American Journal of Epidemiology
Copyright © 2003 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved

Vol. 158, No. 7
Printed in U.S.A.
DOI: 10.1093/aje/kwg206

Effect of Highly Active Antiretroviral Therapy on Time to Acquired Immunodeficiency Syndrome or Death using Marginal Structural Models

Stephen R. Cole[1], Miguel A. Hernán[2], James M. Robins[2,3], Kathryn Anastos[4], Joan Chmiel[5], Roger Detels[6], Carolyn Ervin[7], Joseph Feldman[8], Ruth Greenblatt[9], Lawrence Kingsley[10], Shenghan Lai[1], Mary Young[11], Mardge Cohen[12], and Alvaro Muñoz[1]

[1] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.
[2] Department of Epidemiology, School of Public Health, Harvard University, Boston, MA.
[3] Department of Biostatistics, School of Public Health, Harvard University, Boston, MA.
[4] Lincoln Medical and Mental Health Center, New York, NY.
[5] Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL.
[6] Department of Epidemiology, School of Public Health, University of California, Los Angeles, Los Angeles, CA.
[7] Kenneth Norris Jr. Cancer Hospital, Los Angeles, CA.
[8] Department of Preventive Medicine and Community Health, Health Science Center at Brooklyn, State University of New York, Brooklyn, Brooklyn, NY.
[9] Departments of Medicine and Epidemiology, University of California, San Francisco, San Francisco, CA.
[10] Department of Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.
[11] Division of Infectious Diseases, Georgetown University Hospital, Washington, DC.
[12] Cook County Hospital, Chicago, IL.

Received for publication February 6, 2002; accepted for publication March 10, 2003.

To estimate the net (i.e., overall) effect of highly active antiretroviral therapy (HAART) on time to acquired immunodeficiency syndrome (AIDS) or death, the authors used inverse probability-of-treatment weighted estimation of a marginal structural model, which can appropriately adjust for time-varying confounders affected by prior treatment or exposure. Human immunodeficiency virus (HIV)-positive men and women (n = 1,498) were followed in two ongoing cohort studies between 1995 and 2002. Sixty-one percent (n = 918) of the participants initiated HAART during 6,763 person-years of follow-up, and 382 developed AIDS or died. Strong confounding by indication for HAART was apparent; the unadjusted hazard ratio for AIDS or death was 0.98. The hazard ratio from a standard time-dependent Cox model that included time-varying CD4 cell count, HIV RNA level, and other time-varying and fixed covariates as regressors was 0.81 (95% confidence interval: 0.61, 1.07). In contrast, the hazard ratio from a marginal structural survival model was 0.54 (robust 95% confidence interval: 0.38, 0.78), suggesting a clinically meaningful net benefit of HAART. Standard Cox analysis failed to detect a clear net benefit, because it does not appropriately adjust for time-dependent covariates, such as HIV RNA level and CD4 cell count, that are simultaneously confounders and intermediate variables.

acquired immunodeficiency syndrome; antiretroviral therapy, highly active; causality; confounding factors (epidemiology)

Abbreviations: AIDS, acquired immunodeficiency syndrome; CI, confidence interval; HAART, highly active antiretroviral therapy; HIV, human immunodeficiency virus; HR, hazard ratio; NRTI, nucleoside reverse transcriptase inhibitor; PCP, *Pneumocystis carinii* pneumonia.

In 1997, a randomized trial conducted by AIDS Clinical Trials Group 320 demonstrated that treatment with highly active antiretroviral therapy (HAART) halved the hazard of acquired immunodeficiency syndrome (AIDS) or death

Correspondence to Dr. Stephen R. Cole, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Room E7014, Baltimore, MD 21205 (e-mail: scole@jhsph.edu).

(hazard ratio (HR) = 0.50, 95 percent confidence interval (CI): 0.33, 0.76) in comparison with less potent combination antiretroviral therapy among human immunodeficiency virus (HIV)-infected patients with CD4 cell counts less than 200 cells/mm^3 at randomization (1). Observational studies have been unable to consistently replicate this result using standard statistical methods (e.g., regression or stratification) (2). One study even reported a harmful effect of HAART on time to AIDS or death (adjusted HR = 1.20, 95 percent CI: 1.01, 1.44) (3).

In observational studies, persons who initiate HAART are usually those with poorer values for prognostic biomarkers (i.e., confounding by indication) (4), such as low CD4 cell count and high plasma levels of HIV type 1 (HIV-1) RNA (5). Therefore, one needs to adjust for these time-varying confounders to estimate the effect of HAART on AIDS or death. However, including these time-varying confounders as covariates in standard survival models (e.g., Cox models) may yield an association measure (e.g., hazard ratio) that cannot be interpreted as the overall or net effect of HAART, because current CD4 cell count and HIV RNA level are themselves strongly influenced by past HAART exposure (6–8). In the absence of confounding by other unmeasured factors, the HAART effect in such a model may represent the direct effect of HAART not mediated through CD4 count and HIV RNA. Since it is likely that much of the effect of HAART on AIDS-free survival is mediated by its effect on CD4 count and HIV RNA level, one could expect that such an association measure for HAART would be an underestimate of the net effect of HAART. Below, we estimate the net effect of HAART on AIDS-free survival in prospective observational data using a marginal structural model, which appropriately adjusts for confounding by time-varying factors affected by treatment.

MATERIALS AND METHODS

Data

In this analysis, we utilized information from two ongoing prospective studies of the natural history of HIV infection: the Multicenter AIDS Cohort Study (9) and the Women's Interagency HIV Study (10). The Multicenter AIDS Cohort Study, beginning in 1984, enrolled 5,622 homosexual men in four US cities: Baltimore, Maryland; Chicago, Illinois; Pittsburgh, Pennsylvania; and Los Angeles, California. The Women's Interagency HIV Study, 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. Institutional review boards approved all protocols and informed consent forms, which were completed by study participants in both cohorts. The results presented here are limited to the 1,498 participants who were HIV-positive and AIDS-free and had not initiated HAART prior to the first eligible study visit (see below).

Every 6 months, participants in both studies completed an extensive interviewer-administered questionnaire giving information on antiretroviral treatment and HIV-related symptoms and provided a blood sample for the determination of CD4 cell count and plasma HIV-1 RNA level. The definition of HAART followed the Department of Health and Human Services/Kaiser Panel guidelines (11). HAART was defined as 1) use of two or more nucleoside (or nucleotide) reverse transcriptase inhibitors (NRTIs) in combination with at least one protease inhibitor or one non-NRTI; 2) use of one NRTI in combination with at least one protease inhibitor and at least one non-NRTI; 3) a regimen containing ritonavir and saquinavir in combination with one NRTI and no non-NRTIs; or 4) an abacavir-containing regimen of three or more NRTIs in the absence of both protease inhibitors and non-NRTIs. Combinations of zidovudine and stavudine with either a protease inhibitor or a non-NRTI were not considered HAART. Therapy regimens not classified as HAART were categorized as either monotherapy or combination antiretroviral therapy. Once a participant reported initiation of HAART, he or she was assumed to have remained on HAART for the duration of follow-up. This simplifying assumption correctly classified 94 percent of the observed person-time. An indicator variable for *Pneumocystis carinii* pneumonia (PCP) prophylaxis was constructed using reports of trimethoprim, bactrim, aerosolized pentamidine, and dapsone use. T-cell subsets were determined by immunofluorescence using flow cytometry in laboratories participating in the National Institute of Allergy and Infectious Diseases quality assurance program. Baseline CD4 cell count was modeled in three categories: <200, 200–350, and >350 cells/mm^3. Time-varying CD4 cell count was modeled using a restricted cubic spline with four knots located at the 5th, 35th, 65th, and 95th percentiles.

HIV-1 RNA viral load was quantified using a reverse transcription polymerase chain reaction amplification technique (Roche Molecular Systems, Branchburg, New Jersey). Baseline RNA level was modeled in three categories: <401, 401–10,000, and >10,000 copies/ml. Time-varying RNA level was modeled as an indicator of detection (the detection limit was 400 copies/ml) in concert with a restricted cubic spline with four knots located at the 5th, 35th, 65th, and 95th percentiles for the $\log_{10}$-transformed detected measurements (set to zero for undetected measurements). An indicator variable for the presence of any HIV-related symptoms was constructed using reports of persistent fever, diarrhea, night sweats, and weight loss. Longitudinal data were carried forward from the most recent observed value for the 10 percent of anticipated visits that were missed. Alternate analyses restricted to participants with complete data at baseline or multiply imputed missing baseline data yielded similar results (data not shown).

The outcomes of interest were first diagnosis of clinical AIDS or death from any cause. The 1993 Centers for Disease Control and Prevention clinical conditions criteria were used to define clinical AIDS (12). Therefore, participants with CD4 cell counts less than 200 cells/mm^3 but no clinical conditions were not considered to have clinical AIDS. A description of outcome ascertainment has been published elsewhere (9, 13). Briefly, physician or hospital records were used to confirm reported cases of clinical AIDS in the cohort of men, while in the cohort of women, clinical AIDS was self-reported. Deaths were ascertained using death certifi-

cate abstractions upon notification and national death registry searches.

Each participant contributed a maximum of 13 person-visits of follow-up from the baseline visit (first visit after October 1995) to the last visit at which he or she was seen free of clinical AIDS and alive or the visit before April 2002, whichever came first. Follow-up of participants missing any time-varying characteristic at baseline started at the first subsequent visit at which values were observed.

Marginal structural model

We used a weighted pooled logistic regression model to approximate the parameters of a marginal structural Cox model, as described by Hernán et al. (14, 15). Pooled logistic regression approximates the Cox model well when the risk of events is less than 10 percent per person-time interval (16); herein, the maximum visit-specific risk of AIDS or death was 6 percent.

Time was measured in semiannual visits from the beginning of follow-up and took values (k) from zero (October 1995–April 1996) to 12 (October 2001–April 2002). The subscript i, denoting the subject, is often suppressed, because we assumed that the random vector of data for each subject was drawn independently from an identical distribution. Let $D(k+1)$ be an indicator of first diagnosis of clinical AIDS or death between visits k and $k+1$. Let $X(k)$ be a time-varying indicator of HAART initiation at or before visit k, with $X(-1) \equiv 0$, since the study population was selected to not have HAART exposure prior to the first eligible visit. Let $L(k)$ be a vector of time-varying covariates measured at visit $k-1$, so that $L(k)$ is temporally prior to $X(k)$, with $L(0)$ being the vector of covariates measured at the visit preceding the study period (i.e., the "baseline" visit). For the present analyses, $L(0)$ consisted of age, gender, race, calendar year at study entry, baseline use of (mono- and combination) antiretroviral therapy, and baseline CD4 and RNA categories; $L(k)$ further consisted of CD4 count, RNA level, HIV symptoms, indicators of (mono- and combination) antiretroviral therapy and PCP prophylaxis, and number of days since the prior visit.

For persons who remained AIDS-free, alive, and under follow-up at visit $k+1$, we fit the pooled logistic regression model

$$\text{logit PR}[D(k+1) = 1 | X(k), L(0)] = \beta_0(k) + \beta_1 X(k) + \beta'_2 L(0),$$

where $\beta_0(k)$ is a visit-specific intercept (which we modeled as a restricted cubic spline with four knots at the 5th, 35th, 65th, and 95th percentiles for the number of days since the baseline visit). The contribution of participant i to the calculation at visit k is weighted by $W_i(k)$, which is the product of the estimated inverse probability-of-treatment weight and the inverse probability-of-censoring weight, namely $W_i(k) = W_i^X(k) \times W_i^C(k)$. In the absence of unmeasured confounding, unmeasured informative censoring, and model misspecification, $\exp(\beta_1)$ is a consistent and asymptotically normal estimator of the hazard ratio, which compares the hazard of AIDS or death had everyone initiated HAART at baseline with the hazard had no one initiated HAART during follow-up (8). Therefore, we compared continuous HAART exposure against the collective of no therapy, monotherapy, or combination therapy.

Informally, each participant's inverse probability-of-treatment weight is the inverse of the probability of receiving the treatment history he or she did in fact receive by visit k. Specifically, $W^X(k) = \prod_{j=0}^{k} 1/f[X(j)|\bar{X}(j-1), \bar{L}(j)]$, where $f[\cdot]$ is by definition the conditional density function evaluated at the observed covariate values for a given subject and $\bar{L}(j)$ is the history of time-varying covariates up to time j, including baseline covariates $L(0)$. The approach using the inverse probability-of-treatment weight adjusts for confounding by the variables that are used to create the weights and can be viewed as a generalization of the Horvitz-Thompson estimator (17). Since the inverse probability-of-treatment weight and the inverse probability-of-censoring weight are unknown, we estimate them using the predicted values from pooled logistic models for the probabilities of initiating HAART and of censoring, respectively.

In the absence of unmeasured confounding, unmeasured informative censoring, and model misspecification, weighting creates a pseudo-population in which 1) the probabilities of treatment (i.e., HAART) and censoring are not a function of the time-varying covariates but 2) the effect of HAART on time to clinical AIDS or death is the same as in the original population. Thus, the inverse probability-of-treatment weight effectively removes any association between prior confounding variables and HAART but preserves the relation between HAART and clinical AIDS or death.

A fuller account of the covariate histories (i.e., including covariates measured at $k-2$ and $k-3$), a less restrictive functional form for age, and a broader set of covariates (e.g., white blood, red blood, platelet, CD3, and CD8 cell counts; body mass index (weight (kg)/height (m)^2); an indicator of the last visit having been missed) did not appreciably alter our results. The Hosmer-Lemeshow goodness-of-fit χ^2 value for the final model for the denominator of the weights $W_i^X(k)$ was 23 with 8 degrees of freedom.

To increase the efficiency of our estimator, we stabilized the weights (14, 15). Note that the marginal structural model includes as regressors the baseline variables (age, gender, race, baseline CD4 count, and RNA level) used to stabilize the weights. For computational details and an example of the SAS code, see Hernán et al. (14). Confidence intervals for the inverse probability-of-treatment weight estimators of the marginal structural model are based on robust variance estimates (18) and are conservative (wider than need be) (19, 20). To ensure that we were not being overly conservative in using the robust variance estimate, we compared the conservative confidence intervals with a simple percentile-based nonparametric bootstrap confidence interval calculated from 500 full samples (with replacement) from the observed data.

Note that since baseline covariates $L(0)$ are included in the model, one can also include in the model interaction terms between time-dependent HAART and baseline covariates in order to estimate the hazard ratio at specific levels of the baseline covariates. Specifically, we report on how the effect

TABLE 1. Characteristics of 1,498 human immunodeficiency virus-positive US men and women at study entry and during follow-up, 1995–2002

Characteristic	Subjects (n = 1,498)		Person-years (n = 6,763)	
	No.	%	No.	%
Median age (years)	39 (33, 44)*			
Female gender	992	66		
Caucasian race	561	37		
Antiretroviral therapy				
None	898	60		
Monotherapy	321	21		
Combination therapy	279	19		
HAART†	0	0		
PCP† prophylaxis	394	26	1,639	24
CD4 cell count (cells/mm^3)				
<200	258	17	921	14
200–350	373	25	1,512	22
>350	867	58	4,330	64
Median CD4 cell count	395 (257, 560)		433 (285, 615)	
HIV† RNA level (copies/ml)				
<401	422	28	2,752	41
401–10,000	272	18	1,840	27
>10,000	804	54	2,171	32
Median $\log_{10}$ HIV RNA level‡	4.5 (4.0, 5.0)		4.1 (3.5, 4.7)	
≥1 HIV-related symptom§	396	26	1,642	24

* Numbers in parentheses, quartile cutpoints.

† HAART, highly active antiretroviral therapy; PCP, *Pneumocystis carinii* pneumonia; HIV, human immunodeficiency virus.

‡ Among persons with detectable levels (i.e., ≥401 copies).

§ Symptoms included persistent fever, diarrhea, night sweats, and weight loss.

of HAART is modified by gender and by baseline CD4 cell count categories. Since we are comparing the static regimens "treat always" and "treat never," baseline CD4 count is the CD4 count that subjects would have had at HAART initiation. To our knowledge, this is the first application of marginal structural models with planned exploration of effect modification by baseline covariates. The proportional hazards assumption was not rejected when we estimated the effect of HAART in subperiods (halves) of follow-up time (robust $p = 0.58$) (21).

We also estimated the joint effects of HAART and PCP prophylaxis on time to AIDS or death using a marginal structural model (15). Briefly, we restricted the analysis to the 1,016 (of 1,498) men and women who were naïve to HAART and had not been on PCP prophylaxis during the year prior to study initiation and then estimated inverse probability weights for HAART, PCP prophylaxis, and censoring. The final pooled logistic model was weighted by the product of all three weights. This model included baseline covariates, time-varying HAART and PCP prophylaxis, and their interaction. Using this model, we estimated a pair of hazard ratios for AIDS or death. The first hazard ratio was for the comparison of HAART with no HAART under continuous PCP prophylaxis, while the second was for the comparison of HAART with no HAART under no PCP prophylaxis. All analyses were conducted using SAS, version 8 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

Table 1 presents the data at study entry and averaged over AIDS-free survival time for the 1,498 participants who were followed for up to 6.5 years (median, 5.4 years). During 6,763 person-years of follow-up, 323 incident cases of clinical AIDS and 59 deaths occurred, yielding an incidence of six events per 100 person-years for the combined study endpoint. Of 1,116 censored participants, 857 (77 percent) were still under observation in April 2002. At study entry, the participants had a median age of 39 years; 66 percent were female, and 37 percent were Caucasian. Seventeen percent of the participants had CD4 cell counts less than 200 cells/mm^3 at study entry, while 58 percent had CD4 counts greater than 350 cells/mm^3. Twenty-eight percent of the participants had plasma RNA levels less than 401 copies/ml, while 54 percent had RNA levels greater than 10,000 copies/ml.

TABLE 2. Characteristics associated with initiation of highly active antiretroviral therapy for 1,498 human immunodeficiency virus-positive US men and women, 1995–2002

	HR*,†	95% CI*
Male gender	1.11	0.90, 1.35
Age (per 10 years)	0.97	0.88, 1.07
Caucasian race	1.33	1.09, 1.61
CD4 cell count (cells/mm^3)‡		
200–350 vs. >350	1.76	1.47, 2.12
<200 vs. >350	2.26	1.79, 2.85
HIV* RNA level (copies/ml)‡		
401–10,000 vs. <401	2.24	1.81, 2.77
>10,000 vs. <401	2.75	2.22, 3.41
Antiretroviral therapy‡		
Monotherapy vs. none	2.16	1.63, 2.86
Combination therapy vs. none	4.88	4.11, 5.79
PCP* prophylaxis‡	1.14	0.93, 1.39
≥1 HIV-related symptom‡,§	1.04	0.87, 1.24

* HR, hazard ratio; CI, confidence interval; HIV, human immunodeficiency virus; PCP, *Pneumocystis carinii* pneumonia.

† Adjusted for all variables in the table as well as number of days since the prior visit.

‡ Time-varying characteristic from the prior visit.

§ Symptoms included persistent fever, diarrhea, night sweats, and weight loss.

TABLE 3. Estimated effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death for 1,498 human immunodeficiency virus-positive US men and women, 1995–2002

Model and HAART* use	No. of events	Person-years of follow-up	HR*	95% CI*
Unadjusted				
No HAART	238	3,581	1	
HAART	144	3,182	0.98	0.76, 1.26
Adjusted†			0.81	0.61, 1.07
Weighted†				
No HAART	246	3,586	1	
HAART	125	3,124	0.54	0.38, 0.78‡

* HAART, highly active antiretroviral therapy; HR, hazard ratio; CI, confidence interval.

† Both the adjusted standard model and the weighted marginal structural model accounted for the same set of covariates, namely age, gender, race, calendar year at entry, and baseline CD4 and RNA categories, as well as time-varying CD4 count, RNA level, symptoms related to human immunodeficiency virus, antiretroviral therapy, *Pneumocystis carinii* pneumonia prophylaxis, and number of days since the prior visit. The time-varying covariates were included as regressors in the adjusted standard model only.

‡ Robust 95% confidence interval.

The incidence of HAART initiation was 22 per 100 person-years (918 participants initiated HAART). Hazard ratios and 95 percent confidence intervals for initiation of HAART are presented in table 2. Caucasians initiated HAART at 1.33 times the rate of non-Caucasians (95 percent CI: 1.09, 1.61). Participants with prior-visit CD4 counts less than 200 cells/mm^3 initiated HAART at more than twice the rate of those with CD4 counts greater than 350 cells/mm^3 (HR = 2.26, 95 percent CI: 1.79, 2.85). Participants with RNA levels greater than 10,000 copies/ml at the prior visit were nearly three times more likely to initiate HAART than those with less than 401 copies/ml (HR = 2.75, 95 percent CI: 2.22, 3.41). In addition, use of antiretroviral therapy at the prior visit was strongly associated with an increased rate of HAART initiation.

Table 3 shows various estimates of the hazard ratio for clinical AIDS or death due to HAART. The unadjusted hazard ratio suggested no benefit from HAART (unadjusted HR = 0.98, 95 percent CI: 0.76, 1.26). This apparent lack of treatment benefit is due to the strong confounding by indication for therapy, whereby the sickest participants are the most likely to initiate HAART (see table 2) (5). The hazard ratio adjusting for both baseline characteristics (age, gender, race, calendar year at entry, antiretroviral therapy, and baseline CD4 and RNA categories) and time-varying characteristics (CD4 count, RNA level, HIV symptoms, antiretroviral therapy, PCP prophylaxis, and days since last visit) using a standard time-varying Cox (i.e., pooled logistic) model was 0.81 (95 percent CI: 0.61, 1.07). The hazard ratio from a marginal structural model, which accounted for the same set of covariates, was 0.54 (robust 95 percent CI: 0.38, 0.78). The empirical 95 percent confidence interval obtained by bootstrapping was 0.37, 0.81; this suggests that the robust interval was not noticeably conservative in this example. Analysis with the inverse probability-of-treatment weights trimmed at the first and 99th percentiles produced results similar to those of the untrimmed analysis (HR = 0.62, robust 95 percent CI: 0.44, 0.88). Restricting the analysis to 600 (40 percent of 1,498) participants who were on either monotherapy or combination antiretroviral therapies at study entry yielded a similar hazard ratio, albeit with less precision because of the reduced number of events (HR = 0.51, robust 95 percent CI: 0.29, 0.87).

The effects of HAART were similar among men and women (robust $p = 0.87$). Figure 1 depicts the heterogeneity of the effect of HAART by baseline CD4 cell count. Specifically, HAART had a stronger relative effect on clinical AIDS or death among participants who had lower baseline CD4 cell counts (robust $p < 0.01$). Among those with CD4 counts less than 200 cells/mm^3 at baseline, the hazard ratio was 0.36 (robust 95 percent CI: 0.20, 0.64), while among those with CD4 counts of 200–350 cells/mm^3, the hazard ratio was 0.46 (robust 95 percent CI: 0.27, 0.81). There was no strong benefit from HAART for participants with CD4 counts greater than 350 cells/mm^3 at study entry (HR = 0.82, robust 95 percent CI: 0.54, 1.27).

The subgroup analysis (in 1,016 of 1,498 subjects) carried out to estimate the joint effects of HAART and PCP prophylaxis did not detect a statistically significant interaction

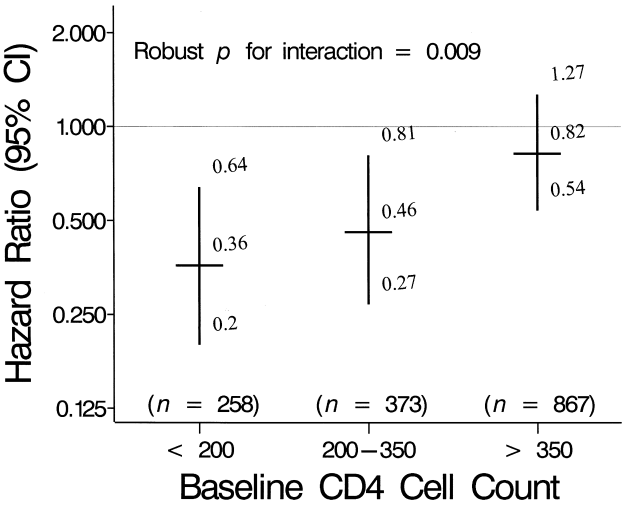

FIGURE 1. Hazard ratio for acquired immunodeficiency syndrome or death due to highly active antiretroviral therapy exposure for 1,498 human immunodeficiency virus-positive US men and women, by category of CD4 cell count (cells/mm^3) at baseline, 1995–2002. Vertical bars, 95% confidence interval (CI).

between these two treatments, but the results suggested that the effect of HAART versus no HAART under continuous PCP prophylaxis was stronger (HR = 0.23, robust 95 percent CI: 0.10, 0.53) than the effect of HAART versus no HAART under no PCP prophylaxis (HR = 0.59, robust 95 percent CI: 0.27, 1.27).

DISCUSSION

Observational cohort studies collecting comprehensive longitudinal data provide a valuable source of information supplementing efficacy measures from randomized trials. In the absence of data from randomized trials, prospective observational data are often the best available evidence for assessment of therapeutic effects. Using a marginal structural model, we estimated that the net hazard of clinical AIDS or death was markedly reduced under continuous exposure to HAART in the Multicenter AIDS Cohort and Women's Interagency HIV studies. In contrast, the adjusted hazard ratio from a standard time-dependent Cox model was attenuated towards 1. This latter estimate is probably a null-biased estimate of the net effect of HAART, because this standard approach effectively excludes causal pathways from HAART to prolonged AIDS-free survival that operate through measured intermediate variables (e.g., HIV RNA level and CD4 cell count). Furthermore, in addition to excluding or blocking intermediary causal pathways, the standard approach can also induce selection bias by implicitly conditioning on variables affected by treatment (22, 23).

Exposure to HAART extended time to clinical AIDS or death, with modification by baseline CD4 count: The effect of HAART on the hazard ratio scale was strongest among persons with baseline CD4 counts less than 200 cells/mm^3. Our primary analysis may be thought of as an attempt to use observational data to simulate the results one would obtain in an intention-to-treat analysis of an unmasked randomized clinical trial. In the stratum where data overlap (i.e., baseline CD4 count less than 200 cells/mm^3), our result is consistent with, albeit stronger than, that of the AIDS Clinical Trials Group 320 randomized trial (1). Our somewhat stronger result may be due to 1) our comparison group's being more heterogeneous (i.e., therapy-naïve, monotherapy and combination therapy) than that in the trial (i.e., combination therapy alone), 2) our duration of follow-up being considerably longer than the trial's, 3) noncompliance with initial randomized assignment in the trial, 4) model misspecification or other uncontrolled sources of bias in our observational analysis, and/or 5) sampling variability. The secondary analysis of the joint effects of HAART and PCP prophylaxis initiation may be thought of as an observational analog to an intention-to-treat analysis of an unmasked 2×2 factorial randomized clinical trial.

Our result is consistent with the findings of Detels et al. (24). They used calendar period as an instrumental variable (25) for HAART exposure in a subset of Multicenter AIDS Cohort Study men for whom seroconversion dates were known (n = 536) and reported a hazard ratio for incident AIDS or death of 0.35 (95 percent CI: 0.20, 0.61) in a comparison of the time period following HAART introduction with the time period of monotherapy. Men initiating HAART in the analysis of Detels et al. were likely to have low CD4 counts, because the sickest individuals were treated with HAART during its inception.

We found no strong beneficial effect of HAART for persons with baseline CD4 counts greater than 350 cells/mm^3. This result differs from the results of Jacobson et al. (26) among Multicenter AIDS Cohort Study participants who initiated HAART at CD4 counts greater than 350 cells/mm^3. Jacobson et al. compared the static regimens "treat always" and "treat never" using historical controls. The present analysis compared the static regimens "treat always" and "treat never" using contemporaneous controls. Jacobson et al. demonstrated notable HIV disease progression in the stratum where baseline CD4 count was greater than 350 cells/mm^3 using the historical comparison group. In contrast, in our analogous stratum, a large proportion of the contemporaneous controls did not demonstrate notable HIV disease progression. Thus, our results in the stratum where baseline CD4 count was greater than 350 cells/mm^3 are probably approximately equal to those from a comparison of the static regimen "treat always" with the dynamic regimen "treat *when* CD4 count is less than 350." Evidence is accumulating on the finding that initiating HAART *while* the CD4 count is greater than 350 cells/mm^3 may not confer additional protection relative to initiating HAART when the CD4 count reaches 350 cells/mm^3 (27–29).

Our hazard ratio estimate can be interpreted as the net effect of HAART only under the assumptions of no unmeasured confounding, no unmeasured informative censoring, and no model misspecification. The foremost assumption may hold approximately, because the most important clinical and laboratory data used by physicians as indications to initiate HAART were collected and used in models for the inverse probability-of-treatment weight (5). Neither the present analyses nor past analyses (14, 15) suggested that

there was notable informative censoring in these data due to measured covariates. Regarding model misspecification, exploration of a broad class of functional forms and summary measures of covariate histories (as described in Materials and Methods) did not appreciably alter our results. However, our 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 bias the estimated hazard ratio in either direction (30). An explicit examination of the sensitivity of our findings to such coarse measurement is warranted. Exploration of the change in a biomarker (e.g., CD4 count) may provide a more sensitive test of the effect of HAART on HIV disease progression than the clinical endpoints used in this analysis (i.e., AIDS or death). This is the topic of ongoing research.

Inverse probability weighting estimation of marginal structural models is an alternative to *g*-estimation of nested structural models or the *g*-computation formula (6). The nonparametric *g*-formula requires low-dimension data and is therefore practical only in select applications. As with any statistical method, marginal structural models have limitations. First, methods based on inverse probability-of-treatment weights make an internal comparison and therefore are valid to the extent that the unexposed group reflects the potential outcomes of the exposed group had they not been exposed (31). While context-specific arguments can be made that external comparison groups may better reflect the potential outcomes of the exposed group, such external comparisons are subject to a similar comparability assumption. Second, marginal structural models, unlike nested structural models, cannot be applied to scenarios where there is a structural probability of 0 or 1 for treatment at a certain level of the covariates. Third, in assessment of the effect of a dynamic treatment regimen (i.e., when interest lies in describing how a time-varying treatment interacts with a time-varying covariate), marginal structural models are less useful than nested structural models (6). Our analysis concentrated on the regimens "treat always" and "treat never." Therefore, our analysis does not directly answer the question of when, with respect to the evolution of CD4 cell count, to initiate HAART. To answer such a question with randomized data, one would conduct a "deferment" trial, wherein, for example, patients with CD4 counts between 200 cells/mm^3 and 350 cells/mm^3 are randomized to immediate treatment with HAART or HAART treatment deferred until the CD4 count crosses below 200 cells/mm^3. In future work using nested structural models and these observational data, we will attempt to answer such questions. We expect that more marked differences between structural and standard methods will be found as an increasing number of epidemiologists become familiar with these novel and appealing quantitative methods.

ACKNOWLEDGMENTS

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 UO1-AI-35042, 5-MO1-RR-00722 (General Clinical Research Center), UO1-AI-35043, UO1-AI-37984, UO1-AI-35039, UO1-AI-35040, UO1-AI-37613, and UO1-AI-35041). The Women's Interagency HIV Study is 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, the National Institute of Dental and Craniofacial Research, the Agency for Health Care Policy and Research, and the Centers for Disease Control and Prevention (grants U01-AI-35004, U01-AI-31834, U01-AI-34994, U01-AI-34989, U01-HD-32632 (National Institute of Child Health and Human Development), U01-AI-34993, U01-AI-42590, and M01-RR00083). (Study websites are located at http://www.statepi.jhsph.edu.)

Dr. Miguel Hernán was supported by National Institutes of Health grant K08-AI-49392, and Dr. James Robins was supported by National Institutes of Health grant R01-AI-32475.

Data were collected by the Multicenter AIDS Cohort Study Investigators and the Women's Interagency HIV Study Collaborative Study Group. Study centers/groups (and Principal Investigators) are as follows: *Multicenter AIDS Cohort Study*—Johns Hopkins Bloomberg School of Public Health (Drs. Joseph B. Margolick and Alvaro Muñoz), Baltimore, Maryland; Howard Brown Health Center and Northwestern University Medical School (Dr. John Phair), Chicago, Illinois; University of California, Los Angeles (Drs. Roger Detels and Beth Jamieson), Los Angeles, California; and University of Pittsburgh (Dr. Charles Rinaldo), Pittsburgh, Pennsylvania; *Women's Interagency HIV Study*—New York City/Bronx Consortium (Dr. Kathryn Anastos); Brooklyn, New York (Dr. Howard Minkoff); Washington, DC, Metropolitan Consortium (Dr. Mary Young); Connie Wofsy Study Consortium of Northern California (Drs. Ruth Greenblatt and Phyllis Tien); Los Angeles County/Southern California Consortium (Dr. Alexandra Levine); Chicago Consortium (Dr. Mardge Cohen); and Data Coordinating Center (Dr. Alvaro Muñoz).

REFERENCES

1. 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.
2. Muñoz A, Gange SJ, Jacobson LP. Distinguishing efficacy, individual effectiveness and population effectiveness of therapies. AIDS 2000;14:754–6.
3. Phillips AN, Grabar S, Tassie JM, et al. Use of observational databases to evaluate the effectiveness of antiretroviral therapy for HIV infection: comparison of cohort studies with randomized trials. EuroSIDA, the French Hospital Database on HIV and the Swiss HIV Cohort Study Groups. AIDS 1999;13:2075–82.
4. Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983;2:267–71.

5. 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.
6. Robins JM. Structural nested failure time models. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. Chichester, United Kingdom: John Wiley and Sons, 1998:4372–89.
7. 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):139S–61S.
8. 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.
9. 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.
10. Barkan SE, Melnick SL, Preston-Martin S, et al. The Women's Interagency HIV Study. WIHS Collaborative Study Group. Epidemiology 1998;9:117–25.
11. Panel on Clinical Practices for Treatment of HIV Infection, US Department of Health and Human Services and Henry J. Kaiser Family Foundation. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Bethesda, MD: AIDSinfo [formerly HIV/AIDS Treatment Information Service], National Institutes of Health, 2000. (World Wide Web URL: http://www.aidsinfo.nih.gov).
12. 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.
13. Hessol NA, Schwarcz S, Ameli N, et al. Accuracy of self-reports of acquired immunodeficiency syndrome and acquired immunodeficiency syndrome-related conditions in women. Am J Epidemiol 2001;153:1128–33.
14. 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.
15. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of non-randomized treatments. J Am Stat Assoc 2001;96:440–8.
16. 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.
17. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 1952; 47:663–85.
18. White HA. Maximum likelihood estimation of misspecified models. Econometrica 1982;50:1–25.
19. Pierce DA. The asymptotic effect of substituting estimators for parameters in certain types of statistics. Ann Stat 1982;10: 475–8.
20. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Am Stat Assoc 1994;89:846–6.
21. Altman DG, De Stavola BL. Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates. Stat Med 1994;13:301–41.
22. Cole SR, Hernán MA. Fallibility in estimating direct effects. Int J Epidemiol 2002;31:163–5.
23. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992;3:143–55.
24. Detels R, Muñoz A, McFarlane G, et al. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study Investigators. JAMA 1998;280:1497–503.
25. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722–9.
26. Jacobson LP, Li R, Phair J, et al. Evaluation of the effectiveness of highly active antiretroviral therapy in persons with human immunodeficiency virus using biomarker-based equivalence of disease progression. Am J Epidemiol 2002;155:760–70.
27. Hogg RS, Yip B, Chan KJ, et al. Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy. JAMA 2001;286:2568–77.
28. Anastos K, Barron Y, Miotti P, et al. Risk of progression to AIDS and death in women infected with HIV-1 initiating highly active antiretroviral treatment at different stages of disease. Arch Intern Med 2002;162:1973–80.
29. Cole SR, Li R, Anastos K, et al. Lead-time adjustment in cohort studies: evaluating when to initiate therapy. (Abstract TuOrB1143). Abstracts of the XIV International AIDS Conference, Barcelona, Spain, July 7–12, 2002. (World Wide Web URL: http://www.aids2002.com).
30. Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol 1985;122:495–506.
31. Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol 2002;31:422–38.