



## Marginal Structural Models for Estimating the Effect of Highly Active Antiretroviral Therapy Initiation on CD4 Cell Count

Stephen R. Cole<sup>1</sup>, Miguel A. Hernán<sup>2</sup>, Joseph B. Margolick<sup>3</sup>, Mardge H. Cohen<sup>4</sup>, and James M. Robins<sup>2,5</sup>

<sup>1</sup> Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

<sup>2</sup> Department of Epidemiology, School of Public Health, Harvard University, Boston, MA.

<sup>3</sup> Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

<sup>4</sup> Cook County Hospital, Chicago, IL.

<sup>5</sup> Department of Biostatistics, School of Public Health, Harvard University, Boston, MA.

Received for publication September 2, 2003; accepted for publication May 4, 2005.

The effect of highly active antiretroviral therapy (HAART) on the evolution of CD4-positive T-lymphocyte (CD4 cell) count among human immunodeficiency virus (HIV)-positive participants was estimated using inverse probability-of-treatment-and-censoring (IPTC)-weighted estimation of a marginal structural model. Of 1,763 eligible participants from two US cohort studies followed between 1996 and 2002, 60 percent initiated HAART. The IPTC-weighted estimate of the difference in mean CD4 cell count at 1 year among participants continuously treated versus those never treated was 71 cells/mm<sup>3</sup> (95% confidence interval: 47.5, 94.6), which agrees with the reported results of randomized experiments. The corresponding estimate from a standard generalized estimating equations regression model that included baseline and most recent CD4 cell count and HIV type 1 RNA viral load as regressors was 26 cells/mm<sup>3</sup> (95% confidence interval: 17.7, 34.3). These results indicate that nonrandomized studies of HIV treatment need to be analyzed with methods (e.g., IPTC-weighted estimation) that, in contrast to standard methods, appropriately adjust for time-varying covariates that are simultaneously confounders and intermediate variables. The 1-year estimate of 71 cells/mm<sup>3</sup> was followed by an estimated continued increase of 29 cells/mm<sup>3</sup> per year (estimated effect at 6 years: 216 cells/mm<sup>3</sup>), providing evidence that the large short-term effect found in randomized experiments persists and continues to improve over 6 years.

acquired immunodeficiency syndrome; antiretroviral therapy, highly active; bias (epidemiology); causality; CD4 lymphocyte count; confounding factors (epidemiology); HIV

Abbreviations: AIDS, acquired immunodeficiency syndrome; CI, confidence interval; GEE, generalized estimating equations; HAART, highly active antiretroviral therapy; HIV, human immunodeficiency virus; IPTC, inverse probability-of-treatment-and-censoring; MSM, marginal structural model.

A declining number of T-lymphocytes expressing the CD4 molecule (CD4 cells) is an important marker of the progression of human immunodeficiency virus (HIV) disease among infected persons (1). Randomized trials conducted in the 1990s indicated a dramatic effect of highly active antiretroviral therapy (HAART) on CD4 cell count (2, 3). The clear evidence of a survival benefit meant that the

HAART trials had to be stopped for ethical reasons. Thus, randomized trial data bearing on the long-term effectiveness of HAART on the evolution of CD4 cell count are neither available nor likely to become available. However, HAART was approved by the US Food and Drug Administration in 1996, and data from the Multicenter AIDS Cohort Study and the Women's Interagency HIV Study are available through

Correspondence to Dr. Stephen Cole, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Room E7640, Baltimore, MD 21205 (e-mail: scole@jhsph.edu).

2002. In this paper, we use these observational data to estimate the effect of HAART on mean CD4 evolution over a period of 6 years.

Estimation of the effect of HAART on CD4 evolution is challenging for the following reason. Current treatment guidelines (4, 5) suggest that physicians use plasma HIV type 1 (HIV-1) RNA level (i.e., viral load) and CD4 cell count to determine the timing of HAART initiation. However, current viral load and CD4 cell count are known predictors of subsequent CD4 cell counts (6). Therefore, to obtain an unconfounded estimate of the total (i.e., direct and indirect) effect of HAART on CD4 cell count, it is necessary to adjust for viral load and CD4 cell count before HAART initiation. A standard approach is to include past viral load and CD4 cell count as time-varying covariates in a regression model for the mean of the current CD4 cell count, conditional on past treatment and confounder history.

Unfortunately, this standard approach fails because evolving viral load and CD4 cell count are strong intermediate variables on the causal pathway from past HAART treatment to current CD4 cell count. For example, the biologic effect of HAART is known to be largely mediated through its effect on viral load (7): HAART dramatically reduces the load of circulating virus by blocking replication at multiple points in the viral life cycle. Thus, the standard approach can, at best, only estimate the relatively small direct effects of past HAART treatment on current CD4 cell count at time  $t$  that are not mediated through the reduction in viral load and the increase in CD4 cell count prior to time  $t$ . Moreover, the standard approach may additionally induce selection bias, because CD4 cell count is affected by previous HAART use (8–10).

However, the above difficulties can be surmounted. Robins (11–13) has developed methods based on marginal and nested structural models to adjust for variables, such as viral load, that are time-varying confounders affected by prior treatment. In a previous report (14), we estimated the total effect of HAART initiation on time to acquired immunodeficiency syndrome (AIDS) or death using a marginal structural Cox model based on observational data from the Multicenter AIDS Cohort Study and the Women's Interagency HIV Study. Here, we use a marginal structural mean model to estimate the effect of HAART on the evolution of CD4 cell counts from 1996 to 2002 in these two ongoing cohort studies.

## MATERIALS AND METHODS

### Study population and measurements

In this analysis, we used information from the Multicenter AIDS Cohort Study (15), which, beginning in 1984, enrolled 5,622 homosexual men in four US cities (Baltimore, Maryland; Chicago, Illinois; Pittsburgh, Pennsylvania; and Los Angeles, California), and the Women's Interagency HIV Study (16), which, beginning in 1994, enrolled 2,628 women in five US cities (New York, New York; Chicago, Illinois; Los Angeles, California; San Francisco, California; and Washington, DC). Every 6 months, participants in both studies completed an extensive interviewer-administered ques-

tionnaire with information on antiretroviral therapy use and provided a blood sample for the determination of CD4 cell count and viral load. Institutional review boards approved all protocols and informed consent forms, which were completed by study participants in both cohorts. Results presented here are limited to the 1,763 men and women who were alive, HIV-positive, and under follow-up in April 1996 when HAART became available.

Each participant contributed a maximum of 12 person-visits beginning with the first semiannual study visit after April 1996 (the baseline visit) and ending with 1) the last visit at which he or she was seen alive, 2) the last visit before the first missing CD4 cell count, or 3) April 2002, whichever came first. For participants who were missing baseline data on any time-varying covariate, baseline was redefined to be the first visit with complete data.

The outcome variable was the number of CD4 cells per cubic millimeter of blood. T-lymphocyte subsets were determined by immunofluorescence using flow cytometry in laboratories participating in the National Institute of Allergy and Infectious Diseases Quality Assurance Program. Specifically, T-cell subsets were measured in purified peripheral blood mononuclear cells or ethylenediaminetetraacetic acid-anticoagulated whole blood by staining with fluorescent dye-conjugated monoclonal antibodies that were specific for CD4 lymphocytes (Becton Dickinson, Mountain View, California) (17).

The effect of exposure to HAART on CD4 cell count was of primary interest. The definition of HAART was based on the US Department of Health and Human Services panel guidelines (4) and has been previously published (14). Typical HAART regimens consisted of two or more nucleoside or nucleotide reverse transcriptase inhibitors in combination with at least one protease inhibitor or one nonnucleotide reverse transcriptase inhibitor. Therapy regimens not classified as HAART were categorized as either monotherapy or combination antiretroviral therapy.

Data on a number of time-fixed and time-varying (i.e., visit-specific) covariates were recorded, including CD4 cell count (in cells/mm<sup>3</sup>), viral load (in copies/ml), and indicators of monotherapy and combination antiretroviral therapy use. Viral load was quantified using a reverse transcription polymerase chain reaction amplification technique (Roche Molecular Systems, Branchburg, New Jersey), which had a lower limit of detection of 400 copies/ml. When there was missing information on time-varying covariates, information was carried forward from the most recent prior observed value.

### Statistical model

Let  $X_{ij}$  be a time-varying indicator of HAART initiation for participant  $i$  on or before the  $j$ th semiannual visit from the start of follow-up (i.e., visit  $j = 0$ ). Let  $L_{ij}$  be the vector of time-varying covariates measured at visit  $j - 1$ , ensuring that  $L_{ij}$  is temporally prior to  $X_{ij}$ , with  $L_{i0}$  being the vector of covariates (time-fixed and time-varying) measured at the visit before baseline. For any time-varying variable, overbars are used to denote the history of that variable up to and including  $j$ ; for instance,  $\bar{L}_{ij} = \{L_{i0}, L_{i1}, \dots, L_{ij}\}$  is the

covariate process for participant  $i$  up to visit  $j$ . The CD4 cell count  $Y_{ij}$  is a component of  $L_{ij}$ . The subscript  $i$ , denoting the participant, ranges from 1 to 1,763, while the subscript  $j$ , denoting the semiannual study visit, ranges from 0 to a maximum of 11.

Consider the piecewise linear spline regression model with a knot at cumulative treatment of 1 year:

$$E[Y_{ij+1}|L_{i0}, \bar{X}_{ij}] = \beta_{0j} + \beta_1' L_{i0} + \beta_2 \text{Cum}_{ij} + \beta_3 (\text{Cum}_{ij} - 1)_+, \quad (\text{model 1})$$

where  $E[\cdot]$  denotes conditional expectation,  $\beta_{0j}$  is a visit-specific intercept,  $\beta_1'$  is the transpose of the column vector of coefficients for the components of the vector  $L_{i0}$ ,  $\text{Cum}_{ij} = \sum_{k=0}^j X_{ik}/2$  is cumulative years of HAART treatment through visit  $j$ , and  $(\text{Cum}_{ij} - 1)_+$  is cumulative treatment minus 1 if cumulative treatment exceeds 1 year and 0 if cumulative treatment is less than or equal to 1 year. Thus, the model states that the mean CD4 cell count increases linearly with cumulative treatment at a slope of  $\beta_2$  until 1 year of cumulative treatment; at that point, the slope changes to  $\beta_2 + \beta_3$ . The baseline covariates  $L_{i0}$  were measured at the semiannual study visit immediately prior to the baseline visit and included indicators of female sex, monotherapy or combination antiretroviral therapy, CD4 cell count in categories ( $<200$ ,  $200-350$ , and  $>350$  cells/mm<sup>3</sup>), and plasma HIV-1 viral load in categories ( $<401$ ,  $401-10,000$ , and  $>10,000$  copies/ml).

For participants who remained alive and under follow-up at visit  $j$ , the above model was fitted using standard generalized estimating equations (GEE) repeated-measures software with an identity working covariance matrix, except that the model was weighted by inverse probability-of-treatment-and-censoring (IPTC) weights. If confounding by unmeasured factors is absent and censoring is ignorable, the IPTC-weighted GEE estimate the parameters of a marginal structural model (MSM) (18). To describe this MSM, let the potential outcome  $Y_{ij+1}(\bar{x})$  be a random variable representing participant  $i$ 's CD4 cell count at visit  $j$  had he or she followed a given therapy history  $\bar{x}$ , rather than his or her observed therapy history  $\bar{X}$ . An average causal effect on a difference scale is the mean of the potential outcomes under the HAART regimen  $\bar{x}$  minus the mean of the potential outcomes under an alternate HAART regimen  $\bar{x}'$ ,  $E[Y_{ij+1}(\bar{x})] - E[Y_{ij+1}(\bar{x}')]$ . The aforementioned IPTC-weighted estimates of model 1 are consistent estimates of the parameters of the marginal structural piecewise linear spline regression model with a knot at cumulative treatment of 1 year:

$$E[Y_{ij+1}(\bar{x})|L_{i0}] = \alpha_{0j} + \alpha_1' L_{i0} + \alpha_2 \text{cum}_{ij} + \alpha_3 (\text{cum}_{ij} - 1)_+, \quad (\text{model 2})$$

where  $\text{cum}_{ij} = \sum_{k=0}^j x_{ik}/2$ . Under model 2, the average causal effect

$$E[Y_{ij+1}(\bar{x})|L_{i0}] - E[Y_{ij+1}(0)|L_{i0}],$$

comparing the mean CD4 cell count under regimen  $\bar{x}$  with the regimen 0, wherein treatment is always withheld, is given by  $\alpha_2 \text{cum}_{ij} + \alpha_3 (\text{cum}_{ij} - 1)_+$ .

Fitting model 1 with IPTC weights gives asymptotically unbiased estimates of the parameters of model 2 under the assumptions of 1) no model misspecification, 2) ignorable censoring, and 3) no unmeasured confounding. This last assumption states that conditional on past measured HAART and covariate histories, current therapy is independent of future potential outcomes (19).

The contribution of participant  $i$  to the calculation at visit  $j$  for model 1 is weighted by an estimate of the IPTC weight  $W_{ij}$ , which is the product of the estimated stabilized inverse probability-of-treatment and inverse probability-of-censoring weights. The stabilized inverse probability-of-treatment weights are

$$W_{ij}^X = \frac{\prod_{k=0}^j f[X_{ik}|\bar{X}_{ik-1}, L_{i0}, \bar{C}_{ik} = 0]}{f[X_{ik}|\bar{X}_{ik-1}, \bar{L}_{ik}, \bar{C}_{ik} = 0]},$$

where  $f[\cdot]$  is the conditional density function evaluated at the observed covariate values for a given participant and  $\bar{L}_{ik}$  includes baseline covariates  $L_{i0}$ . The stabilized inverse probability-of-censoring weights are

$$W_{ij}^C = \frac{\prod_{k=1}^{j+1} \Pr[C_{ik} = 0|\bar{C}_{ik-1} = 0, \bar{X}_{ik-1}, L_{i0}]}{\Pr[C_{ik} = 0|\bar{C}_{ik-1} = 0, \bar{X}_{ik-1}, \bar{L}_{ik-1}]},$$

where  $C_{ik}$  is 1 if participant  $i$  is censored by visit  $k$  and 0 otherwise, and where censoring is defined as the minimum of time to death, first missing CD4 cell count, or administrative censoring in April 2002.

We estimated the denominator of the  $W_{ij}^X$  using a pooled logistic model (20) for the probability of initiating HAART at each visit given baseline and time-varying covariates (18). Specifically, for estimation of the denominator of  $W_{ij}^X$ , the model for the logit of the probability that participant  $i$ , uncensored at visit  $k$ , initiated treatment between visits  $k - 1$  and  $k$  (i.e.,  $X_{ik}$ ) included the baseline regressors  $L_{i0}$  and the subset of the time-varying covariates  $L_{i1}, \dots, L_{ik}$ . This subset consisted of 1) indicator variables for detectable HIV-1 RNA and 2) non-HAART retroviral therapy at  $k - 1$ , 3) CD4 cell count, and 4) detectable log<sub>10</sub> HIV-1 viral load measured at time  $k - 1$ , modeled as restricted cubic splines with four knots located at the 5th, 35th, 65th, and 95th percentiles. For estimation of the denominator  $W_{ij}^C$ , the model for the logit of the probability that participant  $i$  was first censored between  $k$  and  $k + 1$  included the above covariates plus treatment  $X_{ik}$  at  $k$ . The same logit models were used to estimate the numerators of  $W_{ij}^X$  and  $W_{ij}^C$ , except that terms depending on the time-varying covariates  $L_{ik}$ ,  $k > 0$ , were eliminated.

Numerous additional functional forms for the above pooled logistic models were explored (e.g., including covariates measured at times  $j - 2$  and  $j - 3$ ), as well as a broader set of covariates (e.g., age, race, clinical AIDS, body mass index, HIV-related symptoms, receipt of *Pneumocystis carinii* pneumonia prophylaxis, and red blood, platelet, CD3, and CD8 cell counts), but such alternative model specifications did not appreciably alter the results.

The piecewise linear spline form for model 2 was chosen on the basis of an exploratory analysis in which each subsequent year of cumulative exposure was allowed to have its own linear effect. This exploratory analysis clearly showed that the slope of the treatment effect changed at approximately 1 year of cumulative exposure. This would be expected on biologic grounds, since, after sufficient improvement in CD4 cell counts, one would expect homeostatic mechanisms to slow the rate of further increase. In secondary analyses, interactions between HAART and sex and between HAART and baseline CD4 cell count categories, which were suggested by prior research (14), were allowed. In further secondary analyses, the weights were trimmed at the first and 99th percentiles, which is an imperfect method of exploring the impact of influential observations. Trimming weights is not an ideal model-checking procedure, because the extreme weights encode the greatest amount of confounding. Therefore, trimming the weights will typically result in a shift of the estimated effect towards the (biased) unweighted value. Analyses using the natural logarithm of CD4 cell count as the outcome variable provided similar inferences.

All analyses were conducted using the Statistical Analysis System, version 8 (SAS Institute, Inc., Cary, North Carolina). Confidence intervals were based on robust variance estimates (21). For comparison, results from a standard repeated-measures linear model fitted using GEE are presented (22).

### Sensitivity analysis

To explore the possible impact of unmeasured confounding, we performed a sensitivity analysis in which a parameter  $\gamma$  that encodes the degree of confounding by unmeasured factors was varied but not estimated. To do so, we implemented the augmented IPTC-weighted estimators of Robins (23) in a similar fashion as Brumback et al. (24) and Ko et al. (25). Briefly, the observed outcome  $Y_{ij+1}$  is replaced with a bias-adjusted outcome  $Y_{ij+1}(\gamma)$ , defined as

$$Y_{ij+1} - \sum_{k=0}^j \gamma(2X_{ik} - 1) \times f[1 - X_{ik} | \bar{X}_{ik-1}, \bar{L}_{ik}].$$

In words, when a participant is exposed at visit  $j$  (i.e.,  $x_{ij} = 1$ ), one subtracts the product of  $\gamma$  and the conditional probability of being unexposed from the observed outcome, but when a participant is unexposed, one adds the product of  $\gamma$  and the conditional probability of being exposed to the observed outcome. The bias parameter  $\gamma$  in the function represents the average difference (measured in CD4 cells/mm<sup>3</sup>) between the potential outcome in the exposed and the potential outcome in the unexposed, conditional on the covariates. When  $\gamma$  equals 0, there is no unmeasured confounding, and the sensitivity analysis returns the estimate from the standard MSM. When  $\gamma$  is less than 0, the bias-adjusted CD4 count in the exposed exceeds the unadjusted, as is needed to correct for unmeasured confounding (whereby the sickest participants are most likely to receive HAART). For example, with  $\gamma = -50$ , an exposed participant with  $Y_{ij+1} = 310$  cells/mm<sup>3</sup> and a conditional probability of

**TABLE 1. Characteristics of 1,763 HIV-1\*-positive participants from two US cohort studies at study entry, April 1996–April 2002**

Characteristic	Persons (n = 1,763)		Person-years (n = 6,017.5)	
	%	No.	%	No.
Median age (years)	39	(34, 44)†		
Female sex	72	1,270		
Caucasian race	34	595		
HIV-1 treatment				
Antiretroviral therapy	48	842	25	1,476.0
HAART*	0	0	47	2,815.5
HAART initiation	0	0	54	3,258.0
Clinical acquired immunodeficiency syndrome‡	30	521	32	1,908.5
CD4 cell count (cells/mm <sup>3</sup> )				
<200	25	432	18	1,077.5
200–350	23	418	22	1,343.0
>350	52	913	60	3,597.0
Median CD4 cell count (cells/mm <sup>3</sup> )	366	(202, 548)	411	(251, 601)
HIV-1 RNA level (copies/mm <sup>3</sup> )				
<401	24	415	41	2,474.0
401–10,000	22	390	26	1,571.0
>10,000	54	958	33	1,972.5
Median log <sub>10</sub> HIV-1 RNA level (copies/mm <sup>3</sup> )§	4.5	(4.0, 5.0)	4.2	(3.5, 4.8)

\* HIV-1, human immunodeficiency virus type 1; HAART, highly active antiretroviral therapy.

† Numbers in parentheses, interquartile range.

‡ Based on the clinical criteria of the Centers for Disease Control and Prevention (30).

§ Among 1,348 detectable measurements made at baseline and 7,087 made during follow-up.

being unexposed of 0.4 would have  $Y_{ij+1}(\gamma = -50) = 310 + 50(1)0.4 = 330$  cells/mm<sup>3</sup>. When  $\gamma$  is greater than 0, the confounding is in the opposite direction. In the present example, net confounding in this opposite direction is highly unlikely.

### RESULTS

Characteristics of the 1,763 participants at study entry and the 6,017.5 person-years of follow-up are presented in table 1. The average length of follow-up was 4.8 years (standard deviation, 1.7) over the 6-year follow-up period. At study entry, the participants had a median age of 39 years (interquartile range: 34, 44); 72 percent were female, 34 percent were Caucasian, 25 percent had CD4 cell counts below 200 cells/mm<sup>3</sup>, and 54 percent had HIV-1 RNA levels

**TABLE 2. Estimates of the effect of HAART\* on difference in CD4 cell count among 1,763 HIV-1\*-positive participants from two US cohort studies, April 1996–April 2002**

Model	Difference in CD4 cell count (cells/mm <sup>3</sup> )	95% confidence interval
Unweighted, baseline regressors†		
≤1 year	21.5	0.3, 42.8
>1 year	33.0	21.5, 44.6
Unweighted, baseline and time-varying regressors‡		
≤1 year	26.0	17.7, 34.3
>1 year	-0.6	-4.4, 3.1
Weighted, baseline regressors§		
≤1 year	71.0	47.5, 94.6
>1 year	29.0	15.8, 42.3

\* HAART, highly active antiretroviral therapy; HIV-1, human immunodeficiency virus type 1.

† Baseline regressors were sex, use of non-HAART antiretroviral therapy, CD4 cell count, and HIV-1 RNA viral load (in categories).

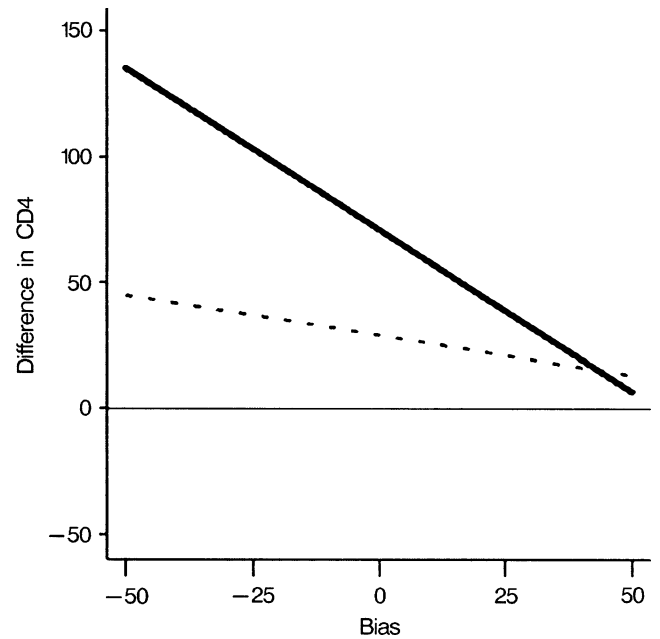
‡ Time-varying regressors were use of non-HAART antiretroviral therapy and splines for CD4 cell count and HIV-1 RNA viral load.

§ Conservative 95% confidence interval.

above 10,000 copies/mm<sup>3</sup>. A total of 599 (34 percent) participants completed follow-up alive; the remaining 1,164 (66 percent) were censored at their first missing CD4 cell count. There were 2,759.5 HAART-free person-years of follow-up, and 1,055 participants (60 percent) initiated HAART, yielding a HAART initiation incidence of 38 per 100 person-years. Among the 1,055 HAART initiators, the average cumulative exposure was 3.1 years (standard deviation, 1.8).

Table 2 shows the estimated difference in average CD4 cell counts per year since HAART initiation for three repeated-measures linear spline models for the response  $Y_{ij+1}$ . The first model was unweighted and only adjusted for the baseline regressors,  $L_{i0}$  (i.e., model 1). The second model was unweighted and adjusted for both the baseline regressors and the same time-varying regressors as those used in the logistic model for the denominator of  $W_{ij}^X$ . The third model used IPTC weights and adjusted for the baseline regressors.

The estimates of parameters  $\beta_2$  and  $\beta_3$  from the weighted model were 71 (95 percent confidence interval (CI): 47.5, 94.6) and -42 (95 percent CI: -71.3, -12.7), respectively. Thus, the weighted model showed a large effect of HAART in the first year after initiation (71 cells/mm<sup>3</sup>) and a smaller effect thereafter (71 - 42 = 29 cells/mm<sup>3</sup> per year). Therefore, the estimated increase in mean CD4 cell count at the maximum follow-up time of 6 years was 216 cells/mm<sup>3</sup>. In a secondary analysis, the estimates were found to be fairly stable after the weighted model was refitted with weights trimmed at the first and 99th percentiles (difference in CD4 cell count: up to 1 year, 68.1 cells/mm<sup>3</sup>, 95 percent CI: 44.7, 91.4; beyond 1 year, 25.3 cells/mm<sup>3</sup>, 95 percent CI: 11.6, 38.9).



**FIGURE 1.** Sensitivity of estimated differences in CD4 cell count from a marginal structural piecewise linear model as a function of bias due to unmeasured confounding, for up to 1 year (solid line) and beyond 1 year (dashed line). Analyses were based on data on 1,763 human immunodeficiency virus type 1-positive participants from two US cohort studies (April 1996–April 2002).

The corresponding estimates from the standard (unweighted) GEE regression model that included baseline and time-varying CD4 cell count and HIV-1 viral load as regressors were 26 cells/mm<sup>3</sup> (95 percent CI: 17.7, 34.3) for the first year and -0.6 cells/mm<sup>3</sup> per year after the first year, which results in an estimated mean CD4 increase at 6 years of 23 cells/mm<sup>3</sup>.

There appeared to be a larger effect of HAART on CD4 cell count within 1 year of initiation among men (92.9 cells/mm<sup>3</sup>, 95 percent CI: 53.9, 131.8) than among women (62.1 cells/mm<sup>3</sup>, 95 percent CI: 36.0, 81.3) (interaction  $p < 0.001$ ), but there was a similar effect of HAART on CD4 cell count by cohort beyond 1 year after initiation (interaction  $p = 0.341$ ). There also appeared to be a larger effect of HAART on CD4 cell count within 1 year of initiation among persons with a lower baseline CD4 cell count (interaction  $p = 0.047$ ). Specifically, among those with a baseline CD4 count below 200 cells/mm<sup>3</sup>, the difference was 105.7 cells/mm<sup>3</sup> (95 percent CI: 80.7, 130.6), while among those with a baseline CD4 count of  $\geq 350$  cells/mm<sup>3</sup>, the difference was 54.1 cells/mm<sup>3</sup> (95 percent CI: 17.8, 90.3). There was no strong evidence for heterogeneity in the effect of HAART on differences in CD4 count beyond 1 year after initiation (interaction  $p = 0.194$ ).

Figure 1 illustrates the sensitivity of the average difference in CD4 cell count due to unmeasured confounding. When the bias parameter  $\gamma$  was -30 cells/mm<sup>3</sup>, reflecting moderately strong residual confounding, the difference in CD4 count within 1 year of HAART initiation increased

from 71 cells/mm<sup>3</sup> to 109.6 cells/mm<sup>3</sup>, while the difference in CD4 count beyond 1 year after HAART initiation increased from 29 cells/mm<sup>3</sup> to 38.5 cells/mm<sup>3</sup>. A  $\gamma$  of  $-30$  represents unmeasured confounding of nearly two thirds the amount of the measured confounding (i.e.,  $30/(71 - 21.5)$ ).

## DISCUSSION

Using a marginal structural piecewise linear model, we estimated that HAART increases the mean CD4 count by 71 cells/mm<sup>3</sup> during the first year after HAART initiation and by 29 cells/mm<sup>3</sup> per year thereafter. The initial effect of HAART appeared greater in the cohort of men and in persons with lower CD4 cell counts at baseline. Estimates of the initial effect of HAART from standard repeated-measures linear models were much smaller regardless of whether time-varying CD4 cell count and HIV-1 viral load were included as regressors, and were presumably biased towards the null because of the presence of time-varying confounders affected by prior treatment (8–10).

In a randomized trial, AIDS Clinical Trials Group 320 (2) previously reported an 82-cell/mm<sup>3</sup> (95 percent CI: 46, 118) difference in CD4 cell count after 40 weeks of HAART treatment as compared with a less potent treatment among participants with prior antiretroviral therapy and CD4 cell counts below 201 cells/mm<sup>3</sup>. In the subgroup with a CD4 cell count less than 200 cells/mm<sup>3</sup> at baseline, a similar estimated 81-cell/mm<sup>3</sup> ( $105.7 \times 40/52 = 81.3$ ) difference at 40 weeks was found using the marginal structural linear spline model.

The present work extends the findings of randomized trials (2, 3) by including participants with baseline CD4 cell counts above 200 cells/mm<sup>3</sup> and extending follow-up to 6 years. In a recent report using an MSM and 2 years of prospective data from the HIV Epidemiology Research Study, Ko et al. (25) reported a 64-cell/mm<sup>3</sup> difference in CD4 count for persons on HAART as compared with those not on HAART in the stratum where baseline CD4 count was below 200 cells/mm<sup>3</sup>. While the effect estimates of Ko et al. appear somewhat muted, the inferential pattern of 1) smaller effects in standard unweighted analyses and 2) a stronger effect at lower baseline CD4 cell counts remains consistent.

These results should be taken in concert with the following limitations. First, the estimates have a causal interpretation only under the assumption of no unmeasured confounding. This assumption probably holds (approximately) here, because the most important clinical and laboratory information used by physicians as indications for HAART was collected and used in the models for the estimation of the weights (26). If the assumption of no unmeasured confounders is correct and the model used to create the treatment weights is correct, then weighting creates a pseudopopulation in which 1) the probability of HAART initiation is not a function of the time-varying covariates (i.e., no confounding exists) but 2) the effect of HAART initiation on CD4 cell count is the same as in the actual study population.

Second, the results are based on the assumption that dropout is ignorable, conditional on measured covariates. Neither

the present analyses nor past analyses (14, 27) suggested that there is notable measured informative censoring in these data. However, a majority of participants were censored at their first missing CD4 cell count. A secondary analysis censoring participants at dropout rather than at first missing CD4 cell count provided a similar estimate but did not significantly improve precision (data not shown). Note that death was treated as a censoring event in the analysis, which may not always be appropriate. Robins and Greenland (28) discuss the pros and cons of this choice.

Third, these results may be sensitive to the relative infrequency of data collection (i.e., 6-month intervals). Misclassification due to this coarse measurement (with respect to time) could have reintroduced some confounding, which could have biased the estimated difference in either direction (29).

Fourth, in these analyses, we assumed that participants remained on HAART after initiation. This assumption was correct for 86 percent of 6,516 post-HAART-initiation person-visits. However, because 14 percent of the visits occurred among persons who had stopped using HAART, the IPTC-weighted analysis did not estimate the effect of continuing HAART therapy versus no HAART therapy. Formally, the IPTC-weighted analysis estimated the “intention-to-treat effect” of HAART therapy versus no HAART therapy in a hypothetical randomized clinical trial in which 1) participants were randomly assigned to begin continuous HAART at different visits, 2) all participants initially complied and began HAART at their assigned visit, and 3) 14 percent later discontinued HAART. If many of the patients in the study discontinued HAART because of toxicity (rather than for nonmedical reasons), it is this intention-to-treat effect, not the effect of continuous HAART usage, that would be the parameter of public health interest. This is because, had we estimated the effect of continuous HAART, we would have been estimating the effect of forcing people to continue therapy even in the presence of toxicity. In fact, detailed information on reasons for discontinuation was not contained in the database. However, on the basis of anecdotal discussions with clinicians, we believe that many participants discontinued therapy because of toxicity, and therefore the analysis presented is the most appropriate. Note that the above discussion implies that the results reported in this paper do not address the effect of discontinuing HAART on CD4 cell count.

Without data from randomized trials that follow patients with widely varying risk profiles for prolonged periods, ongoing prospective observational studies with repeated assessments of exposure and detailed collection of clinical and laboratory information provide the best evidence available for the estimation of risk group-specific, long-term therapeutic effects. These results show, however, that one must be careful to correctly analyze such data. We found that the estimated effect of HAART on CD4 cell count at 6 years based on IPTC-weighted estimation of an MSM was strikingly larger than estimates based on standard GEE analyses. We believe the MSM estimate to be closer to the true effect, both on theoretical grounds (11, 19) and because the MSM analysis, in contrast to standard GEE analyses, successfully reproduced the limited results from a previous randomized trial.

## ACKNOWLEDGMENTS

Dr. Miguel Hernán was partially supported by National Institutes of Health grant K08-AI49392, and Dr. James Robins was partially supported by National Institutes of Health grant R01-AI32457.

Data for this article were collected through the Multi-center AIDS Cohort Study, with centers (Principal Investigators) at the Johns Hopkins Bloomberg School of Public Health (Drs. Joseph B. Margolick and Alvaro Muñoz), the Howard Brown Health Center and Northwestern University Medical School (Dr. John Phair), the University of California, Los Angeles (Drs. Roger Detels and Beth Jamieson), and the University of Pittsburgh (Dr. Charles Rinaldo), and the Women's Interagency HIV Study, with the following centers/consortia (Principal Investigators): the New York City/Bronx Consortium (Dr. Kathryn Anastos), Brooklyn, New York (Dr. Howard Minkoff), the Washington, DC, Metropolitan Consortium (Dr. Mary Young), the Connie Wofsy Study Consortium of Northern California (Dr. Ruth Greenblatt), the Los Angeles County/Southern California Consortium (Dr. Alexandra Levine), the Chicago Consortium (Dr. Mardge Cohen), and the Data Coordinating Center (Dr. Alvaro Muñoz). World Wide Web links for both studies are located at <http://www.statepi.jhsph.edu>.

The Multicenter AIDS Cohort Study is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute (grants U01-AI-35042, 5-MO1-RR-00722 (General Clinical Research Center), U01-AI-35043, U01-AI-37984, U01-AI-35039, U01-AI-35040, U01-AI-37613, and U01-AI-35041). The Women's Interagency HIV Study is also funded by the National Institute of Allergy and Infectious Diseases, with supplemental funding from the National Cancer Institute, the National Institute of Child Health and Human Development, the National Institute on Drug Abuse, and the National Institute of Craniofacial and Dental Research (grants U01-AI-35004, U01-AI-31834, U01-AI-34994, AI-34989, U01-HD-32632 (National Institute of Child Health and Human Development), U01-AI-34993, and U01-AI-42590).

Conflict of interest: none declared.

## REFERENCES

- Muñoz A, Gange SJ. Methodological issues for biomarkers and intermediate outcomes in cohort studies. *Epidemiol Rev* 1998;20:29–42.
- Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* 1997;337:725–33.
- Gulick RM, Mellors JW, Havlir D, et al. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N Engl J Med* 1997;337:734–9.
- Panel on Clinical Practices for the Treatment of HIV Infection, US Department of Health and Human Services. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Bethesda, MD: National Institutes of Health, 2003. ([http://aidsinfo.nih.gov/guidelines/adult/archive/AA\\_071403.pdf](http://aidsinfo.nih.gov/guidelines/adult/archive/AA_071403.pdf)).
- Yeni PG, Hammer SM, Carpenter CC, et al. Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society-USA Panel. *JAMA* 2002;288:222–35.
- Tarwater PM, Margolick JB, Jin J, et al. Increase and plateau of CD4 T-cell counts in the 3(1/2) years after initiation of potent antiretroviral therapy. *J Acquir Immune Defic Syndr* 2001; 27:168–75.
- Palella FJ, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med* 1998;338:853–60.
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;3:143–55.
- Cole SR, Hernán MA. Fallibility in estimating direct effects (with discussion). *Int J Epidemiol* 2002;31:163–5.
- Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
- Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis* 1987;40(suppl 2):139S–61S.
- Robins JM. Structural nested failure time models. New York, NY: John Wiley and Sons, Inc, 1998.
- Robins JM. Marginal structural models. In: 1997 proceedings of the American Statistical Association, Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association, 1998:1–10. (<http://biosun1.harvard.edu/~robins/msm-web.pdf>).
- Cole SR, Hernán MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol* 2003;158:687–94.
- Kaslow RA, Ostrow DG, Detels R, et al. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol* 1987; 126:310–18.
- Barkan SE, Melnick SL, Preston-Martin S, et al. The Women's Interagency HIV Study. WIHS Collaborative Study Group. *Epidemiology* 1998;9:117–25.
- Mellors JW, Muñoz A, Giorgi JV, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med* 1997;126:946–54.
- Hernán MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat Med* 2002;21: 1689–709.
- Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect. *Math Modeling* 1986; 7:1393–512.
- D'Agostino RB, Lee ML, Belanger AJ, et al. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. *Stat Med* 1990; 9:1501–15.
- White HA. A heteroscedasticity-consistent covariance matrix estimator and a direct test of heteroscedasticity. *Econometrica* 1980;48:817–38.
- Diggle PJ, Liang KY, Zeger S. Analysis of longitudinal data. New York, NY: Oxford University Press, 1994.
- Robins JM. Association, causation and marginal structural models. *Synthese* 1999;121:151–79.

24. Brumback BA, Hernán MA, Haneuse SJ, et al. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med* 2004; 23:749–67.
25. Ko H, Hogan JW, Mayer KH. Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models. *Biometrics* 2003;59:152–62.
26. Ahdieh L, Gange SJ, Greenblatt R, et al. Selection by indication of potent antiretroviral therapy use in a large cohort of women infected with human immunodeficiency virus. *Am J Epidemiol* 2000;152:923–33.
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;11:561–70.
28. Robins JM, Greenland S. Comment: causal inference without counterfactuals. *J Am Stat Assoc* 2000;95:431–5.
29. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985;122:495–506.
30. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Morb Mortal Wkly Rep* 1992;41: 1–19.