90}{1.4 + 12.6} = 91 \tag{9}$$

$$\Rightarrow \text{ Causal effect} = 2.0$$

The non-stabilized and stabilized answers are the same because the stabilizing factor $f[A(1)|A(0)]f[A(0)]$ does not depend on $L(1)$; therefore, $(23.4, 54.6)$ is a scalar multiple of $(15, 35)$ and $(1.4, 12.6)$ is a scalar multiple of $(5, 45)$. Note that our IPTW estimate 2.0

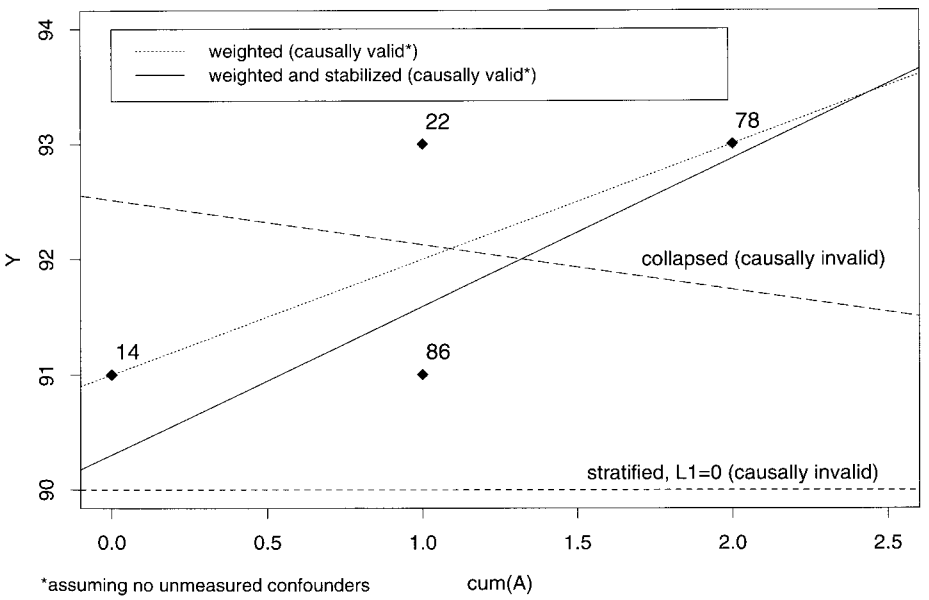

Figure 2. Four regressions.

is the estimate of $\beta_1 + \beta_2 + \beta_3$ in the saturated MSM $E[Y_{\bar{a}}] = \beta_0 + \beta_1 a_0 + \beta_2 a_1 + \beta_3 a_0 a_1$. Indeed, it can be shown that the non-stabilized and stabilized estimates will be the same in any saturated model.

The four consistent estimates for $E[Y_{\bar{a}}]$ from the saturated model for $\bar{a}$ equal to $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ are 91, 91, 93 and 93, respectively. They are classified by $\text{cum}(\bar{A})$ and depicted by diamond-shaped points on the graph in Figure 2. Note in our trial $\text{cum}(\bar{A})$ has three possible values:

$\text{cum}(\bar{A})$	$(A(0), A(1))$
0	$(0,0)$
1	$(0,1)$ or $(1,0)$
2	$(1,1)$

Next to each point is a number representing how many different actual patients contributed information to the estimated causal effect. For instance, the '14' next to the point corresponding to $\text{cum}(\bar{A}) = 0$ represents 5 patients with $(A(0), L(1), A(1)) = (0,1,0)$ and 9 patients with $(A(0), L(1), A(1)) = (0,0,0)$. We next discuss the four regression lines shown in Figure 2.

When the causal effects are summarized by the unsaturated linear MSM model $E[Y_{\bar{a}}] = \beta_0 + \beta_1 \text{cum}(\bar{a})$, the stabilized and non-stabilized weighted estimates of $E[Y_{(1,1)} - Y_{(0,0)}]$ differ, even if the unsaturated model is correctly specified. This difference will only be due to sampling variability. (On the other hand, if model (5) were misspecified, this difference would persist asymptotically and the estimated causal effects would be inconsistent regardless of whether we use stabilized or non-stabilized weights.) Figure 2 graphs results from four regressions of

CD4 count on cumulative exposure cum($\bar{A}$) to zidovudine. They were computed using data from Table I, summarized as follows:

N_O	N_W	N_{SW}	$L(1)$	cum($\bar{A}$)	$E[Y\|\bar{A},\bar{L}]$
63	35	54.6	0	2	90
7	35	15.4	0	1	90
81	45	77.4	0	1	90
9	45	12.6	0	0	90
15	15	23.4	1	2	100
15	15	6.6	1	1	100
5	5	8.6	1	1	100
5	5	1.4	1	0	100

The two causally invalid lines result from unweighted regressions where we either collapse over or stratify by the time-dependent confounder $L(1)$. In both cases, we weight individual subjects equally; this amounts to fitting regression lines to the eight points above with weights determined by the observed N_O and with the response equal to $E[Y|\bar{A}]$ or $E[Y|\bar{A},\bar{L}]$. The collapsed analysis amounts to fitting the regression

$$E[Y|\text{cum}(\bar{A})] = \gamma_0 + \gamma_1 \text{cum}(\bar{A}) \tag{10}$$

which neglects to include, that is, collapses over, $L(1)$. The stratified analysis includes $L(1)$ as a covariate:

$$E[Y|\text{cum}(\bar{A}), L(1)] = \gamma_0 + \gamma_1 \text{cum}(\bar{A}) + \gamma_2 L(1)$$

In the table above, it is readily observed that the least-squares estimates for the parameters γ_0, γ_1 and γ_2 of the stratified analysis will be 90, 0 and 10, respectively. The horizontal line labelled 'stratified' in Figure 2 is the regression line for subjects with $L(1) = 0$ and hence has an intercept of 90.

We next investigate the difference between the non-stabilized and stabilized MSM approaches. Both fit the regression model (10) to the 8 rows of the table above with row specific weights being N_W and N_{SW}, respectively. The non-stabilized MSM regression is algebraically equivalent to the unweighted least-squares fit of (10) to the four diamond-shaped points (0,91), (1,91), (1,93) and (2,93). The stabilized MSM regression is algebraically equivalent to a weighted least-squares fit to the same four points with weights equal to 14, 86, 22 and 78, respectively. Thus, the points $(\text{cum}(\bar{A}), Y) = (1, 91)$ and $(\text{cum}(\bar{A}), Y) = (2, 93)$ have more leverage in the stabilized regression than they do in the non-stabilized analysis. The increased leverage improves efficiency because the two points represent 164 of the 200 patients in the study.

In summary, the simple example demonstrates how the MSM approach can reverse the conclusions of standard regression analyses. Both the stabilized and non-stabilized MSM analyses uncovered a positive effect of zidovudine on CD4 count. On the other hand, the two standard analyses, where we either stratify by or collapse over the time-dependent confounder, falsely suggest a negative or neutral effect.

Table II. Inverse-probability of treatment weighted causal estimates of the causal effect of zidovudine therapy on mean CD4 count in the Multicenter AIDS Cohort Study.

Unweighted estimates*	Parameter	95 per cent CI
Unadjusted	−39.62	−46.69, −32.56
Only baseline covariates	−17.66	−23.44, −11.89
Weighted estimates†	**Parameter**	**95 per cent conservative CI**
Stabilized weights	5.44	−1.77, 12.66
Non-stabilized weights	6.26	−1.39, 13.91

* Non-causal models shown for comparison purposes only. The unadjusted model includes only visit and zidovudine use (yes, no). The model with baseline covariates includes also: age (years), calendar year (1985,1986,1987–89, ⩾1990); CD4; CD8; WBC; RBC; platelets (natural cubic splines); presence of fever; oral candidiasis; diarrhoea; weight loss; oral hairy leukoplakia; herpes zoster (yes, no); zidovudine use, and prophylaxis use.
† Weights estimated as described in the text using data on baseline covariates plus most recent CD4, CD8, WBC, RBC and platelets (splines), and presence of symptoms.

7. DATA ANALYSIS OF THE MACS

Using the unweighted GEE model (1) with CD4 count at visit $t+1$ as the outcome and no covariates other than time t, the estimate of the parameter γ_1 for cumulative zidovudine use was −39.6 (95 per cent confidence interval (CI) −46.7, −32.6). When adding the baseline covariates V to the model, the regression coefficient was −17.7 (95 per cent CI −23.4, −11.9). These confidence intervals are obtained using the robust standard error with $\sigma^2 I$ as the working covariance matrix.

To adjust for time-dependent confounding due to the time-dependent covariates $L(t)$, we estimated the parameters of MSM (5) by calculating a stabilized weight $SW(t) \times SW^{\dagger}(t)$ for each person-visit and then fitting a weighted GEE regression model using $\sigma^2 I$ as the working covariance matrix. Our IPTW causal estimate of the parameter β_1 of MSM (5) was 5.4 (95 per cent conservative CI −1.8, 12.7). The use of non-stabilized weights yielded a similar estimate of β_1 with a slightly wider 95 per cent CI (Table II). The point estimates and robust standard errors for each of the parameters of our MSM are shown in Table III.

The robust standard error is outputted by any standard GEE software that uses the 'sandwich' estimator [10, 11] such as SAS proc genmod [12]. Because this estimator does not take into account the estimation of the weights, it yields an asymptotically conservative confidence interval for our IPTW estimator. Thus, we also estimated this standard error using 1000 bootstrap samples. However, the results were virtually identical: the bootstrap mean estimate of β_1 was 5.4, and its 95 per cent CI was $(-1.9, 12.8)$.

The weights $SW(t) \times SW^{\dagger}(t)$ were estimated by means of four pooled logistic regression models as described in Section 5.3. In two of the models the outcome was 'initiation of zidovudine'. Using the estimated predicted values from each of these models, we calculated two quantities for each person-visit: the probability of each person having his own observed zidovudine history up to visit t given baseline covariates V, and, then, given also the time-varying covariates $L(t)$. Similar models were fit for the outcome 'censoring', after adding zidovudine history as a time-dependent dichotomous variable indicating whether the subject had started zidovudine by visit $t-1$.

Table III. Inverse-probability of treatment weighted estimates of the parameters of the marginal structural model for repeated measures, Multicenter AIDS Cohort Study.

Variable*	Parameter estimate	Robust stand. error	95 per cent conservative CI
Zidovudine	5.44	3.68	−1.77, 12.66
Age	−0.69	1.00	−2.65, 1.26
Year:			
1985	85.12	41.20	4.36,165.87
1986	61.22	43.48	−23.99, 146.4
1987–89	25.71	46.55	−65.53, 116.9
⩾ 1990	0.00		
CD4 (/μl)	0.63	0.18	0.27,0.99
	3.05	3.56	−3.93, 10.02
	−5.06	5.29	−15.43, 5.30
CD8 (/μl)	−0.27	0.11	−0.48, −0.06
	0.99	1.03	−1.02, 3.00
	−1.30	1.65	−4.53, 1.94
WBC (/μl)	0.09	0.04	0.02,0.16
	−1.20	0.66	−2.49, 0.10
	0.38	1.79	−0.20, 3.58
RBC ($\times 10^5$/μl)	−10.77	18.20	−46.44, 24.91
	43.78	64.71	−83.06, 170.6
	−67.94	106.21	−276.1, 140.2
Platelets ($\times 10^3$/μl)	−1.31	0.55	−2.39, −0.22
	9.57	2.55	4.57,14.57
	−19.10	5.05	−29.01, −9.19
Thrush	−59.70	21.11	−101.1, −18.3
Oral leukoplakia	−69.86	44.77	−157.6, 17.89
Weight loss	25.06	38.62	−50.64, 100.8
Herpes zoster	−53.51	33.51	−119.2, 12.17
Diarrhoea	−47.16	50.57	−146.3, 51.95
Fever	0.61	71.28	−139.1, 140.3
Visit	−20.44	2.26	−24.87, −16.02

*All variables measured at baseline except zidovudine (1=ever user, 0 = never user) and visit. The three parameters for CD4, CD8, WBC, RBC and platelets correspond to a linear term (first row) plus two parameters for natural cubic splines with knots at percentiles 1, 27.5, 50, 72.5 and 99 as estimated from SAS macro RCSPLINE by Frank Harrell (http://jse.stat.ncsu.edu:70/ls/software/sas).

Table IV summarizes the empirical distribution at two arbitrary time points (visits 2 and 10) of (a) the numerator and denominator of the stabilized treatment weights $\widehat{SW}(t)$, and (b) the numerator and denominator of the stabilized censoring weights $\widehat{SW}^{\dagger}(t)$. Table V summarizes the empirical distribution at visits 2 and 10 of the stabilized treatment weights $\widehat{SW}(t)$, the non-stabilized treatment weights $\hat{W}(t)$, the stabilized censoring weights $\widehat{SW}^{\dagger}(t)$ and the non-stabilized censoring weights $\hat{W}^{\dagger}(t)$. The non-stabilized weights are a fraction whose denominator is equal to that of the stabilized weights but whose numerator is 1.

The distribution of the stabilized weights $SW(t) \times SW^{\dagger}(t)$ and non-stabilized weights $W(t) \times W^{\dagger}(t)$ is presented in Figures 3 and 4, respectively (a logarithmic transformation was applied for display purposes only). Reading from Table V and Figures 3 and 4, we see that the distribution of $\hat{W}(t)$ is much more variable and skewed than that of $\widehat{SW}(t)$.

Table IV. Distribution of the estimated numerator and denominator of the weights SW and $SW^{\dagger}$ at visits 2 and 10 in the Multicenter AIDS Cohort Study.

	Mean (SD)*	Median (IQR)*	Percentile 1	Percentile 99
Visit 2 ($n = 1186$)				
Probability of having observed zidovudine history:				
given baseline covariates†	0.775(0.242)	0.859(0.136)	0.053	0.978
given time-varying covariates‡	0.808(0.246)	0.907(0.159)	0.006	0.994
Probability of being uncensored:				
given baseline covariates†	0.685(0.096)	0.697(0.124)	0.408	0.858
given time-varying covariates‡	0.691(0.102)	0.704(0.134)	0.383	0.867
Visit 10 ($n = 439$)				
Probability of having observed zidovudine history:				
given baseline covariates†	0.224(0.237)	0.082(0.305)	0.030	0.833
given time-varying covariates‡	0.298(0.292)	0.151(0.463)	0.011	0.947
Probability of being uncensored:				
given baseline covariates†	0.398(0.122)	0.402(0.173)	0.096	0.661
given time-varying covariates‡	0.406(0.125)	0.412(0.186)	0.116	0.705

*SD = standard deviation, IQR = interquartile range.
†Age (years), calendar year (1985,1986,1987–89, $\geqslant$1990); CD4; CD8; WBC; RBC; platelets (natural cubic splines); presence of fever; oral candidiasis; diarrhoea; weight loss; oral hairy leukoplakia; herpes zoster (yes, no); zidovudine use, and prophylaxis use.
‡Baseline covariates plus most recent CD4, CD8, WBC, RBC and platelets (splines), and presence of symptoms.

Table V. Distribution of the estimated weights SW and $SW^{\dagger}$ at visits 2 and 10 in the MACS.

	Mean (SD)*	Median (IQR)*	Percentile 1	Percentile 99
Visit 2 ($n = 1186$)				
SW	1.00(0.60)	0.96(0.07)	0.37	3.69
W	2.49(8.34)	1.10(0.21)	1.01	30.06
$SW^{\dagger}$	0.99(0.04)	0.99(0.04)	0.91	1.14
$W^{\dagger}$	1.49(0.28)	1.42(0.27)	1.06	2.61
Visit 10 ($n = 439$)				
SW	1.02(0.79)	0.79(0.62)	0.16	5.36
W	11.90(20.14)	6.62(12.58)	1.06	90.97
$SW^{\dagger}$	0.98(0.11)	0.98(0.10)	0.72	1.32
$W^{\dagger}$	2.82(1.36)	2.43(1.18)	1.48	8.63

*SD = standard deviation, IQR = interquartile range.

The difference between the unweighted and weighted estimates is an indication of the amount of time-dependent confounding due to the time-dependent covariates $\bar{L}(t)$, which is considerable in the MACS. One might be tempted to adjust for this time-dependent confounding by adding, for example, the variables in $L(t)$ to the unweighted GEE model (1). The point

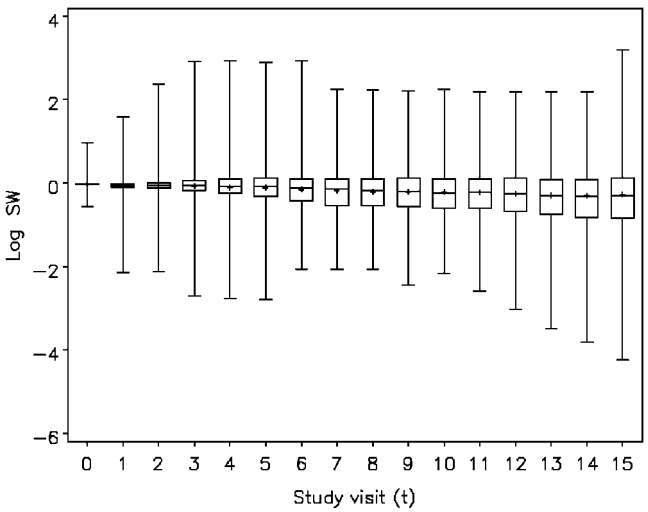

Figure 3. Distribution of stabilized weights at each study visit. The box for each visit shows the location of the mean (+), median (middle horizontal bar) and second and third quartiles (border horizontal bars). Vertical lines extend to the maximum and minimum values.

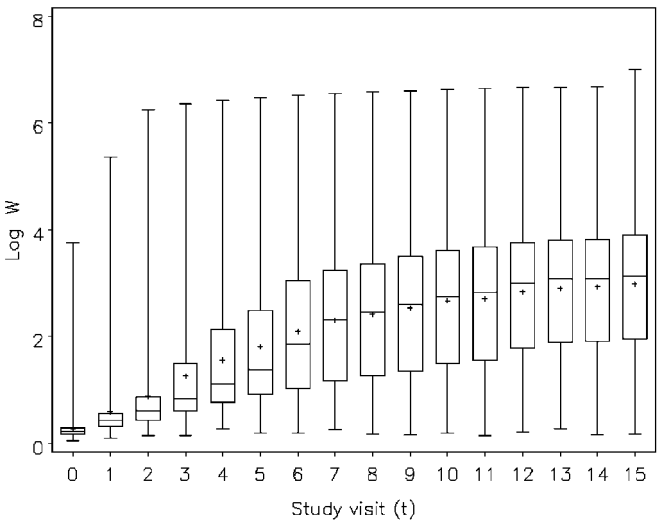

Figure 4. Distribution of non-stabilized weights at each study visit.

estimate of the cumulative zidovudine coefficient from this model was -1.40 (95 per cent CI $-3.11, 0.31$). However, because the covariates in $L(t)$ are affected by earlier treatment, this estimate cannot be causally interpreted as the overall zidovudine effect. It also cannot be interpreted as the direct effect of zidovudine mediated by pathways not through the covariates $L(t)$. Since we have only made the assumption of no unmeasured confounders for the treatment process $A(t)$ and not for the covariate process $L(t)$.

8. DISCUSSION

We have used a marginal structural linear model for repeated measures to estimate the causal effect of zidovudine on CD4 count in the MACS. This method was utilized because standard statistical methods are not appropriate when there exists time-dependent confounding by variables, such as CD4 count, that are affected by previous treatment.

Because of the presence of confounding, the crude GEE estimate for cumulative zidovudine was −39.6 (95 per cent CI −46.7, −32.6), erroneously suggesting a marked decline of CD4 count among zidovudine users. The regression coefficient estimated by the (unweighted) standard model that included only baseline covariates, and that therefore does not adjust for time-dependent confounding, was −17.7 (95 per cent CI −23.4, −11.9), which still suggests a detrimental effect of zidovudine.

In fact, the estimated coefficient β_1 for zidovudine was 5.4 (95 per cent conservative CI −1.8, 12.7) in the weighted analysis that provides, under our assumptions, an unbiased estimate of the expected change in mean CD4 attributable to an additional visit of zidovudine use under the marginal structural model (4). The weighted analysis appropriately adjusts for time-dependent confounders affected by earlier treatment. This result is consistent with those from placebo-controlled randomized trials [25–27] and suggests that, on average, zidovudine users may perhaps undergo a slightly slower decline of their CD4 count as compared to non-users.

MSMs are useful for estimating the effect of prespecified treatment regimes $\bar{a}$ (for example, 'start treatment at visit 3', or 'never start treatment') in the whole population and within levels of the baseline covariates V. However, MSMs are much less useful for estimating the causal effect of dynamic treatment regimes in which treatment in a given visit is decided in part based on a subject's evolving covariate history (for example, 'start treatment when CD4 drops below 200'). To estimate the causal effect of dynamic regimes, which involve interactions between treatment and the time-dependent covariates $L(t)$, structural nested models (SNMs) can be used [3]. An attractive feature of MSMs, compared to SNMs, is their close resemblance to conventional regression models, which renders them both familiar and easy to implement.

The correctness of our causal inferences is dependent on a number of assumptions. First, we assume that the information on visit of zidovudine initiation is accurate. Second, we assume that the measured covariates in $\bar{L}(t)$ are sufficient to adjust for both confounding and for selection bias due to loss to follow-up. This implies that we have available, for each visit t, accurate data recorded in all time-dependent covariates $L(t)$ necessary for the assumption of no unmeasured confounders (assumption (1)) to hold. Unfortunately, as in all observational studies, these two assumptions cannot be tested from the data. In our analysis we assume this goal has been realized, while recognizing that, in practice, this would never be precisely or sometimes even approximately true. Recently, Robins and co-workers have developed extensions of IPTW estimation of MSMs that allow one to evaluate the sensitivity of one's estimates to increasing violation of these fundamental assumptions [28]. Third, we assume that the models for initiation of zidovudine and censoring, given the past, are correctly specified. Fourth, we assume that our MSM for the effect of zidovudine on mean CD4 count, within levels of baseline covariates V, is correctly specified.

A similar set of assumptions (accurate information, no unmeasured confounders, non-informative censoring and no model misspecification) is required to give a causal interpretation to the parameters of standard statistical models used in point-treatment studies.

Furthermore, when studying the effect of a time-dependent treatment like zidovudine, the assumptions of MSMs are less restrictive than those of standard methods: MSMs do not require the absence of time-dependent confounding by variables affected by previous exposure.

ACKNOWLEDGEMENTS

We thank Els Goetghebeur and Donna Pauler for helpful discussions and comments on the manuscript.

REFERENCES

1. 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. *American Journal of Epidemiology* 1987; **126**:310–318.
2. Robins JM. Marginal structural models. In *1997 Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association: Alexandria, VA, 1998; 1–10.
3. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology*: *The Environment and Clinical Trials*, Halloran E, Berry D (eds). Springer-Verlag: New York, 1999; 95–134.
4. Robins JM. Correction for non-compliance in equivalence trials. *Statistics in Medicine* 1998; **17**:269–302.
5. Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *NCHSR*, *Health Services Research Methodology*: *A Focus on AIDS*, Sechrest L, Freeman H, Mulley A (eds). U.S. Public Health Service, 1989; 113–159.
6. Robins JM. Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics 120*, Berkane M (ed.). Springer-Verlag: New York, 1997; 69–117.
7. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period application to the healthy worker survivor effect. *Mathematical Modelling* 1986; **7**:1393–1512 (errata 1987; **14**:917–921).
8. Graham NMH, Zeger SL, Park LP, Vermund SH, Detels R, Rinaldo CR, Phair JP. The effects on survival of early treatment of human immunodeficiency virus infection. *New England Journal of Medicine* 1992; **326**: 1037–1042.
9. Kinloch-de Loes S, Hirschel BJ, Hoen B, Cooper DA, Tindall B, Carr A, Saurat JH, Clumeck N, Lazzarin A, Mathiesen L, Francois R, Antunes F, von Overbeck J, Luthy R, Glauser M, Hawkins D, Baumberger C, Yerly S, Perneger TV, Perrin L. A controlled trial of zidovudine in primary human immunodeficiency virus infection. *New England Journal of Medicine* 1995; **333**:408–413.
10. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
11. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**: 121–130.
12. SAS Institute Inc. *SAS/STAT User's Guide*, *Version 8*. SAS Institute Inc.: Cary, NC, 2000.
13. Robins JM, Greenland S. Comment on 'Causality without counterfactuals'. *Journal of the American Statistical Association* 2000; **95**:431–435.
14. Neyman J. On the application of probability theory to agricultural experiments: essay on principles, Section 9. (1923), translated in *Statistical Science* 1990; **5**:465–480.
15. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics* 1978; **6**: 34–58.
16. Holland PW. Statistics and causal inference (with discussion). *Journal of the American Statistical Association* 1986; **81**:945–970.
17. Robins JM. Addendum to 'A new approach to causal inference in mortality studies with a sustained exposure period application to the healthy worker survivor effect'. *Computers and Mathematics with Applications* 1987; **14**:923–945 (errata 1989; **18**:477).
18. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1952; **47**:663–685.
19. Kalbfleisch JD, Lawless JF. Likelihood analysis of multistage models for disease incidence and mortality. *Statistics in Medicine* 1988; **7**:149–160.
20. Flanders WD, Greenland S. Analytic methods for two-stage case-control studies and other stratified designs. *Statistics in Medicine* 1991; **10**:739–747.
21. Rosenbaum PR. Model-based direct adjustment. *Journal of the American Statistical Association* 1987; **82**: 387–394.

22. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology – Methodological Issues*, Jewell N, Dietz K, Farewell V (eds). Birkhäuser: Boston, MA, 1992; 297–331.
23. Pierce DA. The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Annals of Statistics* 1982; **10**:475–478.
24. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994; **89**:846–866.
25. Volberding PA, Lagakos SW, Koch MA, Pettinelli C, Myers MW, Booth DK, Balfour HH Jr, Reichman RC, Bartlett JA, Hirsch MS, Murphy RL, Hardy WD, Soeiro R, Fischl MA, Bartlett JG, Merigan TC, Hyslop NE, Richman DD, Valentine FT, Corey L, the AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. Zidovudine in asymptomatic human immunodeficiency virus infection. *New England Journal of Medicine* 1990; **322**:941–949.
26. Fischl MA, Richman DD, Hansen N, Collier AC, Carey JT, Para MF, Hardy WD, Dolin R, Powderly WG, Allan JD, Wong B, Merigan TC, McAuliffe VJ, Hyslop NE, Rhame FS, Balfour HH, Spector SA, Volberding P, Petinelli C, Anderson J, the AIDS Clinical Trials Group. The safety and efficacy of zidovudine (AZT) in the treatment of subjects with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection. *Annals of Internal Medicine* 1990; **112**:727–737.
27. Merigan TC, Amato DA, Balsley J, Power M, Price WA, Benoit S, Perez-Michael A, Brownstein A, Kramer AS, Brettler D, Aledort L, Ragni MV, Andes WA, Gill JC, Goldsmith J, Stabler S, Sanders N, Gjerset G, Lusher J, the NHF-ACTG 036 Study Group. Placebo-controlled trial to evaluate zidovudine in treatment of human immunodeficiency virus infection in asymptomatic patients with hemophilia. *Blood* 1991; **78**:900–906.
28. Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Methods in Epidemiology*: *The Environment and Clinical Trials*, Halloran, E, Berry D (eds). Springer Verlag: New York, 1999; 1–92.

