A journey in causal inference: from a complicated algebra to a simple unifying approach

Jonathan Sterne
Department of Population Health Sciences
University of Bristol, UK

It started here…

American Journal of Epidemiology

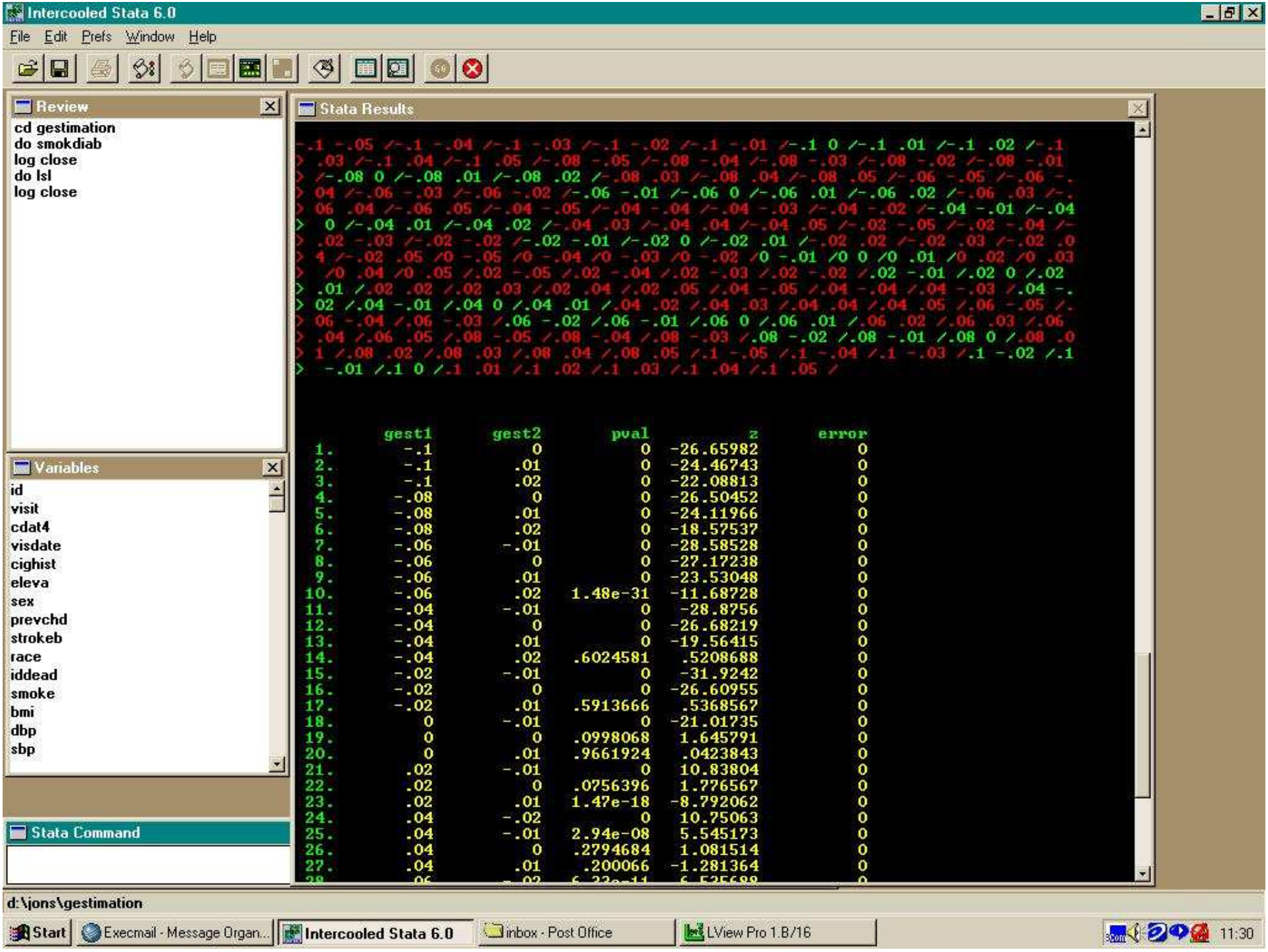
Vol. 148, No. 4
Printed in U.S.A.

G-estimation of Causal Effects: Isolated Systolic Hypertension and Cardiovascular Death in the Framingham Heart Study

Jacqueline C. M. Witteman,[1] Ralph B. D'Agostino,[2] Theo Stijnen,[1] William B. Kannel,[3] Janet C. Cobb,[2]
Maria A. J. de Ridder,[1] Albert Hofman,[1] and James M. Robins[4]

Time-dependent covariates are often both confounders and intermediate variables. In the presence of such covariates, standard approaches for adjustment for confounding are biased. The method of G-estimation allows for appropriate adjustment. Previous studies applying the G-estimation method have addressed effects on all-cause mortality rather than on specific causes of death. In the present study, a method to adjust for censoring by competing risks is presented. The authors used the approach to estimate the causal effect of isolated systolic hypertension on cardiovascular mortality in the Framingham Heart Study, with a 10-year follow-up using data from 1956 to 1970. Arterial rigidity is a major determinant of isolated systolic hypertension and may be a confounder of the relation between isolated systolic hypertension and cardiovascular death. Conversely, isolated systolic hypertension may by itself contribute to stiffening of the vessel wall, and arterial rigidity may therefore also be an intermediate variable in the causal pathway from isolated systolic hypertension to cardiovascular death. While controlling for arterial rigidity and other baseline and time-dependent covariates, isolated systolic hypertension decreased the time to cardiovascular death by 45% (95% confidence interval 3–69). *Am J Epidemiol* 1998;148:390–401.

bias (epidemiology); blood pressure; cardiovascular diseases; epidemiologic methods; follow-up studies; models, statistical; statistics


Intercooled Stata 6.0
File Edit Prefs Window Help
Review
cd gestimation
do smokdiab
log close
do lsl
log close
Variables
id
visit
cdat4
visdate
cighist
eleva
sex
prevchd
strokeb
race
iddead
smoke
bmi
dbp
sbp
Stata Command
Stata Results
-.1 -.05 /-.1 -.04 /-.1 -.03 /-.1 -.02 /-.1 -.01 /-.1 0 /-.1 .01 /-.1 .02 /-.1
> .03 /-.1 .04 /-.1 .05 /-.08 -.05 /-.08 -.04 /-.08 -.03 /-.08 -.02 /-.08 -.01
> /-.08 0 /-.08 .01 /-.08 .02 /-.08 .03 /-.08 .04 /-.08 .05 /-.06 -.05 /-.06 -.
> 04 /-.06 -.03 /-.06 -.02 /-.06 -.01 /-.06 0 /-.06 .01 /-.06 .02 /-.06 .03 /-.
> 06 .04 /-.06 .05 /-.04 -.05 /-.04 -.04 /-.04 -.03 /-.04 -.02 /-.04 -.01 /-.04
> 0 /-.04 .01 /-.04 .02 /-.04 .03 /-.04 .04 /-.04 .05 /-.02 -.05 /-.02 -.04 /-
> .02 -.03 /-.02 -.02 /-.02 -.01 /-.02 0 /-.02 .01 /-.02 .02 /-.02 .03 /-.02 .0
> 4 /-.02 .05 /0 -.05 /0 -.04 /0 -.03 /0 -.02 /0 -.01 /0 0 /0 .01 /0 .02 /0 .03
> /0 .04 /0 .05 /.02 -.05 /.02 -.04 /.02 -.03 /.02 -.02 /.02 -.01 /.02 0 /.02
> .01 /.02 .02 /.02 .03 /.02 .04 /.02 .05 /.04 -.05 /.04 -.04 /.04 -.03 /.04 -.
> 02 /.04 -.01 /.04 0 /.04 .01 /.04 .02 /.04 .03 /.04 .04 /.04 .05 /.06 -.05 /.
> 06 -.04 /.06 -.03 /.06 -.02 /.06 -.01 /.06 0 /.06 .01 /.06 .02 /.06 .03 /.06
> .04 /.06 .05 /.08 -.05 /.08 -.04 /.08 -.03 /.08 -.02 /.08 -.01 /.08 0 /.08 .0
> 1 /.08 .02 /.08 .03 /.08 .04 /.08 .05 /.1 -.05 /.1 -.04 /.1 -.03 /.1 -.02 /.1
> -.01 /.1 0 /.1 .01 /.1 .02 /.1 .03 /.1 .04 /.1 .05 /
gest1 gest2 pval z error
1. -.1 0 0 -26.65982 0
2. -.1 .01 0 -24.46743 0
3. -.1 .02 0 -22.08813 0
4. -.08 0 0 -26.50452 0
5. -.08 .01 0 -24.11966 0
6. -.08 .02 0 -18.57537 0
7. -.06 -.01 0 -28.58528 0
8. -.06 0 0 -27.17238 0
9. -.06 .01 0 -23.53048 0
10. -.06 .02 1.48e-31 -11.68728 0
11. -.04 -.01 0 -28.8756 0
12. -.04 0 0 -26.68219 0
13. -.04 .01 0 -19.56415 0
14. -.04 .02 .6024581 .5208688 0
15. -.02 -.01 0 -31.9242 0
16. -.02 0 0 -26.60955 0
17. -.02 .01 .5913666 .5368567 0
18. 0 -.01 0 -21.01735 0
19. 0 0 .0998068 1.645791 0
20. 0 .01 .9661924 .0423843 0
21. .02 -.01 0 10.83804 0
22. .02 0 .0756396 1.776567 0
23. .02 .01 1.47e-18 -8.792062 0
24. .04 -.02 0 10.75063 0
25. .04 -.01 2.94e-08 5.545173 0
26. .04 0 .2794684 1.081514 0
27. .04 .01 .200066 -1.281364 0
d:\jons\gestimation
Start
Execmail - Message Organ...
Intercooled Stata 6.0
Inbox - Post Office
LView Pro 1.B/16
11:30

Causal models for CVD and risk factors which vary over time

Kate Tilling

Jonathan Sterne

Moyses Szklo

7 December 2000

G-estimation

- Assume that each subject has an underlying survival time - i.e. the time they would have survived had they never been exposed
- conditional on measured history (past and present confounders and past exposure) present exposure is independent of this

  e.g. for 2 individuals with identical histories, the decision to quit smoking does not depend on underlying survival time

G-estimation modeling procedure

- Hypothesize relationship between E and survival

e.g. E multiplies survival by exp(x)

- Estimate underlying survival for all patients
- Model present exposure as function of past history and underlying survival
- choose the x for which exposure is independent of underlying survival

G-estimation modeling procedure

- G-estimated survival ratio

  the ratio of the survival of a person with exposure to that of an identical person with no exposure

- G-estimated hazard ratio

  if survival distribution is Weibull, can convert the g-estimated survival ratio to a hazard ratio

G-estimation of causal effects in longitudinal studies

Jonathan Sterne, George Davey Smith, Yoav Ben-Shlomo

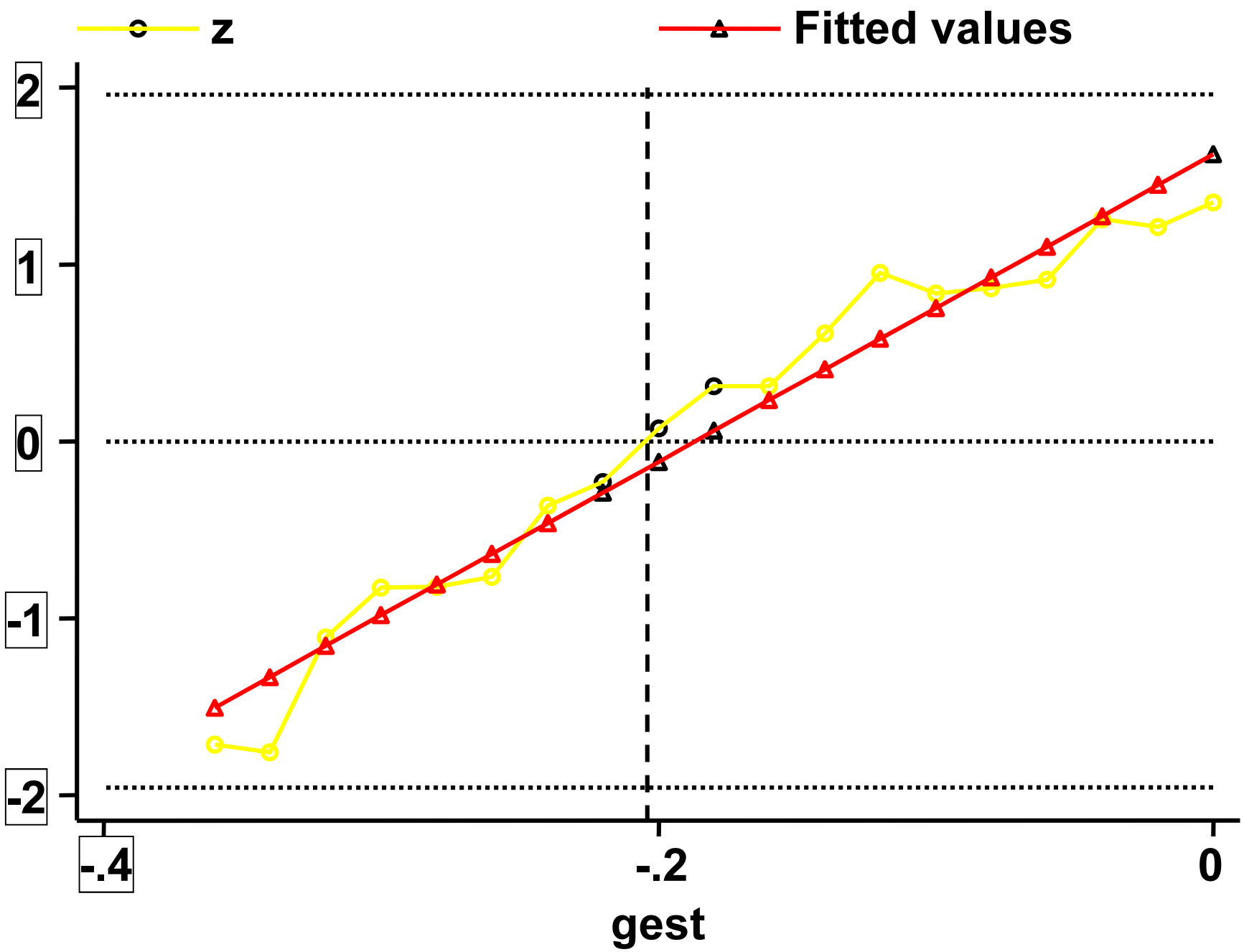
Department of Social Medicine, University of Bristol UK

Kate Tilling

Department of Public Health Sciences, King's College London

13 May 2001

- **`stgest causvar confvars, …`**
- Options
  - Visit(varlist) - indexes examinations
  - Basevis(real) - number of first visit
  - Tcens(varlist) – greatest possible follow up for each person
  - Range(numlist) – specified range for g-estimate
  - Lagconf(varlist) – variables for which lagged effect is to be included
  - Baseconf(varlist) – variables for which baseline effect is to be included
  - Censprob(varlist) – cumulative probability of not being censored, if competing risks are present
  - Idcens(varlist) – indicator variable for censoring
  - Saveres(filename) – save results file
  - Detail – output results of each regression iteration


z
Fitted values
2
1
0
-1
-2
-.4
-.2
0
gest

A covariate is a *time-varying confounder* for the effect of exposure on outcome if:

1. past covariate values predict current exposure
2. past exposure predicts current covariate value
3. current covariate value predicts outcome

Example:

1. obese (*exposed*) people with high blood pressure are advised to lose weight, so are less likely to be obese in future
2. Obesity raises blood pressure
3. High blood pressure is a risk factor for death

Standard survival analyses with time-updated exposure effects will give biased estimates in the presence of time-varying confounding

Results of g-estimation

Exposure	G-estimated ratio (95% CI)
Smoking	0.71 (0.42 to 1.04)
Fibrinogen	0.68 (0.50 to 0.88)
High systolic BP	0.82 (0.66 to 1.04)

21 December 2001

G-estimation of causal effects, allowing for time-varying confounding

Jonathan A C Sterne[1] and Kate Tilling[2]

1. Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR UK. Tel: +44 117 928 7396. Email: jonathan.sterne@bristol.ac.uk.

2. Department of Public Health Sciences, King's College London, Capital House, 42 Weston Street, London SE1 3QD. Tel: +44 20 7848 6629. Email: kate.tilling@kcl.ac.uk.

Abstract

This article describes the **stgest** command, which implements G-estimation (as proposed by Robins) to estimate the effect of a time-varying exposure on survival time, allowing for time-varying confounders.

24 January 2002
Referee comments for *The Stata Journal* on

"G-estimation of causal effects, allowing for time-varying confounding"
by Jonathan Sterne and Kate Tilling

Looking at the 4 criteria set out in the Instructions for Reviewers, I think:

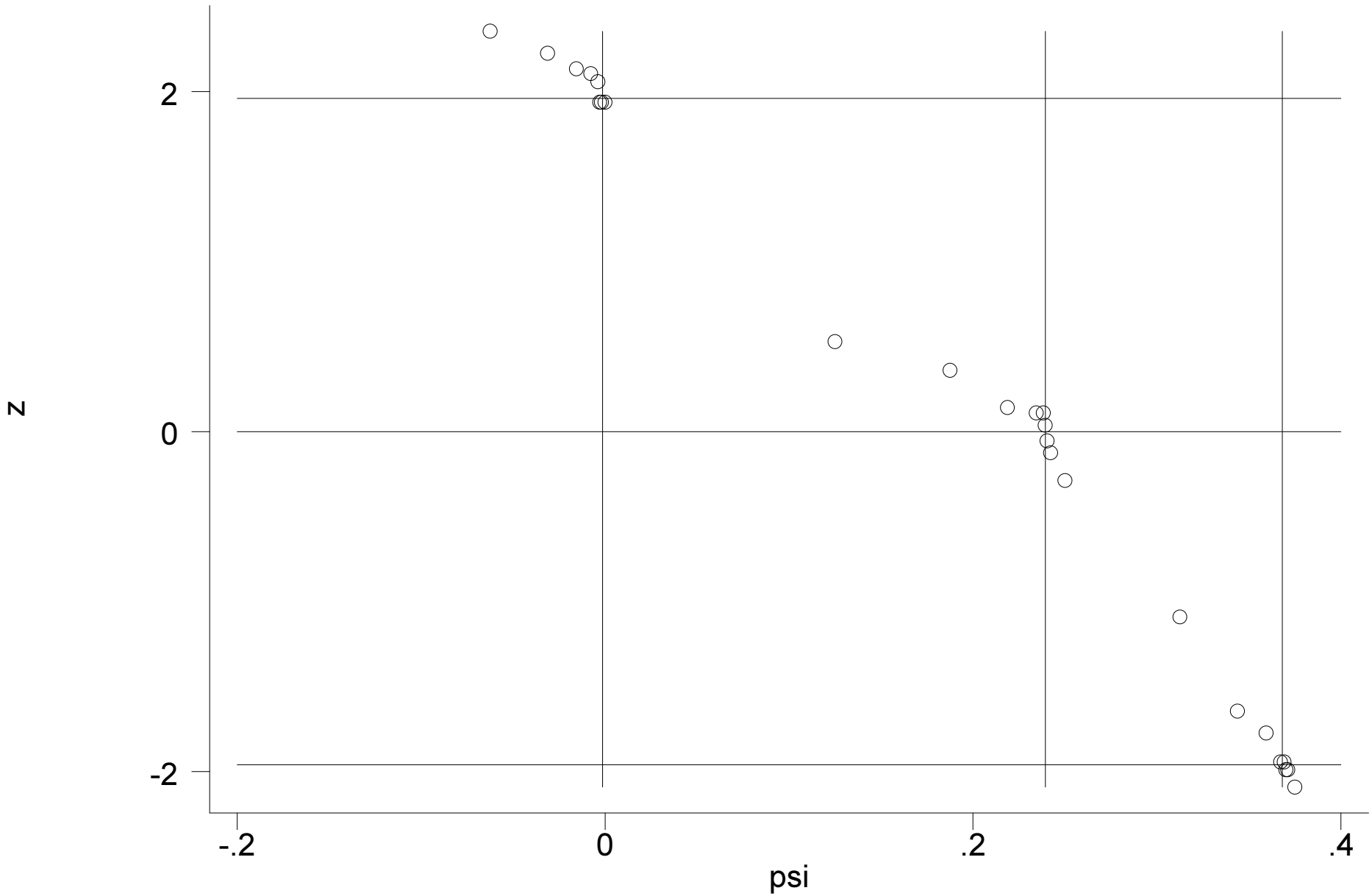
1. the article is definitely of sufficient interest to users to be published;
2. it has enough of a connection to Stata;
3. accuracy from a statistical point of view is not something I am well-placed to judge, but I note that the paper implements a method published in a peer-reviewed journal (but note too point 4)
4. Is the article well-written so that it is easy to understand? Definitely not.

Overall, I like the article, but recommend that it be substantially rewritten in order to (a) clarify the nature of the methods underlying the program, and (b) to widen the appeal beyond a narrow medical statistics audience. The method that they have programmed seems interesting to me, and it would be a shame if the authors' implementation did not receive the circulation that it could.

24 May 2002

APPLICATIONS·OF·G-ESTIMATION·USING·A·NEW·STATA·COMMAND¶
¶
Jonathan·Sterne·(University·of·Bristol·UK)·and·Kate·Tilling·(King's¶
College·London·UK)¶
¶
The·authors·will·present·results·of·causal·modelling·using·"stgest",·a·new·command·for·g-estimation·in·Stata.·Unlike·existing·software,·the·command·can·be·used·as·an·integral·part·of·the·package,·although·running·g-estimation·still·requires·considerable·data·manipulation.¶
Applications·include·risk·factors·for·cardiovascular·disease·using·the·ARIC·and·Caerphilly·cohort·studies,·and·the·effects·of·HAART·in·the·Swiss·HIV·cohort·study.·Future·directions·for·research·will·be·discussed.¶
¶

```
. stgest cursmok Agegrp* fibrin hearta gout highbp diabet chol cholsq
   bpsyst bpdias obese thin,
   visit(visit) firstvis(2)
   lagconf(cursmok fibrin hearta gout highbp diabet chol cholsq bpsyst
   bpdias obese thin)
   baseconf(fibrin hearta gout highbp cursmok chol cholsq diabet bpsyst
   bpdias obese thin)
   lasttime(mienddat) range(-2 2) saveres(caergestsmoknocens) replace

causvar: cursmok
visit: visit
Range: -2 2, rnum: 2
Search method: interval bisection

-2.00 2.00 0.00 1.00 0.50 0.25 0.13 0.19 0.22 0.23 0.24 0.24 0.24 0.24
0.38 0.31 0.34 0.36 0.37 0.37 0.37 0.37 -1.00 -0.50 -0.25 -0.13 -0.06
-0.03 -0.02 -0.01 -0.00 -0.00 -0.00

savres: caergestsmoknocens

G estimate of psi for cursmok: 0.239 (95% CI -0.001 to 0.368)

Causal survival time ratio for cursmok: 0.787 (95% CI 0.692 to 1.001)
```


z
2
0
-2
-.2
0
.2
.4
psi

```
. weibull _t cursmok Agegrp* hearta gout highbp diabet fibrin chol
cholsq bpsyst bpdias obese thin B* L* if visit>=2, dead(_d) t0(_t0) hr

      _t | Haz. Ratio   Std Err      z    P>|z|     [95% Conf. Interval]
 --------+-------------------------------------------------------------
 cursmok |   1.01690   .2083929    0.08   0.935     .6805221    1.519549

(rest of output omitted)

. gesttowb

 g-estimated hazard ratio 1.28 ( 1.00 to  1.47)
```

Future work and (we hope) collaboration

- Implement MSMs in Stata
- Effect of cardiovascular risk factors (e.g. smoking, fibrinogen) and anti-hypertensives in Caerphilly study
- Effect of treatments (e.g. anti-hypertensives, anti-platelet agents) on stroke recurrence using South London Stroke Register

Future work and (we hope) collaboration

- Causal effect of HAART
  - When to start
  - Effect of different drug combinations
  - Will require large collaborations between cohorts
  - Aim to build on an existing collaboration between 13 cohorts involving 12500 patients starting HAART

Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments

Miguel A. Hernán, Babette Brumback, and James M. Robins

Even in the absence of unmeasured confounding factors or model misspecification, standard methods for estimating the causal effect of time-varying treatments on survival are biased when (a) there exists a time-dependent risk factor for survival that also predicts subsequent treatment, and (b) past treatment history predicts subsequent risk factor level. In contrast, methods based on marginal structural models (MSMs) can provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent. MSMs are a new class of causal models whose parameters are estimated using a new class of estimators—inverse-probability-of-treatment weighted estimators. We use a marginal structural Cox proportional hazards model to estimate the joint effect of zidovudine (AZT) and prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of HIV-positive men in the Multicenter AIDS Cohort Study, an observational study of homosexual men. We obtained an estimated causal mortality rate (hazard) ratio of .67 (conservative 95% confidence interval .46–.98) for AZT and of 1.14 (.79, 1.64) for prophylaxis therapy. These estimates will be consistent for the true causal rate ratios when the functional forms chosen for our models are correct and data have been obtained on all time-independent and time-dependent covariates that predict both subsequent treatment and mortality.

KEY WORDS: Causal inference; Confounding; Counterfactual variables; Dependent censoring; Intermediate variables; Semiparametric models. Survival analysis.

1. INTRODUCTION

This article describes the application of marginal structural models (MSMs), a new class of causal models (Robins 1999), to estimate the joint effect of time-dependent nonrandomized treatments, zidovudine (AZT) therapy and prophylaxis therapy, for *Pneumocystis carinii* pneumonia (PCP) on survival among HIV-positive subjects participating in the Multicenter AIDS Cohort Study (MACS), an observational study of homosexual men. The parameters of a MSM can be consistently estimated using a new class of estimators—the inverse-probability-of-treatment weighted estimators. The use of MSMs is an alternative to the semiparametric g-computation algorithm estimator (Robins 1986) and to g-estimation of structural nested models (SNMs) (Robins 1998a).

It is well understood that causal effects can generally be estimated from observational studies only when data on all relevant time-independent and time-dependent confounding factors have been obtained. What is less well known is that standard approaches to confounder control can be biased, even when the causal null hypothesis of no treatment effect is true and there are no unmeasured confounding factors. Specifically, the standard approach to the estimation of the causal effect of a time-varying treatment on survival has been to model the hazard of failure at t as a function of treatment history with a time-dependent proportional hazards model. Robins and colleagues have shown that even in the absence of unmeasured confounding factors or model misspecification, the usual approach may be biased even under the causal null hypothesis, whether or not one adjusts further for the past history of measured covariates in the analysis, when (a) there exists a time-dependent risk factor (say CD4 count and/or PCP history) for survival that also predicts subsequent treatment, and (b) past treatment history predicts subsequent risk factor level (Robins 1986, 1998a; Robins and Greenland 1994). Specifically condition (a) implies that the analysis that does not adjust for covariates is biased because of confounding by CD4 count and/or PCP. Condition (b) implies that the analysis that includes current CD4 count and/or PCP history as a regressor is biased because it adjusts for a variable (CD4 count and/or PCP history) affected by past treatment (see Robins, Greenland, and Hu 1999 for additional details). We show that both conditions (a) and (b) are true in the MACS data. In contrast to standard methods, estimation methods based on MSMs provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent.

2. THE MULTICENTER AIDS COHORT STUDY

The MACS is an ongoing cohort study of more than 5,000 homosexual men from Baltimore, Chicago, Los Angeles, Pittsburgh, and Washington, DC. Study enrollment took place between 1984 and 1991. Follow-up visits are scheduled for every 6 months. During each clinic visit, a structured interview (including questions on demographic variables, therapeutic drugs, and AIDS-related symptoms) is administered, and a physical examination is performed. Blood is collected for a complete blood count, T-cell phenotyping, and assays for HIV-1 antibody. The design and methods of this study have been described previously (Graham et al. 1992). The MACS dataset is available through the National Technical Information Service.

Our analysis concerns two therapies commonly used by HIV-infected patients in the MACS and elsewhere: AZT and prophylaxis for PCP. (In the MACS, aerosolized pentamidine, trimethoprim-sulfamethoxazole, and dapsone were all used as prophylaxis therapy.) AZT temporarily prevents the decline of CD4 lymphocyte count, slows the progression of HIV/AIDS, and, in clinical trials has increased the survival of HIV-infected individuals. PCP is an opportunistic infection that afflicts AIDS patients. Patients may suffer repeated bouts

Miguel Hernán is Instructor, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115 (E-mail: *mhernan@hsph.harvard.edu*). Babette Brumback is Assistant Professor, Department of Biostatistics, University of Washington, Seattle, WA 98195. James Robins is Professor of Epidemiology and Biostatistics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115. This research was supported by NIH grant R01 A-132475.

Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men

Miguel Ángel Hernán,[1] Babette Brumback,[2] and James M. Robins[1,2]

Standard methods for survival analysis, such as the time-dependent Cox model, may produce biased effect estimates when there exist time-dependent confounders that are themselves affected by previous treatment or exposure. Marginal structural models are a new class of causal models the parameters of which are estimated through inverse-probability-of-treatment weighting; these models allow for appropriate adjustment for confounding. We describe the marginal structural Cox proportional hazards model and use it to estimate the causal effect of zidovudine on the survival of human immunodeficiency virus-positive men participating in the Multicenter AIDS Cohort Study. In this study, CD4 lymphocyte count is both a time-dependent confounder of the causal effect of zidovudine on survival and is affected by past zidovudine treatment. The crude mortality rate ratio (95% confidence interval) for zidovudine was 3.6 (3.0–4.3), which reflects the presence of confounding. After controlling for baseline CD4 count and other baseline covariates using standard methods, the mortality rate ratio decreased to 2.3 (1.9–2.8). Using a marginal structural Cox model to control further for time-dependent confounding due to CD4 count and other time-dependent covariates, the mortality rate ratio was 0.7 (95% conservative confidence interval = 0.6–1.0). We compare marginal structural models with previously proposed causal methods. (Epidemiology 2000;11:561–570)

Keywords: counterfactuals, causality, epidemiologic methods, longitudinal data, survival analysis, structural models, confounding, intermediate variables, AIDS.

Marginal structural models (MSMs) can be used to estimate the causal effect of a time-dependent exposure in the presence of time-dependent confounders that are themselves affected by previous treatment.[1,2] The use of MSMs can be an alternative to g-estimation of structural nested models (SNMs).[3]

In our companion paper we describe inverse-probability-of-treatment weighted (IPTW) estimation of a marginal structural logistic model.[4] In this paper, we introduce the marginal structural Cox proportional hazards model, show how to estimate its parameters by inverse-probability-of-treatment weighting, provide practical advice on how to use standard statistical software to obtain the IPTW estimates, and include, as an appendix, the SAS code necessary for the analysis. We use this Cox proportional hazards MSM to estimate the effect of zidovudine on the survival of human immunodeficiency virus (HIV)-positive men enrolled in an observational cohort study, the Multicenter AIDS Cohort Study (MACS). We conclude by comparing methods based on MSMs with previously proposed methods based on g-estimation of SNMs and on the direct estimation of the g-computation algorithm formula.

We now begin by describing the MACS and then summarize why standard methods for survival analysis are not appropriate for estimating the effect of zidovudine on mortality in this cohort.

The Multicenter AIDS Cohort Study and Bias of Standard Methods

Between 1984 and 1991, the MACS enrolled 5,622 homosexual and bisexual men, with no prior acquired immunodeficiency syndrome (AIDS)-defining illness, from the metropolitan areas of Los Angeles, Baltimore-Washington, Pittsburgh, and Chicago. Study participants were asked to return every 6 months to complete a questionnaire, undergo physical examination, and provide blood samples. The design and methods of the MACS have been described in detail elsewhere.[5,6]

We restricted our cohort to HIV-positive men alive in the period during which zidovudine was available for use (that is, after study visit 5; March 1986 through March 1987). Follow-up ended at study visit 21, October 1994, death, or 24 months after the last visit, whichever came

From the Departments of [1]Epidemiology and [2]Biostatistics, Harvard School of Public Health, Boston, MA.

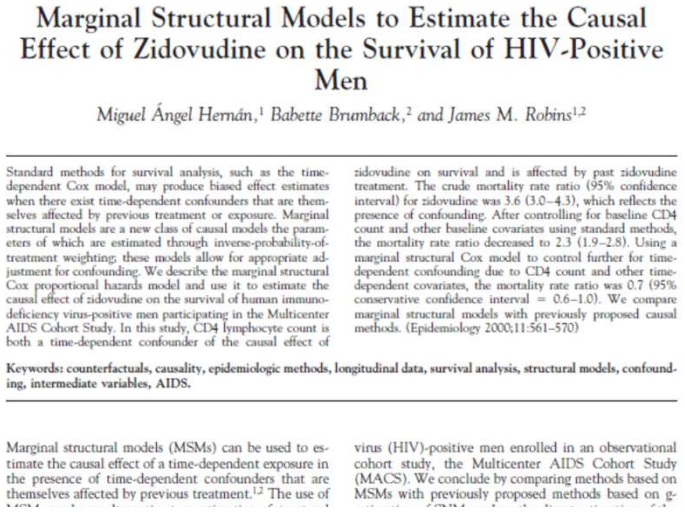
Address correspondence to: Miguel Hernán, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115.

This research was supported by NIH grant R01-A132475.

Submitted March 13, 1999; final version accepted February 28, 2000.

Why use treatment of HIV as an example?

- Extremely strong confounding by indication
  - CD4 count (and other factors) strongly determine start of treatment
  - CD4 count (and the same other factors) are very strongly prognostic
  - HIV cohort studies did an excellent job recording the confounders (the prognostic factors that predicted whether individuals started therapy)



Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men

Miguel Ángel Hernán,[1] Babette Brumback,[2] and James M. Robins[1,2]

Standard methods for survival analysis, such as the time-dependent Cox model, may produce biased effect estimates when there exist time-dependent confounders that are themselves affected by previous treatment or exposure. Marginal structural models are a new class of causal models the parameters of which are estimated through inverse-probability-of-treatment weighting; these models allow for appropriate adjustment for confounding. We describe the marginal structural Cox proportional hazards model and use it to estimate the causal effect of zidovudine on the survival of human immunodeficiency virus-positive men participating in the Multicenter AIDS Cohort Study. In this study, CD4 lymphocyte count is both a time-dependent confounder of the causal effect of zidovudine on survival and is affected by past zidovudine treatment. The crude mortality rate ratio (95% confidence interval) for zidovudine was 3.6 (3.0–4.3), which reflects the presence of confounding. After controlling for baseline CD4 count and other baseline covariates using standard methods, the mortality rate ratio decreased to 2.3 (1.9–2.8). Using a marginal structural Cox model to control further for time-dependent confounding due to CD4 count and other time-dependent covariates, the mortality rate ratio was 0.7 (95% conservative confidence interval = 0.6–1.0). We compare marginal structural models with previously proposed causal methods. (Epidemiology 2000;11:561–570)

Keywords: counterfactuals, causality, epidemiologic methods, longitudinal data, survival analysis, structural models, confounding, intermediate variables, AIDS.

Marginal structural models (MSMs) can be used to estimate the causal effect of a time-dependent exposure in the presence of time-dependent confounders that are themselves affected by previous treatment.[1,2] The use of

virus (HIV)-positive men enrolled in an observational cohort study, the Multicenter AIDS Cohort Study (MACS). We conclude by comparing methods based on MSMs with previously proposed methods based on g-

The crude mortality rate ratio for zidovudine was 3.6 (95% CI 3.0–4.3)… After controlling for baseline CD4 count and other covariates using standard methods, the RR decreased to 2.3 (95% CI 1.9–2.8). Using a marginal structural Cox model, the mortality rate ratio was 0.7 (95% CI 0.6–1.0).

Estimating the causal effect of HAART in the Swiss HIV cohort study

Jonathan Sterne
Department of Social Medicine,
University of Bristol UK

and

CEBU, MCRI,
Royal Children's Hospital Melbourne

Collaborators: Matthias Egger, Miguel Hernán, James Robins, Bruno Ledergerber, Kate Tilling and the Swiss HIV Cohort Study

Marginal structural models for causal inference

- Introduced by Robins *et al.* (1999)
- Stage 1: estimate each subject's probability being treated at each time, using logistic regression
- Stage 2: use these to derive ***inverse probability of treatment weights*** – defined as the inverse of each subject's probability of his or her treatment history at each time

IPT weights

Notation:

$A(k)$ = indicator for treatment at time k

$L(k)$ = value of the vector of risk factors at time k

$\overline{L}(k-1), \overline{A}(k-1)$ = treatment and covariate histories up to time (k-1)

$$iptw_i(t) = \prod_{k=0}^{t} \frac{1}{pr(A(k) = a_i(k) \mid \overline{A}(k-1) = \overline{a}_i(k-1), \overline{L}(k) = \overline{l}_i(k))}$$

Derived by estimating Pr($A(k)$=1) using a pooled logistic regression model (equivalent to a Cox model).

Stabilised weights

Problem: large variation in the iptw weights lead to wide confidence intervals

Solution: stabilised weights

$$sw_i(t) = \prod_{k=0}^{t} \frac{pr(A(k) = a_i(k) \mid \overline{A}(k-1) = \overline{a}_i(k-1), V = v_i)}{pr(A(k) = a_i(k) \mid \overline{A}(k-1) = \overline{a}_i(k-1), \overline{L}(k) = \overline{l}(k))}$$

V = vector of time-independent covariates (included in $L(0)$)

Censoring

Censoring is dealt with in an analogous way:

$$sw_i^*(t) = \prod_{k=0}^{t} \frac{pr(C(k)=0 \mid \overline{C}(k-1)=0, \overline{A}(k-1)=\overline{a}_i(k-1), V=v_i, T>k)}{pr(C(k)=0 \mid \overline{C}(k-1)=0, \overline{A}(k-1)=\overline{a}_i(k-1), \overline{L}(k)=\overline{l}(k), T>k)}$$

Final weight for subject i at time t is:

$$sw_i(t) \times sw_i^*(t)$$

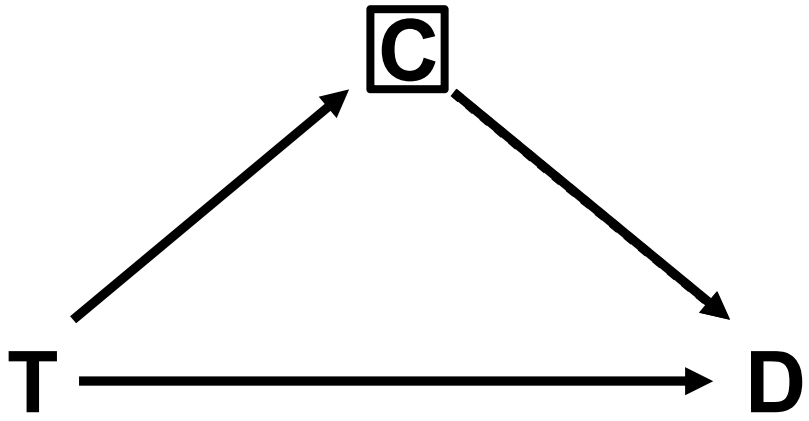
Marginal structural model

- Can be considered to be a *causal* model, in the sense that it compares what happens given your treatment history, to what would have happened in other situations
  - analogous to conducting an RCT each month, among patients still not on HAART
- Assumption: ***no unmeasured confounders***
- **Pooled logistic regression (equivalent to a Cox model), controlling for baseline covariates and baseline hazard, weighted by stabilised weights**

$$\text{logit}\,\Pr[D(t)=1 \mid D(t-1)=0, \overline{A}(t-1), V] = \gamma_0(t) + \gamma_1 A(t-1) + \gamma_2 V$$

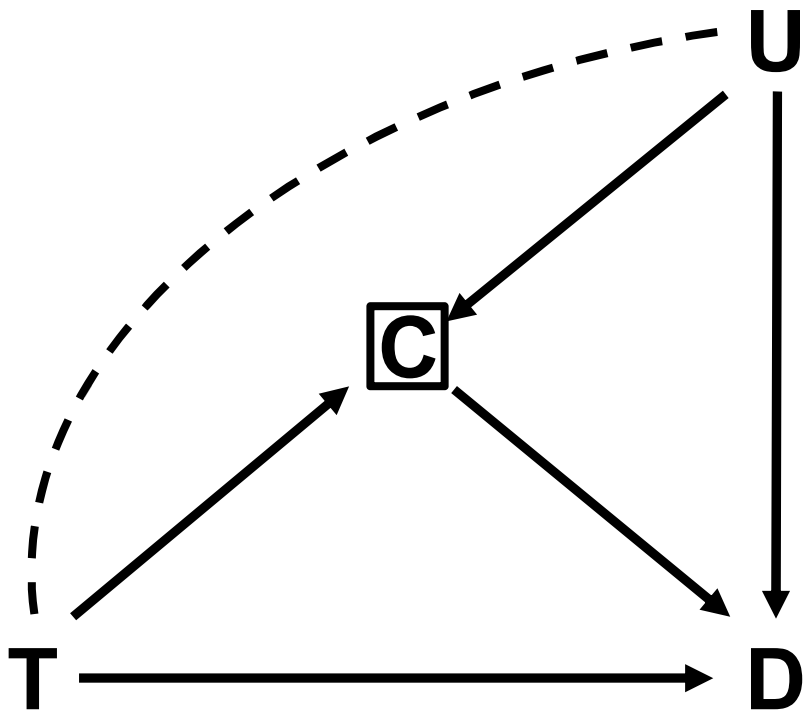
These days I can explain the problem using DAGs

C on the causal pathway



If we control for *C*, we will estimate only the direct effect of *T* on *D*

C on the causal pathway

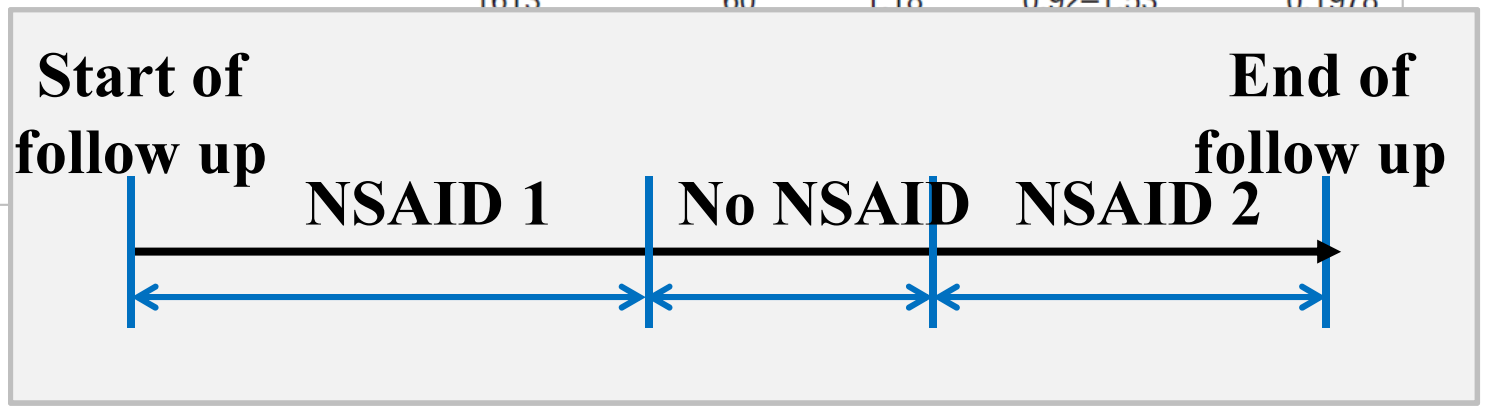


In addition, controlling for *C* may induce confounding

Dealing with time-varying treatments

Cardiovascular Risks of Nonsteroidal Antiinflammatory Drugs in Patients After Hospitalization for Serious

			Reference Nonusers		
	Person-years	Events	IRR	95% CI	*P*
Serious coronary heart disease*					
Nonuser	69 966	2231	1	Reference	
Former	15 604	489	0.95	0.86–1.05	0.3242
Naproxen	1908	49	0.88	0.66–1.17	0.3940
Ibuprofen	1613	60	1.18	0.92–1.53	0.1978
Diclofenac					
Celecoxib					
Rofecoxib					

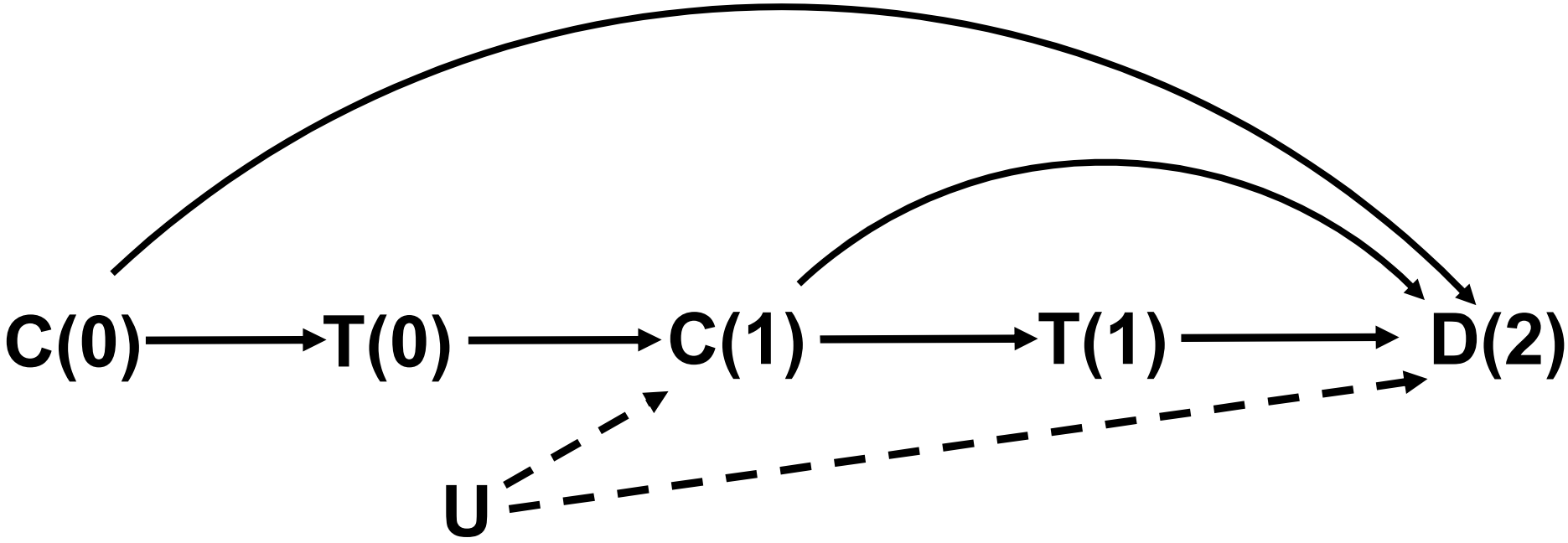


Time-varying confounding
(confounder-treatment feedback)

- Even in the absence of unmeasured confounding factors, standard methods for estimating the causal effect of time-varying treatments on survival are biased when
  - there exists a time-varying risk factor for survival that also predicts subsequent treatment, and
  - past treatment history predicts subsequent risk factor level

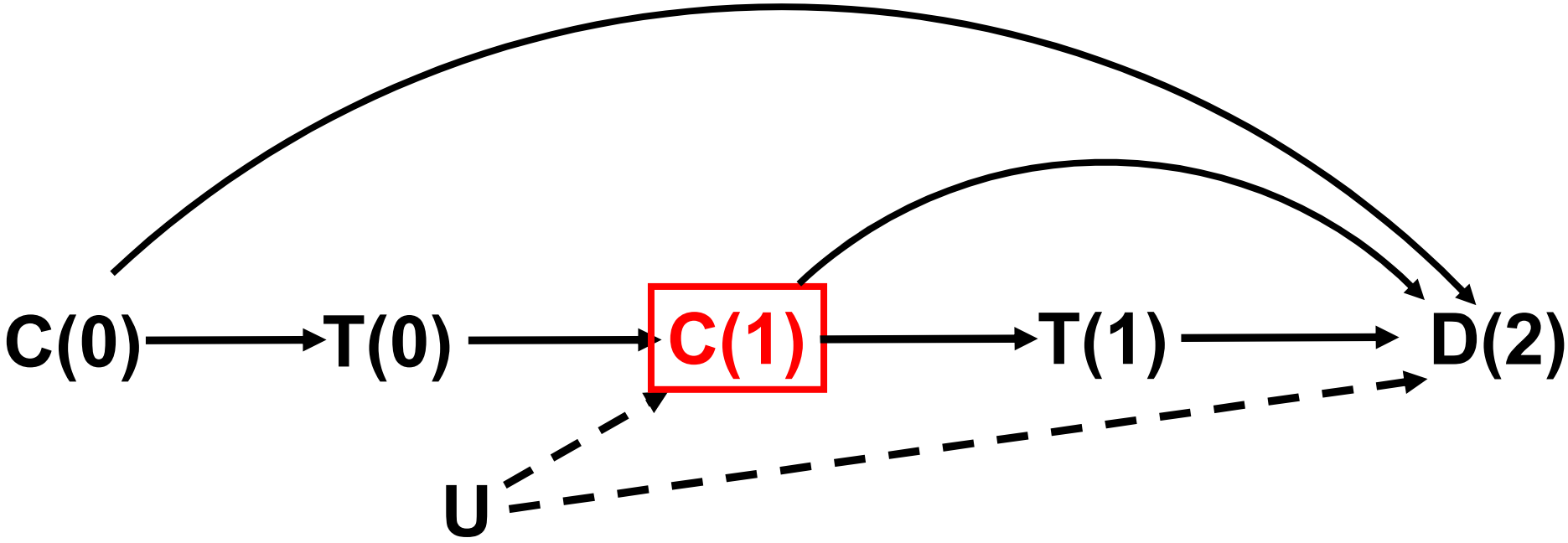
Time-varying confounding
(confounder-treatment feedback)

- Even in the absence of unmeasured confounding factors, standard methods for estimating the causal effect of time-varying treatments on survival are biased when
  - there exists a time-varying risk factor for survival that also predicts subsequent treatment, and
  - past treatment history predicts subsequent risk factor level



Time-varying confounding
(confounder-treatment feedback)

- Even in the absence of unmeasured confounding factors, standard methods for estimating the causal effect of time-varying treatments on survival are biased when
  - there exists a time-varying risk factor for survival that also predicts subsequent treatment, and
  - past treatment history predicts subsequent risk factor level



Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study

*Jonathan A C Sterne, Miguel A Hernán, Bruno Ledergerber, Kate Tilling, Rainer Weber, Pedram Sendi, Martin Rickenbach, James M Robins, Matthias Egger, and the Swiss HIV Cohort Study**

Lancet 2005; 366: 378–84

See **Comment** page 346

*Members listed at end of report

Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK (J A C Sterne PhD, K Tilling PhD, Prof M Egger MD); **Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA** (M A Hernán MD, Prof J M Robins MD); **Division of Infectious Diseases and Hospital Epidemiology, University of Zurich, Zurich, Switzerland** (B Ledergerber PhD, Prof R Weber MD); **Institute for Clinical Epidemiology and Division of Infectious Diseases, Basel University Hospital, Basel, Switzerland** (P Sendi MD); **Data Centre, Swiss HIV Cohort**

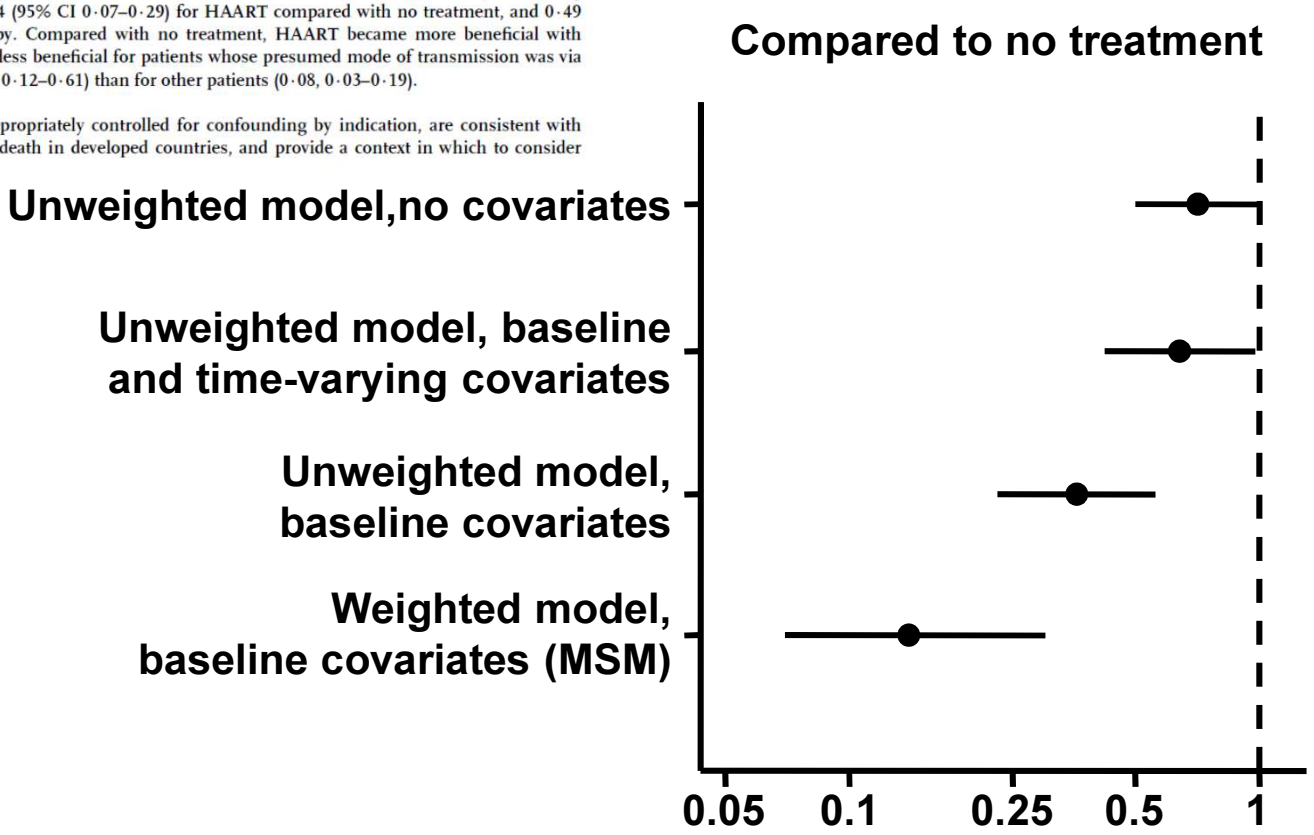
Summary

Background Evidence on the effectiveness of highly active antiretroviral therapy (HAART) for HIV-infected individuals is limited. Most clinical trials examined surrogate endpoints over short periods of follow-up and there has been no placebo-controlled randomised trial of HAART. Estimation of treatment effects in observational studies is problematic, because of confounding by indication. We aimed to use novel methodology to overcome this problem in the Swiss HIV Cohort Study.

Methods Patients were included if they had been examined after January 1996, when HAART became available in Switzerland, were not on HAART, and were free of AIDS at baseline. Cox regression models were weighted to create a statistical population in which the probability of being treated at each time point was unrelated to prognostic factors.

Results Low CD4 counts and increasing HIV-1 viral load were associated with increased probability of starting HAART. Overall hazard ratios were 0·14 (95% CI 0·07–0·29) for HAART compared with no treatment, and 0·49 (0·31–0·79) compared with dual therapy. Compared with no treatment, HAART became more beneficial with increasing time since initiation but was less beneficial for patients whose presumed mode of transmission was via intravenous drug use (hazard ratio 0·27, 0·12–0·61) than for other patients (0·08, 0·03–0·19).

Interpretation Our results, which are appropriately controlled for confounding by indication, are consistent with reported declines in rates of AIDS and death in developed countries, and provide a context in which to consider adverse effects of HAART.



Tutorial in Biostatistics

Statistics in Medicine

Received 22 July 2010, Accepted 30 October 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5686

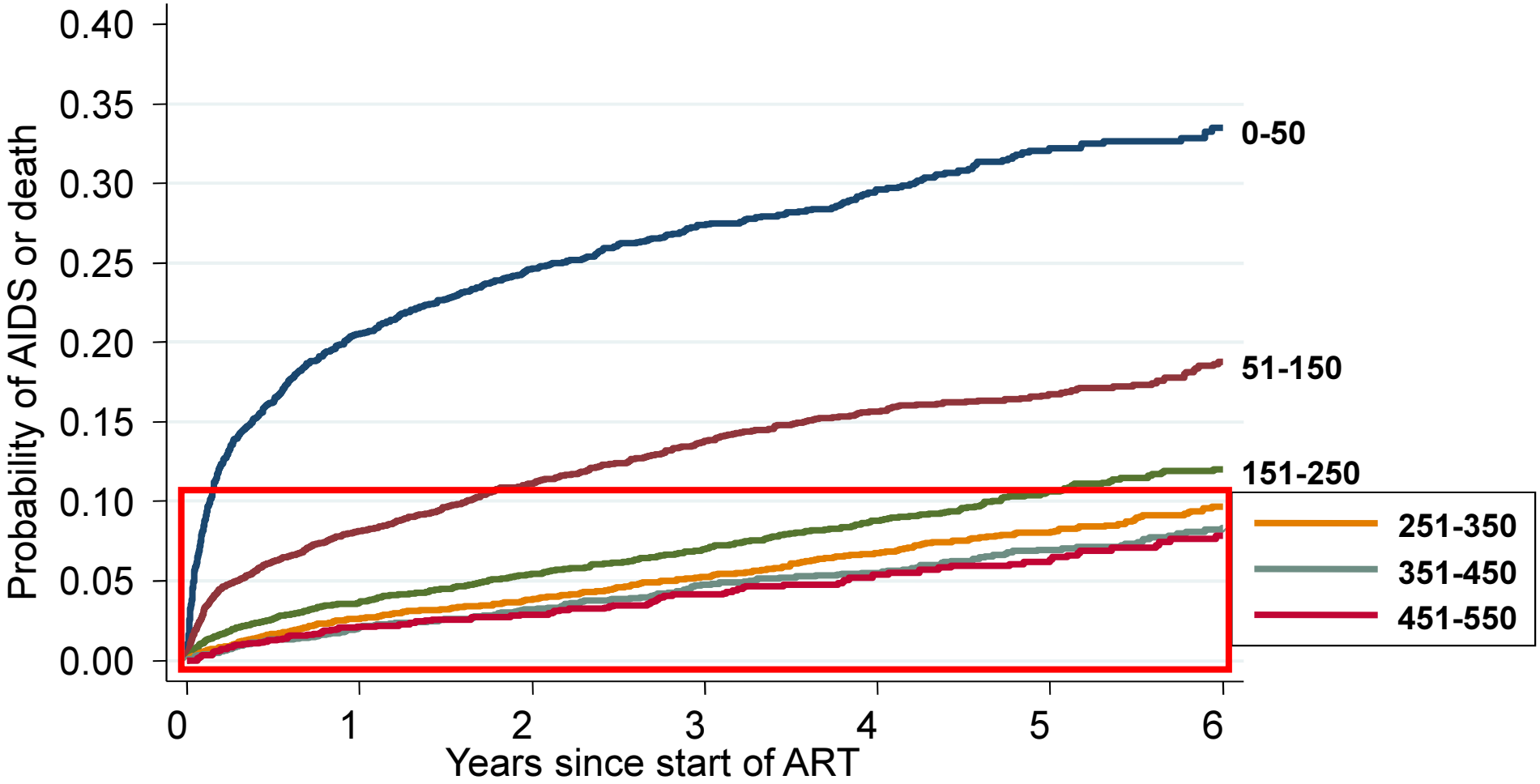
Methods for dealing with time-dependent confounding

R. M. Daniel,[a*†] S. N. Cousens,[a] B. L. De Stavola,[a] M. G. Kenward[a] and J. A. C. Sterne[b]

Longitudinal studies, where data are repeatedly collected on subjects over a period, are common in medical research. When estimating the effect of a time-varying treatment or exposure on an outcome of interest measured at a later time, standard methods fail to give consistent estimators in the presence of time-varying confounders if those confounders are themselves affected by the treatment. Robins and colleagues have proposed several alternative methods that, provided certain assumptions hold, avoid the problems associated with standard approaches. They include the g-computation formula, inverse probability weighted estimation of marginal structural models and g-estimation of structural nested models. In this tutorial, we give a description of each of these methods, exploring the links and differences between them and the reasons for choosing one over the others in different settings.

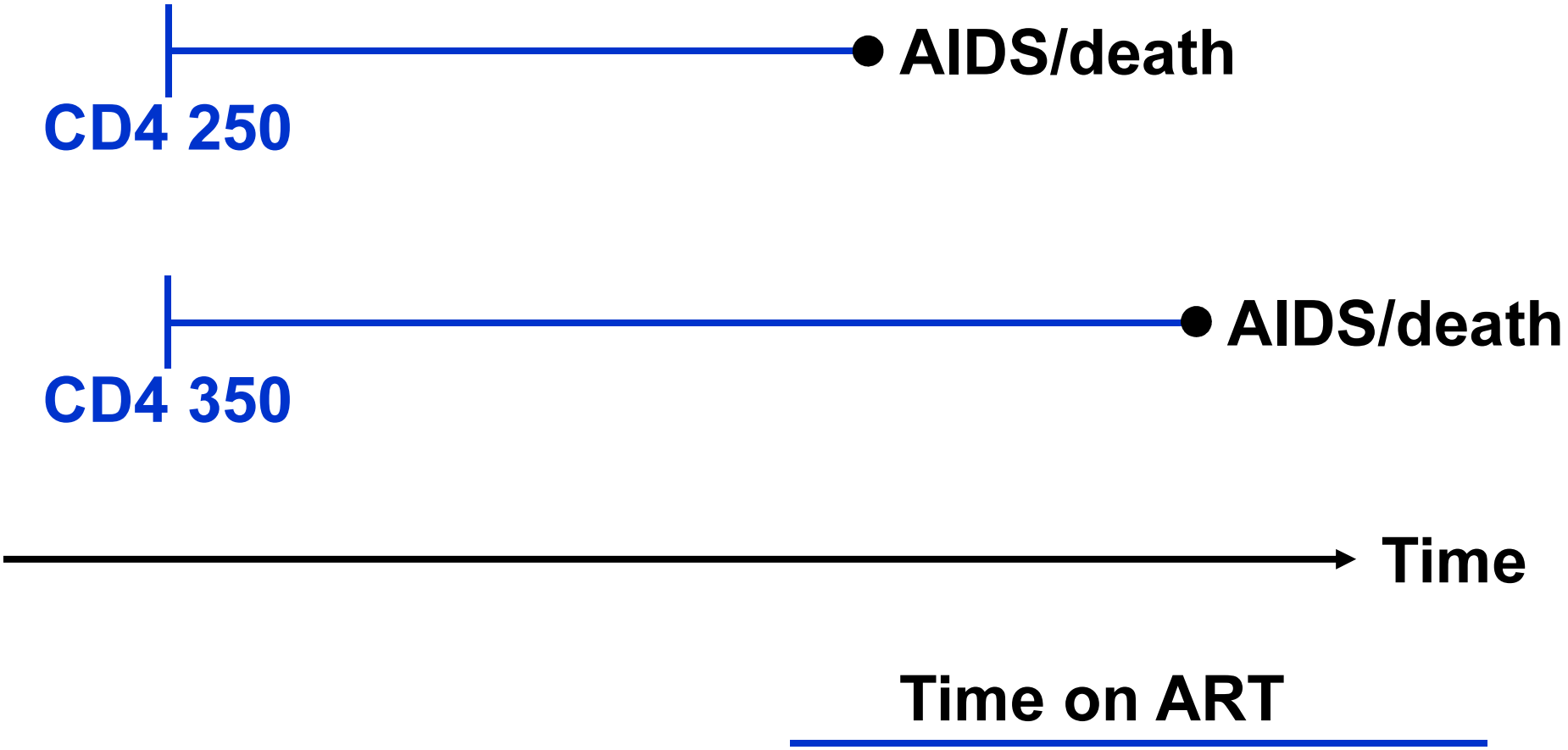
Keywords: time-dependent confounding; g-computation formula; inverse probability weighting; g-estimation; marginal structural model; structural nested model

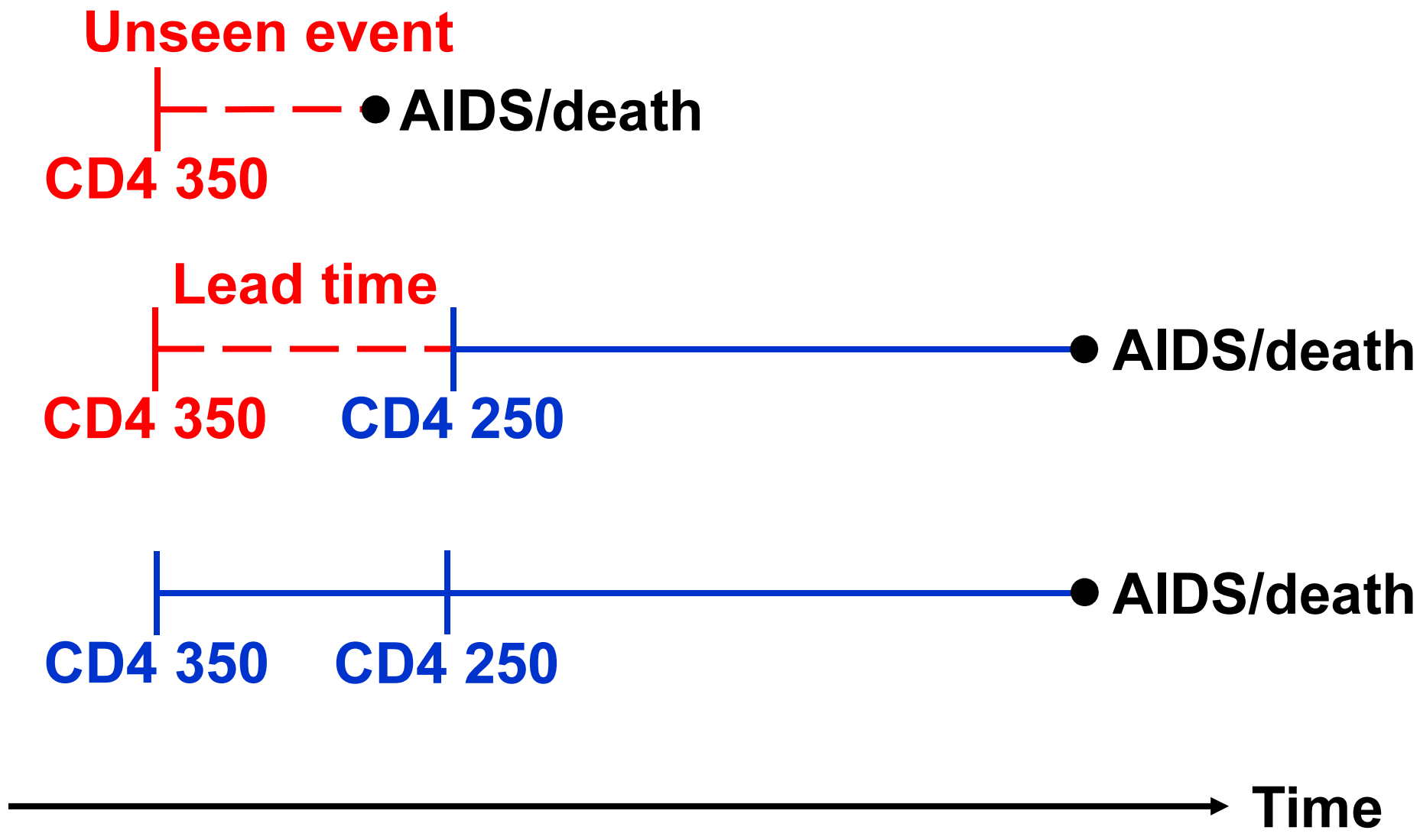
Probability of AIDS or death in ART-naïve AIDS-free non-IDU patients starting cART after 1998



Based on 24,444 patients from 15 cohort studies, 2,366 events in 81,071 person-years of follow up

ART
Cohort Collaboration


AIDS/death
CD4 250
AIDS/death
CD4 350
Time
Time on ART


Unseen event
AIDS/death
CD4 350
Lead time
AIDS/death
CD4 350
CD4 250
AIDS/death
CD4 350
CD4 250
Time
Time not on ART
Time on ART

Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: a collaborative analysis of 18 HIV cohort studies

*When To Start Consortium**

Lancet 2009; 373: 1352–63
Published Online
April 9, 2009
DOI:10.1016/S0140-6736(09)60612-7

See Comment page 1314

*Members listed at end of paper and contributors to each cohort are listed in the webappendix (pp 1–8)

Correspondence to:
Prof Jonathan Sterne,
Department of Social Medicine,
University of Bristol, Canynge Hall, Whatley Road,
Bristol BS8 2PS, UK
jonathan.sterne@bristol.ac.uk

Summary

Background The CD4 cell count at which combination antiretroviral therapy should be started is a central, unresolved issue in the care of HIV-1-infected patients. In the absence of randomised trials, we examined this question in prospective cohort studies.

Methods We analysed data from 18 cohort studies of patients with HIV. Antiretroviral-naive patients from 15 of these studies were eligible for inclusion if they had started combination antiretroviral therapy (while AIDS-free, with a CD4 cell count less than 550 cells per μL, and with no history of injecting drug use) on or after Jan 1, 1998. We used data from patients followed up in seven of the cohorts in the era before the introduction of combination therapy (1989–95) to estimate distributions of lead times (from the first CD4 cell count measurement in an upper range to the upper threshold of a lower range) and unseen AIDS and death events (occurring before the upper threshold of a lower CD4 cell count range is reached) in the absence of treatment. These estimations were used to impute completed datasets in which lead times and unseen AIDS and death events were added to data for treated patients in deferred therapy groups. We compared the effect of deferred initiation of combination therapy with immediate initiation on rates of AIDS and death, and on death alone, in adjacent CD4 cell count ranges of width 100 cells per μL.

Findings Data were obtained for 21 247 patients who were followed up during the era before the introduction of combination therapy and 24 444 patients who were followed up from the start of treatment. Deferring combination therapy until a CD4 cell count of 251–350 cells per μL was associated with higher rates of AIDS and death than starting therapy in the range 351–450 cells per μL (hazard ratio [HR] 1·28, 95% CI 1·04–1·57). The adverse effect of deferring treatment increased with decreasing CD4 cell count threshold. Deferred initiation of combination therapy was also associated with higher mortality rates, although effects on mortality were less marked than effects on AIDS and death (HR 1·13, 0·80–1·60, for deferred initiation of treatment at CD4 cell count 251–350 cells per μL compared with initiation at 351–450 cells per μL).

Interpretation Our results suggest that 350 cells per μL should be the minimum threshold for initiation of antiretroviral therapy, and should help to guide physicians and patients in deciding when to start treatment.

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 APRIL 30, 2009 VOL. 360 NO. 18

Effect of Early versus Deferred Antiretroviral Therapy for HIV on Survival

Mari M. Kitahata, M.D., M.P.H., Stephen J. Gange, Ph.D., Alison G. Abraham, Ph.D., Barry Merriman, M.A., Michael S. Saag, M.D., Amy C. Justice, M.D., Ph.D., Robert S. Hogg, Ph.D., Steven G. Deeks, M.D., Joseph J. Eron, M.D., John T. Brooks, M.D., Sean B. Rourke, Ph.D., M. John Gill, M.B., Ch.B., Ronald J. Bosch, Ph.D., Jeffrey N. Martin, M.D., M.P.H., Marina B. Klein, M.D., Lisa P. Jacobson, Sc.D., Benigno Rodriguez, M.D., Timothy R. Sterling, M.D., Gregory D. Kirk, M.D., Ph.D., Sonia Napravnik, Ph.D., Anita R. Rachlis, M.D., Liviana M. Calzavara, Ph.D., Michael A. Horberg, M.D., Michael J. Silverberg, Ph.D., Kelly A. Gebo, M.D., M.P.H., James J. Goedert, M.D., Constance A. Benson, M.D., Ann C. Collier, M.D., Stephen E. Van Rompaey, Ph.D., Heidi M. Crane, M.D., M.P.H., Rosemary G. McKaig, Ph.D., Bryan Lau, Ph.D., Aimee M. Freeman, M.A., and Richard D. Moore, M.D., for the NA-ACCORD Investigators*

ABSTRACT

BACKGROUND

The optimal time for the initiation of antiretroviral therapy for asymptomatic patients with human immunodeficiency virus (HIV) infection is uncertain.

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Kitahata at the University of Washington,

Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents

December 1, 2009

Developed by the DHHS Panel on
Antiretroviral Guidelines for Adults
and Adolescents – A Working Group of the
Office of AIDS Research Advisory Council (OARAC)

How to Cite the Adult and Adolescent Guidelines:
Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. December 1, 2009; 1-161. Available at http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Accessed (insert date) [insert page number, table number, etc. if applicable]

It is emphasized that concepts relevant to HIV management evolve rapidly. The Panel has a mechanism to update recommendations on a regular basis, and the most recent information is available on the **AIDS*info* Web site** (http://aidsinfo.nih.gov).

From complicated algebra to a simple approach

- What is the randomized trial whose effect we wish to mimic using observational data?

American Journal of Epidemiology

Vol. 183, No. 8
DOI: 10.1093/aje/kwv254
Advance Access publication:

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Bosto
(e-mail: miguel_hernan@post.harvard.edu).

Initially submitted December 9, 2014; accepted for publication September 8, 2015.

Ideally, questions about comparative effectiveness or safety would be answered usin
and conducted randomized experiment. When we cannot conduct a randomized exper
tional data. Causal inference from large observational databases (big data) can be view
a randomized experiment—the target experiment or target trial—that would answer the
the goal is to guide decisions among several strategies, causal analyses of observationa
with respect to how well they emulate a particular target trial. We outline a framework fo
research using big data that makes the target trial explicit. This framework channels co
paring the effects of sustained treatment strategies, organizes analytic approaches, pr
for the criticism of observational studies, and helps avoid common methodologic pitfa

big data; causal inference; comparative effectiveness research; target trial

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 79 (2016) 70–75

Journal of Clinical Epidemiology

Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses

Miguel A. Hernán[a,b,c,*], Brian C. Sauer[d], Sonia Hernández-Díaz[a], Robert Platt[e,f,g], Ian Shrier[g]

[a]*Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA*
[b]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, MA 02215, USA*
[c]*Harvard-MIT Division of Health Sciences and Technology, 77 Massachusetts Avenue, E25-518, Cambridge, MA 02139, USA*
[d]*Salt Lake City Veterans Affairs Medical Center, Division of Epidemiology, Department of Internal medicine, University of Utah, 500 S Foothill Blvd, Salt Lake City, UT 84149, USA*
[e]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada*
[f]*Research Institute of the McGill University Health Centre, 1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada*
[g]*Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, 3755 Côte Ste-Catherine Road, Montreal, Quebec H3T 1E2, Canada*

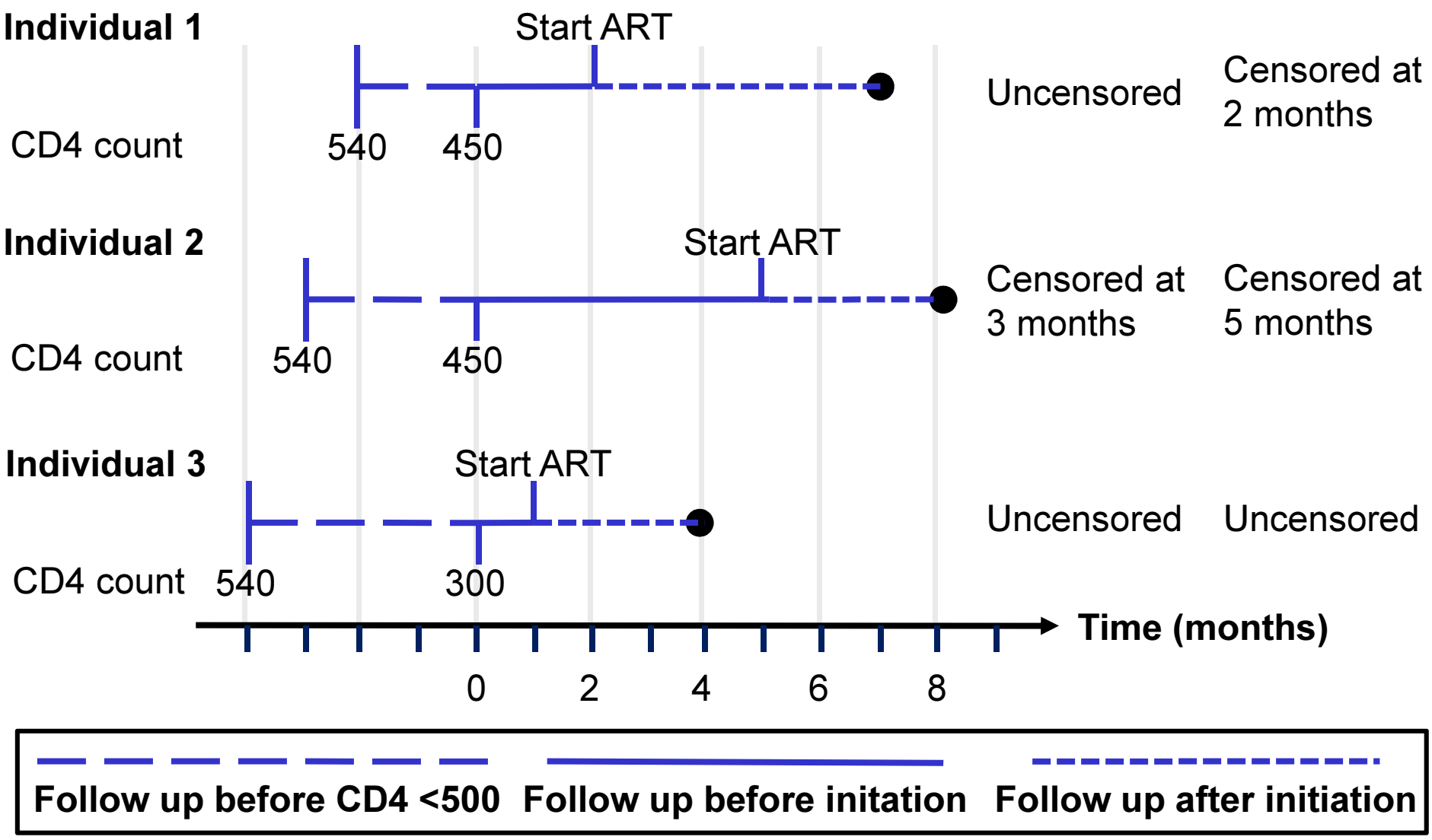
Accepted 23 April 2016; Published online 27 May 2016

Abstract

Many analyses of observational data are attempts to emulate a target trial. The emulation of the target trial may fail when researchers deviate from simple principles that guide the design and analysis of randomized experiments. We review a framework to describe and prevent biases, including immortal time bias, that result from a failure to align start of follow-up, specification of eligibility, and treatment assignment. We review some analytic approaches to avoid these problems in comparative effectiveness or safety research.

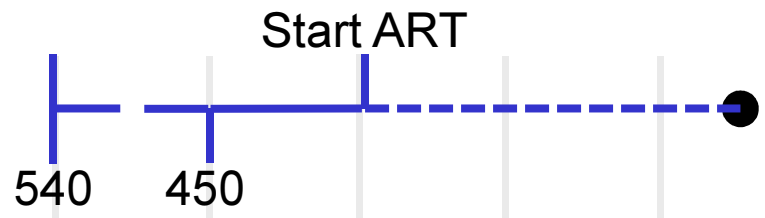
Keywords: Observational; studies; Comparative effectiveness research; Target trial; Time zero; Immortal time bias; Selection bias

Follow-up for three hypothetical individuals



Follow-up for three hypothetical individuals

	Regimen 1 (initiate within 3 months of CD4 <500)	**Regimen 2** (initiate within 3 months of CD4 <350)
Individual 1	Uncensored	Censored at 2 months



Indiv

CD4

Indiv

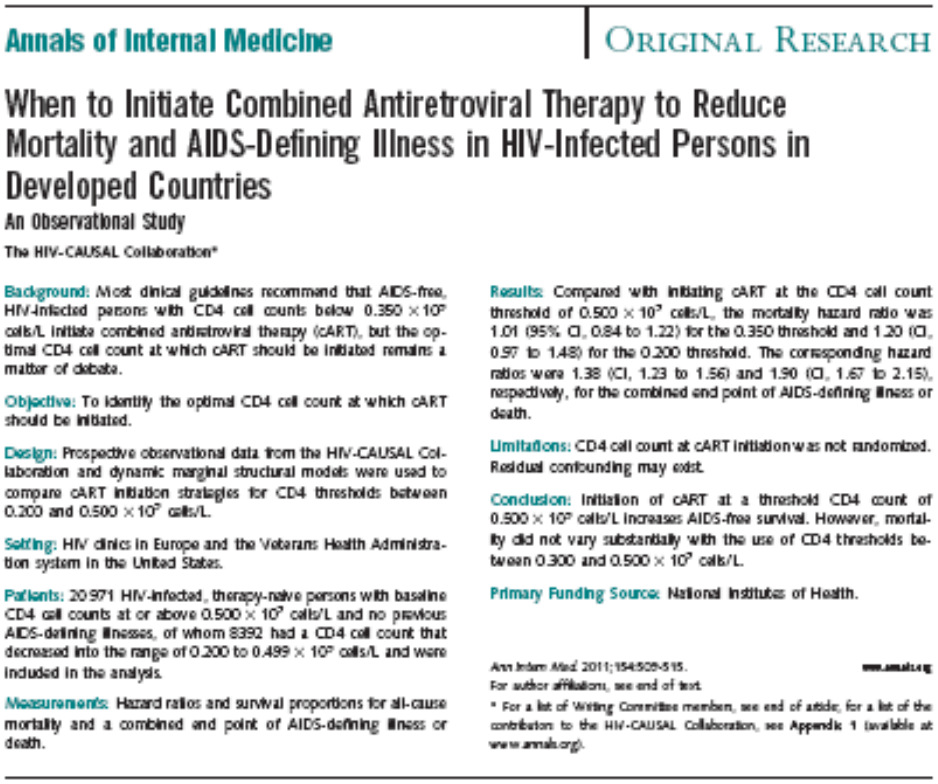
CD4

ed at
ns
ored

To avoid bias, we need to:

- **start follow up for each individual at the time they are eligible for a regimen (as in an RCT)**
- **include individuals' follow up in all the regimen groups with which their follow up is consistent**
- **use inverse probability weighting to correct for the artificial censoring**
- **include all events (regardless of whether they had started ART at the time of the event)**

Follow up before CD4 <500 **Follow up before initation** **Follow up after initiation**



Annals of Internal Medicine | ORIGINAL RESEARCH

When to Initiate Combined Antiretroviral Therapy to Reduce Mortality and AIDS-Defining Illness in HIV-Infected Persons in Developed Countries

An Observational Study

The HIV-CAUSAL Collaboration*

Background: Most clinical guidelines recommend that AIDS-free, HIV-infected persons with CD4 cell counts below 0.350×10^9 cells/L initiate combined antiretroviral therapy (cART), but the optimal CD4 cell count at which cART should be initiated remains a matter of debate.

Objective: To identify the optimal CD4 cell count at which cART should be initiated.

Design: Prospective observational data from the HIV-CAUSAL Collaboration and dynamic marginal structural models were used to compare cART initiation strategies for CD4 thresholds between 0.200 and 0.500×10^9 cells/L.

Setting: HIV clinics in Europe and the Veterans Health Administration system in the United States.

Patients: 20 971 HIV-infected, therapy-naive persons with baseline CD4 cell counts at or above 0.500×10^9 cells/L and no previous AIDS-defining illnesses, of whom 8392 had a CD4 cell count that decreased into the range of 0.200 to 0.499×10^9 cells/L and were included in the analysis.

Measurements: Hazard ratios and survival proportions for all-cause mortality and a combined end point of AIDS-defining illness or death.

Results: Compared with initiating cART at the CD4 cell count threshold of 0.500×10^9 cells/L, the mortality hazard ratio was 1.01 (95% CI, 0.84 to 1.22) for the 0.350 threshold and 1.20 (CI, 0.97 to 1.48) for the 0.200 threshold. The corresponding hazard ratios were 1.38 (CI, 1.23 to 1.56) and 1.90 (CI, 1.67 to 2.15), respectively, for the combined end point of AIDS-defining illness or death.

Limitations: CD4 cell count at cART initiation was not randomized. Residual confounding may exist.

Conclusion: Initiation of cART at a threshold CD4 count of 0.500×10^9 cells/L increases AIDS-free survival. However, mortality did not vary substantially with the use of CD4 thresholds between 0.300 and 0.500×10^9 cells/L.

Primary Funding Source: National Institutes of Health.

Ann Intern Med. 2011;154:509-515. www.annals.org

For author affiliations, see end of text.

* For a list of Writing Committee members, see end of article; for a list of the contributors to the HIV-CAUSAL Collaboration, see Appendix 1 (available at www.annals.org).

The HIV-CAUSAL Collaboration

Annals of Internal Medicine
2011; **154**: 509-515

- Find the optimal CD4 cell count at which to initiate cART
- AIDS or death: Initiation at 500 better than 450 cells/mm^3
- Death alone: similar for initiation at 300-500 cells/mm^3

International Journal of Epidemiology, 2016, 2038–2049
doi: 10.1093/ije/dyv295
Advance Access Publication Date: 31 December 2015
Original article

Methodological insights

Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therap

Writing committee: Lauren E
Margaret T May,[4] Suzanne M
Sophie Abgrall,[6,7] Bryan E She
Giota Touloumi,[11] Georgia Vo
Marie-Anne Vandenhende,[12] F
Hasina Samji,[15] Robert S Hog
Sophie Jose,[18] Julia del Amo,
Benigno Rodríguez,[23] Alessan
Christoph Stephan,[26] Santiag
Jodie L Guest,[28,29,30] Antonella
Richard Moore,[33] Colin NJ Ca
Laurence Meyer,[36] Rémonie S
Heiner C Bucher,[37] Matthias E
Richard Haubrich,[41] Elvin H Ge
Sonia Napravnik,[44] Mari M Kit
Ramón Teira,[46] Amy C Justice
Dominique Costagliola,[49] Jon
Miguel A Hernán[1,5,50] on beha
Collaboration, the Centers for
Clinical Systems, and the HIV-

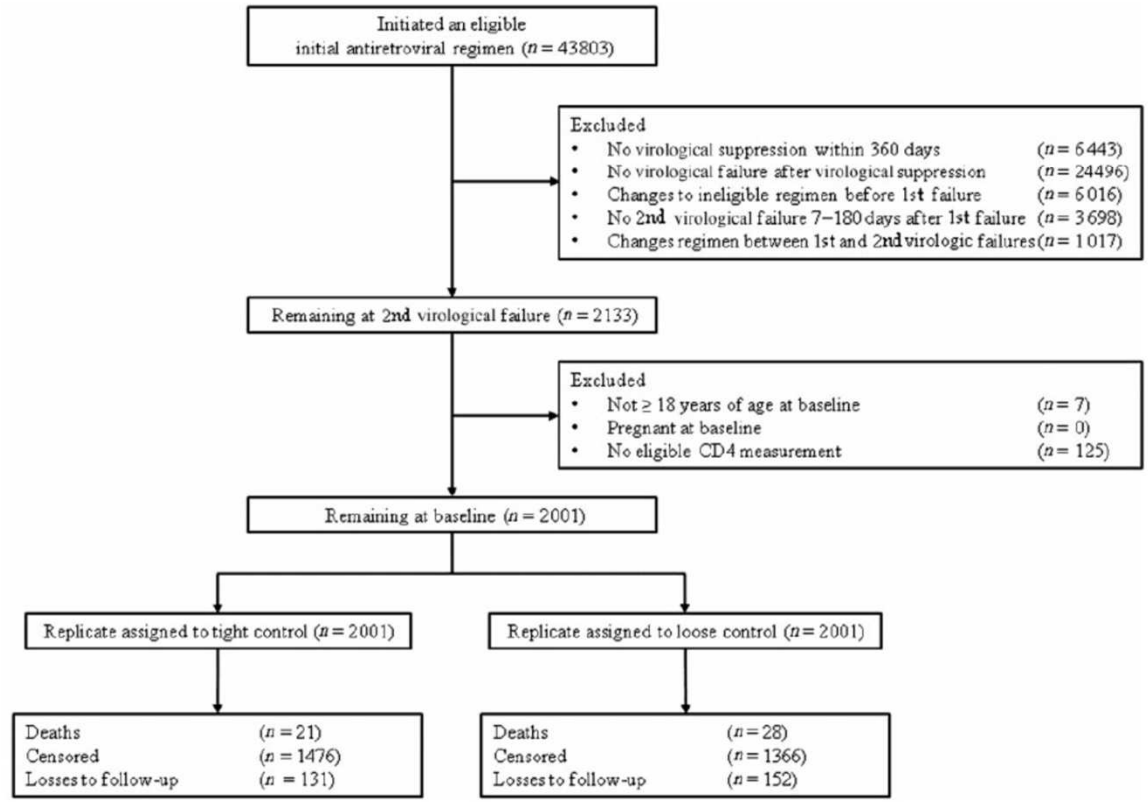
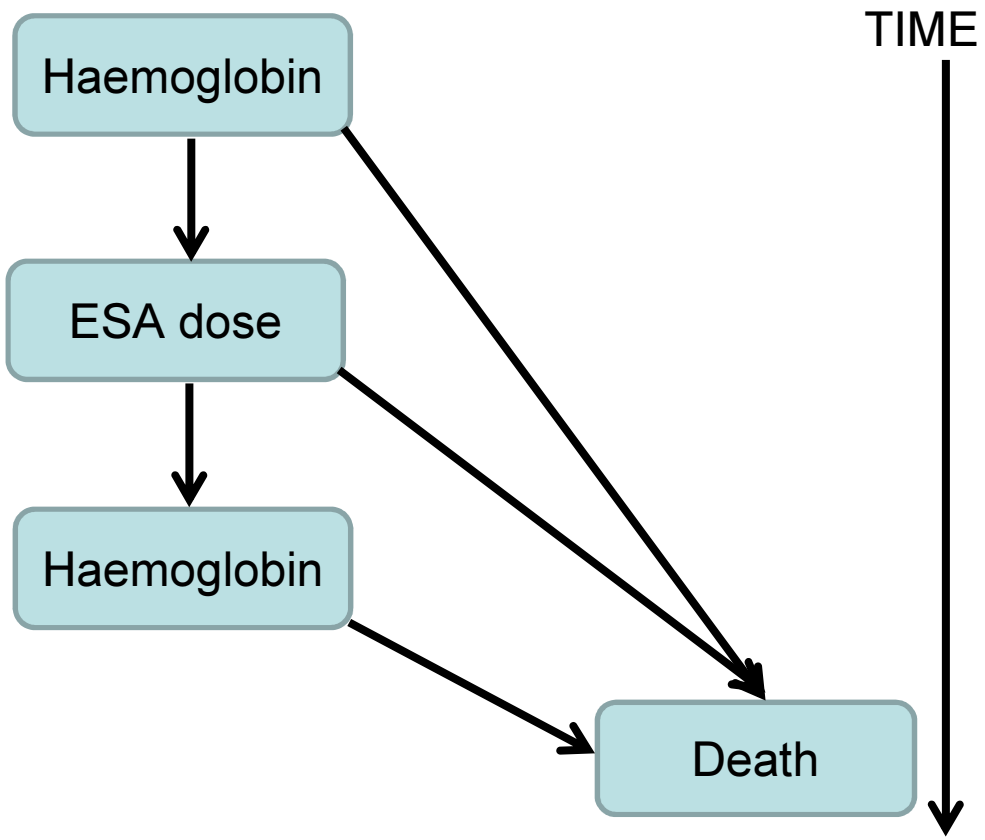


Figure 1. Modified CONSORT flow diagram for the mortality analysis in the ART-CC, the CNICS and the HIV-CAUSAL Collaboration, 2002–12.

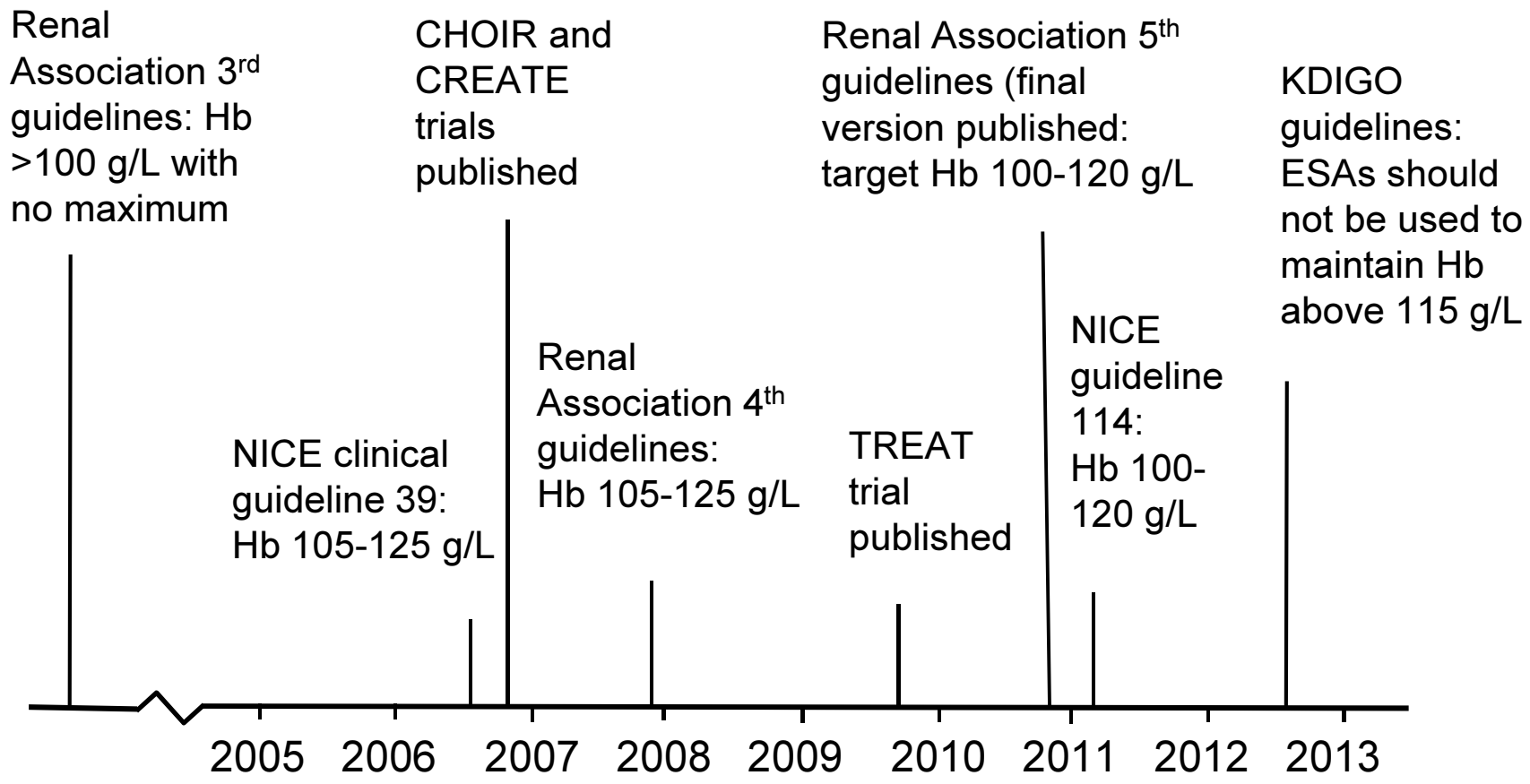
Erythropoiesis-stimulating agent (ESA) therapy for treating anaemia among haemodialysis patients



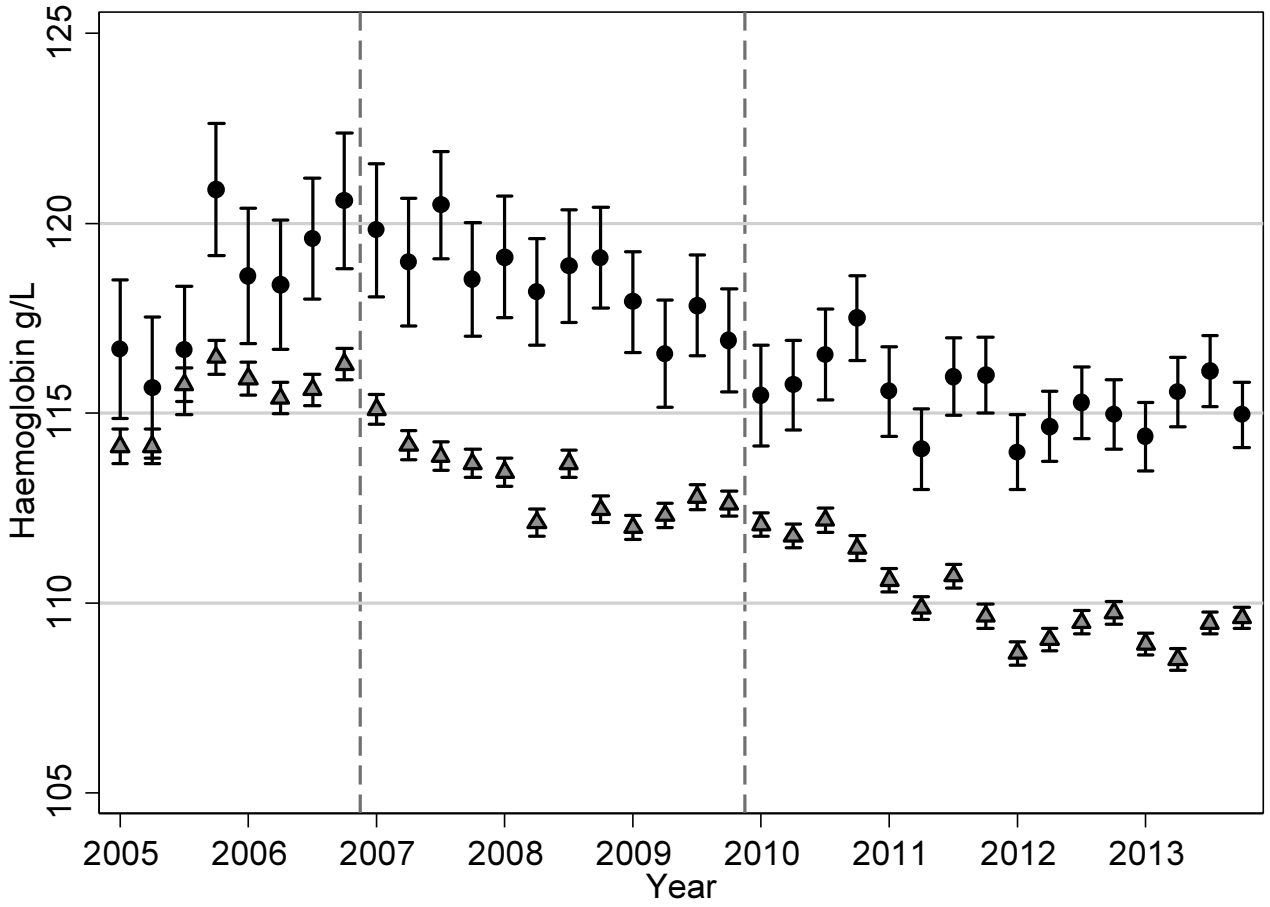
- Anaemia is common in patients with chronic kidney disease (CKD)
- It is measured by blood haemoglobin (Hb) levels
- ESAs with iron supplementation are the main treatment

Timeline of publications

RCTs in patients with CKD not yet on dialysis led to safety concerns over higher Hb targets, because of an increased risk of stroke



Mean Hb levels over time with 95% CIs, in haemodialysis patients



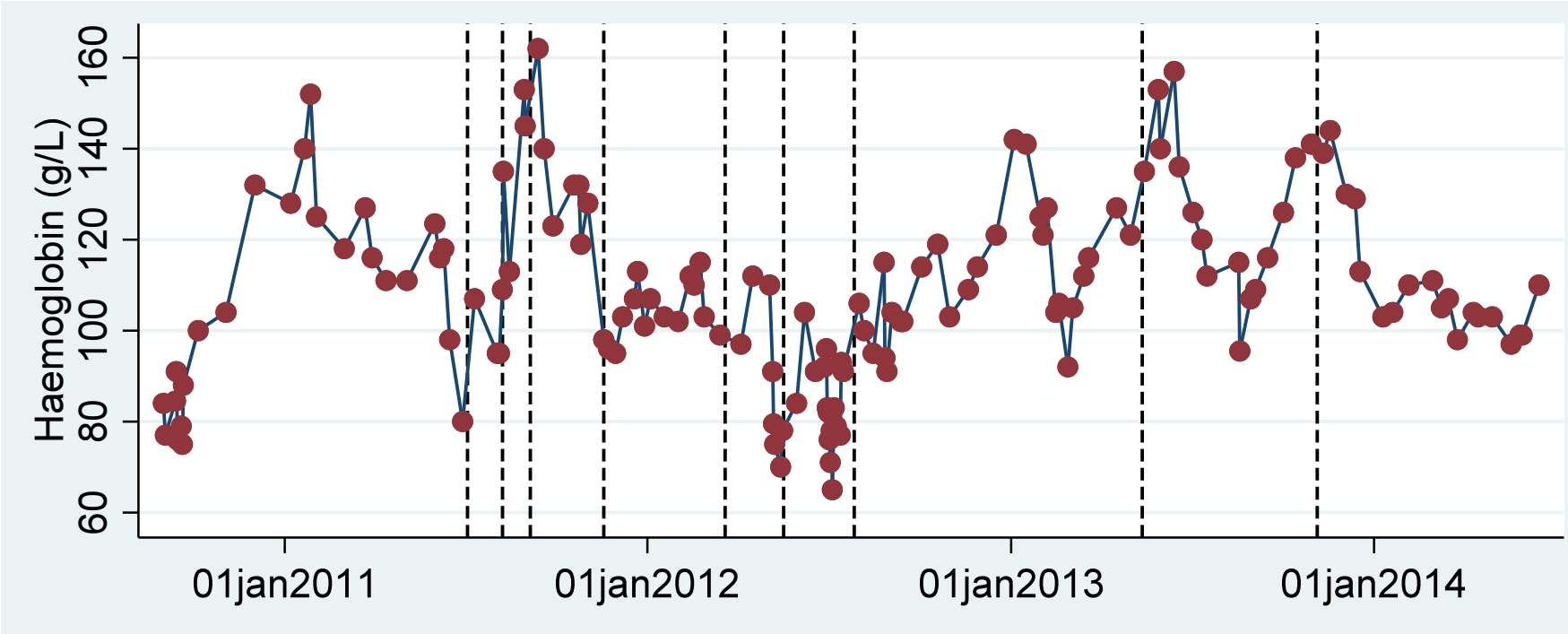
Circles are for patients not receiving ESAs; triangles are for patients receiving ESAs
The dashed vertical lines indicate the publication of the CHOIR and CREATE RCTs (2006) and TREAT (2009)

Motivation

- There is a risk that the current treatment guidelines may prevent patients without major co-morbidity from receiving the maximum benefit from treatment.
- Unlikely to be new RCTs at present.
- Estimating the effect of ESAs on survival in observational studies requires careful measurement of and appropriate adjustment for confounding as a result of time-varying haemoglobin levels and other factors that determine subsequent ESA dose.

Data requirements for this project

- Hb results from every blood test, with dates
- Every ESA dose change, with dates
- Example for the same individual patient as previous slide:



Circles represent Hb and the dashed vertical lines indicate ESA dose changes

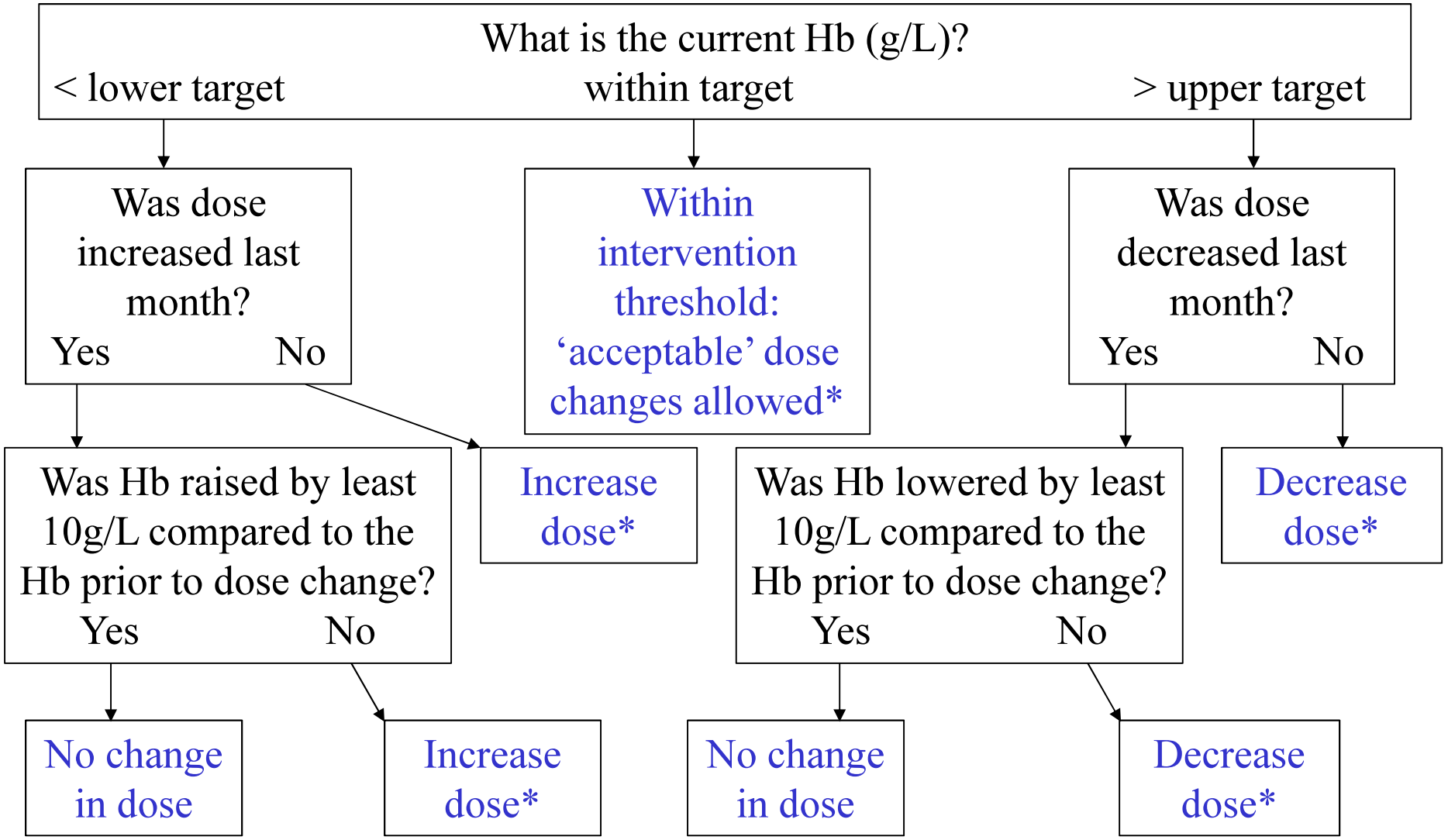
Trial with different Hb targets

- Eligibility criteria: people on haemodialysis for at least 3 months and on EPO
- Exclusions: people who, at the start of their eligibility, have a high ESA dose (≥120 darbepoetin units/week) and low Hb (<80 g/L)
- Comparison groups:

  Group 1: lower target=95 g/L, upper target=115 g/L

  Group 2: lower target=105 g/L, upper target=125 g/L

Protocol



* See separate table for acceptable dose changes

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

For numbered affiliations see end of article.

Correspondence to: J A C Sterne
jonathan.sterne@bristol.ac.uk

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;355:i4919
http://dx.doi.org/10.1136/bmj.i4919

Non-randomised studies of the effects of interventions are critical to many areas of healthcare evaluation, but their results may be biased. It is therefore important to understand and appraise their strengths and weaknesses. We developed ROBINS-I ("Risk Of Bias In Non-randomised Studies - of Interventions"), a new tool for evaluating risk of bias in estimates of the comparative effectiveness (harm or benefit) of interventions from studies that did

such as cohort studies and case-control studies in which intervention groups are allocated during the course of usual treatment decisions, and quasi-randomised studies in which the method of allocation falls short of full randomisation. Non-randomised studies can provide evidence additional to that available from randomised trials about long term outcomes, rare events, adverse effects and populations that are typical of real world practice.[1 2] The availability of linked databases and compilations of electronic health records has enabled NRSI to be conducted in large representative population cohorts.[3] For many types of organisational or public health interventions, NRSI are the main source of evidence about the likely impact of the intervention because randomised trials are difficult or impossible to conduct on an area-wide basis. Therefore systematic reviews addressing the

Conclusions

- Randomized trials provide a reference point for causal inference
- Making causal inferences from observational data requires strong and untestable assumptions
  - To avoid these assumptions, you should conduct a trial
- There is no magical method for making causal inferences
  - Stratification
  - Standardization
  - Regression models
  - Propensity scores
  - Marginal structural models, g-computation, g-estimation, TMLE
- The start point is the trial you'd like to mimic using observational data
  - Specifying the target trial requires discussion with clinical colleagues, and can be surprisingly challenging