



## Original Contribution

# Determining the Effect of Highly Active Antiretroviral Therapy on Changes in Human Immunodeficiency Virus Type 1 RNA Viral Load using a Marginal Structural Left-censored Mean Model

Stephen R. Cole<sup>1</sup>, Miguel A. Hernán<sup>2</sup>, Kathryn Anastos<sup>3</sup>, Beth D. Jamieson<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> Montefiore Medical Center, Bronx, NY.

<sup>4</sup> Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA.

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

Received for publication August 16, 2006; accepted for publication January 17, 2007.

Highly active antiretroviral therapy (HAART) dramatically reduces the load of circulating human immunodeficiency virus type 1 (HIV-1) by blocking replication at multiple points in the viral life cycle, but the long-term effect of HAART on viral load remains unclear. In the Multicenter AIDS Cohort Study and the Women's Interagency HIV Study, 918 HIV-1-infected men and women who were not using antiretroviral therapy were followed for a median of 5.8 years between 1996 and 2005. Follow-up yielded 3,629 person-years of observation, during which 286 (31%) of the participants initiated HAART. A marginal structural left-censored linear model for semiannual repeated assessments of viral load showed a 1.9 log<sub>10</sub> decrease in viral load after HAART initiation as compared with nonuse (95% confidence interval: 1.7, 2.2), which remained stable over the course of follow-up but was stronger among men (interaction  $p < 0.001$ ). This association was attenuated by 10% when the authors ignored the left-censoring of viral load measurements (which comprised 20% of measurements (1,420/7,258)) and attenuated by 57% when the authors adjusted for time-varying covariates in a standard fashion rather than using the marginal structural model. In conclusion, the clinically important protective effect of HAART on dampening viral load appears to be rapid, present at CD4 cell counts greater than 350 cells/mm<sup>3</sup>, and sustained beyond 6 years.

acquired immunodeficiency syndrome; antiretroviral therapy, highly active; bias (epidemiology); causality; confounding factors (epidemiology); HIV-1; viral load

Abbreviations: AIDS, acquired immunodeficiency syndrome; CI, confidence interval; HAART, highly active antiretroviral therapy; HIV-1, human immunodeficiency virus type 1; IPTC, inverse probability-of-treatment-and-censoring; SE, standard error.

Infection with human immunodeficiency virus type 1 (HIV-1) is often monitored by the number of copies of HIV-1 RNA present in circulating plasma (1). The level and change in viral load is an important indicator of HIV-1 disease progression.

Randomized trials conducted in the late 1990s indicated a dramatic ability of highly active antiretroviral therapy

(HAART) to stall HIV-1 disease progression (2–4). Specifically, use of HAART dramatically reduces the load of circulating virus by blocking replication at multiple points in the viral life cycle, thereby allowing immune reconstitution. Due to the success of these trials, randomized evidence bearing on the long-term effectiveness of HAART remains unavailable.

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

Estimating the total effect of HAART on change in viral load from observational data is challenging, for at least three reasons. First, one must measure and account for known time-fixed and time-varying confounders and assume that no strong confounders remain unknown (5, 6). Second, standard adjustment or stratification for known time-varying confounders fails to consistently estimate the total effect of HAART on change in viral load (7) and opens the analysis to the possibility of selection bias (8–10). Third, the distribution of circulating viral load is left-censored because of a limit of detection in the assay used to quantify the number of copies of HIV-1 RNA.

In previous work, we estimated the total (i.e., direct and indirect) effect of HAART on 1) time to acquired immunodeficiency syndrome (AIDS) or death (11) and 2) the evolution of CD4-positive T lymphocyte (CD4 cell) count (12). In the present analysis, we used observational data to estimate the total effect of HAART on changes in mean viral load over a period of more than 8 years. We used inverse probability-of-treatment-and-censoring (IPTC)-weighted estimation of a marginal structural left-censored mean model for repeated measures, which allowed for viral load detection limits and consistent estimation of the total effect of exposure to HAART under the assumptions of no unmeasured confounding, no emigrative selection bias (i.e., no informative censoring), and correct specification of the model used to estimate the IPTC weights.

## MATERIALS AND METHODS

### Study population

This analysis utilized information from the Multicenter AIDS Cohort Study (13) and the Women's Interagency HIV Study (14). Beginning in 1984, the Multicenter AIDS Cohort Study enrolled 5,622 homosexual men in Baltimore, Maryland; Chicago, Illinois; Pittsburgh, Pennsylvania; and Los Angeles, California. Beginning in 1994, the Women's Interagency HIV Study enrolled 2,628 women in 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 questionnaire with information on the use of antiretroviral therapy and provided a blood sample for the determination of CD4 cell count and viral load. Positive enzyme-linked immunosorbent assays with confirmatory Western blots were used to determine HIV-1 seropositivity. Institutional review boards approved all protocols and informed consent forms, which were completed by study participants in both cohorts. The results presented here are based on the 918 men and women who were alive, HIV-1-seropositive, and not using antiretroviral therapy in April 1996 when HAART became available. To facilitate the clinically relevant comparison of HAART versus no antiretroviral therapy, we limited the study population to those men and women who were not using antiretroviral therapy at study entry.

Each participant contributed a maximum of 18 person-visits, beginning with the first semiannual study visit after

April 1996 (the baseline visit) and ending with the last visit in which the participant was seen alive before dropout (defined as two consecutive missed viral load measurements) or the date of analysis in April 2005, 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.

### Outcome ascertainment

The outcome was a  $\log_{10}$  transformation of the number of copies of HIV-1 RNA per milliliter of plasma obtained at each semiannual study visit. Plasma was collected from participants in 10-ml Vacutainer tubes containing heparin or potassium ethylenediaminetetraacetic acid and was stored at  $-70^{\circ}\text{C}$  until testing. Among men, the average interval between collection of blood and freezing of plasma samples was estimated to have been 6 hours, but times were not always recorded (15). Among women, the protocol required an interval of less than 6 hours between collection of blood and freezing of plasma samples. The executive committees of the Multicenter AIDS Cohort Study and the Women's Interagency HIV Study independently selected the types of viral load assays used and the dates of any changes in viral load assays. Viral loads measured after October 1998 (April 1998 for men) were quantified by HIV-1 RNA extracted from 1.0-ml plasma samples using an isothermal nucleic acid sequence-based amplification method for women (bioMérieux, Boxtel, the Netherlands) and a reverse transcription polymerase chain reaction amplification assay for men (Roche Molecular Systems, Branchburg, New Jersey). This assay has a lower detection limit of 80 copies/ml (50 copies/ml for men) and is linear to concentrations as high as  $1.6 \times 10^6$  copies/ml (one copy of HIV-1 RNA is equal to one molecule of HIV-1 RNA) (15). The coefficients of variation were approximately 10 percent. Viral load measurements taken prior to 1998 had lower limits of detection of 4,000 copies/ml (400 copies/ml for men), with some exceptions due to retrospective testing of samples. Plasma HIV-1 RNA levels were measured in laboratories participating in the Virology Quality Assurance Laboratory proficiency testing program of the National Institutes of Health and the National Institute of Allergy and Infectious Diseases.

### Exposure assessment

The effect of HAART use, versus no antiretroviral therapy, on change in viral load was of primary interest. This comparison, rather than a comparison of HAART with non-HAART antiretroviral therapy, is of more current clinical interest, because the most common decision clinicians and recently HIV-1-infected persons face is the initiation of HAART from a state of no antiretroviral therapy (16). Cumulative HAART exposure was approximated by time since HAART initiation. HAART users remained HAART-exposed during 84 percent of their person-time after HAART initiation. The definition of HAART was based on the US Department of Health and Human Services panel guidelines (17) and has been previously published (11). Typical HAART

regimens consisted of two or more nucleoside or nucleotide reverse transcriptase inhibitors in combination with at least one protease inhibitor or one nonnucleoside reverse transcriptase inhibitor.

Time since HAART initiation was classified into four categories: 0, >0–1 year, >1–3 years, and >3–9 years. An exploratory analysis in which each subsequent semiannual visit’s worth of cumulative HAART exposure was allowed to have its own linear effect supported the use of these categories.

**Assessment of covariates**

Data on a number of time-fixed and time-varying covariates were recorded. T-lymphocyte subsets were determined by immunofluorescence using flow cytometry in laboratories participating in the Virology Quality Assurance Laboratory 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) (15). Clinical AIDS-defining illnesses were defined using the 1993 clinical criteria of the Centers for Disease Control and Prevention (18). Missing information on time-varying covariates was replaced with information carried forward from the most recently observed prior value.

**Statistical methods**

We fitted a model for repeated measurements of viral load of the form

$$E[Y_{ij}|V_{i0}, \bar{X}_{ij}] = \mu_{ij} = \beta_0j + \beta_1' V_{i0} + \beta_2 g(\text{Cum}_{ij}), \quad (\text{model 1})$$

where  $E[\cdot]$  denotes conditional expectation,  $Y_{ij}$  is the log<sub>10</sub> viral load for participants  $i = 1-918$  at visits  $j = 0-17$ ,  $V_{i0}$  is a subset of  $L_{i0}$ , the vector of measured baseline covariates,  $\text{Cum}_{ij} = \sum_{k=0}^j X_{ik}/2$  is cumulative amount of time (years) from the first visit in which HAART use was reported to visit  $j$  and  $g(\cdot)$  is a function of  $\text{Cum}_{ij}$ , such as the four-level classification described above,  $\beta_0j$  is a visit-specific intercept,  $\beta_1'$  is the transpose of the column vector of coefficients for the components of the vector  $V_{i0}$ , and  $\beta_2$  is the measure of association between cumulative HAART exposure and log<sub>10</sub> viral load. Model 1 has random errors  $\epsilon_{ij} = Y_{ij} - \mu_{ij}$  with an assumed mean of 0 and normally distributed with variance  $\sigma^2$ . We calculated a robust sandwich estimator of the variance (19) with clustering by participant.

To account for time-varying confounding of HAART initiation and for right-censoring by dropout, initiation of non-HAART antiretroviral therapy, or death, we fitted our repeated-measures model using stabilized IPTC weights of the form  $W_{ij} = W_{ij}^X \times W_{ij}^C$ , where

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

and

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

where  $f[\cdot]$  is the conditional density function evaluated at the observed covariate values for a given participant,  $\bar{L}_{ik-1}$  is the vector of time-varying covariate histories measured up to visit  $k - 1$  (measured past viral load is a component of  $\bar{L}_{ik-1}$ ),  $\bar{L}_{ik-1}$  includes baseline covariates  $L_{i0}$ , and  $C_{ik}$  equals 1 if participant  $i$  is censored because of dropout, death, or initiation of non-HAART antiretroviral therapy by visit  $k$  and 0 otherwise.

The baseline covariates  $L_{i0}$  were measured at the semiannual study visit immediately prior to the baseline visit and included age, sex, category of CD4 cell count (<200, 200–350, 351–500, or >500 cells/mm<sup>3</sup>), and category of viral load (<4,001, 4,001–10,000, or >10,000 copies/ml). The subset of baseline covariates  $V_{i0}$  used to stabilize the IPTC weights included CD4 cell count and viral load categories. The time-varying covariate histories  $\bar{L}_{ik-1}$  were specified as restricted cubic splines with three knots located at the 5th, 50th, and 95th percentiles for CD4 cell count and measured log<sub>10</sub> viral load, both taken at visit  $k - 1$ . We disregarded the left-censoring of measured log<sub>10</sub> viral load because of detection limits when modeling initiation of HAART and right-censoring, because the true unmeasured value of viral load was unknown to both physicians and participants and thus is unlikely to have functioned as a confounder; inclusion of an indicator of detectability did not appreciably alter results.

We estimated the components of  $W_{ij}$  using pooled logistic regression models, as previously described (12), with the exception that we now also censored participants at initiation of non-HAART antiretroviral therapy.

Valid use of IPTC weights requires that there not be a probability of 0 or 1 that participants are exposed at any level of the covariates (20). This assumption was presumably met in theory in our study, as some subjects with high CD4 cell counts and low viral loads initiated HAART while others with low CD4 cell counts and high viral loads did not.

We used a parametric likelihood-based approach to handle left-censoring of viral load measurements due to detection limits. Specifically, we assumed that the residuals from model 1 were mean-0 normal random variables with constant variance  $\sigma^2$ . For participant  $i$  at visit  $j$ , let  $d_{ij} = 1$  if the viral load was detected and 0 otherwise. Further, for a participant with  $d_{ij} = 0$ , let  $Y_{ij}^*$  be the lower limit of detection for viral load, which depends on study and time, as described above in the “Outcome ascertainment” section. Then the contribution to the marginal likelihood for participant  $i$  at visit  $j$  is

$$L_{ij} = \left\{ \Phi \left[ \frac{Y_{ij}^* - \mu_{ij}}{\sigma} \right] \right\}^{w_{ij}(1-d_{ij})} \times \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp \left[ \frac{-(Y_{ij} - \mu_{ij})^2}{2\sigma^2} \right] \right\}^{w_{ij}d_{ij}},$$

where  $\Phi$  is the cumulative distribution function of a standard normal random variable. Detected viral load measurements

( $d_{ij} = 1$ ) contribute to the second term in the likelihood via the normal distribution, while undetected viral load measurements ( $d_{ij} = 0$ ) contribute to the first term in the likelihood as a mass below the detection limit. This specification allows for multiple limits of detection, as occurred in our data because of improvements in the measurement procedure over time. Note that we do not assume that the within-participant normal residuals are independent; rather, we only assume that they are marginally normal with a constant variance. We also present a simplification of this model, labeled “alternate analysis A,” whereby we set all  $d_{ij}$ 's equal to 1 and replace viral loads below the detection limit with one half of the detection limit.

If confounding by unmeasured factors is absent and censoring is ignorable, the IPTC-weighted estimates approximate the parameters of a marginal structural model (21), where the potential outcome  $Y_{ij}(\bar{x})$  is a random variable representing participant  $i$ 's  $\log_{10}$  viral load at visit  $j$  had he or she followed a given therapy history  $\bar{x}$ , rather than the 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  $\bar{x}'$ ,  $E[Y_{ij}(\bar{x})] - E[Y_{ij}(\bar{x}')]$ .

On the basis of prior research (11), interactions between HAART and sex and between HAART and baseline CD4 cell count categories were explored. For comparison, we also fitted model 1 by a standard approach and labeled this “alternate analysis B.” Namely, all  $w_{ij}$ 's were set to 1, but we included the time-varying predictors in the model as regressors. To explore the bias-variance tradeoff due to accounting for potential time-varying confounding, we trimmed (i.e., interval-censored) the IPTC weights from below and above at percentiles  $x$  and  $100 - x$ , respectively, for  $x = 1, 5, 10, 25$ , and  $50$ . To explore the possible impact of emigrative selection bias, we recalculated our primary result under “best”-case and “worst”-case scenarios. The best-case scenario imputed viral loads after dropout until planned study completion as follows: Exposed dropouts were assigned undetectable viral loads and unexposed dropouts were assigned viral loads at the 90th percentile of the viral load distribution (or  $7 \times 10^4$  copies/ml). The worst-case scenario imputed viral loads after dropout until planned study completion as follows: Exposed dropouts were assigned viral loads at the 90th percentile of the viral load distribution and unexposed dropouts were assigned undetectable viral loads.

All analyses were conducted in SAS, version 9 (SAS Institute Inc., Cary, North Carolina), employing the procedure NLMIXED to maximize the weighted likelihood and a user-written macro to calculate the robust variance estimates (19) utilized for all confidence intervals. Details of how the robust variance is calculated and why it is necessary are provided in the Appendix.

## RESULTS

At study entry, the 918 participants were, on average, aged 39 years (standard deviation, 8); 63 percent were women,

and 35 percent were Caucasian (table 1). Thirty-seven percent had a viral load less than 4,001 copies/mm<sup>3</sup>, while 49 percent had a viral load greater than 10,000 copies/mm<sup>3</sup>.

The 918 men and women contributed 3,629 person-years of observation. The average length of follow-up was 5.0 years (standard deviation, 3.4) between 1996 and 2005. A total of 354 (39 percent) participants completed follow-up alive, 78 (9 percent) died during follow-up, 236 (26 percent) were censored at initiation of non-HAART antiretroviral therapy, and the remaining 250 (27 percent) were censored at dropout (i.e., two consecutive missing viral load measurements).

There were 2,389 person-years contributed prior to HAART initiation, with 286 (31 percent) participants initiating HAART during follow-up. This yielded a HAART initiation rate of 12 per 100 person-years (95 percent confidence interval (CI): 11, 13). This rate was notably lower than prior estimates (e.g., 38 per 100 person-years (12)) because it did not allow for HAART initiation after observed use of non-HAART antiretroviral therapy. Among the 286 HAART initiators, average time since HAART initiation over the course of follow-up was 3.1 years (standard deviation, 2.2). The predicted probability of HAART initiation ranged from less than 0.001 to 0.70, with a mean of 0.05, prior to HAART initiation and from 0.006 to 0.86, with a mean of 0.17, at HAART initiation. After stabilization, the IPTC weights ranged from 0.03 to 48, with 1st and 99th percentiles at 0.19 and 4.03. As expected, the rate of HAART initiation increased by a factor of 1.61 (95 percent CI: 1.45, 1.75) for each decrement of 100 CD4 cells/mm<sup>3</sup> and increased by a factor of 2.38 (95 percent CI: 1.89, 3.03) for each additional  $\log_{10}$  viral load, both measured at the visit prior to HAART initiation.

Figure 1 shows a plot of the observed  $\log_{10}$  viral loads by number of years of follow-up, with open circles representing pre-HAART-initiation measurements and filled circles representing post-HAART-initiation measurements. Table 2 shows categorized HAART exposure and the estimated difference in the average  $\log_{10}$  number of copies of viral RNA per ml of blood, with nonusers of antiretroviral therapy designated the reference group, estimated by means of the marginal structural model. Collapsing the data over HAART-exposed groups, there was a clinically meaningful decrease of 1.91  $\log_{10}$  viral load (95 percent CI: 1.65, 2.17) with any HAART exposure as compared with no exposure. This decrease appeared to be sustained through the end of follow-up.

Figure 2 depicts the changes in viral load over cumulative HAART exposure, grouped in 6-month intervals, estimated by means of the marginal structural model. Adjusting only for baseline levels of measured variables using a standard left-censored repeated-measures model yielded attenuated (and slightly more precise) results: For instance, collapsing the data over cumulative HAART exposure groups, the HAART-exposed participants had a decrease of 1.59  $\log_{10}$  viral load (95 percent CI: 1.40, 1.78) as compared with nonusers of antiretroviral therapy.

A marginal structural model that replaced undetectable viral loads with one half of the detection limit (alternate analysis A) yielded a 10 percent ( $[(1.72 - 1.91)/1.91] \times 100$ ) attenuation in the association to a decrease of 1.72

**TABLE 1. Baseline and follow-up characteristics of 918 men and women infected with human immunodeficiency virus type 1, Multicenter AIDS Cohort Study and Women's Interagency HIV Study, 1996–2005**

Characteristic	Baseline (1996) ( <i>n</i> = 918 persons)			Follow-up (1996–2005) ( <i>n</i> = 3,629 person-years)		
	%	Mean	No.	%	Mean	No.
Mean age (years)		39 (8)*			42 (8)	
Female sex	63		575	64		2,311
Caucasian race	35		320	37		1,345
Clinical AIDS†,‡	21		194	29		1,034
Use of HAART†	0		0	38		1,383
Category of CD4 cell count (no. of cells/mm <sup>3</sup> )						
<200	9		84	11		416
200–350	17		160	18		668
351–500	24		223	23		826
>500	49		451	47		1,720
Mean CD4 cell count (no. of cells/mm <sup>3</sup> )		539 (294)			530 (305)	
Category of HIV-1† RNA level (no. of copies/mm <sup>3</sup> )						
<4,001	37		336	55		1,993
4,001–10,000	15		136	11		397
>10,000	49		446	34		1,240
Mean log <sub>10</sub> HIV-1 RNA level§		4.1 (0.88)			3.6 (1.16)	

\* Numbers in parentheses, standard deviation.

† AIDS, acquired immunodeficiency syndrome; HAART, highly active antiretroviral therapy; HIV-1, human immunodeficiency virus type 1.

‡ Based on the 1993 clinical criteria of the Centers for Disease Control and Prevention (18).

§ Among 742 detectable measurements taken at baseline and 5,711 of 7,258 person-visits.

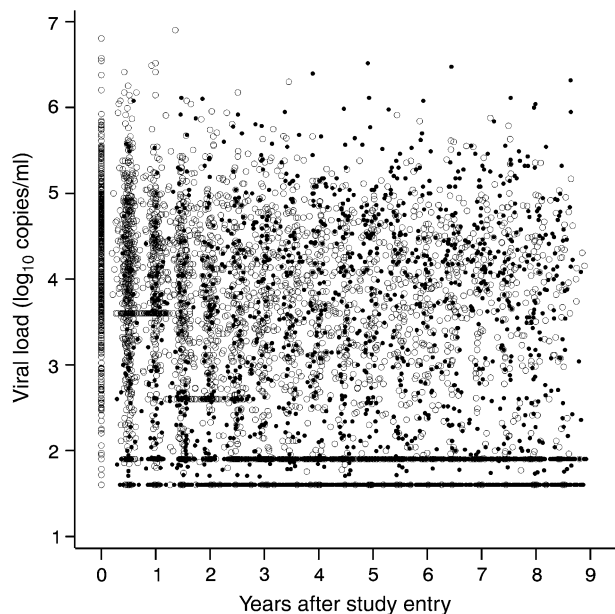
log<sub>10</sub> viral load (95 percent CI: 1.50, 1.95). Alternate analysis B, fitting an unweighted model that adjusted for the same set of time-varying variables, yielded a remarkable 57 percent attenuation in the association to a decrease of 0.83 log<sub>10</sub> viral load (95 percent CI: 0.71, 0.94). However, this attenuated association was not constant over cumulative HAART exposure (>0–1 years: 1.91 log<sub>10</sub> viral load; >1–3 years: 0.31 log<sub>10</sub> viral load; >3 years: 0.30 log<sub>10</sub> viral load). This pattern suggests that this standard method (alternate analysis B) was able to estimate the initial effect of HAART on changes in viral load. However, this standard method was not able to estimate the moderate-to-long-term effect of HAART on changes in viral load. The failure of this standard method was due to adjustment for time-varying consequences of HAART exposure, which probably mediated the effect of HAART exposure on changes in viral load.

The association of HAART exposure with viral load was stronger among men. Specifically, among men, the decrease associated with HAART exposure was 2.46 log<sub>10</sub> viral load (95 percent CI: 2.20, 2.72), while among women the analogous decrease was 1.60 log<sub>10</sub> viral load (95 percent CI: 1.20, 2.00) (interaction *p* < 0.001). The association of HAART with viral load was not notably stronger among persons with a baseline CD4 cell count less than 350 cells/mm<sup>3</sup> (interaction *p* = 0.787).

As expected, trimming the IPTC weights at progressively tighter percentiles moved the marginal structural model point estimate of –1.91 (standard error (SE), 0.134) towards the (assumed biased) baseline-adjusted estimate of –1.59 (SE, 0.098) while decreasing the standard error: Specifically, at censored percentiles of 1 and 99, 5 and 95, 10 and 90, and 25 and 75, the resultant point estimates were –1.80 (SE, 0.110), –1.73 (SE, 0.101), –1.69 (SE, 0.098), and –1.63 (SE, 0.096), respectively. Therefore, the reduction in the standard error achieved by ignoring time-varying confounding was overwhelmed by the bias introduced by time-varying confounding. Under the “best”-case scenario for dropout as described in the “Statistical methods” section, the observed estimate of –1.91 shifted to –2.30, and under the “worst”-case scenario, the observed estimate of –1.91 shifted to –0.01.

## DISCUSSION

Employing a marginal structural left-censored linear model, we estimated that among participants not using antiretroviral therapy, HAART initiation rapidly decreased viral load and HAART initiators sustained an average mean viral load reduction of nearly 2 log<sub>10</sub> copies/ml. This initial



**FIGURE 1.** Observed  $\log_{10}$  viral load among 918 human immunodeficiency virus type 1-infected men and women during 3,629 person-years of follow-up (7,258 person-visits), Multicenter AIDS Cohort Study (men) and Women's Interagency HIV Study (women), 1996–2005. Open circles represent viral load measurements taken prior to initiation of highly active antiretroviral therapy (HAART), and filled circles represent post-HAART-initiation viral load measurements. Horizontal banding due to limits of detection can be seen at  $\log_{10}$  viral loads of 1.7, 1.9, 2.6, and 3.6.

protective effect of HAART appeared to be greater among men. This nearly 2  $\log_{10}$  decrease, from an initial average  $\log_{10}$  viral load of approximately 4  $\log_{10}$  copies/ml to 2  $\log_{10}$  copies/ml or 100 copies/ml, suggests that, on average, persons with continuous HAART use approach the floor of the current viral load detection limit of 50 copies/ml.

An alternate analysis which replaced viral load values below the detection limit with one half of the detection limit rather than allowing such values to be left-censored (alternate analysis A) produced a notable null-bias in the association between HAART initiation and change in viral load. Replacing left-censored viral load measurements with a fixed value, such as one half of the detection limit, results in two problems. First, the variability in the outcome will be underestimated, since the distribution of values below the detection limit has been replaced with a fixed value. Second, the choice of the fixed value is arbitrary and may or may not be an appropriate summary of the expectation of the values below the detection limit. Of course, our assumption that viral load counts below the lowest detection limit follow the assumed normal distribution is not subject to empirical test and could be incorrect.

Another alternate analysis that adjusted for time-varying covariates rather than weighting (alternate analysis B) produced a remarkably strong null-bias in the association between HAART initiation and change in viral load. Adjusting, rather than weighting, for time-varying covariates that are affected by prior exposure precludes the investigator from being able to estimate the total (i.e., direct and indirect) effect of a time-varying exposure (7, 20) and places the analysis at risk for induced selection bias (8–10).

In the AIDS Clinical Trials Group 320 randomized trial, Hammer et al. (2) reported that HAART decreased the mean viral load by 2.2  $\log_{10}$  copies/ml at 24 weeks after therapy initiation ( $p < 0.001$ ) in comparison with less potent therapy among largely male participants with prior antiretroviral therapy experience and CD4 cell counts below 201 cells/ $\text{mm}^3$ . In the present work, we have addressed a timelier clinical question by comparing HAART with no antiretroviral therapy among both men and women with wide-ranging baseline CD4 cell counts who were followed for more than 8 years. We expected to find an effect stronger than that found by Hammer et al. among persons with low baseline CD4 cell counts because of the difference in comparison groups, and we expected to find a weaker effect for persons with higher baseline CD4 cell counts, based on our prior

**TABLE 2.** Cumulative exposure to highly active antiretroviral therapy and difference in  $\log_{10}$  RNA viral load among 918 men and women infected with human immunodeficiency virus type 1, Multicenter AIDS Cohort Study and Women's Interagency HIV Study, 1996–2005

HAART† exposure (years)		No. of observations	%	Difference in $\log_{10}$ HIV-1‡ RNA viral load (no. of copies/ml)*, ‡	95% confidence interval§
Category	Median				
0	0	4,492	62	0¶	
>0–≤1	0.5	566	8	–1.91	–1.68, –2.14
>1–≤3	1.8	916	12	–1.88	–1.59, –2.17
>3–≤9	4.8	1,284	18	–1.96	–1.58, –2.33

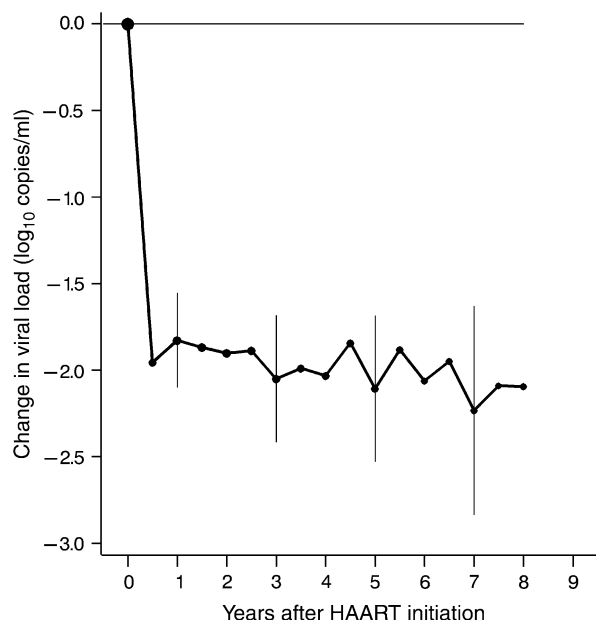
\*  $p$  for trend  $< 0.001$ .

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

‡ Adjusted for sex, age, and baseline CD4 cell count and viral load, as well as time-varying CD4 cell count, viral load, and time (splines were used for the latter two factors).

§ Conservative 95% confidence interval.

¶ Reference category.



**FIGURE 2.** Change in  $\log_{10}$  viral load according to cumulative exposure to highly active antiretroviral therapy (HAART) among 918 human immunodeficiency virus type 1-infected men and women during 3,629 person-years of follow-up (7,258 person-visits), Multicenter AIDS Cohort Study (men) and Women's Interagency HIV Study (women), 1996–2005. Change in  $\log_{10}$  viral load was estimated using a marginal structural model. The solid line connects the 6-month groupings of cumulative HAART exposure, and the black dots show the  $\log_{10}$  numbers of person-visits for particular groupings of cumulative exposure. Horizontal line, reference category (no change); vertical lines, pointwise 95% confidence intervals.

experience. We also anticipated a possible reduction in the effect of HAART with prolonged exposure. We observed no appreciable reduction in the effect over time and an overall effect of approximately 90 percent of that observed by Hammer et al. (2).

A plausible explanation for the observed difference in the strength of association by sex is that the observed stronger association among men may reflect better levels of adherence to HAART rather than a true sex difference in the causal effect of HAART on change in viral load. For example, a recent analysis of Multicenter AIDS Cohort Study and Women's Interagency HIV Study data suggested that 84 percent of men but only 76 percent of women were fully adherent after initiating a HAART regimen (22).

The present results should be interpreted with consideration of the following limitations. First, like all observational analyses, the estimates have a causal interpretation only under the assumption of no unmeasured confounding. This assumption probably holds (approximately) here, since 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 (23). Numerous additional functional forms for the weight models were explored (e.g., longer covariate histories, more flexible splines), as well as a broader set of covariates (e.g., age, race, clinical AIDS, body mass index, HIV-1-related

symptoms, *Pneumocystis jiroveci* pneumonia prophylaxis, and red blood, platelet, CD3, and CD8 cell counts), but such alternative model specifications did not appreciably alter the results. If the assumption of no unmeasured confounders is correct and the model used to create the treatment weights is correctly specified, then weighting creates a pseudopopulation in which the probability of HAART initiation is not a function of the time-varying covariates (i.e., no confounding exists), but the effect of HAART initiation on viral load is the same as in the actual study population.

Second, and like all prospective analyses with right-censoring, the results are based on the assumption that right-censoring is ignorable, conditional on measured covariates. Neither the present analysis nor past analyses (11, 12, 24) suggested that there was notable emigrative selection bias due to variables *measured* in these data. The extent of possible emigrative selection bias due to *unmeasured* variables was demonstrated to be substantial. Death was treated as a censoring event in the present analysis, which may not always be appropriate. Robins and Greenland (25) discuss the pros and cons of this and other choices.

Third, these results may have been 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 (26).

Fourth, in these analyses, we assumed that participants remained on HAART after initiation. This assumption was correct for 84 percent of post-HAART-initiation person-years. The IPTC-weighted analysis was estimating the “intention-to-treat effect” of HAART therapy versus no antiretroviral 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) 16 percent later discontinued HAART. Cole et al. (12) discuss the pros and cons of this choice.

Fifth, we assumed that all viral load measurements below the limit of detection had values greater than 0. This assumption was based on sobering virologic suggestions that HIV-1 harbors itself in multiple organs and is not eradicated even with persistently undetectable (i.e., <50 copies/ml) plasma RNA levels. In other substantive fields, one may believe that the outcome variable is a mixture of  $\pi$  true 0's and  $1 - \pi$  values greater than 0, of which only a percentage are detectable with current laboratory technology (27); in such cases, one may fit a marginal structural model reflecting this mixture.

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 carefully analyze such data. We found that the estimated effect of HAART on viral load based on IPTC-weighted estimation of a marginal structural model was strikingly larger than estimates based on standard analyses

or marginal structural model analyses that ignored left-censoring of viral load.

In conclusion, the observed association of HAART with HIV-1 viral load appears to apply irrespective of baseline CD4 cell count and to be sustained for several years; however, there do not appear to be strong *additional* effects of HAART on reduction of viral load beyond that observed within the first year after therapy initiation.

## ACKNOWLEDGMENTS

The Multicenter AIDS Cohort Study (MACS) 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 (WIHS) 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).

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.

Study centers (and Principal Investigators) for the MACS and the WIHS are as follows: *MACS*—the Johns Hopkins Bloomberg School of Public Health (Joseph B. Margolick, Lisa Jacobson), the Howard Brown Health Center and Northwestern University Medical School (John Phair), the University of California, Los Angeles (Roger Detels, Beth Jamieson), and the University of Pittsburgh (Charles Rinaldo); *WIHS*—the New York City/Bronx Consortium (Kathryn Anastos), Brooklyn, New York (Howard Minkoff), the Washington, DC, Metropolitan Consortium (Mary Young), the Connie Wofsy Study Consortium of Northern California (Ruth Greenblatt), the Los Angeles County/Southern California Consortium (Alexandra Levine), the Chicago Consortium (Mardge Cohen), and the Data Coordinating Center (Stephen Gange). Study websites are located at <http://www.statepi.jhsph.edu>.

Conflict of interest: none declared.

## REFERENCES

- Mellors JW, Rinaldo CR, Gupta P, et al. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996;272:1167–70.
- Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus zidovudine 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 zidovudine, zalcitabine, and didanosine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N Engl J Med* 1997;337:734–9.
- Cameron DW, Heath-Chiozzi M, Danner S, et al. Randomised placebo-controlled trial of zalcitabine in advanced HIV-1 disease. The Advanced HIV Disease Zalcitabine Study Group. *Lancet* 1998;351:543–9.
- Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983;2:267–71.
- Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;15:413–19.
- Robins J. The control of confounding by intermediate variables. *Stat Med* 1989;8:679–701.
- 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.
- 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.
- Cole SR, Hernán MA, Margolick JB, et al. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *Am J Epidemiol* 2005;162:471–8.
- 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, Munoz 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.
- Sterne JA, Hernán MA, Ledergerber B, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet* 2005;366:378–84.
- Panel on Clinical Practices for the Treatment of HIV Infection, National Institutes of Health. 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>).
- 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Recomm Rep* 1992;41:1–19.
- White HA. A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica* 1980;48:817–38.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550–60.
- 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.
- Lazo M, Gange SJ, Wilson TE, et al. Changes in adherence to highly active antiretroviral therapy. (Abstract). Presented at the XVI International AIDS Conference, Toronto, Ontario, Canada, August 13–18, 2006.

23. 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.
24. 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.
25. Robins JM, Greenland S. Comment on: causal inference without counterfactuals. *J Am Stat Assoc* 2000;95:431–5.
26. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985;122:495–506.
27. Chu H, Kensler TW, Munoz A. Assessing the effect of interventions in the context of mixture distributions with detection limits. *Stat Med* 2005;24:2053–67.
28. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
29. Hanley JA, Negassa A, Edwardes MD, et al. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol* 2003;157:364–75.
30. Robins JM. Marginal structural models. In: 1997 proceedings of the Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association, 1998:1–10.
31. White HA. Maximum likelihood estimation of misspecified models. *Econometrica* 1982;50:1–25.

---

## APPENDIX

Robust variance estimates are needed in the present application for two reasons. First, we have repeated measurements of viral load taken in the same person at different

times. The repeated measurements within individuals are likely to be more similar than those between individuals, creating a positive correlation and negating the standard assumption of independent observations. Our robust variance estimates are equivalent to those used for generalized estimating equations (28, 29) with an independent working covariance matrix, and are a general solution to the problem of nonindependent observations. Note in particular that we did not assume that the within-participant normal residuals were independent.

Second, variability in the estimated inverse probability-of-treatment-and-censoring weights in our marginal structural model requires that we use a robust variance estimate to obtain conservative confidence intervals, as described by Robins (30). Robust variance estimates are a general solution to account for misspecified general linear models, as described by White (19, 31).

The weighted likelihood given in the text is maximized on  $N$  participants with  $M$  visits (using the SAS program NLMIXED), and the resulting  $P$  parameter estimates, score matrix, and Hessian matrix are retained. Next, the score contributions for each parameter  $P$  are summed over the  $M$  visits for each of  $N$  individuals. The robust variance is then taken as  $\mathbf{H}^{-1}(\mathbf{X}'\mathbf{X})\mathbf{H}^{-1}$ , where  $\mathbf{H}$  is the Hessian  $P \times P$  matrix (i.e., inverse of the negative Hessian is the standard covariance) and  $\mathbf{X}$  is the  $N \times P$  matrix of summed (over visits) score contributions. Note that  $\mathbf{X}'\mathbf{X}$  can also be written as  $\sum_{i=1}^N S_i S_i'$ , where  $S_i$  is the model-based score for participant  $i$ . Illustrative SAS code is available from the first author upon request.