Marginal Structural Models and Causal Inference in Epidemiology

James M. Robins,[1,2] *Miguel Ángel Hernán*,[1] *and Babette Brumback*[2]

In observational studies with exposures or treatments that vary over time, standard approaches for adjustment of confounding are biased when there exist time-dependent confounders that are also affected by previous treatment. This paper introduces marginal structural models, a new class of causal models that allow for improved adjustment of confounding in those situations. The parameters of a marginal structural model can be consistently estimated using a new class of estimators, the inverse-probability-of-treatment weighted estimators. (Epidemiology 2000;11:550–560)

Keywords: causality, counterfactuals, epidemiologic methods, longitudinal data, structural models, confounding, intermediate variables

Marginal structural models (MSMs) are a new class of causal models for the estimation, from observational data, of the causal effect of a time-dependent exposure in the presence of time-dependent covariates that may be simultaneously confounders and intermediate variables.[1–3] The parameters of a MSM can be consistently estimated using a new class of estimators: the inverse-probability-of-treatment weighted (IPTW) estimators. MSMs are an alternative to structural nested models (SNMs), the parameters of which are estimated through the method of g-estimation.[4–6]

The usual approach to the estimation of the effect of a time-varying exposure or treatment has been to model the probability of disease as a function of past exposure and past confounder history, using analytic methods such as stratified analysis and its parametric analogs (for example, logistic or proportional hazards regression). We will show in sections 4 and 7.1 that these standard approaches may be biased, whether or not one further adjusts for past confounder history in the analysis, when (1) there exists a time-dependent covariate that is a risk factor for, or predictor of, the event of interest and also predicts subsequent exposure, and (2) past exposure history predicts subsequent level of the covariate. We refer to covariates satisfying condition 1 as time-dependent confounders. Conditions 1 and 2 will be true in many observational studies, particularly those in which there is confounding by indication. For example, in a study of the effect of zidovudine (AZT) treatment on mortality among human immunodeficiency virus (HIV)-infected subjects, the time-dependent covariate CD4 lymphocyte count is both an independent predictor of survival and initiation of therapy with AZT and is itself influenced by prior AZT treatment. In a study of the effect of obesity on mortality, the development of clinical cardiac or respiratory disease is an independent predictor of both mortality and subsequent weight loss and is influenced by prior weight gain. Conditions 1 and 2 will always hold when there are time-dependent covariates that are simultaneously confounders and intermediate variables. A more detailed description of the bias of standard methods, as well as several additional epidemiologic examples of time-dependent confounding, has been presented elsewhere.[5,7]

1. Time-Dependent Confounding

Consider a follow-up study of HIV-infected patients. Let A_k be the dose of the treatment or exposure of interest, say AZT, on the k^{th} day since start of follow-up. Let Y be a dichotomous outcome of interest (for example, $Y = 1$ if HIV RNA is not detectable in the blood and is 0 otherwise) measured at end of follow-up on day $K + 1$. Our goal is to estimate the causal effect of the time-dependent treatment A_k on the outcome Y.

Figure 1 is a causal graph that represents our study with $K = 1$. A causal graph is a directed acyclic graph in which the vertices (nodes) of the graph represent variables and the directed edges (arrows) represent direct causal effects.[8] In Figure 1, L_k represents the value on day k of the vector of all measured risk factors for the outcome, such as age, CD4 lymphocyte count, white blood count (WBC), hematocrit, diagnosis of acquired immunodeficiency syndrome (AIDS), and the presence or absence of various symptoms or opportunistic infections such as oral candidiasis. Similarly, U_k represents the value on day k of all unmeasured causal risk factors for Y. Figure 1, *b*, differs from

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

Address correspondence to: James M. Robins, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115.

This research was supported by NIH grant R01-A132475.

Submitted August 4, 1999; final version accepted February 28, 2000.

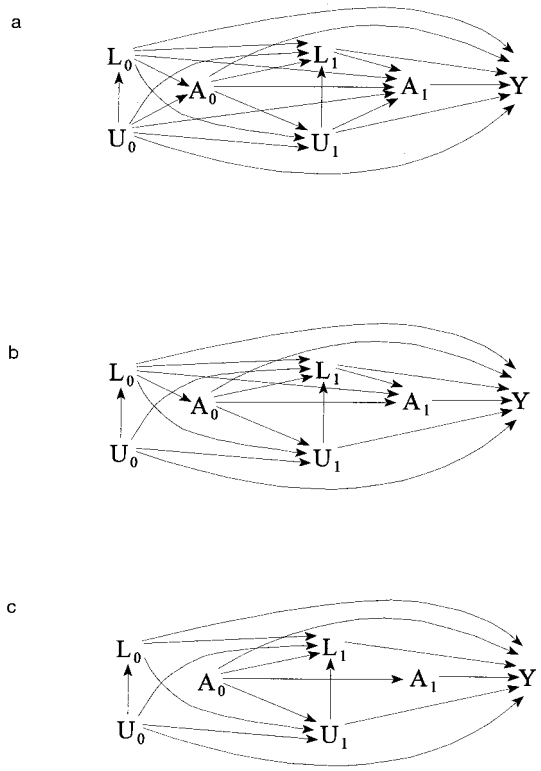

FIGURE 1. Causal graphs for a time-dependent exposure.

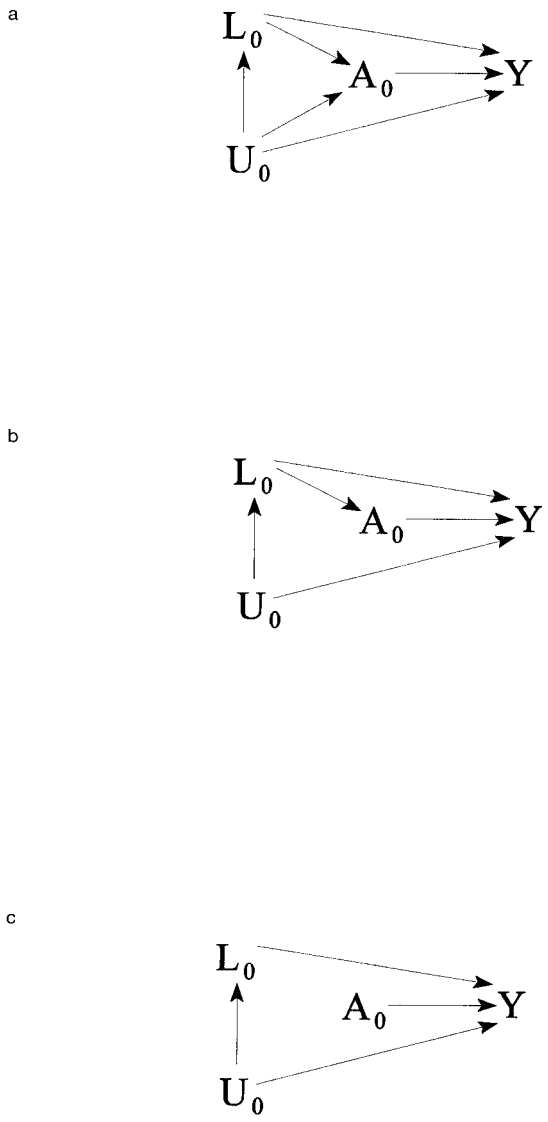

FIGURE 2. Causal graphs for a point exposure A_0.

Figure 1, *a*, only in that the arrows from the unmeasured causal risk factors into the treatment variables have been removed. When, as in Figure 1, *b*, there is no arrow from unmeasured causal risk factors into treatment variables, we say that there are no unmeasured confounders given data on the measured confounders L_k.[9,10] Figure 1, *c*, differs from Figure 1, *a* and *b*, in that none of the risk factors for *Y* (measured or unmeasured) has arrows into any treatment variable. Note, however, that earlier treatment A_0 can causally affect later treatment A_1. When, as in Figure 1, *c*, there is no arrow from any (nontreatment) risk factor into any treatment variable, there is no confounding by either measured or unmeasured factors, in which case we say that treatment is unconfounded.[9,10]

The distinctions drawn above apply equally to more familiar point-treatment studies in which the treatment is not time-dependent. As indicated in Figure 2, a point-treatment study is a special case of the general set-up in which $K = 0$. Figure 2, *a–c*, contains the analogs of Figure 1, *a–c*, for a point-treatment study.

As in any observational study, we cannot determine from the observed data on the L_k, A_k, and *Y* whether there is confounding by unmeasured risk factors. We can only hope that whatever residual confounding there may be due to the U_k is small. Under the untestable assumption that there is no unmeasured confounding given the L_k, we can, however, empirically test from the data whether treatment is unconfounded. Specifically, a sufficient condition for treatment to be unconfounded is that, at each time *k*, among subjects with the same past treatment history $A_0, \ldots, A_{k-1}$, the treatment A_k is unassociated with the past history of measured covariates $L_0, \ldots, L_k$.[9–11] For example, in our point-treatment study, treatment will be unconfounded if A_0 is unassociated with L_0.

2. Counterfactuals in Point-Treatment Studies

We begin with a review of how one would estimate the effect of A_0 on *Y* in the point-treatment study of Figure 2. Suppose treatment A_0 is dichotomous; suppose further that Figure 2, *c*, is the true causal graph, that is, that neither measured nor unmeasured covariates confound the relation between treatment and the outcome. Then the crude risk difference, risk ratio, and odds ratio each measure the causal effect of the treatment A_0 on the outcome *Y*, although on different scales. The crude risk difference is $cRD = pr[Y = 1|A_0 = 1] - pr[Y = 1|A_0 = 0]$, the crude risk ratio is $cRR = pr[Y = 1|A_0 = 1]/pr[Y = 1|A_0 = 0]$, the crude odds ratio is

$cOR = pr[Y = 1|A_0 = 1]pr[Y = 0|A_0 = 0]/\{pr[Y = 1|A_0 = 0]pr[Y = 0|A_0 = 1]\}$, and, for example, $pr[Y = 1|A_0 = 1]$ is the probability that $Y = 1$ among treated subjects ($A_0 = 1$). We assume that the study subjects are a random sample from a large, possibly hypothetical, source population. Probabilities refer to proportions in the source population.

The causal contrasts that correspond to these associational parameters involve counterfactual variables. Specifically, the variable $Y_{a_0=1}$ denotes a subject's outcome if treated and $Y_{a_0=0}$ denote a subject's outcome if left untreated. For a given subject, the causal effect of treatment, measured on a difference scale, is $Y_{a_0=1} - Y_{a_0=0}$. For no subject are both $Y_{a_0=0}$ and $Y_{a_0=1}$ observed. If a subject is treated ($A_0 = 1$), the subject's observed outcome Y equals $Y_{a_0=1}$, and $Y_{a_0=0}$ is unobserved. If $A_0 = 0$, Y equals $Y_{a_0=0}$, and $Y_{a_0=1}$ is unobserved. Let $pr(Y_{a_0=1} = 1)$ and $pr(Y_{a_0=0} = 1)$, respectively, be the probability that $Y_{a_0=1}$ is equal to 1 and $Y_{a_0=0}$ is equal to 1. Then, if treatment A_0 is unconfounded, the crude *RD* equals the causal risk difference $pr[Y_{a_0=1} = 1] - pr[Y_{a_0=0} = 1]$ in the source population. The causal risk difference is the average of the individual causal risk differences $Y_{a_0=1} - Y_{a_0=0}$. Similarly, the crude *RR* equals the causal *RR*, $pr(Y_{a_0=1} = 1)/pr(Y_{a_0=0} = 1)$, and the crude *OR* equals the causal *OR*, $pr(Y_{a_0=1} = 1)pr(Y_{a_0=0} = 0)/\{pr(Y_{a_0=1} = 0)pr(Y_{a_0=0} = 1)\}$. Because of the possibility of effect modification, the population causal parameter need not equal the causal parameter within a stratum of the measured risk factors L_0 even if treatment is unconfounded. Effect modification is considered in section 9.

3. Models for Point-Treatment Studies

The causal *RD*, *RR*, and *OR* can also be expressed in terms of the parameters of the following linear, log linear, and linear logistic models for the two counterfactual probabilities $pr(Y_{a_0=1} = 1)$ and $pr(Y_{a_0=0} = 1)$.

$$pr[Y_{a_0} = 1] = \psi_0 + \psi_1 a_0 \quad (1)$$

$$\log pr[Y_{a_0} = 1] = \theta_0 + \theta_1 a_0 \quad (2)$$

$$logit\ pr[Y_{a_0} = 1] = \beta_0 + \beta_1 a_0 \quad (3)$$

where Y_{a_0} is $Y_{a_0=1}$ if $a_0 = 1$ and Y_{a_0} is $Y_{a_0=0}$ if $a_0 = 0$. Specifically, the causal $RD = \psi_1$, the causal $RR = e^{\theta_1}$, and the causal $OR = e^{\beta_1}$. Models 1–3 are saturated MSMs. They are *marginal* models, because they model the marginal distribution of the counterfactual random variables $Y_{a_0=1}$ and $Y_{a_0=0}$ rather than the joint distribution (that is, models 1–3 do not model the correlation of $Y_{a_0=1}$ and $Y_{a_0=0}$). They are *structural* models, because they model the probabilities of counterfactual variables and in the econometric and social science literature models for counterfactual variables are often referred to as structural.[8,12] Finally, they are *saturated*, because each has two unknown parameters and thus each model places no restriction on the possible values of the two unknown probabilities $pr(Y_{a_0=1} = 1)$ and $pr(Y_{a_0=0} = 1)$. Note that these models do not include covariates, because they are, by definition, models for causal effects on the entire source population; they are not models for observed associations.

The crude *RD*, *RR*, and *OR* can also be expressed in terms of the parameters of the following saturated linear, log linear, and linear logistic models for the observed outcome Y.

$$pr[Y = 1|A_0 = a_0] = \psi_0' + \psi_1' a_0 \quad (4)$$

$$\log pr[Y = 1|A_0 = a_0] = \theta_0' + \theta_1' a_0 \quad (5)$$

$$logit\ pr[Y = 1|A_0 = a_0] = \beta_0' + \beta_1' a_0. \quad (6)$$

These are models for associations observed when comparing subpopulations (defined by levels of treatment) of the source population. The crude *RD* equals ψ_1', the crude *RR* equals $e^{\theta_1'}$, and the crude *OR* equals $e^{\beta_1'}$. The parameters of the associational models 4–6 will differ from the parameters of the MSMs 1–3, except when treatment is unconfounded. Because models 4–6 are models for the observed data, (asymptotically) unbiased estimates of the model parameters can be obtained using standard statistical software (assuming no selection bias or measurement error). When treatment is unconfounded, these same estimates will also be unbiased for the corresponding causal parameters of models 1–3. For example, to fit models 4–6, one could use the general-purpose SAS program Proc Genmod, using the model statement $Y = A_0$ with the outcome Y specified as a binomial variable. To estimate ψ_1', one would specify the identity link; to estimate θ_1', the log link; and for β_1', the *logit* link. Programs analogous to Proc Genmod also exist in other packages, such as S-Plus, Gauss, and Stata.[13–16]

4. No Unmeasured Confounders

Suppose now that treatment is confounded. Then the crude association parameter will not equal the corresponding causal parameter. Similarly, the parameters of the MSMs will fail to equal the parameters of the corresponding observed data models (for example, $\beta_0 \neq \beta_0'$ and $\beta_1 \neq \beta_1'$). Assuming we have no unmeasured confounders given data on measured confounders L_0, unbiased estimates of the causal parameters ψ_1, θ_1, and β_1 can, however, still be obtained using Proc Genmod by performing a weighted analysis. Specifically, using the weight statement (that is, option SCWGT) in Proc Genmod, each subject i is assigned a weight w_i equal to the inverse of the conditional probability of receiving his or her own treatment. That is, $w_i = 1/pr[A_0 = a_{0i}|L_0 = l_{0i}]$, where, for example, l_{0i} is the observed value of the variable L_0 for subject i. The true weights w_i are unknown but can be estimated from the data in a preliminary logistic regression of A_0 on L_0. For example, we might specify the logistic regression model

$$logit\ pr[A_0 = 1|L_0 = l_0] = \alpha_0 + \alpha_1 l_0 \quad (7)$$

where A_0 is AZT treatment, L_0 is, for example, the column vector of covariates with components age, CD4 count, WBC count, hematocrit, and presence of symptoms, and α_1 is a row vector of unknown parameters. We

can then obtain estimates $(\hat{\alpha}_0, \hat{\alpha}_1)$ of (α_0, α_1) using standard logistic regression software. Then, for a subject i with $A_0 = 0$ and $L_0 = l_0$, we would estimate $w_i = 1/pr[A_0 = 0|L_0 = l_{0i}]$ by $1/\{1/[1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 l_{0i})]\} = 1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 l_{0i})$. For a subject with $A_0 = 1$ and $L_0 = l_0$, we would estimate $w_i = 1/pr[A_0 = 1|L_0 = l_{0i}]$ by $\{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1 l_{0i})\}/\exp(\hat{\alpha}_0 + \hat{\alpha}_1 l_{0i}) = 1 + \exp(-\hat{\alpha}_0 - \hat{\alpha}_1 l_{0i})$.

In summary, if there are no unmeasured confounders given data on L_0, one can control confounding (due to L_0) by modifying the crude analysis by weighting each subject i by w_i. The denominator of w_i is informally the probability that a subject had his or her own observed treatment. Thus, we refer to these weighted estimators as IPTW estimators.

Why does this approach work? The effect of weighting in Proc Genmod is to create a pseudopopulation consisting of w_i copies of each subject i. That is, if, for a given subject, $w_i = 4$, the subject contributes four copies of him- or herself to the pseudopopulation. This new pseudopopulation has the following two important properties. First, in the pseudopopulation, unlike the actual population, A_0 is unconfounded by the measured covariates L_0. Second, $pr(Y_{a_0=1} = 1)$ and $pr(Y_{a_0=0} = 1)$ in the pseudopopulation are the same as in the true study population so that the causal *RD*, *RR*, and *OR* are the same in both populations. Hence, it follows that we can unbiasedly estimate the causal *RD*, *RR*, and *OR* by a standard crude analysis in the pseudopopulation. But this is exactly what our IPTW estimator does, because the weights w_i serve to create, as required, w_i copies of each subject. In the Appendix, we present a detailed numerical example to clarify further our IPTW methodology, and we compare our methodology with the propensity score methodology of Rosenbaum and Rubin.[17]

5. Unmeasured Confounding

In the presence of unmeasured confounding factors U_0, one could still unbiasedly estimate the causal risk difference, risk ratio, and odds ratio as above if one used weights $w_i = 1/pr[A_0 = a_{0i}|L_0 = l_{0i}, U_0 = u_{0i}]$ in implementing the analysis in Proc Genmod. Nevertheless, because data on U_0 are not observed, it is not possible to estimate these weights unbiasedly. Indeed, unbiased estimation by any method is impossible in the presence of unmeasured confounding factors without strong additional assumptions.

6. Multilevel Treatment and Unsaturated MSMs

Suppose again that treatment is unconfounded (as represented in Figure 2, *c*) but now A_0 is an ordinal variable representing a subject's daily dose in units of 100 mg of AZT. Possible values of A_0 are 0, 1, . . ., 14, 15. In that case, the number of potential outcomes associated with each subject will be 16. Specifically, we let Y_{a_0} be the value of Y that would have been observed had the subject received dose a_0 rather than the observed dose. Thus, in principle, a subject has a separate counterfactual variable for each of the 16 possible AZT doses a_0. The subject's observed outcome Y is the outcome Y_{a_0} corresponding to the dose a_0 equal to the subject's observed dose. For expositional convenience, we continue to refer to all of the Y_{a_0}'s as counterfactuals, even though for a_0 equal to the observed dose, Y_{a_0} is the factual variable Y. Because there are so many potential outcome variables Y_{a_0}, we can no longer conveniently perform a saturated analysis. Rather, we would usually assume a parsimonious dose-response relationship by specifying a linear logistic MSM such as

$$logit\ pr[Y_{a_0} = 1] = \beta_0 + \beta_1 a_0. \qquad (8)$$

This model says that the probability of success had all subjects been treated with dose a_0 is a linear logistic function of the dose with slope parameter β_1 and intercept β_0, so e^{β_1} is the causal *OR* associated with an increase in AZT dose of 100 mg.

We contrast the MSM model 8 with the following linear logistic association model for the observed data.

$$logit\ pr[Y = 1|A_0 = a_0] = \beta_0' + \beta_1' a_0. \qquad (9)$$

Assuming no selection bias or measurement error, we can unbiasedly estimate the associational parameters β_0' and β_1' by fitting the linear logistic model 9 using a standard logistic regression software package such as SAS Proc Logistic or Proc Genmod. If the treatment is unconfounded, then the parameters of models 8 and 9 are equal. As a consequence, our logistic regression estimate of β_1' is also an unbiased estimate of our causal parameter β_1.

If treatment is confounded by the measured variables L_0, then $\beta_1 \neq \beta_1'$ and our standard logistic regression estimate of β_1' is a biased estimate of the causal parameter β_1 owing to confounding by L_0. However, even when treatment is confounded, if there are no unmeasured confounders given L_0, then one can obtain unbiased estimates of the causal parameter β_1 of model 8 by fitting the logistic model 9 with Proc Genmod if one uses subject-specific weights $w_i = 1/pr(A_0 = a_{0i}|L_0 = l_{0i})$. Again, in practice, w_i is unknown and one must estimate it from the data by specifying a model.

For example, one might specify the following polytomous logistic model:

$$pr[A_0 = a_0|L_0 = l_0] = \exp(\alpha_{0a_0} + \alpha_1 l_0)/\left\{1 + \sum_{j=1}^{15} \exp(\alpha_{0j} + \alpha_1 l_0)\right\},$$

$$a_0 = 1, \ldots, 15;$$

$$pr[A_0 = 0|L_0 = l_0] = 1/\left\{1 + \sum_{j=1}^{15} \exp(\alpha_{0j} + \alpha_1 l_0)\right\} \qquad (10)$$

which can be fit in SAS using Proc Logistic or Genmod to obtain estimates of the parameters α_{01}, α_{02}, . . ., α_{015}, and α_1.

6.1. Stabilized Weights

The probabilities $pr[A_0 = a_{0i}|L_0 = l_{0i}]$ may vary greatly between subjects when components of L_0 are strongly

associated with A_0. This variability can result in extremely large values of the weight w_i for a few subjects. These few subjects will contribute a very large number of copies of themselves to the pseudopopulation and thus will dominate the weighted analysis, with the result that our IPTW estimator will have a large variance and will fail to be approximately normally distributed. If the MSM model is saturated (for example, models 1–3), this variability is unavoidable, because it reflects a lack of information in the data as a result of the confounders L_0 being highly correlated with treatment A_0. For unsaturated MSMs, such as model 8, however, this variability can, to a considerable extent, be mitigated by replacing the weight w_i by the "stabilized weight" $sw_i = pr[A_0 = a_{0i}]/pr[A_0 = a_{0i}|L_0 = L_{0i}]$. To understand the stabilized weight, suppose A_0 was unconfounded so that A_0 and L_0 are unassociated and $pr[A_0 = a_{0i}] = pr[A_0 = a_{0i}|L_0 = l_{0i}]$. Then $sw_i = 1$, and each subject contributes the same weight. When A_0 is confounded, sw_i will not be constant but will vary around the number 1, depending on the subject's value of L_0. sw_i, however, will still tend to be much less variable than w_i. Furthermore, Robins[1,2] shows that, when we use the weight sw_i rather than the weight w_i in Proc Genmod, the estimates of the parameters β of an MSM remain unbiased and, in the case of an unsaturated MSM, will generally be less variable. For saturated MSMs, the variability of our estimate of β will be the same whether we use the stabilized or unstabilized weights.

Of course, $pr[A_0 = a_{0i}]$ and $pr[A_0 = a_{0i}|L_0 = l_{0i}]$ are unknown and must be estimated. $pr[A_0 = a_{0i}|L_0 = l_{0i}]$ can be estimated as described above; $pr[A_0 = a_{0i}]$ can be estimated as the proportion of subjects in the study sample with A_0 equal to a_{0i}. This estimate is equivalent to that obtained by fitting the polytomous model

$$pr(A_0 = a_0) = \exp(\alpha^*_{0a_0}) \Big/ \left[1 + \sum_{j=1}^{15} \exp(\alpha^*_{0j}) \right], \quad a_0 = 1, \ldots, 15$$

$$pr(A_0 = 0) = 1 \Big/ \left[1 + \sum_{j=1}^{15} \exp(\alpha^*_{0j}) \right] \qquad (11)$$

where we place an asterisk on the parameter $\alpha^*_{0a_0}$ to indicate that this parameter will differ from the parameter α_{0a_0} of model 10 when A_0 is confounded.

In Ref 2, Robins introduces augmented IPTW estimators. These estimators are even more efficient than the IPTW estimator that uses stabilized weights but are more difficult to compute.

6.2. Continuous Treatment

Suppose we were able to measure a subject's daily intake of AZT to the nearest tenth of a milligram, so that now AZT is essentially a continuous treatment. Further, for expositional convenience, assume that no subject has AZT dose near 0, and that we can effectively model the distribution of A_0 as normal. Now each individual has an extremely large number of counterfactual outcomes Y_{a_0}. One can still obtain unbiased estimates of the causal parameter β_1 of model 8 by fitting the logistic model 9 with Proc Genmod if one uses the stabilized weights $sw_i = f(a_{0i})/f(a_{0i}|l_{0i})$, where $f(a_0|l_0)$ is the conditional density of the continuous variable A_0 given L_0, and $f(a_0)$ is the marginal density of the continuous variable A_0. To estimate $f(a_0|l_0)$, one might specify that, given L_0, A_0 is normal with mean $\alpha_0 + \alpha_1 L_0$ and variance σ^2. Then unbiased estimates $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\sigma}^2)$ of $(\alpha_0, \alpha_1, \sigma^2)$ can be obtained by ordinary least-squares regression of A_0 on L_0 using, for example, Proc REG in SAS. Then $f(a_{0i}|l_{0i})$ would be estimated by the normal density $(2\pi\hat{\sigma}^2)^{-1/2} \exp\{-[a_{0i} - (\hat{\alpha}_0 + \hat{\alpha}_1 l_{0i})]^2/2\hat{\sigma}^2\}$. To estimate the numerator $f(a_{0i})$ of the stabilized weight sw_i, one might specify that A_0 is normal with mean α^*_0 and variance σ^{*2}. $f(a_{0i})$ could be estimated by the normal density $(2\pi\hat{\sigma}^{*2})^{-1/2} \exp[-(a_{0i} - \hat{\alpha}^*_0)^2/2\hat{\sigma}^{*2}]$ where $\hat{\alpha}^*_0$ is the average of the observed A_0s and $\hat{\sigma}^{*2}$ is their empirical variance. When A_0 is continuous, estimates based on the unstabilized weights $w_i = 1/f(a_{0i}|l_{0i})$ have infinite variance and thus cannot be used.[1,2]

6.3. Confidence Intervals

As described above, we shall estimate the parameters of the MSM 8 by fitting the association model 9 in Proc Genmod using estimates of the stabilized weights sw_i. If we choose the option "repeated" and specify an independence working correlation matrix, the Proc Genmod program will also output a 95% "robust" Wald confidence interval for β_1 given by $\hat{\beta}_1 \pm 1.96\sqrt{var(\hat{\beta}_1)}$, where $var(\hat{\beta}_1)$ is the so-called "robust"[18] or "sandwich" estimator of the variance of $\hat{\beta}_1$. Robins[1,2] shows that the "robust" Wald interval will have coverage probability of at least 95%, although narrower valid intervals can be obtained with some additional programming.[1,2] The ordinary nonrobust model-based Wald confidence interval outputted by most weighted logistic regression programs will not be guaranteed to provide at least 95% coverage and thus should be avoided. Other software packages such as S-Plus, Gauss, and STATA also offer "robust" variance estimators and could be used in place of Proc Genmod.[13,19–21]

7. Time-Dependent Treatments

We now return to the setting of section 1, in which A_k is the dose of treatment AZT on the k^{th} day from start of follow-up and Y is a dichotomous outcome variable measured at end of follow-up on day $K + 1$. Similarly, L_k represents the value on day k of the vector of all measured risk factors for the outcome. Let $\bar{A}_k = (A_0, A_1, \ldots, A_k)$ be the treatment or exposure history through day k and let $\bar{A} = \bar{A}_K$. Define $\bar{L}_k$ and $\bar{L}$ similarly. Let $Y_{\bar{a}}$ be the value of Y that would have been observed had all subjects received dose history $\bar{a} = (a_0, a_1, \ldots, a_K)$ rather than their observed dose history $\bar{A}$. Note that, even if a_k is dichotomous on each day k (that is, on each day a subject is either on or off treatment), there will be 2^K

dose histories $\bar{a}$ and thus 2^K possible counterfactuals, only one of which is observed (that is, is factual). Thus, it may not be possible to estimate a saturated MSM. Therefore, we would generally assume some sort of parsimonious dose-response relationship by specifying a linear logistic MSM such as

$$logit\ pr[Y_{\bar{a}} = 1] = \beta_0 + \beta_1 cum(\bar{a}) \qquad (12)$$

where $cum(\bar{a}) = \Sigma_{k=0}^{K} a_k$ is the cumulative dose through end-of-follow-up associated with the dose history $\bar{a}$.

We contrast the MSM 12 with the following linear logistic association model for the observed data:

$$logit\ pr[Y = 1|\bar{A} = \bar{a}] = \beta'_0 + \beta'_1 cum(\bar{a}). \qquad (13)$$

Assuming no loss to follow-up selection bias or measurement error, we can unbiasedly estimate the parameters β'_1 by fitting the linear logistic model 13 using a standard logistic regression software package. If the treatment is unconfounded, the parameters of models 12 and 13 are equal. As a consequence, our logistic regression estimate of β'_1 is also an unbiased estimate of our causal parameter β_1. If the treatment is confounded, then $\beta_1 \neq \beta'_1$ and our standard logistic regression estimate of β'_1 is a biased estimate of the causal parameter β_1 as a result of confounding by $\bar{L}_k$. When treatment is confounded, however, if there is no unmeasured confounder given the L_k, then one can still obtain unbiased estimates of the causal parameter β_1 of model 12 by fitting the logistic model 13 with the stabilized weights

$$sw_i = \prod_{k=0}^{K} pr(A_k = a_{ki}|\bar{A}_{k-1} = \bar{a}_{(k-1)i})/$$

$$\left\{ \prod_{k=0}^{K} pr(A_k = a_{ki}|\bar{A}_{k-1} = \bar{a}_{(k-1)i}, \bar{L}_k = \bar{l}_{ki}) \right\} \qquad (14)$$

where $\Pi_{k=0}^{K} b_k = b_0 \times b_1 \times b_2 \times \ldots \times b_K$ and $\bar{A}_{-1}$ is defined to be 0. Note in the special case in which $K = 0$ (that is, a point-treatment study), models 12 and 13 reduce to our previous models 8 and 9 and 14 reduces to our previous sw_i. The denominator of sw_i is informally the conditional probability that a subject had his or her own observed treatment history through time K. With time-dependent treatments the variation in the unstabilized weights will often be enormous, with the result that the resulting estimator of β can be highly variable with a markedly non-normal sampling distribution. We therefore strongly recommend the use of stabilized weights.

We emphasize that when treatment is confounded, it is the parameter β_1 of our MSM 12, as opposed to the parameter β'_1 of the association model 13 that is of policy importance. To see why, consider a new subject from the source or target population exchangeable with the N study subjects. We would like to administer to the subject the treatment $\bar{a}$ that minimized probability that he or she has detectable HIV in the serum at the end of follow-up, that is, $pr(Y_{\bar{a}} = 1)$. Thus, for example, if the parameter β_1 of MSM 12 is positive (that is, the probability of having HIV in the blood increases with increasing duration of AZT treatment), we would withhold AZT treatment from our subject. In contrast, the parameter β'_1 of model 13 may be confounded by the association of covariates with treatment. For example, suppose physicians preferentially started AZT on subjects who, as indicated by their prognostic factor history (for example, CD4 count), were doing poorly. Further, suppose that AZT had no causal effect on Y (that is, $\beta_1 = 0$). Then the parameter β'_1 (and thus our estimate from the unweighted logistic regression) will be positive but will have no causal interpretation as the effect of AZT on Y.

7.1. Bias Induced by Controlling for a Variable Affected by Treatment

One might suppose that an alternative approach to controlling confounding by measured covariates is an unweighted logistic model that adjusts for confounder history $\bar{L} \equiv \bar{L}_K$, such as

$$logit\ pr[Y = 1|\bar{A} = \bar{a}, \bar{L} = \bar{l}] = \beta''_0 + \beta''_1 cum(\bar{a}) + \beta''_2 cum(\bar{l}) + \beta''_3 l_k + \beta''_4 l_{k-1} + \beta''_5 l_0$$

where, for simplicity, we here assume L_k consists of a single covariate CD4 count at time k. Nevertheless, even under our assumption of no unmeasured confounders, the parameter β''_1 differs from the parameter β_1 of our MSM. What is worse is that the parameter β''_1 will generally not have a causal interpretation, even if the model for $pr[Y = 1|\bar{A} = \bar{a}, \bar{L} = \bar{l}]$ is correctly specified. This is because $cum(\bar{A})$ depends on a subject's entire treatment history, including A_0, and A_0 may affect the time-dependent covariates L_k and L_{k-1}. Fitting a logistic model that adjusts for a covariate that is both affected by treatment and is a risk factor for the outcome provides an unbiased estimate of the association parameter β''_1 but a biased estimate of the causal parameter β_1. This is true even under the null hypothesis of no direct, indirect, or overall treatment effect (so that β_1 of model 12 equals 0) when, as in Figure 1, a component of L_k (for example, red blood count) and the outcome Y have an unmeasured common cause U_0 (for example, the baseline number of bone marrow stem cells).[5,7,22–26]

To summarize, standard regression methods adjust for covariates by including them in the model as regressors. These standard methods may fail to adjust appropriately for confounding due to measured confounders L_k when treatment is time varying, because (1) L_k may be a confounder for later treatment and thus must be adjusted for, but (2) may also be affected by earlier treatment and thus should not be adjusted for by standard methods. A solution to this conundrum is to adjust for the time-dependent covariates L_k by using them to calculate the weights sw_i rather than by adding the L_k to the regression model as regressors.

8. Estimation of the Weights

We now describe how to estimate the weights sw_i. For simplicity, we again assume that the treatment A_k at each time k is dichotomous. Consider first the denominator of model 14. We begin by estimating the unknown probability $pr[A_k = 1|\bar{A}_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k]$ using a pooled logistic model that treats each person-day as an observation. For example, we might fit the model

$$logit\ pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k] = \alpha_0 + \alpha_1 k + \alpha_2 a_{k-1} + \alpha_3 a_{k-2} + \alpha_4 l_k + \alpha_5 l_{k-1} + \alpha_6 a_{k-1} l_k + \alpha_7 l_0 \quad (15)$$

where, for example, l_k is the vector of CD4 count, WBC, hematocrit, and an indicator for symptoms at time k, and the α_4, α_5, α_6, and α_7 are row vectors. This model says that the probability of being treated on day k depends in a linear logistic fashion on the day k, the previous 2 days' treatment, the current and previous days' covariates, an interaction between yesterday's treatments and today's covariates, and the baseline covariates.

One can fit model 15 using any standard logistic regression program. The numerator probabilities can be estimated similarly, except that, in fitting model 15, we remove the last four terms that are functions of the covariates. That is, we fit the model as follows:

$$logit\ pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}] = \alpha_0^* + \alpha_1^* k + \alpha_2^* a_{k-1} + \alpha_3^* a_{k-2}. \quad (16)$$

For each subject i, we then have our logistic program output the estimated predicted values $\hat{p}_{0i}, \ldots, \hat{p}_{Ki}$ from the fit of model 15, which are maximum likelihood estimates of $pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{(k-1)i}, \bar{L}_k = \bar{l}_{ki}]$. Similarly, we have outputted the predicted values $\hat{p}_{1i}^*, \ldots, \hat{p}_{Ki}^*$ from model 16, which are estimates of the quantities $pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{(k-1)i}]$. Then we estimate sw_i by

$$sw_i = \prod_{k=0}^{K} (\hat{p}_{ki}^*)^{a_{ki}} (1 - \hat{p}_{ki}^*)^{1 - a_{ki}} \Big/ \left\{ \prod_{k=0}^{K} (\hat{p}_{ki})^{a_{ki}} (1 - \hat{p}_{ki})^{1 - a_{ki}} \right\}. \quad (17)$$

For example, $1 - \hat{p}_{ki}^*$ is an estimate of the probability $pr[A_k = a_{ki}|\bar{A}_{k-1} = \bar{a}_{(k-1)i}]$ when $a_{ki} = 0$. The data analyst will need to write a small program to compute sw_i for each subject from the predicted values outputted from the fit of models such as 15 and 16.

Under our assumption of no unmeasured confounders, the resulting estimate of the causal parameter β_1 will be unbiased, provided the model 15 for $pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k]$ is correctly specified. Furthermore, under these same conditions, the 95% robust Wald confidence interval will be guaranteed to cover β_1 at least 95% of the time. The estimate of β_1 will remain unbiased even if the model 16 for $pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}]$ is misspecified.[1,2] Indeed, if model 15 is correct and treatment is confounded, model 16 is guaranteed to be somewhat misspecified because of the noncollapsibility of logistic models.[25]

9. Effect Modification by Pretreatment Covariates

MSMs can be generalized to allow one to include pretreatment covariates. For example, model 12 could be generalized to

$$logit\ pr[Y_{\bar{a}} = 1|V = v] = \beta_0 + \beta_1 cum(\bar{a}) + \beta_2 v + \beta_3 cum(\bar{a})v \quad (18)$$

where V is a component of the vector of measured pretreatment covariates L_0, and β_3 denotes a treatment-covariate interaction. Note in model 18, $\beta_1 + \beta_3 v$ represents the effect of cumulative treatment on a linear logistic scale within level v of the baseline variable V. As our IPTW estimators already automatically adjust for any confounding due to V, the particular subset V of L_0 that an investigator chooses to include in model 18 should only reflect the investigator's substantive interest. For example, a variable V should be included in model 18 only if the investigator both believes that V may be an effect modifier and has greater substantive interest in the causal effect of treatment within levels of the covariate V than in the source population as a whole.

We obtain unbiased estimates of the parameters of model 18 under the assumption of no unmeasured confounders by fitting an association model such as

$$logit\ pr[Y = 1|\bar{A} = \bar{a}, V = v] = \beta_0 + \beta_1 cum(\bar{a}) + \beta_2 v + \beta_3 cum(\bar{a})v \quad (19)$$

using Proc Genmod with the estimated weights sw_i of Eq 17, modified only in that p_k^* is now the estimated predicted value from the fit of a model such as

$$logit\ pr[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}, V = v] = \alpha_0^* + \alpha_1^* k + \alpha_2^* a_{k-1} + \alpha_3^* a_{k-2} + \alpha_4^* v.$$

Elementary epidemiologic textbooks emphasize that effect modification is logically distinct from confounding. Nonetheless, many students have difficulty understanding the distinction, because the same statistical methods (stratification and regression adjustments) are used both for confounder control and detection of effect modification. Thus, there may be some advantage to teaching elementary epidemiologic methods using marginal structural models, because then methods for confounder control (inverse-probability-of-treatment weighting) are distinct from methods for detection of effect modification (adding treatment covariate interaction terms to an MSM).

Finally, an important caveat: MSMs cannot be used to model the interaction of treatment with a time-

varying covariate. For this, structural nested models should be used.[4–6] Therefore, it is not valid to include in the covariate V in model 18 any components of the time-dependent covariate L_k measured at any time $k > 0$.

10. Censoring by Loss to Follow-Up

Heretofore, we have assumed that each study subject is observed until end of follow-up at time $K + 1$. In this section, we allow for censoring by loss to follow-up. Specifically, let $C_k = 1$ if a subject was lost to follow-up by day k and $C_k = 0$ otherwise. We assume that once a subject is lost to follow-up, the subject does not later re-enter follow-up. No new idea is required to account for censoring, provided we conceptualize censoring as just another time-varying treatment. From this point of view, to want to adjust for censoring is only to say that our interest is in the causal effect of the treatment $\bar{A}$ if, contrary to fact, all subjects had remained uncensored, rather than having followed their observed censoring history. Our goal remains to estimate the parameter β_1 of the logistic MSM 12 except now $Y_{\bar{a}}$ refers to a subject's outcome if, possibly contrary to fact, the subject has followed treatment history $\bar{a}$ and has never been censored. Again, we can do so if there are no unmeasured confounders for both treatment and censoring. To formalize this idea, one adds at each time k the variable C_k to the graph in Figure 1 just before L_k and after A_{k-1}. Then, the assumption of no unmeasured confounders for treatment and censoring is that no arrow arising from the unmeasured causal risk factors U goes directly into either C_k or A_k for any k. In that case, the measured covariates $\bar{L}_k$ are sufficient to adjust both for confounding and selection bias due to loss to follow-up.

Again, we can obtain unbiased estimates of the causal parameters β_1 by fitting the linear logistic association model 13 with appropriate weights included. Because the outcome Y is unobserved unless the subject does not drop out, that is, $\bar{C} = (C_0, \ldots, C_{K+1}) = 0$, our weighted logistic regression fit of model 13 is restricted to uncensored subjects. The required subject-specific weight is $sw_i \times sw_i^{\dagger}$, where

$$sw_i^{\dagger} = \prod_{k=0}^{K+1} pr(C_k = 0 | \bar{C}_{k-1} = 0, \bar{A}_{k-1} = \bar{a}_{(k-1)i}) / \left\{ \prod_{k=0}^{K+1} pr(C_k = 0 | \bar{C}_{k-1} = 0, \bar{A}_{k-1} = \bar{a}_{(k-1)i}, \bar{L}_k = \bar{l}_{ki}) \right\}$$

and, in addition, in defining and estimating sw_i, we now add to the right side of each conditioning event in models 14–16 the event $\bar{C}_k = 0$, because otherwise, A_k would not be observed. The unknown probabilities in $sw_i^{\dagger}$ can be estimated using a pooled logistic model that treats each person-day as an observation. Specifically, we fit analogs of models 15 and 16 for *logit* $pr[C_k = 0 | \bar{C}_{k-1} = 0, \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k]$ and for *logit* $pr[C_k = 0 | \bar{C}_{k-1} = 0, \bar{A}_{k-1} = \bar{a}_{k-1}]$. Note that the denominator of the product $sw_i \times sw_i^{\dagger}$ is informally the conditional probability that an uncensored subject had his or her observed treatment and censoring history through time $K + 1$. Thus, we refer to our weighted logistic estimator as an inverse-probability-of-treatment-and-censoring weighted estimator. If we view (A_k, C_k) as a "joint treatment" at time k, then one can informally interpret this denominator as simply the probability that a subject follows his or her own treatment history, which is exactly the interpretation that we had previously in the absence of censoring.

11. Limitations of Marginal Structural Models

It is shown in Ref 2 and Appendix 2 that our IPTW estimators will be biased and thus MSMs should not be used in studies in which at each time k there is a covariate level l_k such that all subjects with that level of the covariate are certain to receive the identical treatment a_k. For example, this circumstance implies that MSMs should not be used in occupational cohort studies. To see why, consider an occupational cohort study in which A_k is the level of exposure to an industrial chemical at time k and $L_k = 1$ if a subject is off work at time k and $L_k = 0$ otherwise. Then all subjects with $L_k = 1$ have $A_k = 0$, because all subjects off work are unexposed. Similarly, in a study of the effect of screening on mortality from cervical cancer, women who have had their cervix operatively removed by time k (which we denote by $L_k = 0$) cannot receive exposure (that is, screening) at that time, so MSMs should not be used. Nevertheless, g-estimation of structural nested models can always be used to estimate exposure effects, even in studies in which MSMs cannot be used. In many studies, such as the analysis of the Multicenter AIDS Cohort Study data described in our companion paper,[27] we believe, based on substantive considerations, the above difficulty does not occur and MSMs are a practical method.

12. Conclusion

We have described how to use MSMs to estimate the causal effect of a time-varying exposure or treatment on a dichotomous outcome. In our companion paper,[27] we extend our results to survival time outcomes and compare and contrast methods based on MSMs to alternative, previously proposed methods, based on g-estimation of structural nested models and on estimation of the g-computation algorithm formula.

References

1. Robins JM. Marginal structural models. In: 1997 Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA: American Statistical Association, 1998;1–10.
2. Robins JM. Marginal structural models *versus* structural nested models as tools for causal inference. In: Halloran E, Berry D, eds. Statistical Models in Epidemiology: The Environment and Clinical Trials. New York: Springer-Verlag, 1999;95–134.

3. Robins JM. Correction for non-compliance in equivalence trials. Stat Med 1998;17:269–302.
4. Robins JM. Structural nested failure time models. In: Andersen PK, Keiding N, section eds. Survival Analysis. In: Armitage P, Colton T, eds. The Encyclopedia of Biostatistics. Chichester, UK: John Wiley and Sons, 1998; 4372–4389.
5. Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients (erratum in Epidemiology 1993;3:189). Epidemiology 1992; 3:319–336.
6. Witteman JCM, D'Agostino RB, Stijnen T, Kannel WB, Cobb JC, de Ridder MAJ, Hofman A, Robins JM. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. Am J Epidemiol 1998;148:390–401.
7. Robins JM. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chron Dis 1987;40(suppl 2):139S–161S.
8. Pearl J. Causal diagrams for empirical research. Biometrika 1995;82:669–688.
9. Pearl J, Robins JM. Probabilistic evaluation of sequential plans from causal models with hidden variables. In: Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 1995;444–453.
10. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37–48.
11. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods: application to control of the healthy worker survivor effect (erratum appear in Math Modelling 1987;14:917–921). Mathematical Modelling 1987;7:1393–1512.
12. Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. Health Services Research Methodology: A Focus on AIDS. Rockville, MD: National Center for Health Services Research, U.S. Public Health Service, 1989;113–159.
13. SAS Institute Inc. SAS/STAT User's Guide Version 8. Cary, NC: SAS Institute, 1999.
14. S-PLUS 4 Guide to Statistics. Seattle: Mathsoft, 1997.
15. GAUSS. Maple Valley, WA: Aptech Systems, 1996.
16. Stata. Stata Statistical Software: Release 6.0. College Station, TX: Stata Corporation, 1999.
17. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
18. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability. Berkeley: University of California Press, 1967;221–233.
19. Carey VJ. YAGS 1.5.1.5: GEE solver for S-plus. In http://biosun1.harvard.edu/~carey/index.ssoft.html, 1998.
20. Källén A. Generalized estimating equations: marginal methods for the analysis of correlated data. In: Riemann Library. http://www.aptech.com, 1999.
21. Stata Corp. Obtaining robust variance estimates. In: Stata Statistical Software: Release 5.0 User's Guide. College Station, TX: Stata Corporation, 1997;235–239.
22. Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by treatment. J R Stat Soc 1984;147:656–666.
23. Greenland S, Neutra RR. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. J Chron Dis 1981;34:433–438.
24. Greenland S, Neutra RR. Control of confounding in the assessment of medical technology. Int J Epidemiol 1981;9:361–367.
25. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol 1987;125:761–768.
26. Robins JM, Greenland S, Hu F-C. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. J Am Stat Assoc 1999;94:687–700.
27. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561–570.
28. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 1952;47:663–685.
29. Rosenbaum PR. Model-based direct adjustment. J Am Stat Assoc 1987;82: 387–394.
30. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell N, Dietz K, Farewell V, eds. AIDS Epidemiology: Methodological Issues. Boston: Birkhäuser, 1992, pp 24–33.
31. Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In: Proceedings of the American Statistical Association, Biopharmaceutical Section, 1993;24–33.

Appendix 1: Example

We will analyze the data in Table A1 under the assumption of no unmeasured confounders given L_0. For convenience, we shall ignore sampling variability and thus the distinction between parameters of the source population and their empirical estimates. Under the assumption of no unmeasured confounder, $pr(Y_{a_0=1} = 1)$ is a weighted average of the L_0-stratum-specific risks among the treated with weights proportional to the distribution of L_0 in the entire study population. That is, $pr(Y_{a_0=1} = 1)$ is given by

$$\sum_{l_0} pr[Y = 1|A_0 = 1, L_0 = l_0]pr[L_0 = l_0] \quad \text{(A1)}$$

where the sum is over the possible values of L_0.[17] We refer to Eq. A1 as the L_0-standardized risk in the treated. Calculating from Table A1, we obtain that $pr(Y_{a_0=1} = 1) = 0.32$. Similarly, $pr(Y_{a_0=0} = 1)$ is the L_0-standardized risk in the untreated,

$$\sum_{l_0} pr[Y = 1|A_0 = 0, L_0 = l_0]pr[L_0 = l_0]$$

which, from Table A1, is 0.64. It follows that the causal risk difference, risk ratio, and odds ratio are −0.32, 0.50, and 0.26. Note that these differ from the crude parameters computed from Table A2. Thus, $\psi_1 = -0.32$, $\theta_1 = \log 0.50$, and $\beta_1 = \log 0.26$ in models 1–3 differ from the parameters $\psi_1' = -0.40$, $\theta_1' = \log .044$, and $\beta_1' = \log 0.18$ of models 4–6.

As is well known, the causal risk difference and causal risk ratio are also equal to weighted averages of the stratum-specific risk differences and risk ratios. For example, the causal *RD* equals the standardized risk difference (*SRD*) where

$$SRD = \sum_{l_0} RD_{l_0} pr[L_0 = l_0]$$

and $RD_{l_0} = pr[Y = 1|A_0 = 1, L_0 = l_0] - pr[Y = 1|A_0 = 0, L_0 = l_0]$ is the risk difference in stratum l_0.

TABLE A1. Observed Data from a Point-Treatment Study with Dichotomous Treatment A_0, Stratified by the Confounder L_0

	$L_0 = 1$		$L_0 = 0$	
	$A_0 = 1$	$A_0 = 0$	$A_0 = 1$	$A_0 = 0$
Y = 1	108	24	20	40
Y = 0	252	16	30	10
Total	360	40	50	50

TABLE A2. Crude Data from the Point-Treatment Study of Table A1

	$A_0 = 1$	$A_0 = 0$
Y = 1	128	64
Y = 0	282	26
Total	410	90

TABLE A3. Inverse Probability of Treatment Weights w and Composition of the Pseudopopulation in a Point-Treatment Study

L_0	A_0	Y	N Observed Population	$pr(A_0\|L_0)$	w	N Pseudo Population
1	1	1	108	0.9	1.11	120
1	1	0	252	0.9	1.11	280
1	0	1	24	0.1	10	240
1	0	0	16	0.1	10	160
0	1	1	20	0.5	2	40
0	1	0	30	0.5	2	60
0	0	1	40	0.5	2	80
0	0	0	10	0.5	2	20

Indeed, the usual way to estimate the causal *RD* is to calculate the *SRD*. Our IPTW method is an alternative approach to estimation of the causal *RD* that, in contrast to the approach based on calculating the *SRD*, appropriately generalizes to unsaturated MSMs in longitudinal studies with time-varying treatments, as discussed in section 7.

Table A3 displays the data from the study in a different format. In particular, it gives the number of subjects with each of the possible combinations of l_0, a_0, and y, as well as the weight $w = 1/pr[A_0 = a_0|L_0 = l_0]$ associated with each. The final column of the table represents the number of subjects in the weighted pseudopopulation for each combination of (l_0, a_0, y). Note that the weights w_i need not be whole numbers or sum to 1. As a consequence, the number of subjects in the pseudopopulation can be greater than the number in the actual population. Tables A4 and A5 display the data from the pseudopopulation in the same format as Tables A1 and A2. It can be seen that L_0 and A_0 are unassociated in the pseudopopulation, which implies that the treatment is unconfounded. Furthermore, the lack of association between L_0 and A_0 implies that in the pseudopopulation, the L_0-standardized risk in the treated equals the crude risk $pr(Y = 1|A_0 = 1) = 0.32$ and the L_0-standardized risk in the untreated equals the crude risk $pr(Y = 1|A_0 = 0) = 0.64$. Furthermore, the crude risk in the treated pseudopopulation equals the L_0-standardized risk in the treated actual population and thus equals $pr(Y_{a_0=1} = 1)$. Similarly, the crude risk in the untreated pseudopopulation equals the L_0-standardized risk in the untreated true population and thus equals $pr(Y_{a_0=0} = 1)$. It follows that, under the assumption of no unmeasured confounder given L_0, the crude risk difference, risk ratio, and odds ratio in the pseudopopulation equal the causal risk difference, risk ratio, and odds ratio in the actual population. Finally, an IPTW analysis in Proc Genmod estimates a crude parameter of the pseudopopulation and thus a causal parameter of the actual population.

TABLE A4. Pseudopopulation Created by Inverse Probability of Treatment Weighting from a Point-Treatment Study with Dichotomous Treatment A_0, Stratified by the Confounder L_0

	$L_0 = 1$		$L_0 = 0$	
	$A_0 = 1$	$A_0 = 0$	$A_0 = 1$	$A_0 = 0$
Y = 1	120	240	40	80
Y = 0	280	160	60	20
Total	400	400	100	100

TABLE A5. Crude Data from the Pseudopopulation of Table A4

	$A_0 = 1$	$A_0 = 0$
Y = 1	160	320
Y = 0	340	180
Total	500	500

Relation to Propensity Score and Horvitz-Thompson Methods

Rosenbaum and Rubin[17] refer to the probability $p_i = pr[A_0 = 1|L_0 = l_{0i}]$ that subject i would receive treatment as the propensity score. Note that IPTW weight w_i is not simply the inverse of the propensity score. Specifically, although w_i is the inverse of the propensity score for treated subjects, it is the inverse of $1 - p_i$ for untreated subjects. Rosenbaum and Rubin[17] showed that, under the assumption of no unmeasured confounders, one can control for confounding due to measured covariates in a point-treatment study with a dichotomous treatment by regarding the propensity score as the sole confounder. Because the propensity score p_i is a continuous covariate, however, they suggested that, in practice, one either approximately match treated with untreated subjects on the propensity score or stratify (that is, subclassify) on the basis of propensity score quintiles. Even when there are no unmeasured confounders and the propensity score is unbiasedly estimated, Rosenbaum and Rubin's[17] approach, unlike our approach, suffers from the potential for substantial residual confounding due to the inability to obtain sufficiently close matches or to uncontrolled intrastratum confounding. More importantly, Rosenbaum and Rubin's[17] propensity score methods, in contrast to our IPTW methods, do not generalize straightforwardly to studies with nondichotomous or time-dependent treatments or exposures.

In the special case of a dichotomous time-independent treatment, our IPTW estimator is essentially equiv-

TABLE A6. Observed Data from a Point-Treatment Study in Which Eq A2 Holds

	$L_0 = 1$		$L_0 = 0$	
	$A_0 = 1$	$A_0 = 0$	$A_0 = 1$	$A_0 = 0$
$Y = 1$	0	240	20	40
$Y = 0$	0	160	30	10
Total	0	400	50	50

alent to estimating $pr(Y_{a_0=0} = 1)$ and $pr(Y_{a_0=1} = 1)$ separately among the untreated ($a_0 = 0$) and treated ($a_0 = 1$) using the Horvitz-Thompson estimator[28] from the sample survey literature.[29] Robins and Rotnitzky[30] and Robins[31] proposed generalizations of the Horvitz-Thompson estimator that could be used to estimate the parameters of a saturated MSM model with a time-varying treatment. Our IPTW estimators are further generalizations that allow the estimation of nonsaturated MSMs with both time-independent and time-dependent treatments.

Appendix 2

Bias of Inverse-Probability-of-Treatment Weighted Estimators in the Setting of Section 11

Consider a new study population for which $pr(Y_{a_0=1} = 1)$ and $pr(Y_{a_0=0} = 1)$, and therefore the causal risk difference, are the same as for the population in Tables A1–A5. The observed data for the new population, however, given in Table A6, differs from the observed data for the population studied in Tables A1–A5. Specifically, Table A6 differs from the observed data in Table A1 only in that no subject with $L_0 = 1$ receives treatment $A_0 = 1$, that is,

$$pr(A_0 = 1|L_0 = 1) = 0, \qquad (A2)$$

and thus represents the type of study discussed in section 11. We will show that when Eq A2 holds the IPTW estimator of the causal risk, difference is now biased.

In Table A6, $pr(Y_{a_0=0} = 1)$ is again 0.64, the L_0-standardized risk in the untreated. Nevertheless, the L_0-standardized risk in the treated $pr(Y = 1|A_0 = 1, L_0 = 0)pr(L_0 = 0) + pr(Y = 1|A_0 = 1, L_0 = 1)pr(L_0 = 1)$ cannot be computed from the data in Table A6, because there is no subject with history $(A_0 = 1, L_0 = 1)$, rendering $pr(Y = 1|A_0 = 1, L_0 = 1)$ uncomputable. Similarly, the *SRD* is not computable, because the stratum-specific risk difference is undefined in the stratum $L_0 = 1$. Thus, $pr(Y_{a_0=1} = 1)$ and the causal risk difference are not computable from the data in Table A6, although we know by assumption that they are still equal to the previous values 0.32 and −0.32. Table A7 displays the data in Table A6 in the format of Table A3. Tables A8 and A9 display the stratified and crude data for the pseudopopulation constructed from the last column of Table A7. Note that the *SRD* in Table A8 for the pseudopopulation is undefined. The pseudopopulation crude *RD* from Table A9 is −0.24, which differs from the true causal risk difference $\psi_1 = -0.32$. As discussed previously, however, it is the crude *RD* in the pseudopopulation that our IPTW estimate of the parameter ψ_1 in the MSM $pr(Y_{a_0} = 1) = \psi_0 + \psi_1 a_0$ actually estimates. We conclude that our MSM estimate is biased for the causal risk difference ψ_1.

TABLE A7. Inverse Probability of Treatment Weights *w* and Composition of the Pseudopopulation in a Point-Treatment Study in Which Eq A2 Holds

L_0	A_0	Y	N Observed Population	$pr(A_0\|L_0)$	*w*	N Pseudo Population
1	1	1	0	0	∞	0*
1	1	0	0	0	∞	0*
1	0	1	240	1	1	240
1	0	0	160	1	1	160
0	1	1	20	0.5	2	40
0	1	0	30	0.5	2	60
0	0	1	40	0.5	2	80
0	0	0	10	0.5	2	20

* If N = 0 in the observed data, then, regardless of the weight value, there is nobody to be reweighted, so N = 0 in the pseudopopulation too.

TABLE A8. Pseudopopulation Created by Inverse Probability of Treatment Weighting from a Point-Treatment Study in Which Eq A2 Holds

	$L_0 = 1$		$L_0 = 0$	
	$A_0 = 1$	$A_0 = 0$	$A_0 = 1$	$A_0 = 0$
$Y = 1$	0	240	40	80
$Y = 0$	0	160	60	20
Total	0	400	100	100

TABLE A9. Crude Data from the Pseudopopulation of Table A8

	$A_0 = 1$	$A_0 = 0$
$Y = 1$	40	320
$Y = 0$	60	180
Total	100	500