STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:749–767 (DOI: 10.1002/sim.1657)

Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures

Babette A. Brumback[1,*,†], Miguel A. Hernán[2], Sebastien J. P. A. Haneuse[3] and James M. Robins[2,4]

[1]*Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095, U.S.A.*
[2]*Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115, U.S.A.*
[3]*Department of Biostatistics, University of Washington, Seattle, Washington 98121, U.S.A.*
[4] *Department of Biostatistics, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115, U.S.A.*

SUMMARY

Robins introduced marginal structural models (MSMs) and inverse probability of treatment weighted (IPTW) estimators for the causal effect of a time-varying treatment on the mean of repeated measures. We investigate the sensitivity of IPTW estimators to unmeasured confounding. We examine a new framework for sensitivity analyses based on a nonidentifiable model that quantifies unmeasured confounding in terms of a sensitivity parameter and a user-specified function. We present augmented IPTW estimators of MSM parameters and prove their consistency for the causal effect of an MSM, assuming a correct confounding bias function for unmeasured confounding. We apply the methods to assess sensitivity of the analysis of Hernán *et al.*, who used an MSM to estimate the causal effect of zidovudine therapy on repeated CD4 counts among HIV-infected men in the Multicenter AIDS Cohort Study. Under the assumption of no unmeasured confounders, a 95 per cent confidence interval for the treatment effect includes zero. We show that under the assumption of a moderate amount of unmeasured confounding, a 95 per cent confidence interval for the treatment effect no longer includes zero. Thus, the analysis of Hernán *et al.* is somewhat sensitive to unmeasured confounding. We hope that our research will encourage and facilitate analyses of sensitivity to unmeasured confounding in other applications. Copyright © 2004 John Wiley & Sons, Ltd.

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

1. INTRODUCTION

In this paper, we present sensitivity analyses to unmeasured confounding assuming a marginal structural model (MSM) [1] for repeated measures. We re-analyse the results of Hernán,

*Correspondence to: Babette A. Brumback, Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90095, U.S.A.
†E-mail: brumback@ucla.edu

Contract/grant number: NIH Grants AI-29168, AT-00251 and AI-49392

Received December 2002
Accepted June 2003

Copyright © 2004 John Wiley & Sons, Ltd.

Brumback, and Robins (HBR) [2], who use an MSM to estimate the causal effect of zidovudine therapy on repeated CD4 counts among HIV-infected men in the Multicenter AIDS Cohort Study (MACS) [3]. For validity, HBR's analysis requires: (1) correctness of the MSM and the models for probability of treatment initiation conditional on measured confounders and probability of dropout conditional on past treatment and measured confounders, and (2) correctness of HBR's assumption of no additional confounding due to unmeasured factors given the measured time-independent and dependent confounders.

In practice, we expect neither of these assumptions to precisely hold. The statistical literature on sensitivity to model specification (assumption 1) represents several topics, e.g. goodness-of-fit diagnostics and generalized classes of models. Sensitivity to sufficiency of measured confounders (assumption 2) has been addressed by several authors, including Cornfield *et al.* [4], Greenland [5], Lin *et al.* [6], Rosenbaum [7], and Schlesselman [8], but not within the context of MSMs. Recently, Robins [9, 10] proposed a formal method for sensitivity analysis to unmeasured confounding of inverse probability-of-treatment (IPTW) estimators for MSM parameters. In this paper, we present and explain the method, apply it to the MACS analysis of HBR, and prove that the resulting estimator for the causal parameter is consistent when assumption (1) and a non-identifiable user-supplied confounding bias function quantifying the magnitude of unmeasured confounding are both correct.

The paper is organized as follows. In Section 2 we give a brief summary of HBR's MSM analysis of the MACS and describe our notation. Section 3 presents and explains the method for sensitivity analysis and contrasts various bias functions for unmeasured confounding. Section 4 presents the augmented IPTW estimators proposed by Robins [9, 10] for estimating MSM parameters assuming a given model for the unmeasured confounding; it also contains a consistency result proved in the appendix. In Section 5 we apply the sensitivity analysis to the MACS and discuss implications for re-interpreting the results of HBR. Section 6 concludes with a discussion.

2. MSM ANALYSIS OF THE MACS

2.1. 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 [3, 11].

HBR analysed the cohort of HIV-positive men alive in the period during which zidovudine was available for use (i.e. after the study visit 5; March 1986–March 1987). Follow-up ended at study visit 21, October 1994, death, or first missed visit, whichever came first. The analysis includes the $n = 1486$ men (contributing 9752 visits) who attended the first two eligible visits between visits 5 and 21 while HIV-positive, who did not have an AIDS-defining illness, and who were not on antiretroviral therapy before the first eligible visit.

We align the data by study entry; let $Y_i(t)$ be subject i's CD4 count at his tth eligible visit, and $A_i(t)$ be one if he was on zidovudine at visit t, and zero otherwise, $t = 0, \ldots, T \equiv 16$. We assume zidovudine is a monotone treatment; i.e. for the duration of the study it is never

Copyright © 2004 John Wiley & Sons, Ltd.

discontinued once begun. Each subject's data is assumed to be drawn independently from a distribution common to the N subjects. We suppress subject-specific notation, writing $Y(t)$ for $Y_i(t)$. Let V be a vector of time-independent baseline covariates measured at the first eligible visit. 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, diarrhea for 2 weeks, weight loss of $\geqslant 4.5$ kg, oral hairy leukoplakia, or herpes zoster. Let $L(t)$ be the vector of time-varying covariates recording CD4, CD8, WBC, RBC, platelets, and presence of symptoms measured at visit t, $t=1,\ldots,T-1$. Let $L(0)$ be a vector of baseline covariates that contains V. $L(t)$ includes $Y(t)$, and conventionally, $L(t)$ is measured prior to deciding $A(t)$. To denote the censoring process, let $C(t)=0$ if a subject remains in the study beyond time t, 1 otherwise. Then the outcome $Y(t)$ is observed if and only if $C(t)=0$. Subjects' data are regarded as censored at the first missed visit. Let overbars represent a covariate history, e.g. $\bar{A}(t)=\{A(u);u=0,\ldots,t\}$ 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 denote statistical independence, e.g. $A \amalg B|C$ means A is conditionally independent of B given C.

2.2. The potential outcomes

Let $\bar{a}=(a(0),\ldots,a(T))$, represent a possible treatment regimen, where $a(t)$ is 1 if the regimen prescribes treatment at visit t and 0 otherwise, and let $\bar{a}(t)$ represent the treatments prescribed through visit t. Note $\bar{a}=\bar{a}(T)$. Because the treatment is monotone, we need only consider $J=17$ possible treatment regimens—$\bar{a}_0$: begin zidovudine at or after visit 0 but before visit 1, $\bar{a}_j$: begin at or after visit j but before visit $j+1$, $j=1,\ldots,15$, and $\bar{a}_{16}$: begin at or after visit 16. For our purposes, $\bar{a}_{16}$ is equivalent to never starting zidovudine, as we do not consider outcomes measured past visit 16.

Associated with each $\bar{a}$ are the potential or counterfactual outcomes $Y_{\bar{a}}(t)$, which denote a subject's outcome Y at time t had, possibly contrary to the fact, the subject followed treatment plan $\bar{a}$. The subject's observed outcome $Y_i(t)$ is assumed to equal the potential outcome $Y_{\bar{a}}(t)$ when $\bar{a}$ is equal to the treatment regimen that the subject actually followed. That is, $Y(t)\equiv Y_{\bar{A}}(t)$, where $\bar{A}$ is the subject's observed treatment history and $Y_{\bar{a}}(t)=Y_{\bar{A}}(t)$ for $\bar{a}=\bar{A}$.

2.3. The marginal structural model

We model the marginal mean of the potential outcomes for the entire population, assuming no censoring, as

$$E[Y_{\bar{a}}(t)|V]=g(\bar{a}(t-1),V;\beta)$$

$$g(\bar{a}(t-1),V;\beta)=\beta_0+\beta_1\,\mathrm{cum}[\bar{a}(t-1)]+\beta_2 t+\beta_3' V \tag{1}$$

for $t\in 1,\ldots,16$, where $\mathrm{cum}[\bar{a}(t-1)]\equiv\sum_{k=0}^{t-1}a(k)$ and β_3 is a parameter vector.

Copyright © 2004 John Wiley & Sons, Ltd.

2.4. The conditional probabilities of receiving treatment and remaining uncensored

We use logistic regression to model the probability that an individual uncensored as of visit k and untreated as of visit $k-1$ receives treatment at visit k conditional on the history of measured confounders, i.e. for $k=1,\ldots,15$,

$$\text{logit}\, P[A(k)=1|\bar{A}(k-1)=0,\ \bar{L}(k),\bar{C}(k)=0;\theta]=\theta_0(k)+\theta_1'V+\theta_2'L(k) \tag{2}$$

Note, $P[A(k)=1|\,A(k-1)=1,\ \bar{A}(k-2),\ \bar{L}(k)]\equiv 1$ because we assume zidovudine is a monotone treatment.

We also use logistic regression to model the probability of remaining uncensored through visit k conditional on the history of measured confounders, i.e. for $k=0,\ldots,15$,

$$\text{logit}\, P[C(k+1)=0|C(k)=0,\ \bar{A}(k),\ \bar{L}(k);\gamma]=\gamma_0(k)+\gamma_1'V+\gamma_2'L(k) \tag{3}$$

2.5. The measured confounders

The measured confounders used to estimate β_1 based on outcomes at visit t are contained in $\bar{L}(t-1)$. HBR assume that given $\bar{L}(t-1)$, there are no additional unmeasured confounders for estimating the causal effect of $\bar{A}(t-1)$ on $Y(t)$, i.e. that

$$Y_{\bar{a}}(t)\amalg A(k)|\bar{A}(k-1),\bar{L}(k),\quad \forall\bar{a}\text{ and }t>k. \tag{4}$$

This assumption would be true if all prognostic factors for $Y_{\bar{a}}(t)$ 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, the assumption would be suspect if $\bar{L}(k)$ did not contain history of CD4 count. Equation (4) is also referred to as the assumption of sequential randomization.

In general, the measured baseline confounders are not sufficient to estimate the effect of a time-varying treatment, and thus time-dependent confounders are also incorporated into the analysis. In this paper, we assess the sensitivity of HBR's analysis to the assumption of sufficient measured time-dependent confounders; i.e. we reanalyse the data assuming (4) does not hold.

We will continue to assume, as do HBR, that the censoring process is ignorable given $\bar{L}(t-1)$ and $\bar{A}(t-1)$; i.e. $\forall t$

$$Y_{\bar{a}}(j);j\geqslant t\amalg C(t)|C(t-1)=0,\ \bar{A}(t-1),\bar{L}(t-1) \tag{5}$$

2.6. IPTW (inverse probability of treatment weighted) estimation

Model (1) cannot be fit by standard regression methods because it involves counterfactual random variables. However, one can get consistent estimates by instead fitting an ordinary regression with specially constructed weights. The inverse probability of treatment weighted (IPTW) estimator of β_1 is the weighted least squares estimator of η_1 in the repeated measures regression

$$E[Y(t)|\bar{A}(t-1),\bar{L}(t-1)]=g(\bar{A}(t-1),V;\eta)$$

$$g(\bar{A}(t-1),V;\eta)=\eta_0+\eta_1\,\text{cum}[\bar{A}(t-1)]+\eta_2 t+\eta_3'V,\quad t=1,\ldots,16$$

Copyright © 2004 John Wiley & Sons, Ltd.

with weights

$$W(t-1)=\prod_{k=0}^{t-1}\frac{f[A(k)|\bar{A}(k-1),\bar{C}(k-1)=0,V]Pr[C(k+1)=0|C(k)=0,\ \bar{A}(k),V]}{f[A(k)|\bar{A}(k-1),\bar{L}(k-1),\bar{C}(k-1)=0;\theta]Pr[C(k+1)=0|C(k)=0,\bar{A}(k),\bar{L}(k);\gamma]} \tag{6}$$

where we define $A(-1)=\bar{a}(-1)=0$ for all $\bar{a}$, and subjects do not contribute further to the analysis once censored. We estimated the conditional probabilities in the denominator using logistic regression with models (2) and (3). We used similar logistic regression models without the $\bar{L}(k)$ term to estimate the conditional probabilities in the numerator. The functional forms of the models (2) and (3) must be correctly specified in order for the IPTW estimator of β_1 to be consistent. In contrast, the numerator models may be misspecified.

Estimates are computed using standard software for repeated measures that accommodates weighted observations and a misspecified within-person working covariance matrix. We use SAS Proc Genmod with the repeated and scale weights options, and specify an independent working covariance matrix.

When (4) does not hold, unmeasured confounding may bias the IPTW estimator. The next section formalizes the problem, and Section 4 presents augmented IPTW estimators as its solution.

3. SENSITIVITY ANALYSIS FOR UNMEASURED CONFOUNDING

3.1. Confounding at baseline only

For intuition, we introduce the method in terms of a univariate end-of-study outcome Y (CD4 count at visit 16), a binary treatment A administered at baseline (zidovudine status at visit 0), and measured baseline confounders V. In this setting, there are two potential outcomes per patient, Y_0 following no zidovudine at baseline and Y_1 following initiation of zidovudine at baseline. Thus $U=(Y_0,Y_1)$, albeit unmeasured, would be a sufficient set of confounders for estimating the parameters of (1), because were U observed, model (1) would resolve into an ordinary regression expressed in terms of measured outcomes.

The method proposed by Robins [9, 10] begins by quantifying the unmeasured confounding through

$$c(a,v)=E[Y_a|A=a,V=v]-E[Y_a|A=1-a,V=v]. \tag{7}$$

This function represents, among the subgroup of men with $V=v$, the average difference in potential outcomes Y_a between those treated with $A=a$ and those treated with $A=1-a$. With $c(a,v)\neq 0$, the sufficient measured confounders assumption $Y_a \amalg A|V$, $a=0,1$ does not hold. When $c(a,v)\equiv 0$, we say there is no unmeasured confounding for the mean outcome. The condition $c(a,v)\equiv 0$ is sufficient for the IPTW estimator of β to be consistent.

Ignoring unmeasured confounders for the mean will typically bias the estimated causal effect. For example, estimating

$$\theta=E[Y_1-Y_0|V=v] \tag{8}$$

with the naive estimand

$$\theta_{\text{naive}}=E[Y|A=1,V=v]-E[Y|A=0,V=v] \tag{9}$$

Copyright © 2004 John Wiley & Sons, Ltd.

leads to a bias of

$$c(1,v)P[A=0|V=v] - c(0,v)P[A=1|V=v] \tag{10}$$

because

$$\begin{aligned}
E[Y_1|V=v] &= E[Y_1|A=1,V=v]P[A=1|V=v] + E[Y_1|A=0,\ V=v]P[A=0|V=v] \\
&= E[Y_1|A=1,V=v]P[A=1|V=v] + (E[Y_1|A=1,V=v] - c(1,v))P[A=0|V=v] \\
&= E[Y_1|A=1,V=v] - c(1,v)P[A=0|V=v] \\
&= E[Y|A=1,V=v] - c(1,v)P[A=0|V=v]
\end{aligned}$$

and similarly, $E[Y_0|V=v] = E[Y|A=0,V=v] - c(0,v)P[A=1|V=v]$.

If $c(1,v)>0$ but $c(0,v)<0$, then on average, treated individuals will have higher potential outcomes to both treatment and no treatment than untreated individuals; i.e. healthier men are treated. On the other hand, $c(1,v)<0$ but $c(0,v)>0$ suggests confounding by indication [12]; i.e. unhealthier men are treated.

When $c(a,v)>0$, the potential outcome to treatment with a is better, on average, among those who choose it than among those who do not. Thus, the observed treatment allocation is beneficial relative to the alternative which reverses treatment assignment for everyone. Likewise, when both $c(1,v)<0$ and $c(0,v)<0$, the observed treatment allocation is undesirable relative to the alternative which reverses treatment assignment for everyone.

We next consider six functional forms for $c(a,v)$, which we later generalize for time-dependent confounding and repeated outcomes. The forms are

$$\begin{aligned}
c_1(a,v) &= \alpha(2a-1) \qquad c_2(a,v) = \alpha \qquad c_3(a,v) = \alpha a \\
c_4(a,v) &= \alpha(2a-1)v_1 \quad c_5(a,v) = \alpha v_1 \quad c_6(a,v) = \alpha a v_1
\end{aligned}$$

where v_1 is a real-valued function of the baseline covariates in v. Accompanying each one is a scale parameter α which quantifies the magnitude and direction of unmeasured confounding.

For unmeasured confounding as in c_1, the causal effect is identical in the treated and untreated groups, i.e. $E[Y_1-Y_0|A=1,V=v] = E[Y_1-Y_0|A=0,V=v]$. When $\alpha>0$, the potential outcomes to treatment and to no treatment tend to be better among the treated than the untreated. Thus, the treated individuals tend to be healthier than the untreated, and θ_{naive} is biased upwards by α. When $\alpha<0$, the untreated individuals tend to be healthier than the treated, and the naive estimator is biased downwards by α.

Under c_2, $E[Y_1 - Y_0|A=1,V=v] = E[Y_1 - Y_0|A=0,V=v] + 2\alpha$, i.e. the causal effect differs between the treated and untreated. In this case, the unmeasured confounders induce effect-measure modification. When $\alpha>0$ ($\alpha<0$), the observed treatment allocation is beneficial (harmful) relative to the alternative which reverses treatment assignment for everyone. The naive estimator is biased by $\alpha(P[A=0|V=v] - P[A=1|V=v])$. For v such that $P[A=0|V=v] = P[A=1|V=v] = 0.5$, the naive estimator is unbiased. When $P[A=0|V=v] > P[A=1|V=v]$, i.e. when proportionally more people are untreated than treated within the v-stratum, the naive estimator is biased upwards for $\alpha>0$ and downwards for $\alpha<0$. When

Copyright © 2004 John Wiley & Sons, Ltd.

proportionally more people are treated than untreated, the naive estimator is biased downwards for $\alpha > 0$ and upwards for $\alpha < 0$.

Under $c_3, E[Y_1 - Y_0|A=1, V=v] = E[Y_1 - Y_0|A=0, V=v] + \alpha$, i.e. when $\alpha > 0$ ($\alpha < 0$) the causal effect is higher (lower) in the treated subgroup, and the naive estimator is biased upwards (downwards) by $\alpha P[A=0|V=v]$. The mean of Y_1 is α units higher in the treated, whereas the mean of Y_0 is equal in the treated and untreated.

The functions c_4, c_5, and c_6 model interactions with the baseline covariate v_1, e.g. CD4 count. Since CD4 counts are positive, the functions possess greater magnitude for higher as opposed to lower baseline CD4 counts. For example, c_4 with $\alpha > 0$ suggests that treated individuals with high baseline counts tend to be much healthier at baseline than untreated individuals with high baseline counts. The treated individuals with low baseline counts also tend to be healthier at baseline than their untreated counterparts, but this difference is less pronounced.

In the specific context of the HBR example, we regard it as a priori unlikely that $c(1, v)$ and $c(0, v)$ have different signs as this implies that within levels of V the treated have better prognosis under treatment and the untreated have better prognosis under no treatment; in general, individuals with good prognosis under antiretroviral therapy (ART) also tend to have good prognosis under no such therapy as compared to individuals with poor prognosis under ART. Thus, for the example of this paper, we regard c_2, c_3, c_5, and c_6 as less plausible than c_1 and c_4. Nonetheless, because the functional form of $c(a, v)$ is not identified we will examine, in a sensitivity analysis, all six of the above possibilities.

Note, as discussed by Robins [9], if one wants to allow for the possibility that the sharp null hypothesis that $Y_1 = Y_0$ holds for all individuals in the population, then one must choose a function anti-symmetric in a such as c_1 and c_4, i.e. such that $c(1, v) = -c(0, v)$.

For a user-specified choice of $c(a, v)$, e.g. c_1 with $\alpha = 50$, we can obtain unbiased estimates of (8) by replacing each observed outcome Y with the adjusted outcome $Y^\alpha = Y - c(A, V)$ $f[1 - A|V]$ and recalculating the naive estimand as

$$\bar{\theta}^\alpha = E[Y^\alpha|A=1, V] - E[Y^\alpha|A=0, V]$$

Note that θ^α effectively subtracts the bias term (10) from θ_{naive}. Consider for example c_1 with $\alpha > 0$, which implies that the treated are healthier than the untreated. To counteract bias, we adjust the observed outcome of a treated individual downward by subtracting $\alpha P(A=0|V)$, and that of an untreated individual upward by adding $\alpha P(A=1|V)$.

3.2. Time-dependent confounding and repeated measures

To perturb the sufficient measured confounders assumption (4)

$$Y_{\bar{a}}(t) \amalg A(k) | \bar{A}(k-1), \bar{L}(k), \ \forall \bar{a} \text{ and } t > k$$

Robins [9, 10] generalized (7). For $A(k)$ dichotomous the generalization is

$$\begin{aligned} c(t, k, \bar{a}(t-1), \bar{l}(k)) = {} & E[Y_{\bar{a}}(t)|\bar{A}(k) = \bar{a}(k), \bar{L}(k) = \bar{l}(k)] \\ & - E[Y_{\bar{a}}(t)|A(k) = 1 - a(k), \ \bar{A}(k-1) = \bar{a}(k-1), \ \bar{L}(k) = \bar{l}(k)] \quad (11) \end{aligned}$$

Copyright © 2004 John Wiley & Sons, Ltd.

for $t>k$ and $\bar{a}(k-1)$ such that $P[A(k)=a(k)|\bar{A}(k-1)=\bar{a}(k-1),\bar{L}(k)=\bar{l}(k);\theta]$ exists and is bounded away from 0 and 1; otherwise, $c\equiv 0$. Conditionally on the history of confounders and treatments observed at each visit k, this function quantifies the effect of unmeasured confounding on the mean of future repeated outcomes at $t=k+1,\ldots,T$.

By specifying a variety of forms for $c(t,k,\bar{a}(t-1),\bar{l}(k))$, we can conduct a sensitivity analysis of the MACS MSM analysis to several types of unmeasured confounding. Here, we generalize the functional forms of Section 3.1 to

$$c_1=\alpha(2a(k)-1) \quad c_2=\alpha \quad c_3=\alpha a(k)$$

$$c_4=\alpha(2a(k)-1)l_1(k) \quad c_5=\alpha l_1(k) \quad c_6=\alpha a(k)l_1(k)$$

where $l_1(k)$ is a real-valued function of the covariates in $\bar{l}(k)$. Note, for monotone treatments such as zidovudine in our MACS example, $c(t,k,\bar{a}(t-1),\bar{l}(k))\equiv 0$ for $k>\text{init}(\bar{a})$, where $\text{init}(\bar{a})\equiv \min_k(a(k)=1)$.

For example, when unmeasured confounding is of type C_1 and subjects never stop therapy once begun, the mean difference in potential outcomes at t to treatment with $\bar{a}$ between individuals actually treated with $\bar{a}(k)$ and those treated with $\bar{a}(k-1)$ followed by $1-a(k)$, conditional on observing $\bar{L}(k)=\bar{l}(k)\bar{A}(k-1)=0$, is $-\alpha$ when $a(k)=0$ and α when $a(k)=1$. The difference is 0 when $a(k)=a(k-1)=1$. Supposing that $\bar{a}$ prescribes initiation at visit $j<t$, the difference is $-\alpha$ for $k<j,\alpha$ for $k=j$, and 0 for $t>k>j$. Thus, $\alpha=50$ means that, given available information through time j, the CD4 counts at $t>j$ of individuals initiating treatment at j are 50 units higher on average than CD4 counts of individuals initiating treatment later or never would have been, had the late initiators also begun at j. Note, this does not have to imply that actual CD4 counts at j are lower in the late initiators.

4. AUGMENTED IPTW ESTIMATION

The method of adjustment proposed in Section 3.1 generalizes as follows for time-dependent confounding and repeated measures. First, compute the adjusted outcome

$$Y^{\alpha}(t)\equiv Y(t)-\sum_{k=0}^{t-1}c(t,k,\bar{A}(t-1),\bar{L}(k))f[1-A(k)|\bar{A}(k-1),\bar{L}(k);\theta] \tag{12}$$

and then, recompute the IPTW estimator of β_1 with $Y^{\alpha}(t)$ in place of $Y(t)$, $t=1,\ldots,T$. In the appendix, we prove that subject to weak regularity conditions the resulting *augmented* IPTW estimator is consistent for β_1, assumption (1) of Section 1, and correct specification of $c(t,k;\bar{A}(t-1),\bar{L}(k))$ for each t,k such that $t>k$ and $t=1,\ldots,T$.

For monotonic treatments such as zidovudine in our example, the adjustment to $Y(t)$ simplifies as follows. For patients who initiate treatment at visit $m\leqslant t-1$, the sum in (12) reduces to

$$\sum_{k=0}^{m-1}c(t,k,\bar{A}(t-1),\bar{L}(k))P[A(k)=1|\bar{A}(k-1)=0,\bar{L}(k)]+c(t,m,\bar{A}(t-1),\bar{L}(m))$$

$$\times P[A(m)=0|\bar{A}(m-1)=0,\bar{L}(m)] \tag{13}$$

Copyright © 2004 John Wiley & Sons, Ltd.

whereas for those initiating treatment at $m \geqslant t$, it remains

$$\sum_{k=0}^{t-1} c(t,k,0,\bar{L}(k))P[A(k)=1|\bar{A}(k-1)=0,\bar{L}(k)] \tag{14}$$

For example, with unmeasured confounding of type $c_3 = \alpha a(k)$, (13) equals $\alpha P[A(m)=0|\bar{A}(m-1)=0,\bar{L}(m)]$, whereas (14) equals 0, because under c_3, measured confounders are sufficient for estimating potential outcomes following no zidovudine treatment; i.e. unmeasured causes of $Y_{\bar{a}}(t)$ do not influence treatment prior to visit init($\bar{a}$).

4.1. Estimating standard errors

When the weights at (6) are estimated based on a correct parametric model, the IPTW estimator of a MSM parameter assuming no unmeasured confounding is more efficient than its counterpart based on known weights. However, the AIPTW estimator of a MSM parameter in the presence of quantified unmeasured confounding does not necessarily possess this property, because the likelihood does not factorize into separate components, one containing β_1 of (1) and the other θ of (2). Therefore, in Section 5 we compare bootstrap standard errors, based on resampling subjects with replacement and re-estimating the weights, with naive standard errors, correct for known weights but incorrect for estimated weights. Another option would be to compute a sandwich estimator (see p. 422 of Reference [13]) that accounts not only for an incorrectly specified working covariance matrix but also for estimated weights.

5. APPLICATION TO THE MACS

Tables I and II summarize sensitivity to unmeasured confounding of the MSM analysis of HBR and an inverse probability of censoring (IPCW) analysis that adjusts for censoring and baseline covariates but fails to adjust for confounding of treatment by the measured time-dependent covariates. Thus one would expect a much greater degree of unmeasured confounding in the IPCW analysis than in the HBR MSM analysis. The IPCW analysis is almost identical to the MSM analysis, but with weights

$$W_{\mathrm{IPCW}}(t-1) = \prod_{k=0}^{t-1} \frac{\Pr[C(k+1)=0|C(k)=0,\bar{A}(k),V]}{\Pr[C(k+1)=0|C(k)=0,\bar{A}(k),\bar{L}(k);\gamma]}$$

substituted for $W(t-1)$ at (6). This yields a consistent estimator of β_1 in (1) under assumptions almost identical to those of the MSM analysis, but with

$$Y_{\bar{a}}(t) \amalg A(k)|\bar{A}(k-1),V \quad \forall \bar{a} \text{ and } t>k.$$

substituted for (4); i.e. variables in V are assumed sufficient for estimating the causal effect of $\bar{A}(t)$ on $Y(t+1)$. The augmented IPTW estimator is almost identical to the one used for the MSM sensitivity analysis, but again with weights $W_{\mathrm{IPCW}}(t-1)$ rather than $W(t-1)$.

In both tables, sensitivity to unmeasured confounding of types $c_1,\ldots,c_6$ as specified in Section 3.2 is reported for several values of the scale parameter α. In c_4—c_6, $l_1(k)$ is the indicator variable that is 1 is CD4 count at baseline (time 0) exceeds 500 and is zero otherwise. The first row of each table includes results of the original MSM or IPCW analysis for comparison.

Copyright © 2004 John Wiley & Sons, Ltd.

Table I. Sensitivity of the MSM analysis to unmeasured confounding.

Sensitivity function	α	Effect estimate	Naive SE	Bootstrap SE	Nominal effect direction
—		5.4	3.7	3.8	0
c_1	200	−42.0	3.9	4.7	−
	100	−18.3	3.5	3.8	−
	50	−6.4	3.5	3.7	0
	−50	17.3	4.0	4.2	+
c_2	500	15.0	8.3	11.5	0
	200	9.3	5.2	6.3	0
	100	7.4	4.4	4.9	0
	50	6.4	4.0	4.3	0
	−50	4.5	3.5	3.6	0
	−500	−4.1	5.5	8.9	0
c_3	200	−16.4	3.8	4.0	−
	100	−5.5	3.7	3.9	0
	50	0.0	3.7	3.9	0
	−50	10.9	3.7	3.8	+
c_4	200	−20.9	3.4	3.8	−
	100	−7.7	3.2	3.3	−
	50	−1.1	3.4	3.4	0
	−50	10.9	3.7	4.4	+
c_5	500	9.2	8.4	9.5	0
	200	6.9	5.1	5.7	0
	100	6.2	4.3	4.6	0
	50	5.8	3.9	4.2	0
	−50	5.1	3.5	3.6	0
	−500	1.7	6.3	7.0	0
c_6	200	−7.0	3.5	3.5	−
	100	−0.8	3.6	3.6	0
	50	2.3	3.6	3.7	0
	−50	8.6	3.8	4.0	+

The columns contain augmented IPTW estimates of β_1, naive standard errors computed as in HBR assuming the weights are known rather than estimated, and bootstrap standard errors based on resampling subjects with replacement. The naive asymptotic standard errors for the original IPTW and IPCW estimators tend to be conservative, but for the augmented IPTW estimators they may be larger or smaller than standard errors computed using a sandwich estimator of variance that accounts for estimating the weights. We also report nominal effect directions, $-, +$, or 0, based on rejection regions of a 0.05-level two-sided hypothesis test of no treatment effect. In both tables, these results are unaltered when bootstrap standard errors are substituted for naive standard errors.

In Table I, the effect estimate assuming no unmeasured confounding is 5.4 with an SE of 3.7. The bootstrap SE of 3.8 is not much larger; the discrepancy with asymptotic theory,

Copyright © 2004 John Wiley & Sons, Ltd.

Table II. Sensitivity of the IPCW analysis to unmeasured confounding.

Sensitivity function	α	Effect estimate	Naive SE	Bootstrap SE	Nominal effect direction
—		−18.0	2.9	2.9	−
c_1	50	−26.9	2.9	2.8	−
	−50	−9.1	3.0	3.0	−
	−100	−0.2	3.1	3.1	0
	−200	17.6	3.3	3.3	+
	−300	35.4	3.6	3.7	+
c_2	50	−19.7	3.1	3.0	−
	−50	−16.3	2.8	2.8	−
	−100	−14.7	2.7	2.7	−
	−200	−11.3	2.6	2.6	−
	−300	−7.9	2.6	2.7	−
	−500	−1.2	2.8	3.2	0
c_3	50	−23.3	3.0	2.9	−
	−50	−12.7	2.9	2.9	−
	−100	−7.4	2.9	2.8	−
	−200	3.1	2.8	2.8	0
	−300	13.7	2.8	2.9	+
c_4	50	−22.7	2.9	2.8	−
	−50	−13.3	3.0	2.9	−
	−100	−8.7	3.1	3.0	−
	−200	0.7	3.3	3.2	0
	−300	10.0	3.6	3.6	+
c_5	50	−19.3	3.0	3.0	−
	−50	−16.7	2.8	2.8	−
	−100	−15.3	2.8	2.7	−
	−200	−12.7	2.7	2.7	−
	−300	−10.0	2.8	2.7	−
	−500	−4.7	3.1	3.1	0
c_6	50	−21.0	3.0	2.9	−
	−50	−15.0	2.9	2.9	−
	−100	−12.0	2.8	2.8	−
	−200	−6.0	2.9	2.9	−
	−300	0.0	3.0	2.9	0
	−500	12.0	3.2	3.2	+

which would entail an SE smaller than 3.7, suggests our sample is too small for large sample theory to precisely hold. The nominal effect direction is zero. For each confounding function c_1 through c_6, we searched for magnitudes of α yielding positive or negative directions. For c_1, c_3, c_4, and c_6, negative values of α moderate in size (e.g. $\alpha = -50$) produce a positive nominal effect, whereas larger positive values (e.g. $\alpha = 100$ or 200) are needed to produce a negative nominal effect. For c_2 and c_5, even improbable values of α such as $+/-500$ are not consistent with a positive or negative treatment effect.

Copyright © 2004 John Wiley & Sons, Ltd.

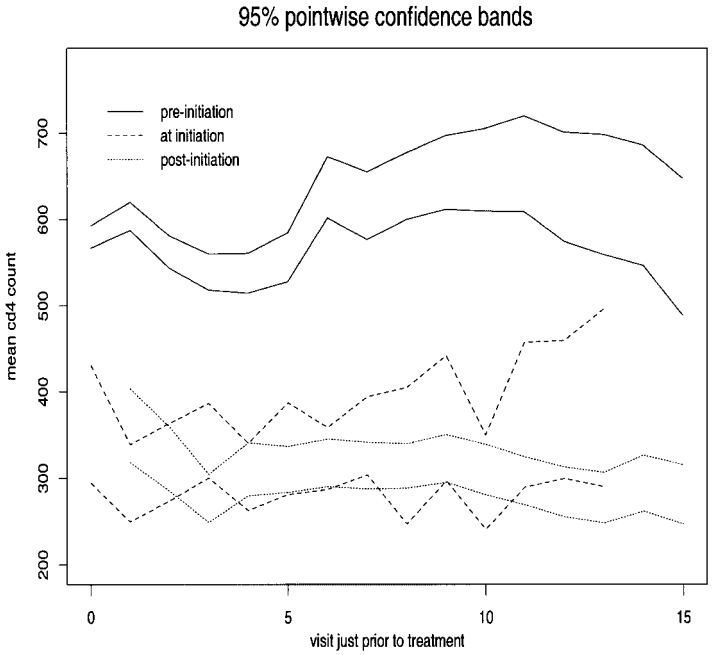

Figure 1. 95 per cent pointwise confidence bands for mean CD4 count by visit number and treatment status.

In Table II, the effect estimate assuming no time-dependent confounding and no unmeasured baseline confounding is -18.0 with naive and bootstrap SE's of 2.9. For all confounding functions except c_1, α must exceed -100 before the negative nominal effect disappears, and even for c_1, $\alpha = -50$ is insufficient.

Figure 1 presents 95 per cent pointwise confidence bands based on a crude unweighted analysis for mean CD4 counts stratified by visit and treatment status (pre-initiation, at initiation, and post-initiation); Table III lists the number of participants in each stratum. Recall that CD4 count is measured just prior to treatment decision at each visit. Suppose for a moment that initiating zidovudine at t did not affect outcomes at $t+1$; then, we might accept the observed difference in mean CD4 counts between initiates and pre-initiates at t as a rough approximation to the counterfactual difference at visit $t+1$. Figure 1 suggests the difference is nearly constant across t and equal to -250. Now the results of randomized trials have shown that the effect of AZT on mean increase in CD4 count is relatively small (less than 25). This leads us to believe that α's of -200 or -300 are plausible in Table II. Indeed, neglected time-dependent confounding of this magnitude enables AIPTW effect estimates in Table II to agree with the IPTW estimate of HBR, which equals 5.4.

In both tables, but especially Table I, results for c_2 and c_5 are unusual in that plausible magnitudes of confounding do not alter the nominal effect direction. However, as argued earlier, it is unlikely that c_2 or c_5 are plausible models for the effect of unmeasured confounders in this application.

Copyright © 2004 John Wiley & Sons, Ltd.

Table III. Number of participants by visit number and treatment status.

Visit	Pre-initiation	At initiation	Post-initiation
0	1883	62	NA
1	1256	90	140
2	916	93	177
3	685	87	223
4	504	70	264
5	400	59	292
6	294	74	294
7	231	40	336
8	202	14	349
9	169	18	310
10	136	19	284
11	119	11	263
12	99	10	237
13	83	9	219
14	70	1	205
15	59	1	156

The SE's for c_2 and c_5 in Table I increase with increasing $|\alpha|$. Figure 2 illustrates this graphically with a comparison of Tables I and II results for c_1 and c_2. Effect estimates and 95 per cent pointwise confidence bands are plotted versus the scale parameter α. The results demonstrate that effect modification by unmeasured confounders, as induced by c_2, combined with IPTW balancing of measured confounders within treatment groups, leads to increasingly variable effect estimates for large $|\alpha|$. The observations concerning variability reflect the fact that under c_1 the difference between the estimate $\hat{\beta}_1(\alpha) - \hat{\beta}_1(\alpha = 0)$ is a nonrandom function of α. However, in Table I, under c_2 and c_5, $\hat{\beta}_1(\alpha) - \hat{\beta}_1(\alpha = 0)$ depends on the product of α and a function of the estimator $\hat{\theta}$ of the parameter θ of the model (2) for the treatment process, resulting in an increase in the variance of $\hat{\beta}_1(\alpha)$ with $|\alpha|$ due to the variability in $\hat{\theta}$. In contrast, no estimate of the treatment process is required to construct Table II and thus variability does not increase with $|\alpha|$. For all analyses, effect estimates are linear in α because we use a linear MSM. Effect estimates are typically nonlinear in α for nonlinear MSMs.

6. DISCUSSION

We used a non-identified structural model to quantify unmeasured confounding in a marginal structural model analysis of repeated measures. Because we estimated the parameter of a linear MSM, we used a linear model for the unmeasured confounding. Extensions to nonlinear MSMs will sometimes pair better with a nonlinear model for unmeasured confounding, as discussed by Robins *et al.* [14].

We used our methodology to roughly estimate the magnitude of confounding due to the measured time-dependent covariates. Section 5 illustrates this by comparing an IPTW analysis to an IPCW analysis that effectively sets coefficients of time-dependent confounders in the PT model to zero. When the coefficients of the time-dependent confounders are nonzero,

Copyright © 2004 John Wiley & Sons, Ltd.

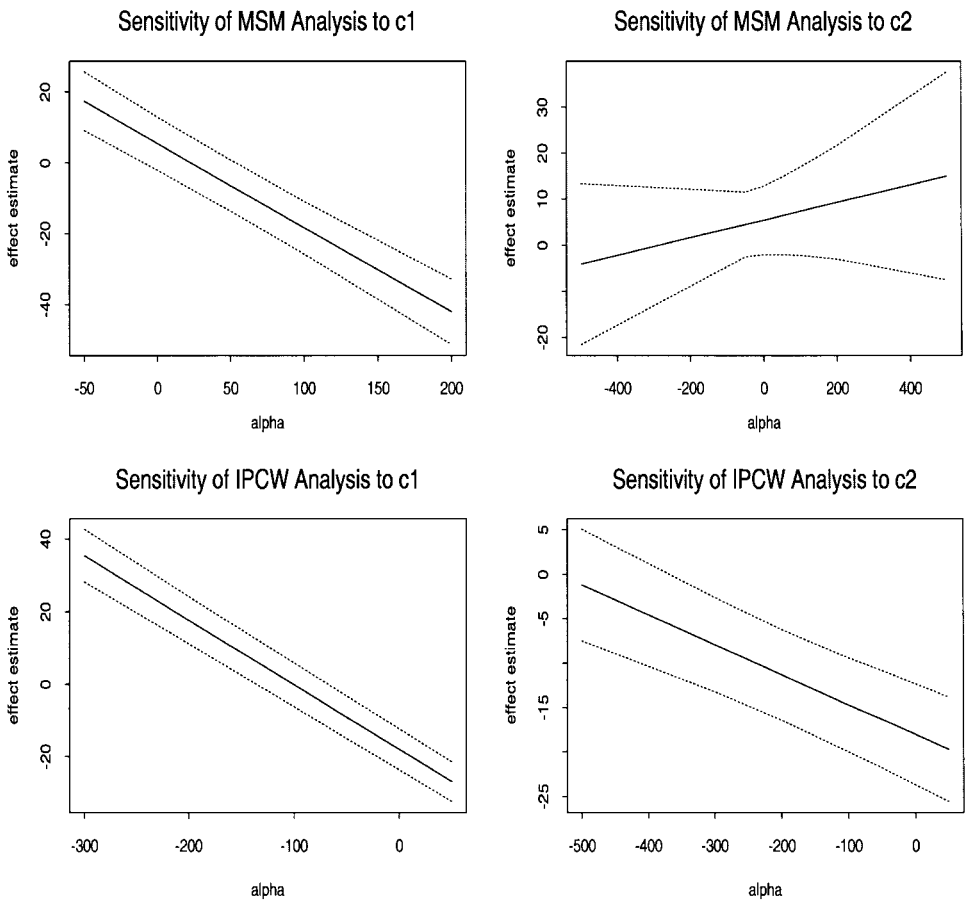

Figure 2. Graphical presentations of the sensitivity analyses based on c_1 and c_2. Effect estimates (solid line) and 95 per cent pointwise confidence bands (dotted line) are graphed versus the scale parameter, α.

neglected measured time-dependent confounders function just like unmeasured confounders. We varied the sensitivity parameter in the IPCW analysis to find values that reproduced the IPTW estimate of HBR. Assuming no unmeasured confounders, this allows us to quantify neglected time-dependent confounding.

We quantified the net confounding on the mean of the outcome $Y_{\bar{a}}$ through the selection bias function c without any reference to the unmeasured common causes U of treatment and the outcome, which are the source of this confounding. A large body of previous work, originating with Cornfield [4] on sensitivity analysis in causal inference models with time-independent treatments has tried to directly model the effect of these unmeasured causes U. In such a sensitivity analysis, one varies the association of U with the outcome Y (within levels of treatment and measured confounders) and the association of U with the treatment (within levels of measured confounders) [8, 15, 6]. In contrast, in our approach, we simply

Copyright © 2004 John Wiley & Sons, Ltd.

model the association of the mean of the counterfactual outcome variable with the treatment (within levels of measured confounders). The advantage of our approach is that (i) there are fewer sensitivity parameters to vary, and (ii) the (essentially impossible) decision as to whether to view U as univariate or multivariate, continuous or discrete is done away with. A link between the two approaches is that the vector of 17 counterfactual outcomes $Y_{\bar{a}}$ can be considered the ultimate unmeasured confounder U. This reflects the fact that, given the counterfactuals and treatment, other unmeasured covariates are superfluous, since the observed outcome variable is a deterministic function of the treatment and the counterfactual outcome.

In our opinion, the unmeasured confounder U approach should be preferred to our counterfactual approach in circumstances where (i) U represents a known confounder (e.g. cigarette smoking) that for logistical reasons was not measured in a particular study, and furthermore, (ii) there exists reasonable historical knowledge about the magnitude of association of U with both the outcome (conditional on treatment and measured confounders) and the treatment (conditional on measured confounders). In contrast, when U is to represent all possible unmeasured factors, we believe that is substantively easier for subject-matter experts to give their opinions about the plausible magnitude of the association of the mean of the counterfactual outcome with treatment than about the modality of any unmeasured confounders U (e.g. continuous or discrete, single or multidimensional) and the strength of associations of such confounders with treatment and the outcome. Furthermore, we have seen that our counterfactual approach leads to extremely simple computations that can be carried out with standard software. In contrast, as discussed by Lin *et al.* [6], there can be formidable computational difficulties associated with the approach based on positing an unmeasured covariate U.

Competitors to sensitivity analysis as means for summarizing uncertainty due to confounding by unmeasured factors about causal effects in observational studies include formal Bayesian inference and computing bounds for the causal effect. In observational studies the bounds are, in general, too wide to be very useful since they tend always to include the null hypothesis of no treatment effect. We view a Bayesian analysis as complementary to a sensitivity analysis. A sensitivity analysis accurately reports what we can learn from the data in a statement of the form 'if this is the degree of confounding due to unmeasured factors, then this is what we can conclude from the data'. If a decision has to be made, we need to go further, and a natural direction would be to place a prior distribution on the nonidentified parameter α (and indeed on the functional form of c) as well as on the other parameters of the model. Thus, we regard our sensitivity analysis as pre-processing for a full Bayesian analysis. Mathematical details are described in Section 11 of Reference [9], although the discussion there is restricted to rather simple settings because of technical problems with implementing nonparametric Bayesian procedures.

Topics for future research include calibrating unmeasured confounding for reliable input by subject-matter experts with limited statistical background. Though we have investigated six types of unmeasured confounding in this paper, further application would explore other types and add insight into our current methods. For example, we question the utility of functions like c_2, c_3, c_5, and c_6, which induce effect modification by the unmeasured confounder. For nonzero values of α, these functions overrule the sharp null hypothesis of no treatment effect in any of the individuals under study. On the other hand, they quantify an implication of the real possibility that doctors are treating patients either correctly or incorrectly via intuition not captured by measured variables. Brumback *et al.* [16] introduced a structural nested

Copyright © 2004 John Wiley & Sons, Ltd.

mean model to quantify effect modification by treatment indication as a function of measured variables; the relationship between their method and our sensitivity analysis of the IPCW results (Table II), useful for assessing the magnitude of neglected confounding due to measured variables, is of further interest.

In summary, our analysis offers a new perspective on previous results. We find the analysis of HBR to be somewhat sensitive to unmeasured confounding. We hope that our research will encourage and facilitate analyses of sensitivity to unmeasured confounding in more general applications.

APPENDIX

We prove the theorem for uncensored data, in which case the conditional probabilities of remaining uncensored used to compute the weights $W(t-1)$ at (6) are all equal to one. It is straightforward to show that the AIPTW estimator remains consistent for right censored data due to loss-to-followup, assuming ignorability as in (5).

Theorem
Assuming (i) consistency and stability of the potential outcomes, i.e.

$$Y(t) = Y_{\bar{a}_j}(t) | A(0) = a_j(0), \ldots, A(t-1) = a_j(t-1), \quad \forall t \in 1, \ldots, T \quad \text{and} \quad \forall j \in 0, \ldots, J$$

(ii) correct specification of the MSM (1), (iii) correct specification of the association model (2), (iv) correct specification of the unmeasured confounding function $c(t, k, \bar{a}(t-1), \bar{l}(k))$, and (v) ordinary statistical regularity conditions, the AIPTW estimator of β_1 that solves

$$U(\beta) \equiv \sum_{i=1}^{N} \sum_{t=0}^{T-1} b_t(\bar{A}_i(t), V_i) W_i(t) (Y_i^a(t+1) - g(\bar{A}_i(t), V_i; \beta)) = 0$$

where $b_t(\bar{A}(t), V)$ is a user-specified function of $(\bar{A}(t), V)$ (e.g. in this paper we set $b_t(\bar{A}(t), V) = (1, \text{cum}(\bar{A}(t)), V')'$), is consistent for β_1 of MSM (1).

We use the following lemma:

Lemma
For $t = 1, \ldots, T$,

$$\begin{aligned} & E[Y_{\bar{a}}(t) | V] \\ & = \int_{l(0)} \cdots \int_{l(t-1)} (E[Y_{\bar{a}}(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)] \\ & \quad - \sum_{k=0}^{t-1} (c(t, k, \bar{a}(t-1), \bar{l}(k)) P[1 - a(k) | \bar{A}(k-1) = \bar{a}(k-1), \bar{L}(k) = \bar{l}(k); \theta])) \\ & \quad \times \prod_{k=0}^{t-1} P[l(k) | \bar{A}(k-1) = \bar{a}(k-1), \bar{L}(k) = \bar{l}(k), V; \theta] \, \mathrm{d}l(k). \end{aligned} \tag{A1}$$

Copyright © 2004 John Wiley & Sons, Ltd.

Proof of Lemma
For $0 \leqslant m < t,\ 0 < t \leqslant T$,

$$
\begin{aligned}
&E[Y_{\bar{a}}(t)|\bar{A}(m-1) = \bar{a}(m-1), \bar{l}(m)] \\
&\quad = E[Y_{\bar{a}}(t)|\bar{A}(m) = \bar{a}(m), \bar{l}(m)] - c(t, m, \bar{a}(t-1), \bar{l}(m))P[1 - a(m)|\bar{A}(m-1) = \bar{a}(m-1), \bar{l}(m)] \\
&\quad = \int_{l(m+1)} (E[Y_{\bar{a}}(t)|\bar{A}(m) = \bar{a}(m), \bar{l}(m+1)] \\
&\qquad - c(t, m, \bar{a}(t-1), \bar{l}(m))P[1 - a(m)|\bar{A}(m-1) = \bar{a}(m-1), \bar{l}(m)]) \\
&\qquad\quad P[l(m+1)|\bar{A}(m) = \bar{a}(m), \bar{l}(m)]\, \mathrm{d}l(m+1) \\
&\qquad = \int_{l(m+1)} (E[Y_{\bar{a}}(t)|\bar{A}(m+1) = \bar{a}(m+1), \bar{l}(m+1)] \\
&\qquad\quad - \sum_{k=0}^{1} c(t, m+k, \bar{a}(t-1), \bar{l}(m+k)) \\
&\qquad\quad \times P[1 - a(m+k)|\bar{A}(m+k-1) = \bar{a}(m+k-1), \bar{l}(m+k)]) \\
&\qquad\quad \times P[l(m+1)|\bar{A}(m) = \bar{a}(m), \bar{l}(m)]\, \mathrm{d}l(m+1) \\
&\qquad = \int_{l(m+1)} \cdots \int_{l(t-1)} (E[Y_{\bar{a}}(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{l}(t-1)] \\
&\qquad\quad - \sum_{k=0}^{t-m-1} c(t, m+k, \bar{a}(t-1), \bar{l}(m+k)) \\
&\qquad\quad \times P[1 - a(m+k)|\bar{A}(m+k-1) = \bar{a}(m+k-1), \bar{l}(m+k)]) \\
&\qquad\quad \times \prod_{k=1}^{t-m-1} P[l(m+k)|\bar{A}(m+k-1) = \bar{a}(m+k-1), \bar{l}(m+k)]\, \mathrm{d}l(m+k)
\end{aligned}
$$

where the first equality follows from the double expectation theorem and the definition of c at (11), the second from another application of the double expectation theorem, the third from substituting the first into the second with $m+1$ in place of m, and the fourth by iteration. Fixing $m = 0$ and taking the conditional expectation of both sides over $P[L(0)|V]$ proves (A1).

Copyright © 2004 John Wiley & Sons, Ltd.

Proof of Theorem
Rewriting (A1) as

$$E[Y_{\bar{a}}(t)|V]$$
$$= \int_{l(0)} \cdots \int_{l(t-1)} \left(E[Y_{\bar{a}}(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)] - \sum_{k=0}^{t-m-1} c(t,k,\bar{a}(t-1),\bar{l}(k)) \right)$$
$$W(t-1)P[\bar{L}(t-1) = \bar{l}(t-1)|\bar{A}(t-1) = \bar{a}(t-1), V] \prod_{k=0}^{t-1} \mathrm{d}l(k)$$

leads to

$$E[Y_{\bar{a}}(t)|V] = E[W(t-1)Y^{\alpha}(t)|\bar{A}(t-1) = \bar{a}(t-1), V]$$

by the consistency assumption (i) and the definition of $Y^{\alpha}(t)$ at (12). Thus, $E(U(\beta)|\bar{A}(t-1) = \bar{a}(t-1), V) = 0$, for all $\bar{a}$, and so $U(\beta)$ is an unbiased estimating equation. It follows that the AIPTW estimator of β_1 is consistent for β_1 of MSM (1), by standard asymptotic theory of estimating equations [17].

ACKNOWLEDGEMENTS

This work was supported by NIH Grants AI-29168, AT-00251 and AI-49392 and UW New Investigator Award 62-4773.

REFERENCES

1. 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.
2. Hernán MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine* 2002; **21**:1689–1709.
3. 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.
4. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* 1959; **22**: 173–203.
5. Greenland S. Sensitivity analysis versus Bayesian analysis of possible uncontrolled confounding and response bias in evaluating the relation of magnetic fields to childhood leukemia. *Journal of the American Statistical Association* 2003, in press.
6. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**:948–963.
7. Rosenbaum PR. *Observational Studies*. Springer-Verlag: New York, 1995.
8. Schlesselman JJ. Assessing effects of confounding variables. *American Journal of Epidemiology* 1978; **108**:3–8.
9. Robins JM. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models, Sections 6–11. In *Statistical Models in Epidemiology*: *The Environment and Clinical Trials*, Halloran ME, Berry D (eds). Springer-Verlag: New York, 1999; 1–92.
10. Robins JM. Association, causation, and marginal structural models. *Synthese* 1999; **121**:151–179.
11. 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.

Copyright © 2004 John Wiley & Sons, Ltd.

12. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. *Annals of Internal Medicine* 1979; **90**:247–261.
13. Bickel PJ, Doksum KA. *Mathematical Statistics*: *Basic Ideas and Selected Topics*, vol. I. Prentice-Hall: New Jersey, 2001; 422.
14. Robins JM, Greenland S, Hu FC. Rejoinder to discussion of estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association* 1999; **94**:708–712.
15. Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of Royal Statistical Society B* 1983; **45**:212–218.
16. Brumback B, Greenland S, Redman M, Kiviat N, Diehr P. The intensity-score approach to adjusting for confounding. *Biometrics* 2002; **59**:274–285.
17. Newey WK, McFadden D. *Large Sample Estimation and Hypothesis Testing. Hand-book of Econometrics*, vol. 4. Amsterdam: Elsevier, 1994.

Copyright © 2004 John Wiley & Sons, Ltd.