

ESTIMATING THE EFFECT OF ZIDOVUDINE ON KAPOSI'S SARCOMA FROM OBSERVATIONAL DATA USING A RANK PRESERVING STRUCTURAL FAILURE-TIME MODEL

MARSHALL M. JOFFE¹*, DONALD R. HOOVER², LISA P. JACOBSON², LAWRENCE KINGSLEY³, JOAN S. CHMIEL⁴, BARBARA R. VISSCHER⁵ AND JAMES M. ROBINS⁶

¹ *Division of Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 602 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, U.S.A.*

² *Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, 624 N. Broadway, Baltimore, MD 21205, U.S.A.*

³ *Departments of Infectious Diseases and Microbiology and of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15213, U.S.A.*

⁴ *Department of Preventive Medicine, Northwestern University Medical School Biometry Section, Suite 1104, 680 N. Lake Shore Drive, Chicago, IL 60611-4402, U.S.A.*

⁵ *Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90095, U.S.A.*

⁶ *Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115-6096, U.S.A.*

SUMMARY

Researchers commonly express scepticism about using observational data to estimate the effect of a treatment on an outcome the treatment is intended to affect. In this paper, we consider using data from the Multicenter AIDS Cohort Study (MACS) to determine whether zidovudine prevents the development of Kaposi's sarcoma among HIV-positive gay men. Several methodologic issues common to observational data characterized the study: information on potentially important confounders was missing at some study visits; investigators did not always know the time of changes in treatment level, nor the value of confounders at that time, and the censoring process depended strongly on time-varying covariates related to outcome. We describe application to our data of Robins' paradigm for defining, modelling and estimating the effect of a time-varying treatment and show how to modify his approach to deal with the methodologic issues we have mentioned. Further, we demonstrate that relative risk regression is less well equipped to deal with these issues. We compare our results to the findings from randomized trials, and conclude that observational studies may sometimes be useful in evaluating the effect of treatment on an intended outcome. © 1998 John Wiley & Sons, Ltd.

* Correspondence to: Marshall M. Joffe, Division of Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, 602 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, U.S.A. E-mail: mjoffe@cceb.upenn.edu.

Contract grant sponsor: National Institute of Allergy and Infectious Diseases,

Contract grant numbers: U01-A1-35042, 5-MO-RR-00722(GCRC), U01-A1-35043, U01-A1-37984, U01-A1-35039, U01-A1-35040, U01-A1-37613, U01-A1-35041.

Contract grant sponsor: National Cancer Institute

Contract grant numbers: CA 09142

1. INTRODUCTION

In a recent series of papers,^{1–4} Robins proposed a novel approach for using observational data to estimate the causal effect of a time-varying treatment on an outcome of interest. This approach involves a new class of models for treatment effect, the structural nested failure time models (SNFTMs), and a new estimation method, G-estimation. The approach has seldom been used in applications. Mark and Robins⁵ used the approach to estimate the effect of smoking cessation; Robins and Greenland⁶ used it as a component of an analysis of a clinical trial of zidovudine.

We report here an analysis using this approach to estimate the effect of zidovudine (AZT) on Kaposi's sarcoma (KS) from an observational study of men infected with the human immunodeficiency virus (HIV). Because of the nature of our data, we had to deal with several methodologic problems and issues in our analysis:

1. We were missing information on important confounders for some study subjects at some points during their follow-up.
2. Many subjects in our study either died without developing KS or were lost to follow-up before the end of the study.
3. There were several time-varying covariates that were strong predictors of subsequent treatment and of disease risk. Some of these covariates were also strongly affected by previous treatment.
4. For many study subjects, it was unclear whether treatment preceded disease or vice versa.
5. Changes in treatment generally occurred between visits, times for which concurrent confounder information was unavailable.

We believe that these issues are frequently of concern in observational studies, especially when the goal of the study is to estimate the effect of a treatment on an outcome it is intended to influence.

The goal of this paper is to show that, with some modifications to be described, structural nested failure-time models and G-estimation provide an appropriate and practical framework for dealing with these issues. To do this, we discuss the relevant methodologic issues in detail, show how to use SNFTMs and G-estimation to deal with these problems, and propose, apply, and justify adaptations to the approach to deal with issues 4 and 5. To our knowledge, this is the first report of application of this approach to observational data in which issues 1–3 are prominent. Recent clinical trials of AZT provide a 'gold standard' against which to compare our findings. We show that careful attention to the principles we outline improves the consistency of our observational estimates with the results of clinical trials. We also illustrate some difficulties in applying standard relative risk regression to our data.

Section 2 discusses what is known about the association between AZT and KS. Section 3 describes the Multicenter AIDS Cohort Study (MACS), the source of our data, and Section 4 provides appropriate structure and notation for the data. Section 5 defines the causal effect of a time-varying treatment on a failure-time outcome and Section 6 discusses a model for such effects. Section 7 discusses conditions that permit valid estimation of the model parameters, while Section 8 shows how to estimate these parameters. Section 9 presents estimates obtained using the MACS data. In Section 10 we describe problems that arise when treatment changes at times for which covariate information is unavailable and discuss how we dealt with these problems. Sections 11 and 12 outline how we dealt with uncertainty concerning times when treatment changes. Sections 13 and 14 describe how to deal with censoring by competing causes. Section 15

discusses problems with relative risk estimation from the MACS. The paper concludes with a discussion about the use of observational data for estimating treatment effects.

2. AZT AND KAPOSÍ'S SARCOMA

Kaposi's sarcoma (KS) continues to be a common condition among HIV-positive gay men.^{7,8} KS may involve many sites: skin manifestations produce unsightly lesions; pulmonary involvement causes severe respiratory symptoms.^{9,10} Further, some approaches to treatment (for example, the cancer chemotherapeutic drugs vinblastine and bleomycin) of KS themselves produce substantial morbidity.⁹ Any treatment that prevents or postpones the occurrence of KS could potentially improve the quality of life of many people infected with HIV.

Zidovudine (AZT) has been the primary chemotherapeutic agent used to slow the progression of HIV-associated immune deficiency and to delay the onset of associated clinical manifestations. Many randomized trials¹¹⁻²¹ and several observational studies²²⁻²⁸ have examined the effects of AZT. With justification, reports of these studies have concentrated on aggregate measures of AZT effect, using as primary endpoints death, the initial diagnosis of the acquired immunodeficiency syndrome (AIDS), the development of specified symptoms, or some measure of quality of life.²⁹⁻³¹ However, there has been relatively little discussion of the effect of antiretroviral agents on specific AIDS-defining illnesses, including KS. Presumably, this is in part a consequence of the small number of cases of KS except in one large, recent study.¹¹

Although there has been little discussion, several randomized clinical trials of AZT provide data that bear on our question. Table I lists the randomized clinical trials of AZT from which data on KS are available. We used published reports to abstract the number of cases of KS and estimate the amount of person-time in each arm of the various trials. Table I provides a summary.

Using a Poisson regression model for the placebo-controlled randomized trials, we estimated that the rate of KS in the AZT treatment group was 0.64 times the rate in the placebo group; the 95 per cent confidence interval for the rate ratio was 0.44 to 0.93. These data provide little indication of heterogeneity of the rate ratio across studies. However, given the small number of cases of KS in most of the studies, one would expect tests of heterogeneity to have little power. The studies themselves were heterogeneous in several ways: the dose of AZT used; the stage of HIV disease at outset, and the proportion in the 'placebo' group who used AZT at some point during follow-up.

An observational analysis of the effect of AZT on KS may add to our knowledge in two ways. First, large observational studies, such as the MACS, can add additional precision to estimates of AZT effect. This role is reduced by the several potential biases (some discussed below) that may afflict observational studies but not randomized ones. A second role of such analyses is to evaluate different methods for causal inference from observational data. Availability of the more reliable randomized trial results aids this evaluation. This report concentrates on the second role.

3. THE MULTICENTER AIDS COHORT STUDY

The Multicenter AIDS Cohort Study (MACS) is an observational study of a cohort of homosexual and bisexual men in the United States. Between April 1984 and March 1985, the study enrolled 4954 homosexual and bisexual men from four metropolitan areas: Los Angeles; Baltimore-Washington; Pittsburgh, and Chicago. Study participants were asked to return semi-annually for follow-up visits. For each visit, subjects were asked to complete a detailed interview.

Table I. Randomized trials of AZT; KS cases

Trial	Dose (mg/day)	Treatment group		Placebo group	
		Cases	Person-years	Cases	Person-years
<i>Placebo-controlled studies</i>					
Concorde ¹¹	1500	27	2717	41	2702
Mulder <i>et al.</i> ¹⁸	1000	3	178	1	193
Cooper <i>et al.</i> ¹²	1000	2	875	3	895
Hamilton <i>et al.</i> ¹⁷	1500	6	385	7	395
Fischl <i>et al.</i> ¹⁵	1200	2	322	3	330
Volberding <i>et al.</i> ²¹	1500	1	448	5	502
	500	3	479		
Fischl <i>et al.</i> ¹³	1500	6	48	10	44
<i>Additional studies of different doses</i>					
Nordic Medical Research Councils ¹⁹	1200	13	247		
	800	13	250		
	400	12	253		
Fischl <i>et al.</i> ¹⁴	1500	40	554		
	600	26	570		
				This group received 1200 mg/day for the 1st 4 weeks	

Rate ratio from Poisson regression of placebo-controlled studies

Point estimate	95% CI
0.68	0.47, 0.97

Likelihood-ratio test for heterogeneity of rate ratio:

Deviance	D.F.	<i>p</i> -value
3.05	6	0.80

Further, at each visit, subjects underwent a physical examination and provided blood for haematologic, serologic and virologic study. The MACS used active surveillance to obtain medical records to confirm KS diagnoses.^{10,32} Of the initial cohort, 1809 were seropositive to HIV at study entry, and 414 additional subjects had seroconverted by visit 16 (September 1991–May 1992). Other references^{10, 24, 25, 32–34} provide additional details on the rationale behind and the methods of data collection used in the MACS.

A quirk in the history of the HIV epidemic motivated further restriction of the study population. AZT was approved by the Food and Drug Administration for general use in 1987 (generally after visit 5) and so was not generally available before that time. Because of this, follow-up for our analysis began at visit 5. Of the 2223 subjects infected with HIV during the study period, 1847 attended at least one visit between visits 5 and 17 while HIV-positive, and so were eligible for inclusion in the analysis.

4. THE OBSERVABLE DATA

A proper analysis of the MACS data requires accounting for the relationships between changing levels of treatment and covariates and between these quantities and the time of KS. Because of the important role time plays, we discuss briefly how to deal with time, together with notation for our variables of interest.

We use two time scales in the rest of this report. One scale reflects the discreteness of the process by which the MACS obtained treatment and covariate information, while the other reflects the continuous nature of changes in treatment, covariates, and disease status.

We distinguish between our treatment of interest, AZT, other variables of interest, to which we subsequently refer as covariates, and the outcome of interest, KS. The MACS obtained treatment and covariate information from each individual i at each study visit k , $k = 1, \dots, 17$. Denote by $L_{i,k}$ the vector of time-dependent covariates recorded for subject i at visit k . $L_{i,1}$ includes time-independent covariates. Beginning at visit 6, each subject reported on use of AZT since the previous visit. Let $Z_{i,k}$ be an indicator of AZT use between visits k and $k + 1$; $Z_{i,k} = 1$ if a subject reported AZT use at visit $k + 1$ and $Z_{i,k} = 0$ otherwise. Overbars denote the history of any time-dependent process; for example $\bar{L}_{i,k} \equiv (L_{i,1}, \dots, L_{i,k})$ is the history of the recorded covariate process to visit k .

This discrete, visit-based time scale provides a convenient way to describe how treatment and covariate information was recorded. More adequate representation of the underlying covariate, treatment and disease processes of interest requires a continuous time scale.

We define the zero point of the time scale as the start of follow-up. This is the date of the first visit at which subjects might have used AZT and were at risk for KS. If a subject was seropositive at visit 5 and did not skip study visit 5, this was visit 5. Otherwise, it was the first visit after visit 5 that the subject attended and at which he was seropositive.

We define other events in reference to this zero point. One unit on this scale is 180 days, roughly the planned time between visits. The time of visit k , $t_{i,k}$, is defined as the length of time visit k is after the start of follow-up; for a subject seropositive at visit 5 who returns for visit 6 180 days later, $t_{i,6} = 1$. We denote by T_i the time (on this scale) that subject i developed KS. For a time-dependent process, we use parentheses to denote the value of that process at a particular time; thus, $L_i(t)$ is the value of the covariate vector for subject i at time t . For now, assume that at each time $t \in (t_k, t_{k+1}]$ a subject's treatment status $Z_i(t)$ is available; Sections 11 and 12 discuss inference when, as in the MACS, this assumption is violated.

Although the MACS is still ongoing, this analysis considered only events occurring before visit 17 (late 1992). Consequently, the date of KS was not observed for many subjects who have subsequently developed or will later develop KS. Let C_i be a potential censoring time; subjects who develop KS before this time are recorded as cases in our study, while others are censored at this time. In this analysis, C_i is the time of visit 17 for subject i , measured with respect to the zero point. Some subjects missed visit 17. For these subjects, we can compute an approximate date visit 17 would have occurred had the subject continued to attend visits regularly. Suppose a subject's last visit to the MACS was study visit k . We take C_i to be $182 \times (17 - k)$ days after $t_{i,k}$, the time of visit k . Note that C_i is known approximately at the beginning of the study; we consider it a pretreatment covariate and include it in $L_{i,1}$. C_i is useful for the estimation procedures developed in Section 8. We defer discussion of and notation for censoring by death and loss to follow-up to Section 13.

5. THE EFFECT OF A TIME-VARYING TREATMENT

Following Robins and coauthors,¹⁻⁵ we define the causal effect of a possibly time-varying treatment in terms of latent or potential failure-time variables. For each individual i we assume there exists a set of possibly unobserved failure times $U_{i,v}$. $U_{i,v}$ is the time that subject i would fail (develop KS) under a *regime* or *generalized treatment* v , under which he receives his observed level

of treatment to time v and receives no treatment thereafter. We observe $U_{i,v}$ if regime v and regime T_i (the regime under which a subject receives his observed treatment level to the observed failure time T_i and no treatment thereafter) are identical to T_i .

Causal effects are comparisons of the latent failure-time variables $U_{i,v}$. One such effect is a comparison of the observed failure time $T_i \equiv U_{i,T_i}$ with the potential failure time $U_{i,v}$. This comparison is the effect for individual i of following his observed treatment regime for the last $T_i - v$ units of time.

For notational convenience, write $U_i \equiv U_{i,0}$ and $U_{i,k} \equiv U_{i,t_{i,k}}$. U_i is the time subject i would fail if he never received treatment; we sometimes refer to it as that subjects' prognosis. $U_{i,k}$ is the time subject i would fail if he received his observed treatment through the time of visit k and no treatment thereafter.

6. A MODEL FOR TREATMENT EFFECT

Because of the large variety of patterns of treatment followed and of patterns of time-varying confounders L , non-parametric identification of the causal effects of AZT is impractical even using large databases such as the MACS.³⁵⁻³⁷ Models can make estimation practical in such settings.^{4,37,38}

To model the effect of AZT on the time to KS, we use a generalization of the one parameter accelerated failure-time model with time varying covariates.^{1-5,39} The deterministic or strong form of this model allows one to compute the potential failure time U_i from: (i) the observed failure time T_i ; (ii) the observed treatment history $\bar{Z}_i(T_i)$; and (iii) a scalar parameter Ψ_0 representing the causal effect of each bit of AZT use on failure:

$$U_i = \int_0^{T_i} \exp\{Z_i(t)\Psi_0\} dt. \tag{1}$$

Under this model, each month of AZT use adds $1 - \exp(\Psi_0)$ months to the time to KS of people using AZT. Consider two subjects with identical treatment histories; according to the model, the subject who actually failed first would also have failed first had, possibly contrary to fact, both of them never been treated. This is a rank preserving structural nested failure-time model (RPSNFTM), because the ranks of failure times are preserved under different treatment regimes.¹⁻⁵

There are some indications in the literature that the effects of AZT may be limited in time;^{11,40,41} if true, model (1) is too restrictive. Let S_i denote the time that a subject first uses AZT; for subjects who never use AZT, we write $S_i = \infty$. Suppose that, for the first M months after initiation of AZT, the effect of each month of AZT is to add $1 - \exp(\Psi_0)$ months to the time to KS; each month after the first M months adds $1 - \exp(\gamma_0)$ months to this time. More precisely, the corresponding generalization of (1) is

$$U_i = \int_0^{T_i} \exp[Z_i(t)\{l(t - S_i \leq M)\Psi_0 + l(t - S_i > M)\gamma_0\}] dt \tag{2}$$

where $l(r) = 1$ if r is true, 0 otherwise. Equation (2) preserves the rank-preserving characteristics of (1). For our analysis, we took $M = 30$ months (= 5 units). We concentrate on the scalar parameter Ψ_0 in (2) rather than the vector parameter $\{\Psi_0, \gamma_0\}$ because of computational considerations and because, even if (2) were the true model, concentrating on Ψ_0 allows more

direct comparison of our results with those of randomized trials, because of the shorter length of follow-up in those trials.

In this context, the effect of treatment received more than 30 months after AZT initiation (that is, after $S_i + M$) is a nuisance parameter. We consider now a generalization of the model for AZT effect (2) which allows the effect of treatment received after 30 months to remain completely unspecified. Such a model implies that the potential failure time U_i can be calculated from the parameter of interest Ψ_0 and observable quantities only when failure is observed within M months of treatment initiation; when failure is observed later, only a minimum potential failure time can be calculated, that is

$$U_i = \int_0^{T_i} \exp(Z_i(t)\Psi_0) dt, \quad T_i \leq S_i + M$$

$$U_i > \int_0^{S_i + M} \exp(Z_i(t)\Psi_0) dt, \quad T_i > S_i + M. \quad (3)$$

Section 8 uses (3) as a basis for estimation of Ψ_0 ; we call (3) a semi-parametric rank preserving structural nested failure-time model (SRPSNFTM).

Models (1), (2) and (3) make the strong non-interaction assumptions that the effect of treatment varies with neither measured nor unmeasured covariates. Robins and coauthors^{4,42} provide a class of failure-time models, the structural nested failure-time models, which includes the rank preserving form as a subclass but allows the magnitude of the treatment effect to depend on unmeasured factors and measured covariates. The G-estimation procedures described below apply to this more general class of models.

7. CONDITIONS FOR VALID ESTIMATION OF Ψ_0

This section discusses conditions sufficient to identify Ψ_0 from observational data such as the MACS. We first present conditions which one can use to justify causal interpretation of standard analyses. Because these conditions are too restrictive for the MACS, we consider weaker and more plausible conditions. Section 8 explains how to estimate Ψ_0 under these weaker conditions.

We present first an assumption implicitly used when investigators assign causal interpretation to observational estimates of treatment effect; Robins^{1-3,5,35-37,43,44} makes this assumption precise. Let $A \perp\!\!\!\perp B \mid C$ mean that A is independent of B , given C . We assume that at each visit k , treatment assignment does not depend on the latent failure time $U_{i,k}$, given recorded treatment and covariate histories to visit k ($\bar{Z}_{i,k-1}$ and $\bar{L}_{i,k}$); that is

$$Z_{i,k} \perp\!\!\!\perp U_{i,k} \mid \bar{Z}_{i,k-1}, \bar{L}_{i,k}, T_i > t_{i,k}. \quad (4)$$

(4) would be true if perceived prognosis ($U_{i,k}$) did not affect treatment decisions at visit k except through the measured variables $\bar{Z}_{i,k-1}$ and $\bar{L}_{i,k}$. Robins^{1-4,6,35-37,43,44} calls this assumption *no unmeasured confounders*. Others have termed this or similar assumptions *maintained comparability*⁵ or *strong ignorability of treatment assignment*.⁴⁵

This assumption was not fully plausible in the MACS. At each visit, the MACS took care to measure information on major predictors of AZT initiation and KS (that is confounders): CD4 count; the development of specific symptoms, and the development of AIDS. Unfortunately, some potentially major determinants of quitting AZT (for example, nausea) were not recorded. Inasmuch as these determinants predict the latent failure times $U_{i,k}$, assumption (4) is incorrect.

Let $\bar{Z}_{i,k-1} = 0$ mean that subject i received no treatment with AZT to visit k . For the MACS, a weaker and more plausible assumption than (4) is that there were no unmeasured confounders for treatment initiation, that is

$$Z_{i,k} \perp\!\!\!\perp U_{i,k} \mid \bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k} \tag{5}$$

We can replace $U_{i,k}$ in (5) with U_i , because for subjects who received no treatment before visit k (that is for $\bar{Z}_{i,k-1} = 0$), $U_{i,k} = U_i$.

We would believe that (5) is true if we believed that we had recorded and included in $\bar{L}_{i,k}$ accurate information on all factors that predict prognosis U_i and treatment initiation (confounders) between visits k and $k + 1$. Suppose that, at each visit, we rolled dice to determine which subjects who had not yet started AZT should start using it; we would then believe (5). Although we would seldom believe (5) to be exactly true in observational studies, inclusion in $\bar{L}_{i,k}$ of more covariates thought to be confounders (predictors of prognosis and treatment initiation) makes it more plausible that (5) approximates the truth. An attempt to make (5) approximate the truth was our primary consideration in variable (confounder) selection in our data analysis; Section 9 provides more details.

8. G-ESTIMATION OF Ψ_0

Our definition of and model for causal effects and our assumption about treatment assignment involve the latent variable U_i . To estimate the effect of treatment, we must translate statements involving this often unobservable quantity into statements involving only observable ones. This section describes how to do this.

We begin by defining observable quantities about which we can make testable predictions. Let Ψ denote a value we hypothesize for Ψ_0 . Define $U_i(\Psi)$ as subject i 's prognosis under our model (3) if the hypothesized value Ψ is correct, that is, $U_i(\Psi_0) = U_i$. We may write

$$U_i(\Psi) = \int_0^{T_i} \exp(Z_i(t)\Psi) dt, \quad T_i \leq S_i + M$$

$$U_i(\Psi) > \int_0^{t_{i,0} + M} \exp(Z_i(t)\Psi) dt, \quad T_i > S_i + M. \tag{6}$$

Were we able to follow all study subjects until they develop KS (this is true if $C_i = \infty$ and $T_i < \infty$ for all i) and were the effect of AZT on time to KS independent of past AZT use (that is, $M = \infty$), we could calculate $U_i(\Psi)$ from observable quantities for all subjects.

Our assumption of no unmeasured confounders for treatment initiation (5) allows us to make testable predictions about $U_i(\Psi)$. Under our model for treatment effect (3), we rewrite this assumption as $Z_{i,k} \perp\!\!\!\perp U_i(\Psi_0) \mid \bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k}$. Consequently,

$$Z_{i,k} \perp\!\!\!\perp U_i(\Psi) \mid \bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k} \tag{7}$$

is true when $\Psi = \Psi_0$, so we test hypotheses $\Psi_0 = \Psi$ by testing (7). Under the assumptions of the last paragraph, all quantities in (7) are observable, so testing (7) is straightforward in principle.

In the MACS, as in most observational studies, $U_i(\Psi)$ is not always observable. We seek to generalize our approach based upon testing the conditional independence in (7). To do this, we define an observable function of $U_i(\Psi)$ which we can substitute for $U_i(\Psi)$ in (7); this function must

be conditionally independent of treatment initiation given covariate history when (5) holds and $\Psi = \Psi_0$.

We consider how to find a function of $U_i(\Psi)$ that is always observable given only the conditioning in (5) and (7) (that is, $\bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k}$). Under our suppositions, we could substitute this function for $U_i(\Psi)$ in (7) to test a hypothesis $\Psi_0 = \Psi$. Consider a subject i who at the time of visit $k(t_{i,k})$ has not yet started AZT. For a subject who starts AZT therapy at visit k , we can calculate $U_i(\Psi)$ only if: (i) failure is observed (that is, it occurs before the potential censoring time C_i , so $T_i < C_i$), and (ii) it occurs within M units of the initiation of therapy (that is, if $T_i < t_{i,k} + M$). $C_{i,k}^* \equiv \min(C_i, t_{i,k} + M)$ is then the latest time for which we can calculate the putative potential failure time $U_i(\Psi)$ for subject i from the observed data using (6).

One might consider computing a modified maximum follow-up time $X_{i,k} \equiv \min(T_i, C_{i,k}^*)$. One could then compute the modified maximum follow-up time $X_{i,k}(\Psi) \equiv t_{i,k} + \int_{t_{i,k}}^{X_{i,k}} \exp\{Z_i(t)\Psi\} dt$, which is an observable quantity. Unfortunately, one cannot simply replace $U_i(\Psi)$ with $X_{i,k}(\Psi)$ in (7), because it is not true in general that the correctly modified follow-up time $X_{i,k}(\Psi_0)$ is conditionally independent of treatment assignment. (Hazard-based alternatives for estimating Ψ_0 , such as those described in reference 46, will also fail in the presence of time-varying confounders.)

To see this, consider comparing subject i with a subject i' , identical to subject i except that he receives no AZT at k . Suppose that AZT in fact lengthens the time to KS ($\Psi_0 < 0$), so failure time $T_{i'} (> C_i)$ remains unobserved for the treated subject i , but is observed and less than C_i for the comparison subject i' . Then the transformed follow-up time for the treated subject ($X_{i,k}(\Psi_0)$) will be less than the transformed failure or follow-up time for the comparison subject ($X_{i',k}(\Psi_0)$) and so the random variable $X_{i,k}(\Psi_0)$ is dependent on treatment assignment at k .

Fortunately there are observable functions of the modified follow-up time $X_i(\Psi)$ which are independent of treatment assignment and so can replace $U_i(\Psi)$ in (7). Let $C_{i,k}^*(\Psi) \equiv t_{i,k} + \min((C_{i,k}^* - t_{i,k}), (C_{i,k}^* - t_{i,k})\exp(\Psi))$ be a modified potential censoring time; $C_{i,k}^*(\Psi)$ is the greatest value of $U_i(\Psi)$ which is observable no matter what treatment regime subject i follows after visit k . Given the conditioning event in (5) and (7) (that is, $\bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k}$; recall that C_i is part of $\bar{L}_{i,k}$), $C_{i,k}^*(\Psi)$ is fixed. Let $X_{i,k}^*(\Psi) \equiv \min(U_i(\Psi), C_{i,k}^*(\Psi))$ be an alternative modified follow-up time, and let $\Delta_{i,k}(\Psi) \equiv I(U_i(\Psi) \leq C_{i,k}^*(\Psi))$ be a modified censoring indicator; the vector $\{X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi_0)\}$ is an observable function of $U_i(\Psi)$. For both untreated subject i' , who was observed to fail, and for treated subject i , who was censored, the modified follow-up time $X_{i,k}^*(\Psi_0)$ is the same and equals $C_{i,k}^*(\Psi_0)$. Similarly, for subjects whose potential failure times U_i are less than $C_{i,k}^*(\Psi_0)$, the modified follow-up times $X_{i,k}^*(\Psi_0)$ equal U_i no matter what treatment is received after visit k . Thus, $X_{i,k}^*(\Psi_0)$ is a variable fixed at visit k , unaffected by treatment received after k ; the assumption of no unmeasured confounders for treatment initiation implies that it is independent of treatment received at k given other variables fixed at k (including $C_{i,k}^*(\Psi_0)$). Similar arguments hold for the modified censoring indicator $\Delta_{i,k}(\Psi_0)$ and for the vector $\{X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi_0)\}$, so conditional independence $(Z_{i,k} \perp\!\!\!\perp \{X_{i,k}^*(\Psi_0), \Delta_{i,k}(\Psi_0)\} | \bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k})$ follows from the assumption of no unmeasured confounders for treatment initiation (5). Consequently, tests of

$$Z_{i,k} \perp\!\!\!\perp \{X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi)\} | \bar{Z}_{i,k-1} = 0, \bar{L}_{i,k}, T_i > t_{i,k} \quad (8)$$

under (5) are tests of hypotheses $\Psi_0 = \Psi$ about the parameter of interest.

The approach sketched above depends on knowing the potential censoring times C_i for all subjects, even those whose failure was observed and so were not censored. Most familiar

estimation procedures for censored survival data (for example, partial likelihood estimation of the proportional hazards model⁴⁷) do not use information on the potential censoring times even when available; however, there are other methods which use this information.^{48–50} This information contained in the potential censoring times allows estimation of the effect of a treatment regime or plan even when treatment changes for some subject-intervals i, k are not ignorable; this is not true for familiar regression approaches to modelling survival data as a function of a time-varying treatment (for example, reference 48; see also Section 15) and for approaches which combine information from these modelling approaches with information of the effect of the time-varying treatment on intermediate variables.^{37, 51, 52}

In practice the potential censoring time C_i was known only approximately and so the method used to assign its values (Section 4) for the analysis was somewhat arbitrary. It is thus important to examine the robustness of the inference to different ways of assigning these values. One may replace the potential censoring times C_i with another set of values C_i^+ with the following characteristics: (i) a failure that occurred before C_i^+ would be observed (this will be true if $C_i^+ \leq C_i$); (ii) C_i^+ can reasonably be treated as a variable fixed for each individual before the beginning of follow-up. Inference then follows from the same arguments sketched above, replacing C_i by C_i^+ throughout in the derivation of modified potential censoring and follow-up variables. As Robins and others^{1, 3–5} point out, there are efficiency advantages in taking C_i^+ to be the largest permissible value (that is, $C_i^+ = C_i$). We examined the robustness of inference to this point and found that choice of an alternative value (for example, March 6 1992, the first date any subject in our study had a visit 17) had little effect on inference (Table IV).

We provide one approach to testing (8). Consider a correctly specified logistic model for treatment initiation:

$$\text{logit}\{p(Z_{i,k} = 1 | \bar{L}_{i,k}, \bar{Z}_{i,k-1} = 0, T_i > t_{i,k})\} = \alpha_k + w(\bar{L}_{i,k})\beta \tag{9}$$

where $w(\cdot)$ is a known function and α_k is a visit-specific intercept term. Because the visit number k is implicitly included in the covariate history $\bar{L}_{i,k}$, this formulation allows the relation of any component of that history (for example, CD4 count at k) to treatment initiation at k to vary with k . This is a model for initiating treatment between visits k and $k + 1$; we sometimes refer to this as a model for AZT initiation *at* visit k . Expand the logistic model to include some function $g(X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi), \bar{L}_{i,k})$:

$$\begin{aligned} \text{logit}\{p(Z_{i,k} = 1 | \bar{L}_{i,k}, \bar{Z}_{i,k-1} = 0, T_i > t_{i,k}, X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi))\} \\ = \alpha_k + w(\bar{L}_{i,k})\beta + g\{X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi), \bar{L}_{i,k}\}\zeta. \end{aligned} \tag{10}$$

Under the assumption of no unmeasured confounders for treatment initiation (5), $\zeta = 0$ if $\Psi = \Psi_0$. Large-sample (for example, score) tests of $\zeta = 0$ in (10) under these assumptions are then tests of $\Psi_0 = \Psi$. This score test is asymptotically equivalent to Rosenbaum's⁵³ likelihood ratio test when the following conditions obtain: (i) treatment has no putative effect (that is, $\Psi = 0$); (ii) failure is always observed so, $X_{i,k}^*(0) = T_i$; (iii) the function $g(\cdot)$ used is $g\{X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi), \bar{L}_{i,k}\} = X_{i,k}^*(\Psi)$, and (iv) there is only a single treatment assignment time k . Section 12 describes and justifies choice of the function $g\{X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi), \bar{L}_{i,k}\}$.

We obtain point and interval estimates of Ψ_0 by inverting the score test of $\zeta = 0$. Robins^{1, 2, 4} termed the estimation procedure sketched in this selection G-estimation, because it estimates the effect of a generalized treatment or regime.

Table II. Alternative logistic models for AZT initiation

	Models for treatment assignment				
	1. 'Crude' analysis	2. Control for CD4	3. Full model	4. Add observed changes after visit k	5. Use adjusted change in CD4
Estimates of Ψ_0					
Point estimate(s)	> 1.45	0.30	0.23	0.06	- 0.01
95% confidence limits	0.71, ∞	0.01, 0.76	- 0.18, 0.75	- 0.42, 0.55	- 0.52, 0.33
Variable	Point estimates (standard error) of coefficients in logistic model for AZT initiation				
1. Intercept (α_5)	- 9.2 (0.5)	- 8.1 (0.6)	- 10.5 (0.8)	- 9.6 (0.8)	- 9.1 (0.8)
Visit k -visit 5 ($\alpha_k - \alpha_5$):					
2. Visit = 6	1.3 (0.2)	1.5 (0.2)	1.9 (0.3)	1.9 (0.3)	1.8 (0.3)
3. Visit = 7	1.8 (0.2)	2.0 (0.2)	2.6 (0.3)	2.7 (0.3)	2.7 (0.3)
4. Visit = 8	1.9 (0.2)	2.0 (0.2)	2.8 (0.3)	2.9 (0.3)	2.9 (0.3)
5. Visit = 9	2.0 (0.2)	2.0 (0.2)	2.9 (0.3)	3.0 (0.3)	3.0 (0.3)
6. Visit = 10	1.9 (0.2)	2.0 (0.2)	3.0 (0.3)	3.1 (0.3)	3.1 (0.3)
7. Visit = 11	2.9 (0.2)	3.3 (0.2)	4.4 (0.3)	4.5 (0.3)	4.5 (0.3)
8. Visit = 12	2.3 (0.2)	2.8 (0.2)	3.9 (0.3)	4.1 (0.3)	4.2 (0.3)
9. Visit = 13	2.1 (0.2)	2.6 (0.3)	3.9 (0.3)	4.0 (0.4)	4.1 (0.4)
10. Visit = 14	1.6 (0.3)	2.4 (0.3)	3.7 (0.4)	3.8 (0.4)	3.8 (0.4)
11. Visit = 15	1.8 (0.3)	2.3 (0.3)	3.7 (0.4)	4.0 (0.4)	3.9 (0.4)
12. Visit = 16	1.4 (0.3)	2.0 (0.3)	3.5 (0.4)	3.7 (0.4)	3.6 (0.4)
13. $\ln(t_{k+1} - t_k)$	0.99 (0.09)	1.2 (0.1)	1.25 (0.11)	0.99 (0.13)	0.91 (0.12)
CD4 count at visit k :					
14. CD4/100		0.00 (0.41)	0.90 (0.46)	0.91 (0.48)	0.84 (0.47)
15. $(\text{CD4}-100)^+/100^*$		- 1.41 (0.58)	- 2.22 (0.62)	- 2.12 (0.64)	- 2.06 (0.64)
16. $(\text{CD4}-200)^+/100$		0.89 (0.36)	0.79 (0.37)	0.75 (0.38)	0.73 (0.38)
17. $(\text{CD4}-300)^+/100$		0.03 (0.22)	0.00 (0.22)	- 0.07 (0.23)	- 0.05 (0.23)
18. $(\text{CD4}-400)^+/100$		- 0.17 (0.16)	- 0.06 (0.16)	- 0.08 (0.17)	- 0.11 (0.17)
19. $(\text{CD4}-700)^+/100$		0.56 (0.16)	0.49 (0.16)	0.48 (0.16)	0.52 (0.16)
Other patient characteristics at visit k :					
20. Oral hairy leukoplakia			0.61 (0.18)	0.54 (0.18)	0.52 (0.19)
21. Diarrhoea			0.38 (0.20)	0.36 (0.20)	0.32 (0.20)
22. Fatigue			0.52 (0.13)	0.56 (0.13)	0.55 (0.13)
23. Thrush			- 0.16 (0.28)	- 0.21 (0.28)	- 0.21 (0.28)
24. Weight loss			0.45 (0.20)	0.37 (0.20)	0.34 (0.20)
25. AIDS			0.93 (0.43)	0.91 (0.44)	0.78 (0.44)
26. $\sqrt{\ln(\text{number male sex partners} + 1)/10}$			- 0.41 (0.87)	- 0.42 (0.89)	- 0.43 (0.89)
27. Birthdate/365.25			- 0.016 (0.006)	- 0.015 (0.006)	- 0.015 (0.006)
28. Study centre: Los Angeles versus other			0.15 (0.08)	0.17 (0.09)	0.15 (0.09)
29. Thrush \times CD4/100			0.12 (0.07)	0.13 (0.07)	0.12 (0.07)
30. $(k - 9) \times \sqrt{(\text{CD4}/100)}$			0.05 (0.03)	0.03 (0.04)	0.03 (0.04)
31. $(k - 9)^2 \times \sqrt{(\text{CD4}/100)}$			- 0.009 (0.009)	- 0.003 (0.009)	- 0.002 (0.009)
32. $(k - 9) \times \text{CD4}/100$			- 0.17 (0.03)	- 0.19 (0.03)	- 0.19 (0.03)

Table II. Alternative logistic models for AZT initiation

	Models for treatment assignment				
	1. 'Crude' analysis	2. Control for CD4	3. Full model	4. Add observed changes after visit <i>k</i>	5. Use adjusted change in CD4
Estimates of Ψ_0					
Point estimate(s)	> 1.45	0.30	0.23	0.06	- 0.01
95% confidence limits	0.71, ∞	0.01, 0.76	- 0.18, 0.75	- 0.42, 0.55	- 0.52, 0.33
Variable	Point estimates (standard error) of coefficients in logistic model for AZT initiation				
33. $V28 \times (k - 9)$			- 0.07 (0.28)	- 0.07 (0.03)	- 0.07 (0.03)
34. AIDS \times (oral hairy leukoplakia + diarrhoea + fatigue + weight loss)			0.40 (0.32)	0.57 (0.32)	0.58 (0.32)
35. AIDS \times thrush			- 1.20 (0.75)	- 1.35 (0.76)	- 1.34 (0.76)
36. Thrush $\times (k - 9)$			- 0.18 (0.06)	- 0.18 (0.06)	- 0.18 (0.06)
37. Thrush $\times (k - 9)^2$			0.04 (0.02)	0.04 (0.02)	0.04 (0.02)
Development of new conditions between visits <i>k</i> and <i>k</i> + 1:					
38. Diarrhoea				0.74 (0.22)	0.71 (0.22)
39. Fatigue				1.19 (0.16)	1.12 (0.17)
40. Herpes zoster				0.59 (0.22)	0.64 (0.22)
41. Night sweats				0.21 (0.40)	0.15 (0.40)
42. $V41 \times CD4/100$				0.08 (0.11)	0.09 (0.11)
43. AIDS				0.51 (0.33)	0.52 (0.34)
44. $V43 \times CD4_k/100$				0.33 (0.11)	0.27 (0.11)
45. $(CD4_{i,k}(0; A, B) - CD4_k)/100$				0.99 (1.04)	- 6.22 (1.16)
46. $V45^2$				32.8 (6.1)	9.9 (6.5)
47. $V45 \times (V38 + V39 + V40 + V41)$				3.26 (1.65)	3.03 (1.68)

For all variables involving CD4 on the natural (square root) scale, the actual estimates are 1/100 (1/10) times the numbers stated in the table. $CD4_{i,k}(0; A, B)$ defined as in (11). In model 5, $A = 1, B = 0$ (that is, $CD4_{i,k}(0; A, B) = CD4_{i,k+1}$); in model 6, $A = 40, B = 0.3$.

*The notation $(X)^+$ means take the positive part of *X*, that is, $(X)^+ = X$ if $X \geq 0$; $(X)^+ = 0$ otherwise. Thus variables 14–19 model the effect of CD4 on AZT initiation as a linear spline with five knots (at 100, 200, 300, 400 and 700)

9. MODELS FOR TREATMENT INITIATION

This and the next section present a series of logistic models (9) for AZT initiation of increasing sophistication and complexity. The complexity reflects increasing concern with controlling every source of confounding and so satisfying the conditions discussed in Section 7. As the models increasingly reflect this concern, the associated observational estimates of treatment effect become more consistent with the results of the randomized studies discussed in Section 2. Table II presents coefficient estimates from each successive logistic model for treatment initiation together with associated estimates of Ψ_0 .

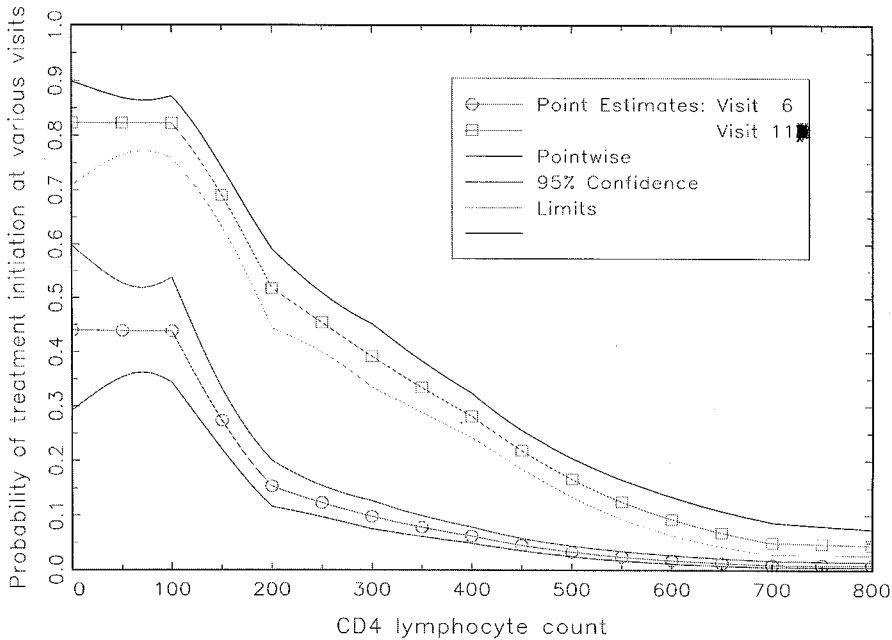


Figure 1. Six-month probability of AZT initiation following visits 6 and 11 as a function of CD4 count, based on model 2

The first and each successive model for AZT initiation include separate terms α_k comparing the rate of initiation in the interval following each visit k with the rate following visit 5. This was because patterns of AZT initiation did not change smoothly over time (see Table II, column 1). At study outset, AZT use was low. After levelling off for a few visits, the rate of AZT initiation increased sharply at visit 11, corresponding to the publication of results from a randomized trial suggesting a benefit for AZT in subjects with less advanced disease.²¹ Rates of AZT initiation decreased slowly thereafter. Our model (9) also includes a term for the logarithm of the time between visit k and visit $k + 1$, because the probability of AZT initiation increases with the length of time between study visits.

Inasmuch as the first model for treatment assignment included no time-independent or time-dependent covariates, estimates of Ψ_0 based on this model are conceptually equivalent to 'crude' or unadjusted estimates. The point estimate (95 per cent confidence interval) for Ψ_0 is > 1.45 (0.77, ∞), that is, AZT appears strongly to cause KS. As we can see by comparison with the results from the randomized trials, and as we shall see by comparison with estimates adjusted for covariates, these estimates are highly confounded.

The next model added CD4 count at visit k to the model for treatment initiation between visits k and $k + 1$. Because CD4 count is such a strong predictor of AZT use, and because the relation does not appear linear, we allowed the relation of CD4 count and the (logit of the) probability of AZT initiation to vary in a piecewise linear fashion. Lower CD4 counts are associated with increased probability of AZT initiation; Figure 1 sketches the relation of CD4 and

the probability of AZT initiation at different study visits. Using this model for AZT initiation, the point (interval) estimate of Ψ_0 is 0.31 (0.02, 0.78).

The next model included many more variables as predictors of AZT initiation between visits k and $k + 1$. We sacrificed model parsimony and interpretability to the goal of faithfully reflecting the assumption of no unmeasured confounders for treatment initiation (5). Under this model, the point (interval) estimate of Ψ_0 was 0.24 (−0.19, 0.60). The confidence interval, and to some degree, the point estimate, are more consistent with randomized trials results than the models considered previously.

10. COVARIATE CHANGES BETWEEN VISITS

Despite an extensive search for covariates to include in the model for treatment initiation, our control of confounding was incomplete. This was due in part to changes in covariate levels between visits. This section elaborates on the last assertion and shows how to modify the models of the last section to adjust for some of the remaining confounding.

It is plausible that treatment initiation between visits k and $k + 1$ depended not only on the value the covariates assumed at visit k but also on their values between visits k and $k + 1$. These values, unavailable to the MACS, were available to physicians and subjects making decisions about AZT use. Inasmuch as changes in these covariates also served as prognostic indicators, such changes are confounders; failure to account for these changes provides incomplete control of confounding.

We attempted to control for confounding by changes in covariates by adding covariate changes between visits k and $k + 1$ to the previous logistic model. The variables added include: change in CD4 count between visits k and $k + 1$, and the development of AIDS and new symptoms during that interval. The development of new symptoms and new AIDS diagnoses is associated with increased probability of AZT initiation. The estimate of Ψ_0 using the new logistic model is 0.07 (−0.34, 0.57).

AZT use affects subsequent levels of CD4 count^{11–13, 15, 17, 18, 21} and may prevent or delay the subsequent development of the specified symptoms^{11, 12, 15, 18} or AIDS.^{12, 13, 15, 17, 21} Inclusion of these variables in the logistic regression model for treatment initiation amounts to controlling for variables that have been affected by treatment. It is well known that controlling for such variables can lead to bias^{1–4, 35–37, 43, 44, 54–58}.

Fortunately, the direction of the bias is known. A numerical example, in which CD4 count is the only source of confounding and AZT does not affect KS, illustrates the bias.

Table III presents the data from one stratum of this hypothetical study, defined by covariates at the start of follow-up; for the purpose of the example, follow-up continues only until the next follow-up visit, six months later. This stratum includes only subjects who, at their initial visit to the study, never used AZT and have CD4 counts of 300/mm³. During the following three months, CD4 count declines by 25/mm³ in half of the cohort, and remains unchanged in the other half; the group in which CD4 count declines has a worse prognosis (that is, a higher probability of developing KS) than the group with no change. At about 3 months, these individuals visit their physicians, who measure their CD4 counts. The only criterion upon which physicians base their decision whether to prescribe AZT is decline in CD4 count; among people with the same CD4 count at 3 months, groups of subjects starting AZT are comparable with those not so doing. Nonetheless, subjects whose CD4 count declined are 4 times more likely to receive AZT than those whose counts stayed the same. Subsequently, AZT use raises CD4 count by 50/mm³ among

Table III. A hypothetical study of AZT use

CD4 count, visit 1/2	Number of subjects	AZT initiation, visits 1/2	Number of subjects	CD4 count, visit 1	Cases of KS, visit 1	KS risk
275	500	Yes	400	300	80	0.2
		No	100	250	20	0.2
300	500	Yes	100	350	10	0.1
		No	400	300	40	0.1

Stratify on CD4 count, visit 1

CD4 count visit1	AZT use, visit 1	Number of subjects	Cases of KS, visit 1	Ks risk
250	No	100	20	0.2
300	Yes	400	80	0.2
	No	400	40	0.1
350	Yes	100	10	0.1

Do not stratify on CD4

AZT use	Number of subjects	Cases of KS, visit 1	KS risk
No	500	60	0.12
Yes	500	90	0.18

users over the next several weeks. The underlying rate of decline in CD4 continues unchanged over the subsequent 3 months, so that, at 6 months, subjects' CD4 counts are the sum of: (a) their initial counts; (b) AZT effect, and (c) double their decline in CD4 over the first three months. At this point (6 months), subjects return for a follow-up visit to the study site, where investigators assess current CD4 counts and AZT use. At the final visit at 6 months, the study accurately determines who has developed KS. All subjects who develop KS do so after 3 months, no subjects die or are otherwise lost to follow-up, and no subjects alter their use of AZT from that begun at 3 months time. CD4 counts at 3 months are unavailable to study investigators.

Consider stratifying on CD4 count at visit 1. An approach sometimes taken is to include the values of CD4 count at visit 1 in $L_{i,0}$ and to estimate the stratum-specific risk differences $p(T_i \leq 1 | Z_{i,0} = 1, L_{i,0}) - p(T_i \leq 1 | Z_{i,0} = 0, L_{i,0})$. Only the 'stratum' defined by CD4 = 300 at visit 1 has both treated and untreated subjects; comparison is possible for this stratum. For this stratum, the risks in the treated and untreated are $80/400 = 0.2$ and $40/400 = 0.1$, respectively, yielding a risk difference of 0.1. AZT appears harmful, because AZT has made some subjects with poor prognosis have covariate levels indicative of good prognosis. Note that if one fails to control for changes in CD4 count, AZT also appears harmful (Table III); both estimates are then biased in the same direction.⁵⁴

AZT also may affect the levels of other covariates (development of new symptoms or AIDS) so as to make subjects on AZT appear at lower risk for KS than if they had not taken AZT. In consequence, control for changes in observed covariate levels between visits k and $k + 1$ also makes AZT appear more harmful, and estimates of Ψ_0 are biased upward.

Table IV. Estimates of Ψ_0 using different analytic approaches

		Estimates of Ψ_0			
		Modified estimators (using $g_{i,k}^{(2*)}(\Psi)$)		Unmodified estimators (using $g_{i,k}^{(2)}(\Psi)$)	
		Point estimate	95% confidence interval	Point estimate	95% confidence interval
Assumptions about AZT effect on CD4 count (A, B in (11))					
A	B				
0	0	0.06	– 0.42, 0.55		
20	20	0.02	– 0.49, 0.35		
40	30	– 0.01	– 0.52, 0.33		
60	40	– 0.14	– 0.56, 0.31		
Assumptions about time of AZT initiation between visits k and $k + 1$					
In subjects developing KS between k and $k + 1$ (θ in (12))	In subjects not developing KS between k and $k + 1$ (ρ in (13))				
0%	1	– 0.01	– 0.49, 0.32	– 0.28	– 0.59, 0.05
25%	0.5	0.00	– 0.49, 0.33	– 0.08	– 0.42, 0.28
	1	– 0.01	– 0.52, 0.33	– 0.09	– 0.46, 0.35
	2	– 0.01	– 0.54, 0.33	– 0.04	– 0.44, 0.38
50%	1	– 0.07	– 0.54, 0.35	0.11	– 0.33, 0.62
100%	1	– 0.02	– 0.52, 0.37	0.53	0.11, > 1.25
Assumptions about changes in AZT use after last visit: $M1\%$ of subjects treated at last visit and $M0\%$ of subjects untreated at last visit change AZT level					
$M1$	$M0$				
0	0	– 0.01	– 0.52, 0.33		
2.5	5	– 0.03	– 0.54, 0.34		
5	10	– 0.03	– 0.52, 0.38		
How potential censoring time C_i determined?					
As described in					
Section 4		– 0.01	– 0.52, 0.33		
6 March 1992		– 0.01	– 0.53, 0.33		

Following Rosenbaum⁵⁴ and an earlier report from the MACS,⁵⁹ we attempted to adjust for part of this bias. We tried to recover the value of CD4 count that we would have observed at visit $k + 1$ had subjects not started AZT therapy after visit k ; we denote this variable by $CD4_{i,k}(0)$. Were we successful, this variable would, in principle, be unaffected by AZT treatment, and by controlling for it we would control for a pretreatment prognostic indicator. Our model for the effect of AZT initiation on CD4 at visit $k + 1$ is

$$CD4_{i,k}(0; A, B) = CD4_{i,k+1} - \min(A, B \times CD4_{i,k+1}) \tag{11}$$

where A and B are parameters. If AZT has no effect on CD4 count, $A = B = 0$, and $CD4_{i,k}(0; A = 0, B = 0) = CD4_{i,k+1}$, the observed value. Thus, the estimates presented above based on observed changes in CD4 are a special case of inference based on $CD4_{i,k}(0; A, B)$. $A = B = 0$ does not represent reasonably the effect of AZT; we need to use values for A and B more consistent with AZT's known effect. Based on randomized trials,^{11,13,15,17,21} we assumed that $A = 40$, $B = 0.3$ more closely approximates AZT's effect on CD4. We consider the meaning of this model (11) with these parameter values for subjects initiating AZT between visits k and $k + 1$. Under the model, subjects with CD4 at visit $k + 1$ above $133/\text{mm}^3$ would have had CD4 counts $40/\text{mm}^3$ lower had they not started using AZT; other subjects would have had CD4 counts 30 per cent below their observed level ($CD4_{i,k+1}$). Using $CD4_{i,k}(0; 40, 0.3)$ in variables 45–47 in the logistic model produces an estimate for ψ_0 of -0.01 ($-0.52, 0.33$). Using these values for A and B , the estimated value of ψ_0 is lower, as expected from our discussion, and more in line with the results of the randomized trials. Table IV reports estimates under different values of A and B consistent with published results.

11. TIMING OF TREATMENT CHANGES

In the MACS, the dates of change in AZT use were unavailable for large numbers of subjects (information on date of diagnosis of KS is generally available). This section considers inference for ψ_0 when some of the times of treatment changes are unknown but a correct model or prior is available for the distribution of these times. Section 12 considers how to make inference insensitive to this choice of prior. This section concentrates on uncertainty in the time of AZT initiation, because inference is crucially based on correct assumptions (that is, (5)) about this (see Sections 7 and 8). The results presented in Sections 9 and 10 relied on the methods outlined in these sections.

We used simple models for the time of AZT initiation. Consider a subject who reported initiating AZT between visits k and $k + 1$, but did not develop KS during that interval. We assumed that the cumulative distribution function of S_i , the time of AZT initiation, followed the relation

$$F_S(s | T_i > t_{i,k+1}, t_{i,k} < S_i \leq t_{i,k+1}) = \left[\frac{s - t_{i,k}}{t_{i,k+1} - t_{i,k}} \right]^\theta \quad (12)$$

where θ is some positive constant; if $\theta = 1$, this is a uniform distribution on the interval $[t_{i,k}, t_{i,k+1}]$. Now consider a subject who developed KS during the same interval as he initiated AZT therapy. We assumed that the probability he started treatment before he developed KS was ρ . For subjects who started AZT before KS, we assumed the probability of AZT initiation followed a uniform distribution on the interval $[t_{i,k}, T_i]$; for subjects who started AZT after KS, the distribution of AZT initiation time is of no concern. Write the model for AZT initiation time as

$$p_S(s < T_i | t_{i,k} < T_i \leq t_{i,k+1}, t_{i,k} < S_i \leq t_{i,k+1}) = \rho \quad (13)$$

$$F_S(s | t_{i,k} < S_i \leq T_i \leq t_{i,k+1}) = \left[\frac{t - t_{i,k}}{T_i - t_{i,k}} \right].$$

The analyses of the previous sections assumed that $\rho = 0.25$, $\theta = 1$.

We used multiple imputation^{60–62} to estimate ψ_0 under our model for initiation time ((12) and (13)). For each choice of parameters in (12) and (13), we imputed times of AZT initiation $j = 10$ times; except as noted, we imputed values only for subjects who did not report a date for AZT initiation. For the j th imputed data set, $j = 1, \dots, 10$, denote by $S^j(\Psi)$ and $I^j(\Psi)$ the values of the score test statistic for $\zeta = 0$ in (10) computed under the assumption $\psi_0 = \psi$ and the corresponding information statistic. We calculate our overall score test of each hypothesis $\psi_0 = \psi$ as

$$S(\Psi) = \left(\frac{1}{J}\right) \sum_{j=1}^J S_j(\Psi). \tag{14}$$

We calculate the variance of $S(\Psi)$ and the associated p -value as in Little and Rubin.⁶⁰

It is uncertain which, if any, choice of parameters in (12) and (13) provides a reasonable approximation to the unknown distribution of AZT initiation times. Consequently, it is desirable to choose an analytic method under which inference is insensitive to untestable assumptions about the time of treatment initiation. Table IV (column 1) presents estimates obtained under different choices for parameters in (12) and (13). Inference was relatively insensitive to such choices; Section 12 provides reasons for this. The technical details of Section 12 are not necessary to understand the remainder of the article and so may be skipped by those so inclined.

12. CHOICE OF ESTIMATING FUNCTION

Section 8 described how to use arbitrary functions $g(\cdot, \cdot, \cdot)$ of the observables $X_{i,k}^*(\Psi)$, $\Delta_{i,k}(\Psi)$, and $\bar{L}_{i,k}$, in tests of hypotheses $\Psi_0 = \Psi$. This section provides additional details concerning choice of the function $g(\cdot, \cdot, \cdot)$ in (10) and a brief justification for our choice. Estimates of ψ_0 discussed previously used the approach of this section.

We used three criteria to choose a function $g(\cdot, \cdot, \cdot)$: efficiency; feasibility, and insensitivity to assumptions about temporal ordering. Let the superscript (l) denote alternative choices for functions $g(\cdot, \cdot, \cdot)$. Let $G_{i,k}^{(l)}(\Psi) \equiv g_{i,k}^{(l)}(X_{i,k}^*(\Psi), \Delta_{i,k}(\Psi), \bar{L}_{i,k})$ represent a particular arbitrary choice. Let

$$G_{i,k}^{(1)}(\Psi) \equiv \Delta_{i,k}(\Psi) [E\{Z_i(X_{i,k}^*) | X_{i,k}^*(\Psi), \bar{L}_{i,k}, \Delta_{i,k}(\Psi) = 1, \bar{Z}_{i,k-1} = 0, Z_{i,k} = 1\} - E\{Z_i(X_{i,k}^*) | X_{i,k}^*(\Psi), \bar{L}_{i,k}, \Delta_{i,k}(\Psi) = 1, \bar{Z}_{i,k-1} = 0, Z_{i,k} = 0\}]. \tag{15}$$

Note that $G_{i,k}^{(1)}(\Psi)$ is a function of $X_{i,k}^*(\Psi)$, $\Delta_{i,k}(\Psi)$, and $\bar{L}_{i,k}$. Efficiency concerns^{1,4,63} motivate the choice of $G_{i,k}^{(1)}(\Psi)$.

Because these expectations are unknown, we cannot calculate $G_{i,k}^{(1)}(\Psi)$. However, we may estimate $E\{Z_i(X_{i,k}^*) | X_{i,k}^*(\Psi), \bar{L}_{i,k}, \Delta_{i,k}(\Psi) = 1, \bar{Z}_{i,k-1} = 0, Z_{i,k} = j\}$, where $j = 0, 1$, by using linear logistic models for the treatment level at failure ($Z_i\{X_{i,k}^*(\Psi)\}$); these are models for $Z_i\{X_{i,k}^*(\Psi)\}$ in the subset of subjects who are not censored in the Ψ -transformed data (that is, $\Delta_{i,k}(\Psi) = 1$), who have not used AZT before visit k , and who receive level j of AZT at k . Both models (for $j = 0, 1$) included coefficients for the following functions of modified failure time $X_{i,k}^*(\Psi)$ and covariate history $\bar{L}_{i,k}$: MACS visit (ordinal; that is k), and r, r^2, r^3, r^4 , where $r \equiv \ln\{X_{i,k}^*(\Psi) + 1\}$. We substituted fitted values for $E[Z_i(X_{i,k}^*) | X_{i,k}^*(\Psi), \bar{L}_{i,k}, \Delta_{i,k}(\Psi) = 1, \bar{Z}_{i,k-1} = 0, Z_{i,k} = 1], j = 0, 1$, in (15) to give the function $G_{i,k}^{(2)}(\Psi)$.

We consider now how to modify an arbitrary function $G_{i,k}^{(l)}(\Psi)$ to make it insensitive to assumptions about the unknown time of AZT initiation. Let $p_{i,k}$ be the fitted conditional probability in (9) of initiating AZT between visits k and $k + 1$, given covariate history; that is,

$p_{i,k} \equiv 1/(1 + \exp(-(\hat{\alpha}_k + w(\bar{L}_{i,k})\hat{\beta})))$, where hats denote estimates of the parameters in (9). We use the score test of $\zeta = 0$ in our logistic model (10) to test $\Psi_0 = \Psi$ and write the score as

$$\sum_i \sum_k (Z_{i,k} - p_{i,k}) G_{i,k}^{(l)}(\Psi) \quad (16)$$

where the sum is over all subject-visits $\{i, k\}$ included in the logistic model for treatment initiation (10). Suppose that $\Psi_0 = 0$, and consider testing that hypothesis. For subjects who initiate AZT and develop KS between visits k and $k + 1$, but for whom the ordering of these events is uncertain, it is unclear what value to assign to $Z_{i,k}$. In general, the value of the score (16) is sensitive to how $Z_{i,k}$ is imputed. However, it is possible to modify $G_{i,k}^{(l)}(\Psi)$ so that the value of the score is insensitive to the method of imputation. Let $G_{i,k}^{(l^*)}(\Psi) = G_{i,k}^{(l)}(\Psi) I(X_{i,k}(\Psi) \geq \max(1, \Psi))$; recall that one unit in the scale for $X_{i,k}(\Psi)$ is 180 days, roughly the planned gap between visits k and $k + 1$. Now, wherever $Z_{i,k}$ in (16) is uncertain, $G_{i,k}^{(l^*)}(\Psi) = 0$, so that the sum in (16) is little affected by different methods of imputation. Consequently, inference becomes less sensitive to misspecification of models for AZT initiation. The estimates presented in previous sections used the modified function $G_{i,k}^{(2^*)}(\Psi)$ and so were insensitive to specification of models ((12) and (13)) for AZT initiation (Table IV, column 1). In contrast, estimation using the unmodified function $G_{i,k}^{(2)}(\Psi)$, derived as above, is highly sensitive to specification of these models (Table IV, column 2).

13. CENSORING BY COMPETING CAUSES

In the MACS, many subjects die or are lost to follow-up before the end of the study (C_i), and so the variables $X_{i,k}^*(\Psi)$ and $\Delta_{i,k}(\Psi)$ are not always observable. Thus, the methods sketched above are not feasible. We lump together death and loss to follow-up under the term 'competing causes'. This section outlines modifications in the methods of the last section necessary to deal with censoring by competing causes. These modifications were used to obtain the results presented in previous sections.

Let $\tau_i = 1$ if subject i remains under observation until the event of interest occurs (at T_i) or until planned study termination (at C_i); let $\tau_i = 0$ otherwise (i censored by a competing cause). Let D_i represent the date of censoring by a competing cause; for subjects not censored by a competing cause, $D_i = \infty$. Let Y_i be the minimum of failure time, potential censoring time C_i , and the time of censoring by competing causes; that is, $Y_i = \min(T_i, C_i, D_i)$. Let $\lambda_D(t|\cdot)$ represent the hazard of censoring at t given the events denoted by ' \cdot '. Suppose that we identify sufficient predictors of censoring and KS to make the censoring process ignorable, that is, given a subject's treatment and covariate history, time to KS if untreated does not predict the censoring process. We formalize this as

$$\lambda_D(t|\bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t_{i,k}, t_{i,k} \leq t < t_{i,k+1}, U_i) = \lambda_D(t|\bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t_{i,k}, t_{i,k} \leq t < t_{i,k+1}). \quad (17)$$

Let

$$K_i \equiv \exp\left(-\int_0^{Y_i} \lambda_D(t|\bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t, t_{i,k} \leq t < t_{i,k+1}) dt\right).$$

We modify our approach by substituting $G_{m,i}^{(j)*}(\Psi) = \tau_i G_{m,i}^{(j)}(\Psi)/K_i$ for $G_{m,i}^{(j)}(\Psi)$ in score tests of $\zeta = 0$ in the logistic regression model for treatment initiation.⁴

We provide here an informal rationale for this approach. Under conditions stronger than (17), K_i is the conditional probability that subject i survived until Y_i without being censored, given

subject i 's treatment and covariate history. Under those conditions, if subject i is observed to fail at Y_i and $K_i = 0.1$, then subject i would need to count for $1/K_i = 10$ subjects (himself and 9 others who were censored prior to Y_i). Although under the weaker assumption (17), K_i is not the probability of surviving to Y_i without being censored, the approach remains valid.⁶⁴

Because the hazard $\lambda_D(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t, t_{i,k} \leq t < t_{i,k+1})$ is not known, we must estimate it (as $\lambda_D(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, Y_i > t, t_{i,k} \leq t < t_{i,k+1})$). For computational reasons, we chose to use a fully parametric exponential-Poisson regression model for the hazards instead of the Cox proportional hazards model with time-varying covariates suggested in Robins *et al.*⁴ Let $q(\cdot, \cdot)$ be an arbitrary function of its arguments, and let β_{Pois} be a vector parameter. Write the exponential-Poisson model as

$$\lambda_D(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t_{i,k}, t_{i,k} \leq t < t_{i,k+1}; \beta_{\text{Pois}}) = \exp(q(\bar{L}_{i,k}, \bar{Z}_{i,k})\beta_{\text{Pois}}). \tag{18}$$

We use fitted values of the hazard,

$$\hat{\lambda}^D(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, Y_i > t, t_{i,k} \leq t < t_{i,k+1}) = \lambda_D(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, Y_i > t, t_{i,k} \leq t < t_{i,k+1}; \hat{\beta}_{\text{Pois}}),$$

to calculate

$$\hat{K}_i \equiv \exp\left(-\int_0^{\hat{\tau}_D} (t | \hat{L}_{i,k}, \bar{Z}_{i,k}, T_i > t, t_{i,k} \leq t < t_{i,k+1}) dt\right).$$

Robins³ shows that substituting estimates \hat{K}_i for the true value K_i leads to asymptotically conservative confidence intervals (as long as one uses the robust estimator^{42, 65} of the variance of the score statistic (16) from the logistic regression model (10). Robins^{3, 42} discusses analytic corrections that produce confidence intervals which have asymptotically nominal coverage.

14. MODELS FOR THE CENSORING PROCESS

This section presents alternative models used for the censoring process. It also presents estimates of the AZT effect on KS under each alternative.

We chose variables for modelling the censoring process so as to make assumption (17) plausible. This criterion for variable selection mirrors that used in the model for AZT initiation (Sections 8–10). Because causal interpretation of estimates depends on the correctness of (17), making the censoring process ignorable is more important than fitting a parsimonious model.

Table V presents estimates of coefficients in the model for censoring by competing causes. Using this model to calculate \hat{K}_i , the point and interval estimates for Ψ_0 were -0.01 and $(-0.51, 0.33)$. The estimates of Ψ_0 presented in previous sections used this model for censoring.

We have lumped under the title *censoring by competing causes* the separate causes death and loss to follow-up. This aggregation may obscure important differences in the effect of various predictors on each separate type of censoring and so lead to misspecification of the model for censoring. To examine the potential impact of this, we fit separate models for censoring by death and by other loss to follow-up. Denote the cause specific hazard functions as $\lambda_{Dl}(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t, t_{i,k} \leq t < t_{i,k+1})$, where $l = 1$ refers to censoring by death and $l = 2$ refers

Table V. Coefficients estimates in models for censoring by competing causes and associated inference for ψ_0

Variable	Estimates of Ψ_0		
	1. Single model for all types of censoring	Models for censoring 2. Separate models for different types of censoring 2a. Death	2b. Other loss to follow-up
Point estimate	- 0.01		- 0.06
95% confidence interval	- 0.52, 0.33		- 0.52, 0.34
Variable	Estimates of Poisson regression coefficients		
1. Intercept	- 2.97 (0.14)	- 4.62 (0.23)	- 3.55 (0.21)
2. Visit \geq 11	- 0.04 (0.12)	- 0.27 (0.16)	0.30 (0.18)
3. Visit \geq 13	0.47 (0.13)	0.25 (0.19)	0.55 (0.19)
4. Visit \geq 16	- 0.71 (0.21)	- 0.49 (0.26)	- 1.06 (0.35)
5. $\sqrt{\text{CD4 (centred)}/10}$	- 0.074 (0.076)	- 0.37 (0.17)	0.24 (0.13)
6. $V5^2$	0.13 (0.05)	0.23 (0.09)	- 0.25 (0.11)
7. AIDS at visit k	1.37 (0.16)	2.59 (0.26)	- 0.36 (0.42)
8. New AIDS in interval after k		2.96 (0.38)	- 1.45 (1.17)
9. $V8 \times V5$	- 0.01 (0.02)	0.03 (0.02)	0.04 (0.06)
10. $\ln(\text{haematocrit (centred)})$	- 0.32 (0.57)	- 1.96 (0.08)	1.44 (1.00)
11. $V10^2$	8.80 (2.02)	6.35 (2.99)	10.31 (6.30)
12. $V10^3$	8.16 (7.03)	16.85 (10.06)	0.67 (18.20)
13. $V10^4$	- 9.93 (10.58)	6.31 (11.74)	- 54.09 (52.68)
14. $\sqrt{\ln(\text{number male sex partners} + 1)/10}$	- 0.19 (0.07)	- 0.36 (0.10)	0.07 (0.11)
15. Race (other versus white)	0.21 (0.11)	- 0.01 (0.16)	0.51 (0.16)
16. AZT use	- 0.23 (0.10)	- 0.19 (0.14)	- 0.25 (0.16)
17. Oral hairy leukoplakia	0.09 (0.14)	0.05 (0.17)	0.35 (0.26)
18. Fatigue	- 0.02 (0.13)	0.08 (0.18)	- 0.04 (0.22)
19. Night sweats	0.13 (0.20)	0.10 (0.25)	0.05 (0.39)
20. $V3 \times V8$	- 0.66 (0.24)	- 0.41 (0.27)	0.85 (0.83)
21. $V8 \times V16$	- 0.41 (0.20)	- 0.49 (0.23)	- 0.38 (0.78)
22. $V10 \times V18$	- 1.52 (0.57)	- 0.83 (0.68)	- 4.36 (1.73)
23. $V10 \times V19$	0.40 (0.74)	0.28 (0.87)	4.42 (2.87)
24. Employment (full-time versus less)	- 0.25 (0.09)	- 0.23 (0.11)	- 0.25 (0.14)
25. Other symptoms: diarrhoea; fever; herpes zoster; thrush; weight loss (any versus none)	0.29 (0.10)	0.33 (0.13)	0.19 (0.17)

to censoring by other causes. We fit separate exponential-Poisson models (18) for the cause-specific hazard function; Table V presents estimates of coefficients in these models. Redefine K_i as

$$K_i \equiv \exp \left(- \int_0^{Y_i} \{ \lambda_{D1}(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t, t_{i,k} \leq t < t_{i,k+1}) + \lambda_{D2}(t | \bar{L}_{i,k}, \bar{Z}_{i,k}, T_i > t, t_{i,k} \leq t < t_{i,k+1}) \} dt \right)$$

then substitute estimates of the cause-specific hazards to obtain \hat{K}_i which we used in Section 13 to obtain estimates of Ψ_0 . The estimates of Ψ_0 obtained using \hat{K}_i as redefined were $-0.06(-0.52, 0.34)$, not terribly different from those obtained by fitting only a single model for the censoring process.

15. RELATIVE RISK REGRESSION AND TEMPORAL ORDERING

Relative risk regression is the most common approach to estimating treatment effects in observational studies such as the MACS; as such, it may be viewed as a competitor to the G-estimation approach we have outlined. Previous sections showed how our approach could successfully deal with several methodologic issues. This section briefly outlines inadequacies in the way that standard approaches to relative risk regression handle some of these issues.

In observational studies, a typical approach is to specify a model for the hazard as a function of covariates and treatment. Let $\lambda_T(t|\cdot)$ denote the hazard of failure at t given the conditioning. A general model specification is

$$\ln(\lambda_T(t | \bar{L}_i(t), \bar{Z}_i(t))) = g(\bar{L}_i(t), t) + w(\bar{Z}_i(t)) \beta_r. \tag{19}$$

β_r is a log relative-risk parameter that measures the association between treatment and outcome; we would like to know when estimates of β_r have a causal interpretation. The most common choices for $w(\bar{Z}_i(t))$ are: (i) $w(\bar{Z}_i(t)) = Z_i(t)$, that is, using current treatment in the relative-risk regression, and (ii) $w(\bar{Z}_i(t)) = \int_0^t Z_i(u) du$, that is, using cumulative treatment in the relative-risk regression. Checkoway *et al.*⁶⁶ and other texts discuss alternative choices of $w(\cdot)$. We concentrate on measures of current treatment, because this is the most frequent approach to effect estimation in non-randomized (see, for example, references 24, 25 and 67) and randomized (see, for example, reference 21) studies. Robins^{1,36,37} and Robins *et al.*⁴ discuss difficulties in use of cumulative exposure in similar settings.

We examined the effect of AZT under the same assumptions about timing of treatment initiation described in Section 11. Inference was highly sensitive to these assumptions. Table VI presents maximum-likelihood estimates of β_r in an exponential-Poisson regression model for rates. One extreme assumption is that subjects who started AZT between visits k and $k + 1$, did not report a date for AZT initiation, and who developed KS between visits k and $k + 1$ invariably started AZT first before KS. Under this assumption, AZT appears to cause KS. Under the opposite assumption, that such subjects invariably started AZT after developing KS, AZT appears to prevent KS or have little effect. Intermediate assumptions yield intermediate estimates of AZT effect. Under the assumption maintained in previous sections, that 25 per cent of subjects initiating AZT between visits k and $k + 1$ and also developing KS in that interval started AZT first, AZT appears to have little effect; the point estimate (confidence interval) for β_r was $-0.07(-0.39, 0.25)$.

Table VI. Relative risk (exponential-Poisson) regression model for Kaposi's sarcoma: estimates of AZT effect (β_1)

Assumptions about the times of change in AZT treatment	Point estimate	95% confidence interval
Proportion of subjects initiating AZT and developing KS between visits k and $k + 1$ who initiate AZT first (θ in (12)) (among subjects not reporting date of initiation)		
0%	- 0.07	- 0.39, 0.25
25%	0.02	- 0.29, 0.33
50%	0.21	- 0.16, 0.59
100%	0.51	0.17, 0.84
Proportion of subjects initiating AZT and developing KS between visits k and $k + 1$ who initiate AZT first (θ in (12)) (among all subjects)		
25%	- 0.14	- 0.50, 0.23
100%	0.56	0.24, 0.88
Use 'standard' approach: if treatment changes between visits k and $k + 1$, it changes at end of interval; disregard stated AZT initiation dates		
	- 0.11	- 0.41, 0.19

We included terms in the model for: study visit; CD4 count; haematocrit; diagnosis of AIDS; employment status; use of alpha interferon; number of sexual partners; insertive rimming; race; date of birth; study centre, and number of symptoms

An approach often taken in observational studies is to assume the level of treatment recorded at visit k remains the same until visit $k + 1$.^{4, 24, 25} When we applied this approach to the MACS data, the estimate of log relative risk was $- 0.11$ ($- 0.41, 0.19$). Investigators typically justify this approach by arguing (rightly) that developing KS can influence AZT use and its recall; consequently, estimators that use information on treatment gathered at visit $k + 1$ may be severely biased. Nonetheless, use of inaccurate measures of treatment level can produce biased estimates of a treatment's effect.

There are additional difficulties with relative risk estimation. Estimating the effect of treatment received in one short interval on outcome during the same interval requires that treated and untreated subjects be comparable except for treatment received during that interval (4). In general, one would want to consider treatment received during earlier intervals ($\bar{Z}_{i,k-1}$) as a potential confounder needing control (which is seldom done in practice.) This might be accomplished by letting $w(\cdot)$ in (19) be a vector function, for example, $w(\bar{Z}_{i,k}) = [Z_{i,k}, W^+(\bar{Z}_{i,k-1})]^T$, where $W^+(\bar{Z}_{i,k-1})$ is a known function of previous treatment for example, cumulative treatment to visit $k - 1$. Then, even when (4) holds and the relative risk model (19) is correctly specified, the regression coefficient for current treatment, but not that for previous treatment, has causal interpretation.⁵

Suppose further that the more plausible assumption of no unmeasured confounders for treatment initiation (5) rather than for all treatment assignment (4) holds. This restricts us to considering in the estimation only intervals following visits k for which a subject had not previously been treated ($\kappa: \bar{Z}_{i,k} = 0$). Under this restriction, inference becomes even more sensitive to assumptions about the time of AZT initiation (data not shown).

16. DISCUSSION

This analysis used observational data to estimate the effect of AZT on the incidence of KS. Such use of observational data to estimate the effect of a treatment on an outcome it is intended to affect is controversial.^{68, 69} One reason is that physicians and patients often use intermediate predictors of the outcome to make treatment decisions,^{58, 68} creating strong confounding that may be difficult to control. The associations of AZT, KS and other covariates are typical of observational studies of such outcomes. Clinicians use AZT to delay the progression of HIV infection and use intermediate measures of progression to make treatment decisions. KS is a component of such progression and is strongly associated with other measures of disease progression, such as CD4 count.^{32, 63, 70} Investigators often use randomized experiments to estimate intended effects; when conducted well, such trials are justifiably regarded as 'gold standard'. Nonetheless, it is sometimes impractical or unethical to conduct randomized trials, and sometimes clinicians and patients may desire preliminary answers before results from randomized trials become available. Thus, it is important to determine if and when observational analyses can reliably provide useful evaluations of treatment effect. This section discusses what our investigation suggests about these issues.

We have outlined Robins' approach to characterizing and estimating the effect of a time-varying treatment. Further, we have shown how to modify his approach to deal with the methodologic issues relevant to our data. Our analysis has shown that implementation of this approach is practical.

We consider here the role of our assumptions in our data analysis. Perhaps most critical is the role of our treatment assignment assumption (5). Miettinen⁶⁸ has argued that the subtleness of indications for treatment will result in uncontrollable confounding and the failure of such assumptions as (5); consequently, investigators should refrain from observational evaluations of the effect of a treatment on an intended outcome. We seldom expect our assumptions to be exactly correct in this or any observational study. Nonetheless, our experience in analysing the MACS data has made us somewhat more sanguine about the possibility of using observational data for such purposes. We provide here the basis for this conclusion.

There are two possible justifications for analysing observational data when a no-unmeasured-confounders assumption such as (5) is not exactly correct. First, we may believe that the assumption is approximately correct and the resultant bias unimportant.⁷¹ We subject such belief to partial empirical investigation by comparing estimates obtained under assumption (5) to estimates from summarizing the randomized trials. We take the randomized trials results, which suggest that AZT has a weak to moderate effect in preventing KS, to be more accurate; we do this because of the ever-present worry about unmeasured confounders. Nonetheless, there are other reasonable explanations for divergence between estimates from observational and randomized studies. For example, standard intent-to-treat analyses purport to measure effectiveness while our observational analysis aims to estimate efficacy (in which case one might expect to find stronger associations in the observational studies). In addition, case definition in our analysis was different than in many of the randomized trials. Thus, randomized studies are an imperfect 'gold standard' for observational analyses, and in some settings one might expect even valid estimates from observational studies to diverge from those obtained from randomized trials.^{72, 73} Sampling variability may also reasonably explain the remaining differences between the observational and randomized trial estimates.

Despite these caveats, the concern with the possibility of unmeasured confounding makes the comparison of observational and randomized studies instructive. This concern is heightened by the strength of measured confounding in our study; the 'crude' point estimate of Ψ_0 is greater than 1.45, and AZT wrongly appears strongly to cause KS. Applying the approach sketched in Section 10 for controlling confounding yields a point estimate of -0.01 ; the approach appears to have removed most of the confounding. Empirical evaluation of whether (5) has been satisfied and confounding controlled, even approximately, is inconclusive.

Second, suppose that residual confounding is substantial, as is consistent with the data. We reasonably may speculate about its direction. Most determinants of AZT initiation we have found are indicators of poor prognosis (that is, increased risk of KS). We suspect that unmeasured determinants of AZT initiation would exhibit the same property, as apparent increased risk of disease progression seems to be an important basis for doctors and patients to initiate AZT. Inasmuch as this is the case, estimates of Ψ_0 are too high, and AZT may have a stronger effect in preventing KS than these estimates indicate.

Further, one may tailor assumptions about reasons for treatment changes to reflect more accurately the processes we believe operative in the study population; doing this may reduce the bias created by inability to measure accurately the subtle reasons for treatment changes. We have presented two complementary approaches for doing this. Section 10 discusses an approach to controlling for confounding by unmeasured changes in a covariate between visits. We would seldom expect such simple deterministic models for the effect of treatment on a covariate as (11) to be exactly correct. Nonetheless, analytic corrections based on the model made our estimates of treatment effect more consistent with the randomized trials results. It will be of interest to investigate the utility of such simple methods when our model is not exactly correct and to develop alternative approaches when simple methods are not available, such as when the covariate affected by treatment is dichotomous. Robins⁴² has proposed a complementary approach for examining the sensitivity of inference about Ψ_0 to unmeasured confounding. In addition, Section 7 shows how to modify assumptions about treatment assignment and then use G-estimation methods to estimate treatment effects under these more reasonable assumptions.

Sections 11 and 12 discuss an approach to inference when the time of change in treatment level is not known. This approach appears successfully to minimize sensitivity of inference to untestable assumptions about this time. However, more investigation is warranted.

There is another reason for a more optimistic view of the potential utility of observational analyses. Unlike randomized trials designed primarily to examine the effect of a treatment, the MACS was set up to investigate many issues about the natural history of HIV infection.³³ Consequently, investigators did not design their data collection instruments primarily to minimize the methodologic difficulties involved in estimating the effect of AZT. In situations in which the investigators' main goal is to estimate treatment effects, investigators could reduce the potential for bias by (i) accurately recording dates of change in treatment status, and (ii) recording information from all subjects regarding all important determinants of treatment status. In the MACS, treatment was not provided by the investigator, which led to some of the uncertainties discussed above regarding treatment and covariates. Studies in which the providers of medical care (for example, physicians, Health Maintenance Organizations) also record information on treatment, covariates, and outcomes have the potential for greater accuracy regarding these matters. Of course, such designs may heighten the potential for observer bias well-known to epidemiologists, statisticians, and other clinical researchers.^{57, 69}

Observational analyses such as ours are not the equivalent of randomized trials. However, where randomized trials are unfeasible or unethical, analyses of observational data may be necessary. Given appropriate data and proper analytic methods, observational analyses can sometimes provide useful evidence concerning the intended effect of a treatment. We hope that our investigation will spur further investigation of the validity and utility of observational approaches.

ACKNOWLEDGEMENTS

The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute and the Agency for Health Care Policy and Research: U01-A1-35042; 5-MO-RR-00722 (GCRC); U01-A1-35043; U01-A1-37984; U01-A1-35039; U01-A1-35040; U01-A1-37613; U01-A1-35041. Dr. Joffe was also supported in part by a Cancer Epidemiology Training Grant from the National Cancer Institute (CA 09142).

Data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS) with centres (Principal Investigators) at the Johns Hopkins School of Hygiene and Public Health (Alfred Saah, Alvaro Muñoz, Joseph Margolick); Howard Brown Health Center and Northwestern University Medical School (John Phair); University of California, Los Angeles (Roger Detels, Janis V. Giorgi); and University of Pittsburgh (Charles Rinaldo).

The Multicenter AIDS Cohort Study (MACS) includes the following:

Baltimore. The Johns Hopkins University School of Hygiene and Public Health: Alfred J. Saah (Principal Investigator); Haroutune Armenian; Homayoon Farzadegan; Neil Graham; Nancy Kass; Joseph Margolick; Justin McArthur; Ellen Taylor.

Chicago. Howard Brown Health Center and Northwestern University Medical School: John P. Phair (Principal Investigator); Joan S. Chmiel; Bruce Cohen; Maurice O'Gorman; Daina Variakojis; Jerry Wesch; Steven M. Wolinsky.

Los Angeles. University of California, UCLA Schools of Public Health and Medicine: Roger Detels (Principal Investigator); Barbara R. Visscher; Janice P. Dudley; John L. Fahey; Janis V. Giorgi; Andrew Kaplan; Oto Martinez-Maza; Eric N. Miller; Hal Morgenstern; Parunag Nishanian; John Oishi; Jeremy Taylor; Harry Vinters; Jerome Zack.

Pittsburgh. University of Pittsburgh, Graduate School of Public Health: Charles R. Rinaldo (Principal Investigator); Roger Anderson; James T. Becker; Phalguni Gupta; Monto Ho; Lawrence Kingsley; John Mellors; Oliver Ndimbie; Sharon Riddler; Anthony Silvestre; Sharon Zucconi.

Data Coordinating Center. The Johns Hopkins University School of Hygiene and Public Health: Alvaro Muñoz (Principal Investigator); Cheryl Enger; Stephen Gange; Donald R. Hoover; Lisa P. Jacobson; Robert Lyles; Steven Piantadosi; Sol Su.

NIH. National Institute of Allergy and Infectious Diseases: Lewis Schragar, Project Officer; National Cancer Institute: Daniela Seminará.

The authors also thank Sander Greenland for helpful discussions and suggestions and a referee, whose comments led to improvements in the manuscript's content and presentation.

REFERENCES

1. Robins, J. 'Estimation of the time-dependent accelerated failure time model in the presence of confounding factors', *Biometrika*, **79**, 321–334 (1992).

2. Robins, J. M. 'The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies', in Sechrest, L., Freeman, H. and Mulley, A. (eds), *Health Service Research Methodology: A Focus on AIDS, NCHSR, U.S. Public Health Service*, 1989, pp. 113–159.
3. Robins, J. M. 'Analytic methods for estimating HIV-treatment and cofactor effects', in Ostrow, D. G. and Kessler, R. (eds), *Methodological Issues of AIDS Behavioral Research*, Plenum, New York, 1993, pp. 213–287.
4. Robins, J. M., Blevins, D., Ritter, G. and Wulfsohn, M. 'G-estimation of the effect of prophylaxis therapy for pneumocystic carinii pneumonia on the survival of AIDS patients', *Epidemiology*, **3**, 319–366 (1992).
5. Mark, S. D. and Robins, J. M. 'Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model', *Statistics in Medicine*, **12**, 1605–1628 (1993).
6. Robins, J. M. and Greenland, S. 'Adjusting for differential rates of prophylaxis therapy for PCP in high-versus low-dose AZT treatment arms in an AIDS randomized trial', *Journal of the American Statistical Association*, **89**, 737–749 (1994).
7. Jacobson, L. P., Munoz, A., Fox, R., Phair, J. P., Dudley, J., Orams, G. I., Kingsley, L. A., Polk, B. F. and The Multicenter AIDS Cohort Study, 'Incidence of Kaposi's sarcoma in a cohort of homosexual men infected with the human immunodeficiency virus type 1', *Journal of Acquired Immune Deficiency Syndromes*, **3**, (Suppl. 1), S24–S31 (1990).
8. Kaldor, J. M., Tindall, B., Williamson, P., Elford, J. and Cooper, D. A. 'Factors associated with Kaposi's sarcoma in a cohort of homosexual and bisexual men', *Journal of Acquired Immune Deficiency Syndromes*, **6**, 1145–1149 (1993).
9. American Foundation for AIDS Research. *AIDS/HIV Treatment Directory*, Vol. 6, 1993.
10. Hoover, D. R., Black, C., Jacobson, L. P., Martinez-Maza, O., Semnara, D., Saah, A., Von Roenn, J., Anderson, R. and Armenian, H. K. 'Epidemiologic analysis of Kaposi's sarcoma as an early and later AIDS outcome in homosexual men', *American Journal of Epidemiology*, **138**, 266–278 (1993).
11. Concorde Coordinating Committee. 'Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection', *Lancet*, **343**, 871–881 (1994).
12. Cooper, D. A., Gatell, J. M., Kroon, S., Clumeck, N., Millard, J., Goebel, F. -D., Bruun, J. N., Stingl, G., Melville, R. L., Gonzalez-Lahoz, J., Stevens, J. W., Fiddian, A. P. and The European-Australian Collaborative Group, 'Zidovudine in persons with asymptomatic HIV infection and CD4 + cell counts greater than 400 per cubic millimeter', *New England Journal of Medicine*, **329**, 297–303 (1993).
13. Fischl, M. A., Richman, D. D., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., Leedom, J. M., Groopman, J. E., Mildvan, D., Schooley, R. T., Jackson, G. G., Durack, D. T., King, D. and The AZT Collaborative Working Group, 'The efficacy of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex: a double-blind, placebo-controlled trial', *New England Journal of Medicine*, **317**, 185–191 (1987).
14. Fischl, M. A., Parker, C. B., Pettinelli, C., Wulfsohn, M., Hirsch, M. S., Collier, A. C., Antoniskis, D., Ho, M., Richman, D. D., Fuchs, E., Merigan, T. C., Reichman, R. C., Gold, J., Steigbigel, N., Leoung, G. S., Rasheed, S., Tsiatis, A. and The AIDS Clinical Trials Group, 'A randomized controlled trial of a reduced daily dose of zidovudine in patients with the acquired immunodeficiency syndrome', *New England Journal of Medicine*, **323**, 1009–1014 (1990).
15. Fischl, M. A., Richman, D. D., Hansen, N., Collier, A. C., Carey, J. T., Para, M. F., Hardy, W. D., Dolin, R., Powderly, W. G., Allan, J. D., Wong, B., Merigan, T. C., McAuliffe, V. J., Hyslop, N. E., Rhame, F. S., Balfour, H. H., Spector, S. A., Volberding, P., Pettinelli, C., Anderson, J. and the AIDS Clinical Trials Group. 'The safety and efficacy of zidovudine (AZT) in the treatment of subjects with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection: a double-blind, placebo-controlled trial', *Annals of Internal Medicine*, **112**, 727–737 (1990).
16. Fischl, M. A., Richman, D. D., Causey, D. M., Grieco, M. H., Bryson, Y., Mildvan, D., Laskin, O. L., Groopman, J. E., Volberding, P. A., Schooley, R. T., Jackson, G. G., Durack, D. T., Andrews, J. C., Nusinoff-Lehrman, S., Barry, D. W. and The AZT Collaborative Working Group. 'Prolonged zidovudine therapy in patients with AIDS and advanced AIDS-related complex', *Journal of the American Medical Association*, **262**, 2405–2410 (1989).

17. Hamilton, J. D., Hartigan, P. M., Simberkoff, M. S., Day, P. L., Diamond, G. R., Dickinson, G. M., Drusano, G. L., Egorin, M. J., George, W. L., Gordin, F. M., Hawkes, C. A., Jensen, P. C., Klimas, N. G., Labriola, A. M., Lahart, C. J., O'Brien, W. A., Oster, C. N., Weinhold, K. J., Wray, N. P., Zolla-Pazner, S. B. and The Veterans Affairs Cooperative Study Group on AIDS Treatment. 'A controlled trial of early versus late treatment with zidovudine in symptomatic human immunodeficiency virus infection', *New England Journal of Medicine*, **326**, 437–443 (1992).
18. Mulder, J. W., Cooper, D. A., Mathiesen, L., Sandstrom, E., Clumeck, N., Gatell, J. M., French, M., Donovan, B., Gray, F., Yeo, J. M., Lange, J. M. A. and The European-Australian Collaborative Group. 'Zidovudine twice daily in asymptomatic subjects with HIV infection and a high risk of progression to AIDS: a randomized, double-blind placebo-controlled study', *AIDS*, **8**, 313–321 (1994).
19. Nordic Medical Research Councils' HIV Therapy Group. 'Double blind dose-response study of zidovudine in AIDS and advanced HIV infection', *British Medical Journal*, **304**, 13–17 (1992).
20. Richman, D. D., Fischl, M. A., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., Leedom, J. M., Groopman, J. E., Mildvan, D., Hirsch, M. S., Jackson, G. G., Durack, D. T., Nusinoff-Lehrman, S. and The AZT Collaborative Working Group. 'The toxicity of azidothymidine (AZT) in the treatment of patients with AIDS and AIDS-related complex: a double-blind, placebo-controlled trial', *New England Journal of Medicine*, **317**, 192–197 (1987).
21. Volberding, P. A., Lagakos, S. W., Koch, M. A., Pettinelli, C., Myers, M. W., Booth, D. K., Balfour, H. H., Reichman, R. C., Bartlett, J. A., Hirsch, M. S., Murphy, R. L., Hardy, W. D., Soeiro, R., Fischl, M. A., Bartlett, J. G., Merigan, T. C., Hyslop, N. E., Richman, D. D., Valentine, F. T., Corey, L. and The AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. 'Zidovudine in asymptomatic human immunodeficiency virus infection: a controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter', *New England Journal of Medicine*, **322**, 941–949 (1990).
22. Buira, E., Gatell, J. M., Miro, J. M., Batalla, J., Zamora, L., Mallolas, J., Aznar, E., Soriano, E. and San Miguel, J. G. 'Influence of treatment with zidovudine (ZDV) on the long-term survival of AIDS patients', *Journal of Acquired Immune Deficiency Syndromes*, **5**, 737–742 (1992).
23. Creagh-Kirk, T., Doi, P., Andrews, E., Nusinoff-Lehrman, S., Tilson, H., Hoth, D. and Barry, D. W. 'Survival experience among patients with AIDS receiving zidovudine', *Journal of the American Medical Association*, **260**, 3009–3015 (1988).
24. Graham, N. M. H., Zeger, S. L., Park, L. P., Phair, J. P., Detels, R., Vermund, S. H., Ho, M., Saah, A. J. and The Multicenter AIDS Cohort Study. 'Effect of zidovudine and *Pneumocystis carinii* pneumonia prophylaxis on progression of HIV-1 infection to AIDS', *Lancet*, **338**, 265–269 (1991).
25. Graham, N. M. H., Zeger, S. L., Park, L. P., Vermund, S. H., Detels, R., Rinaldo, C. R. and Phair, J. P. 'The effects on survival of early treatment of human immunodeficiency virus infection', *New England Journal of Medicine*, **326**, 1037–1042 (1992).
26. Moore, R. D., Hidalgo, J., Sugland, B. W. and Chaisson, R. E. 'Zidovudine and the natural history of the acquired immunodeficiency syndrome', *New England Journal of Medicine*, **324**, 1412–1416 (1991).
27. Vella, S., Giuliano, M., Pezzotti, P., Agresti, M. G., Tomino, C., Florida, M., Greco, D., Moroni, M., Visco, G., Milazzo, F., Giannelli, F., Angarano, G., Ortona, L., Zanussi, C. and The Zidovudine Evaluation Group. 'Survival of zidovudine-treated patients with AIDS compared with that of contemporary untreated patients', *Journal of the American Medical Association*, **267**, 1232–1236 (1992).
28. Saah, A. J., Hoover, D. R., He, Y., Kingsley, L. A., Phair, J. P. and The Multicenter AIDS Cohort Study. 'Factors influencing survival after AIDS: report from the Multicenter AIDS Cohort Study (MACS)', *Journal of Acquired Immune Deficiency Syndromes*, **7**, 287–295 (1994).
29. Gelber, R. D., Lenderking, W. R., Cotton, D. J., Cole, B. F., Fischl, M. A., Goldhirsch, A., Testa, M. A. and The AIDS Clinical Trials Group. 'Quality-of-life in a clinical trial of zidovudine therapy in patients with mildly symptomatic HIV infection', *Annals of Internal Medicine*, **116**, 961–966 (1992).
30. Lenderking, W. R., Gelber, R. D., Cotton, D. J., Cole, B. F., Goldhirsch, A., Volberding, P. A. and Testa, M. A. 'Evaluation of the quality of life associated with zidovudine treatment in asymptomatic human immunodeficiency virus infection', *New England Journal of Medicine*, **330**, 738–743 (1994).
31. Wu, A. W., Mathews, W. C., Brysk, L. T., Atkinson, J. H., Grant, I., Abramson, I., Kennedy, C. J., McCutchan, J. A., Spector, S. A. and Richman, D. D. 'Quality of life in a placebo-controlled trial of zidovudine in patients with AIDS and AIDS-related complex', *Journal of Acquired Immune Deficiency Syndromes*, **3**, 683–690 (1990).

32. Armenian, H. K., Hoover, D. R., Rubb, S., Metz, S., Kaslow, R., Visscher, B., Chmiel, J., Kingsley, L. and Saah, A. 'Composite risk score for Kaposi's sarcoma based on a case-control and longitudinal study in the Multicenter AIDS Cohort Study (MACS) population', *American Journal of Epidemiology*, **138**, 256–265 (1993).
33. Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., Rinaldo, C. R. and The Multicenter AIDS Cohort Study. 'The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants', *American Journal of Epidemiology*, **126**, 310–318 (1987).
34. Graham, N. M. H., Zeger, S. L., Kuo, V., Jacobson, L. P., Vermund, S. H., Phair, J. P., Detels, R., Rinaldo, C. R., Saah, A. J. and The Multicenter AIDS Cohort Study. 'Zidovudine use in AIDS-free HIV-1-seropositive homosexual men in the Multicenter AIDS Cohort Study (MACS), 1987–1989', *Journal of Acquired Immune Deficiency Syndromes*, **4**, 267–276 (1991).
35. Robins, J. 'A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect', *Mathematical Modelling*, **7**, 1393–1512 (1986).
36. Robins, J. 'A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods', *Journal of Chronic Diseases*, **40**, 139S–161S (1987).
37. Robins, J. 'The control of confounding by intermediate variables', *Statistics in Medicine*, **8**, 679–701 (1989).
38. Greenland, S. 'Modeling and variable selection in epidemiologic analysis', *American Journal of Public Health*, **79**, 340–349 (1989).
39. Cox, D. R. and Oakes, D. *Analysis of Survival Data*, Chapman and Hall, London, 1984.
40. Lipsky, J. J. 'Concorde lands', *Lancet*, **343**, 866–867 (1994).
41. Volberding, P. A., Lagakos, S. W., Grimes, J. M., Stein, D. S., Balfour, H. H., Reichman, R. C., Bartlett, J. A., Hirsch, M. S., Phair, J. P., Mitsuyasu, R. T., Fischl, M. A., Soeiro, R. and the AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. 'The duration of zidovudine benefit in persons with asymptomatic HIV infection: prolonged evaluation of protocol 019 of the AIDS Clinical Trials Group', *Journal of the American Medical Association*, **272**, 437–442 (1994).
42. Robins, J. M. 'Structural nested failure time models', in Keiding, N. and Andersen, P. K. (eds), *The Encyclopedia of Biostatistics, Vol. Survival Analysis*, Wiley, Chichester, 1998.
43. Robins, J. M. 'Errata for "A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect"', *Computers and Mathematics with Applications*, **14**, 917–921 (1987).
44. Robins, J. 'Addendum to "A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect"', *Computers and Mathematics with Applications*, **14**, 923–945 (1987).
45. Rosenbaum, P. R. and Rubin, D. B. 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, **70**, 41–55 (1983).
46. Robins, J. M. and Tsiatis, A. A. 'Semiparametric estimation of an accelerated failure time model with time-dependent covariates', *Biometrika*, **79**, 311–319 (1992).
47. Cox, D. R. 'Regression models and life-tables (with discussion)', *Journal of the Royal Statistical Society, Series B*, **34**, 187–220 (1972).
48. Oakes, D. 'A note on the Kaplan–Meier estimator', *American Statistician*, **47**, 39–40 (1993).
49. Kaplan, E. L. and Meier, P. 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association*, **53**, 457–481 (1958).
50. Berkson, J. and Gage, R. P. 'Survival curve for cancer patients following treatment', *Journal of the American Statistical Association*, **47**, 501–515 (1952).
51. Pearl, J. 'Causal diagrams for empirical research', *Biometrika*, **83**, 669–690 (1995).
52. Arjas, E. and Eerola, M. 'On predictive causality in longitudinal studies', *Journal of Statistical Planning and Inference*, **34**, 361–386 (1993).
53. Rosenbaum, P. R. 'Conditional permutation tests and the propensity score in observational studies', *Journal of the American Statistical Association*, **79**, 565–574 (1984).
54. Rosenbaum, P. R. 'The consequences of adjustment for a concomitant variable that has been affected by the treatment', *Journal of the Royal Statistical Society, Series A*, **147**, 656–666 (1984).

55. Mark, S. D. and Robins, J. M. 'A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial', *Controlled Clinical Trials*, **14**, 79–97 (1993).
56. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
57. Rothman, K. J. *Modern Epidemiology*, Little, Brown and Co., Boston, 1986.
58. Greenland, S. and Neutra, R. 'Control of confounding in the assessment of medical technology', *International Journal of Epidemiology*, **9**, 361–367 (1980).
59. Munoz, A., Schragger, L. K., Bacellar, H., Speizer, I., Vermund, S. H., Detels, R., Saah, A. J., Kingsley, L. A., Seminara, D. and Phair, J. P. 'Trends in the incidence of outcomes defining acquired immunodeficiency syndrome (AIDS) in the Multicenter AIDS Cohort Study: 1985–1991', *American Journal of Epidemiology*, **137**, 423–438 (1993).
60. Little, R. J. A. and Rubin, D. B. *Statistical Analysis With Missing Data*, Wiley, New York, 1987.
61. Rubin, D. B. and Schenker, N. 'Multiple imputation in health-care databases: an overview and some applications', *Statistics in Medicine*, **10**, 585–598 (1991).
62. Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York, 1987.
63. Joffe, M. M. 'Estimation of the effect of a treatment on a failure-time outcome using structural nested failure-time models and G-estimation', Ph.D. dissertation, University of California, Los Angeles, 1994.
64. Robins, J. M. and Rotnitzky, A. 'Recovery of information and adjustment for dependent censoring using surrogate markers', in Jewell, N. P., Dietz, K. and Farewell, V. (eds), *AIDS Epidemiology: Methodological Issues*, Birkhauser, Boston, 1992, pp. 297–331.
65. White, H. 'Maximum likelihood estimation of misspecified models', *Econometrica*, **50**, 1–25 (1982).
66. Checkoway, H., Pearce, N. E. and Crawford-Brown, D. J. *Research Methods in Occupational Epidemiology*, Oxford, New York, 1989.
67. Lavori, P. W., Dawson, R. and Mueller, T. B. 'Causal estimation of time-varying treatment effects in observational studies: application to depressive disorder', *Statistics in Medicine*, **13**, 1089–1100 (1994).
68. Miettinen, O. S. 'The need for randomization in the study of intended effects', *Statistics in Medicine*, **2**, 267–271 (1983).
69. Pocock, S. J. *Clinical Trials: A Practical Approach*, Wiley, Chichester, 1983.
70. Goedert, J. J., Biggar, R. J., Melbye, M., *et al.* 'Effect of T4 count and cofactors on the incidence of AIDS in homosexual men infected with human immunodeficiency virus', *Journal of the American Medical Association*, **257**, 331–334 (1987).
71. Greenland, S. 'Confounding in epidemiologic studies', *Biometrics*, **45**, 1309–1310 (1989).
72. Goetghebeur, E. J. T. and Pocock, S. J. 'Statistical issues in allowing for noncompliance and withdrawal', *Drug Information Journal*, **27**, 837–845 (1993).
73. Newcombe, R. G. 'Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur', *Statistics in Medicine*, **7**, 1179–1186 (1988).