Introduction to Causal Inference and Causal Data Science

Instrumental Variables, Mediation and Complex Longitudinal Settings

Bas Penning de Vries

Topics of lecture

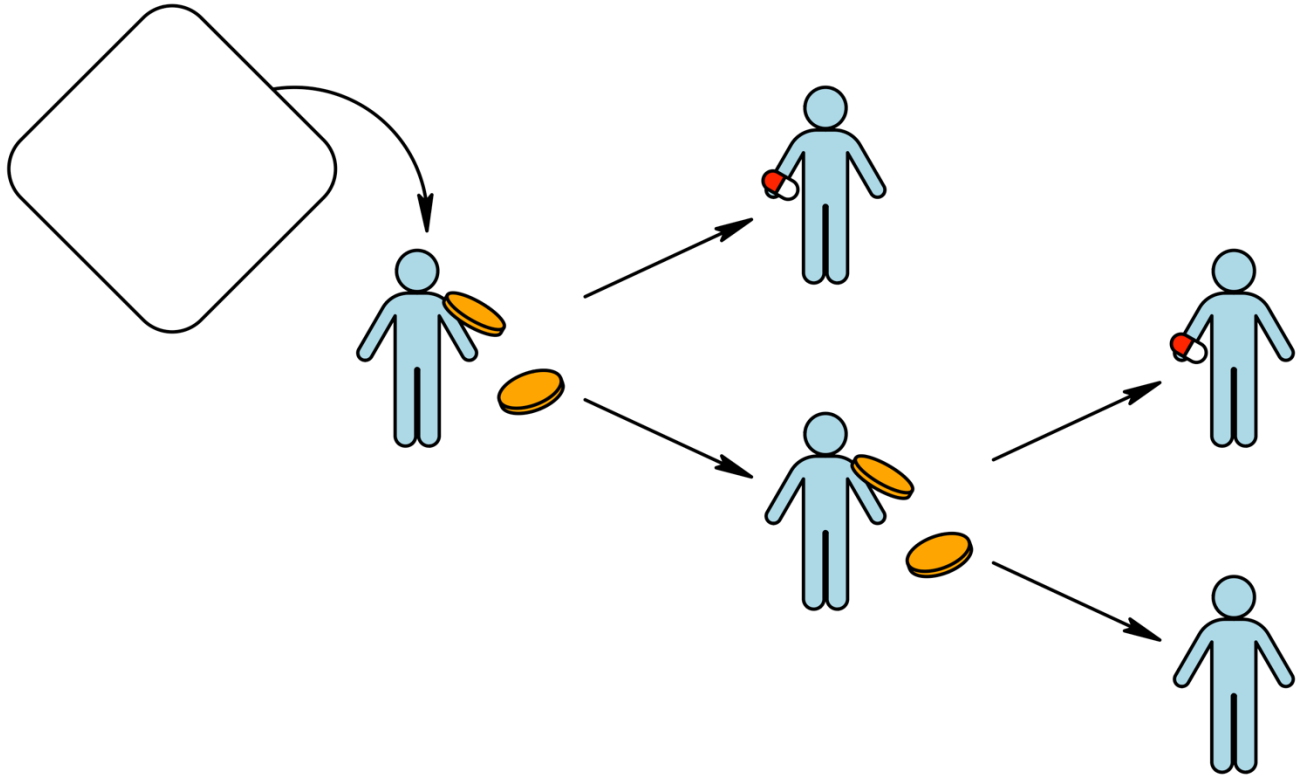
- A **taxonomy of estimands** on time-varying treatments (incl. intention-to-treat and per-protocol effects)
- Why **traditional methods fail** for inference about time-varying treatments in the presence of treatment-covariate feedback
- **G-methods** for time-varying treatments
- Typical estimands in **mediation analysis**
- Why **traditional approach** to mediation **analysis fail**
- **G-methods** overcome limitations of traditional approaches
- **Instrumental variable** analysis with **non-exchangeable treatment groups**

Statin-cancer example revisited

Previous studies implicitly compared long-term statin users versus non-users – don't necessarily answer questions like …

- What would be my 10-year cancer risk if – possibly contrary to fact – I would start statin treatment now? And what if I wouldn't?
- What would be my 10-year cancer risk if – possibly contrary to fact – I would start statin treatment now and **adhered** to it? And what if I wouldn't start now *or* in the future?

Intention-to-treat versus per-protocol effect



Inference about time-varying treatments

If treatment/exposure is time-varying, there are many possible causal contrasts …

Single versus multiple-point interventions

Single-point (baseline) intervention

- Eg: *assign*/*initiate* versus withhold drug treatment at baseline
- Individuals are allowed to deviate (*intention-to-treat*)
- Randomisation at baseline only

Multiple-point (joint) intervention

- Eg: sustained/daily/weekly/monthly drug use versus continuous non-use (*per-protocol*)
- Randomisation at multiple points

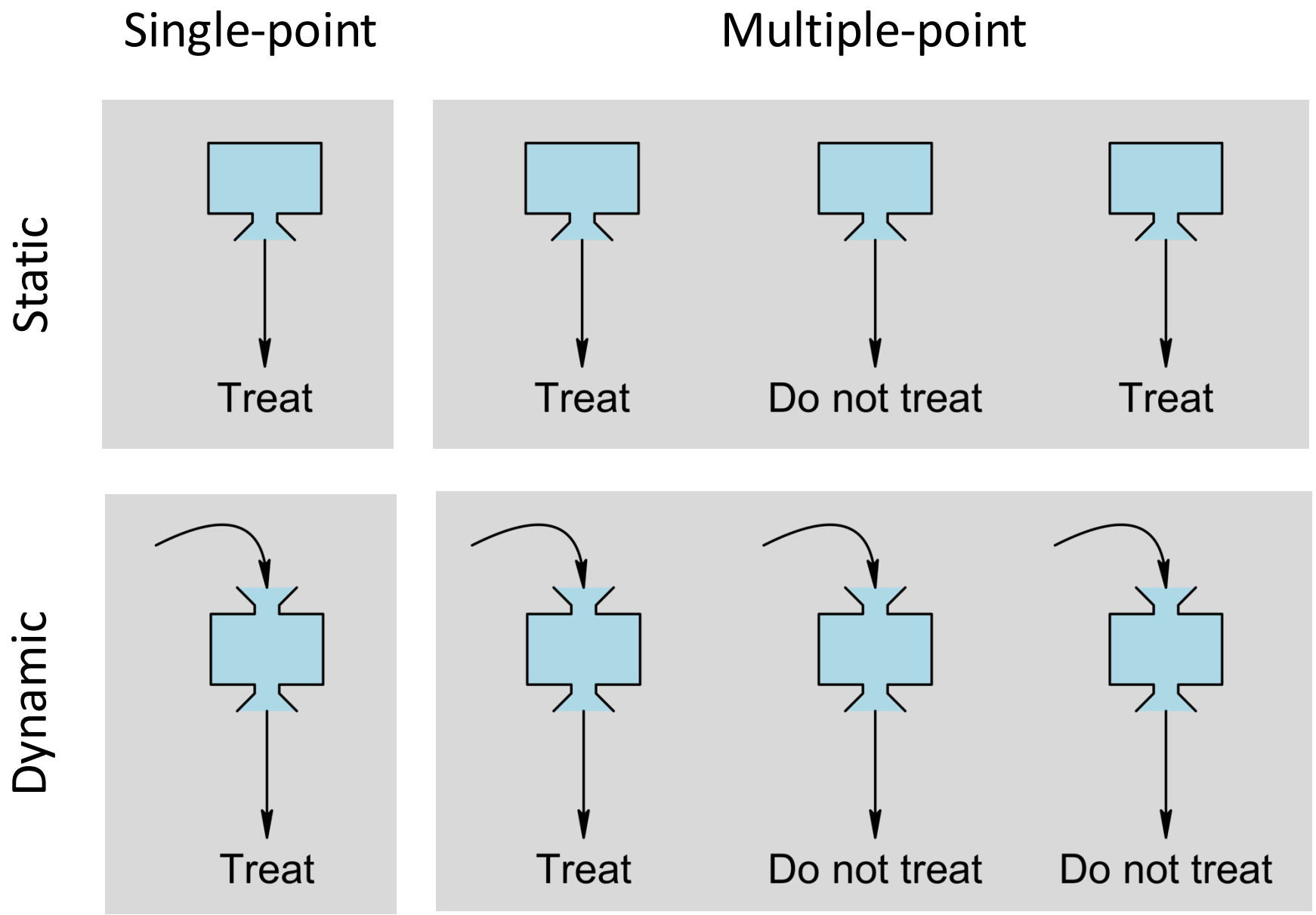
Static versus dynamic interventions

Static treatment rule/regime/protocol/…

- … assigns the same treatment option to everyone
- Eg: assign versus withhold treatment regime at baseline (intention-to-treat)
- Eg: always treat versus never treat (per-protocol)

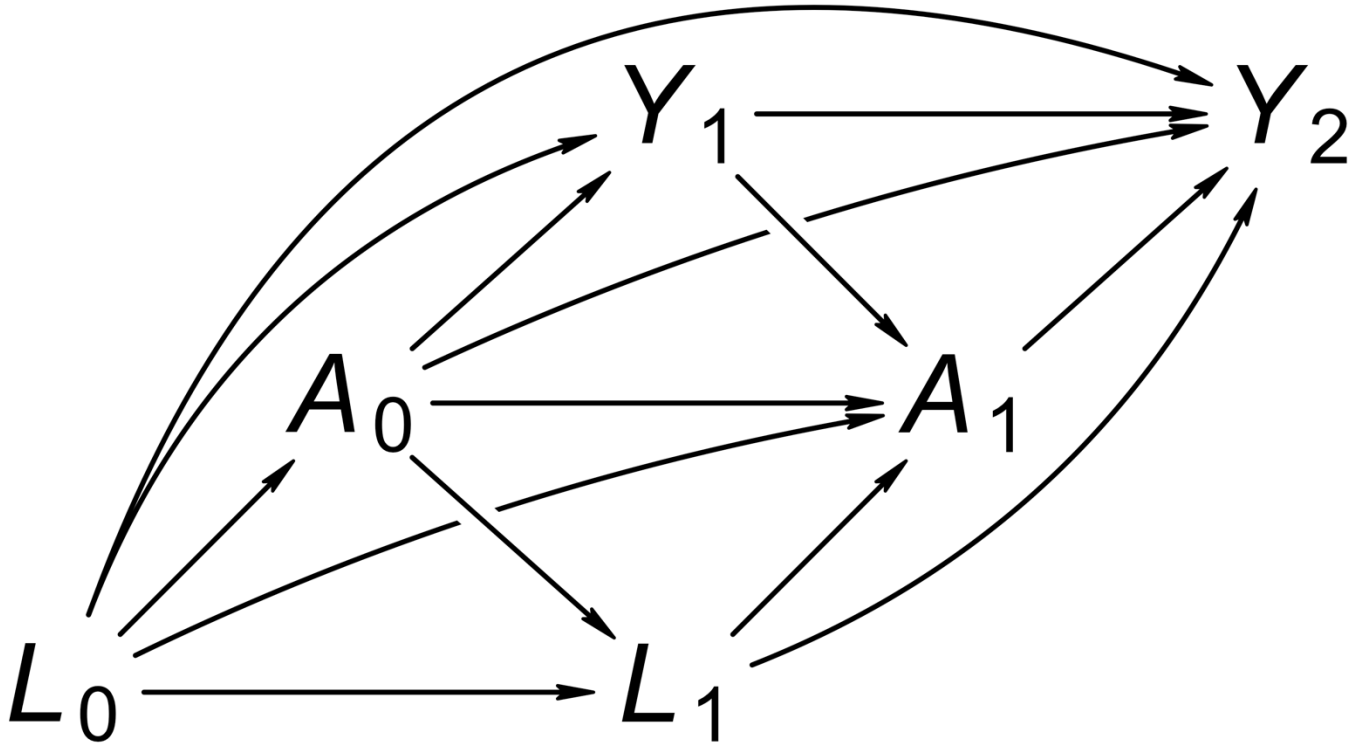
Dynamic (individualised) treatment rule

- … assigns treatment based on the then-available information
- Eg: choose dose depending on baseline covariates
- Eg: start when blood marker first drops below threshold
- Eg: stop when toxicity occurs


Single-point
Multiple-point
Static
Dynamic
Treat
Treat
Do not treat
Treat
Treat
Treat
Do not treat
Do not treat

Which of the following are examples of a single-point intervention?

A. *Initiate* an exercise program consisting of 30 minutes per day
B. *Initiate* a program of daily exercise and *adhere* to it for at least 3 months
C. *Initiate* 10 mg/day atorvastatin and *continue* for five years or *until* the development of a contraindication
D. *Issue* a single prescription for daily atorvastatin use within a given period of time



Treatment-covariate feedback

Adjust for L_1 if we want to know the effect of always- versus never treatment?

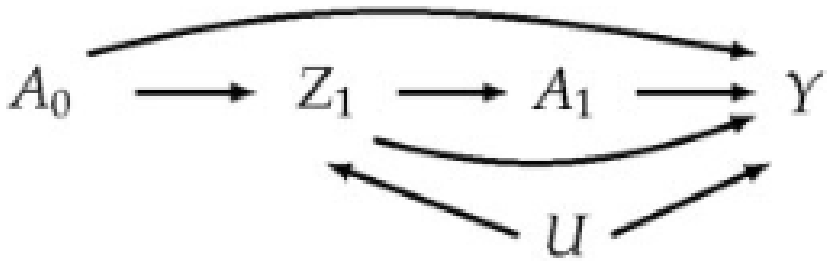


Figure 1. Causal diagram representing the relation between anti-retroviral treatment at time 0 (A_0), HIV viral load just prior to the second round of treatment (Z_1), anti-retroviral treatment status at time 1 (A_1), the CD4 count measured at the end of follow-up (Y), and an unmeasured common cause (U) of HIV viral load and CD4.

International Journal of Epidemiology, 2017, Vol. 46, No. 2

Treatment-covariate feedback

Adjust for Z_1 (HIV viral load just prior to second round of treatment) if we want to know the effect of always- versus never anti-retroviral treatment?

Treatment-covariate feedback

- Traditional methods (multivariable regression modelling) are not suited to deal with time-varying confounding when it is affected by past treatment (treatment-covariate feedback)

- Methods that can handle treatment-covariate feedback and adjusting for time-varying confounding:
  - Inverse probability weighting (IPW)
  - G-computation
  - G-estimation

IPW for sustained treatments

- Let $Y_1^{a_0}, Y_2^{a_0,a_1}$ be counterfactual Y_1, Y_2 if all A_0, A_1 were a_0, a_1
- Target: $\Pr(Y_1^{a_0} = Y_2^{a_0,a_1} = 0)$ – survival probability if …?
- Same assumptions but slightly different form – eg, sequential conditional **exchangeability**:
  $Y_1^{a_0}, Y_2^{a_0,a_1}$ independent of
  - independent of A_0 given L_0
  - independent of A_1 given L_0, L_1, $A_0=a_0$, $Y_1=0$
- Idea: *patients who do not deviate from protocol compensate for "similar" patients who do*
- Weights are time-varying

Clone-censor-weight approach

- ‘Clone’ each patient once for each treatment regimen of interest
- ‘Censor’ each clone when their person-time is no longer consistent with the corresponding treatment regimen
- ‘Weight’ the remaining person-time by the inverse probability of being censored

When to Start Treatment? A Systematic Approach to the Comparison of Dynamic Regimes Using Observational Data*

Lauren E. Cain, James M. Robins, Emilie Lanoy, Roger Logan, Dominique Costagliola, and Miguel A. Hernán

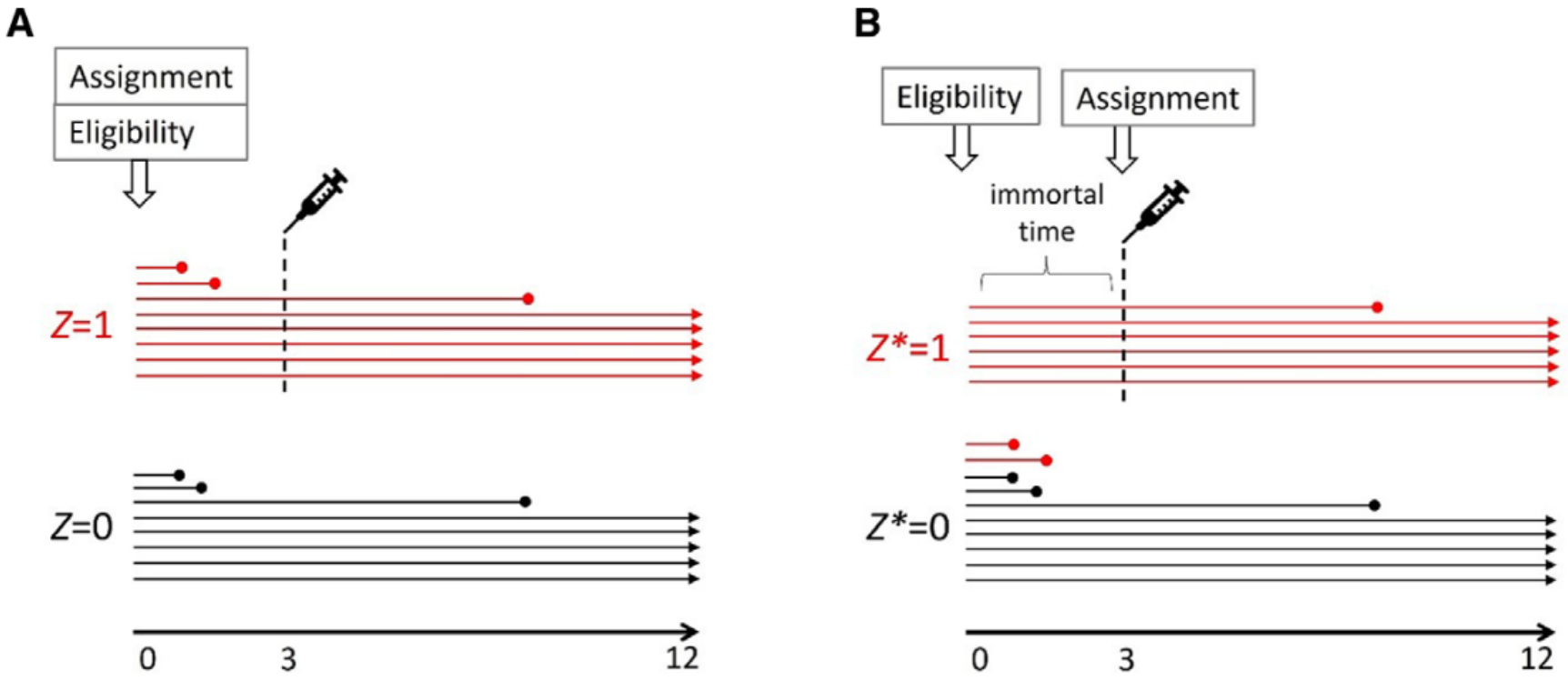
Abstract

Dynamic treatment regimes are the type of regime most commonly used in clinical practice. For example, physicians may initiate combined antiretroviral therapy the first time an individual's recorded CD4 cell count drops below either 500 cells/mm^3 or 350 cells/mm^3. This paper describes an approach for using observational data to emulate randomized clinical trials that compare dynamic regimes of the form "initiate treatment within a certain time period of some time-varying covariate first crossing a particular threshold." We applied this method to data from the French Hospital database on HIV (FHDH-ANRS CO4), an observational study of HIV-infected patients, in order to compare dynamic regimes of the form "initiate treatment within m months after the recorded CD4 cell count first drops below x cells/mm^3" where x takes values from 200 to 500 in increments of 10 and m takes values 0 or 3. We describe the method in the context of this example and discuss some complications that arise in emulating a randomized experiment using observational data.

Table 1: Six hypothetical individuals who follow multiple regimes in the class "initiate treatment within m months after the recorded CD4 cell count first drops below x cells/mm^3" where x takes the values 200 to 500 in increments of 10 and m takes values 0 and 3.

Individual	Time (months)	CD4 cell count	Treatment (1: yes, 0: no)	No. of regimes followed (range of x)	
				$m = 0$	$m = 3$
1	0	352	1	15 (360-500)	15 (360-500)
1	1	380	1	15 (360-500)	15 (360-500)
1	2	273	1	15 (360-500)	15 (360-500)
1	3	273	1	15 (360-500)	15 (360-500)
1	4	198	1	15 (360-500)	15 (360-500)
2	0	352	0	16 (200-350)	31 (200-500)
2	1	380	1	0	15 (360-500)
2	2	273	1	0	15 (360-500)
2	3	273	1	0	15 (360-500)
2	4	198	1	0	15 (360-500)
3	0	352	0	16 (200-350)	31 (200-500)
3	1	380	0	16 (200-350)	31 (200-500)
3	2	273	1	8 (280-350)	23 (280-500)
3	3	273	1	8 (280-350)	23 (280-500)
3	4	198	1	8 (280-350)	23 (280-500)

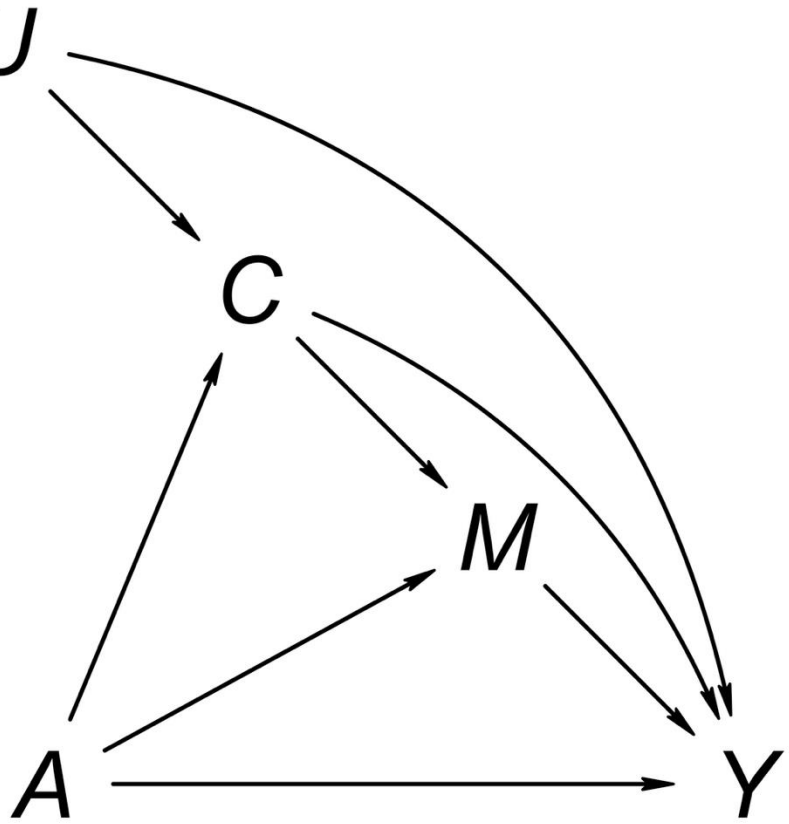
In the absence of confounding by unmeasured factors, as formalized in the strengthened identifiability conditions of exchangeability, positivity, and consistency described by Robins and Hernán (2008) we can eliminate the bias introduced by the artificial censoring if we weight each replicate of an individual by the individual's time-varying, unstabilized inverse probability weight



The clone-censor-weight approach is especially well suited to prevent immortal time bias (eg, due to misclassification as per figure B)!

Concluding remarks

- IPW/clone-censor-weighting (and other g-methods), but not traditional methods, are suited to handle time-varying confounding affected by past treatment (feedback)
- As with IPW for time-fixed confounding, default standard error estimators of many software packages are not appropriate for weighted regressions, because the weights are falsely assumed to reflect actual (independent!) observation frequencies
- IPW can (and sometimes need) be combined with marginal structural modelling (Robins et al., Epidemiology, 2000;11:550-560)

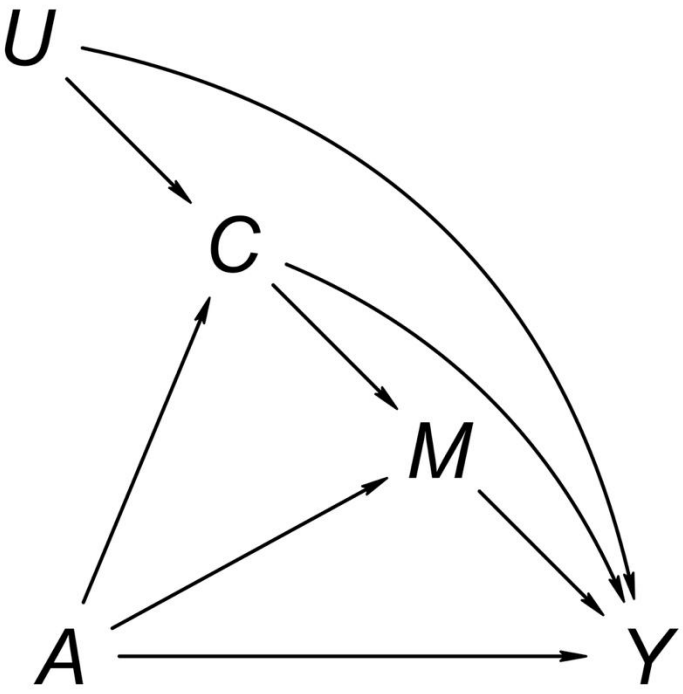


Mediation

What part of the effect “goes via” or is “mediated by” a third variable?

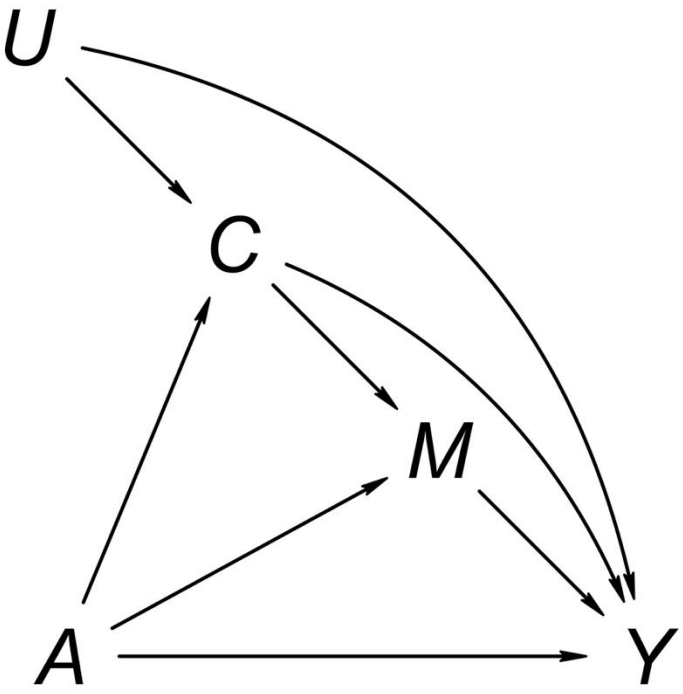
Examples

- What part of the effect of cognitive behavioural therapy (CBT) on depression symptoms is **mediated by** antidepressant use?
- To what extent is the effect of genetic variants on incident lung cancer **mediated by** smoking behaviour?
- To what extent is the effect of maternal smoking on infant mortality **mediated by** birth weight?



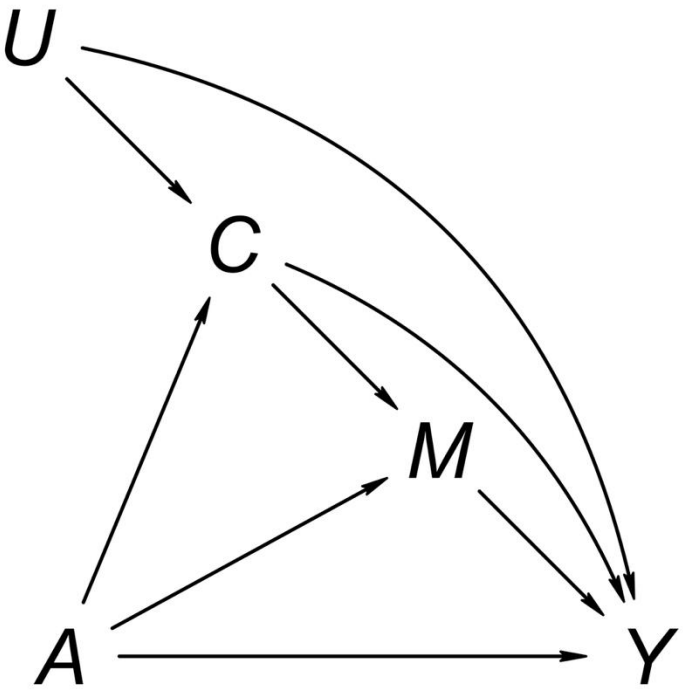
Traditional approach

- Exposure: CBT (*A*); outcome: depression symptoms (*Y*); mediator: antidepressant use (*M*)?
- Regress *Y* on *A* and *M* to estimate part of effect ‘not through’ *M*
- But what happens if we condition on *M*?



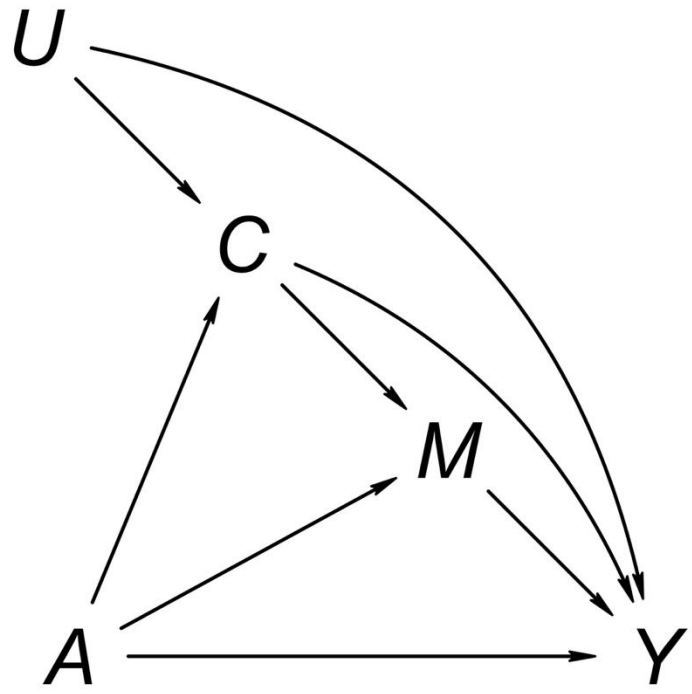
Why the traditional approach fails

- Assumes no unmeasured confounding between *M* and *Y*
  (not guaranteed in trials where A is randomly assigned)
  Else: **collider-stratification (bias)**!



Why the traditional approach fails

- Assumes common causes of *M* and *Y* (eg, *C*) do not also mediate the effect of *A* on *Y*, other than through *M*
  Else: adjusting for *C* would '**wash away**' part of indirect effect



Why the traditional approach fails

- Assumes no interaction between *A* and *M*
- And more … (depending what it is we want to know exactly)

Mediation analysis and causal inference on time-varying treatments

- "The study of causal **mediation** can be seen as a **special case** of causal inference with **time-varying treatments**"

- "Rather than having a single treatment that takes different values over time, in mediation analysis we have two different variables—the treatment of interest and the mediator—at different times"

Hernán and Robins, 2020, *Causal Inference: What if*

Modern counterfactual outcomes perspective on mediation analysis

Let's clarify what we want to know, the **estimand**! The counterfactual outcomes framework gives us a language to do this …

Notation:

- Treatment (e.g., CBT): A
- Mediator (e.g., antidepressant use): M
- Mediator had treatment A been a: M^a
- Outcome (e.g., depression symptoms): Y
- Outcome had A and M been a and m: $Y^{a,m}$
- Outcome had A been a: Y^{a,M^a}

Effect decomposition and estimand

Mediation analyses presume some sort of **decomposition** of the **total** effect into **mediated** (or **indirect**) and **direct** effects …

Total effect = $\mathrm{E}[Y^{1,M^1} - Y^{0,M^0}]$

$= \mathrm{E}[Y^{1,M^1} - Y^{1,M^0} + Y^{1,M^0} - Y^{0,M^0}]$

$= \mathrm{E}[Y^{1,M^1} - Y^{1,M^0}] + \mathrm{E}[Y^{1,M^0} - Y^{0,M^0}]$

= **natural indirect effect** + **natural direct effect**

Effect decomposition and estimand

Mediation analyses presume some sort of **decomposition** of the **total** effect into **mediated** (or **indirect**) and **direct** effects …

Total effect = $\mathrm{E}[Y^{1,M^1} - Y^{0,M^0}]$

= $\mathrm{E}[Y^{1,m} - Y^{1,m}]$ + …

= **controlled direct effect** (m) + …

Controlled versus natural direct effects

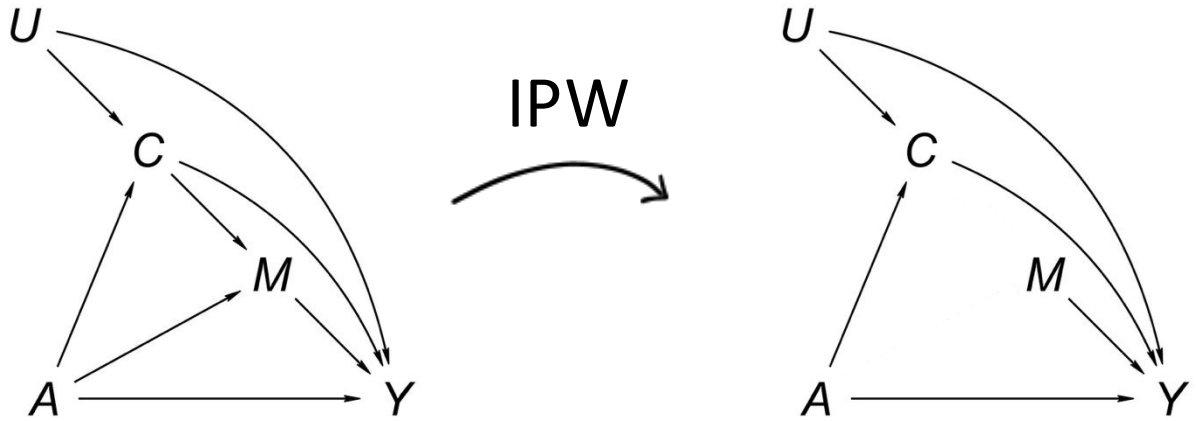
- Different interpretations
  - NDE: ‘how much of the effect “flows” through the mediator path?’
  - CDE: ‘how much of the effect would be removed by fixing the mediator to a constant?’
- For policy, CDE typically most relevant (directly)
- NDE and CDE can differ substantially – e.g., when treatment does not affect *M* yet (additively) interacts with *M*:
  - NDE = total effect
  - CDE(*m*) may differ from total effect for some *m*
- NDE and CDE are equal under strong assumptions

Cross-world assumptions

- Traditional approach (assuming not measured covariates) can be viewed as starting with two models:
  - One for the mediator:
    $M^a = \beta_0 + \beta_1 a + \epsilon$, with mean-zero ϵ independent of A
  - One for the outcome:
    $Y^{a,m} = \alpha_0 + \alpha_1 a + \alpha_2 m + \varepsilon$, with mean-zero ε independent of A,ϵ
- Under these models, CDE and NDE coincide
- But they imply a **cross-world independence**:
  $Y^{1,m}$ and M^0 are independent!
- These counterfactuals don't live in the same world: one lives in a world where A is set to 1 and the other where it is set to 0!

Identification of controlled direct effects

- For identification of CDE, we don’t need cross-world assumptions
- Typical identification strategies that overcome limitations of traditional approach: g-computation, IPW
- Idea of IPW:



Identification of controlled direct effects

Suppose that, in addition to consistency and positivity, the following **exchangeability** conditions are met for binary *A,M*:

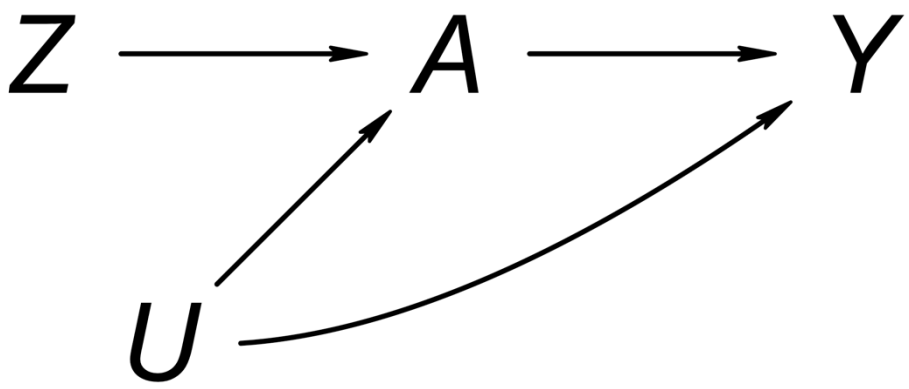
- $Y^{a,m}$ independent of *A*
  (**no unmeasured exposure-outcome confounding**)
- $Y^{a,m}$ independent of *M* given *C* and *A=a*
  (**no unmeasured mediator-outcome confounding** given *C* and *A*)

Then, $E[Y^{a,m}] = E_{\text{pseudopopulation}}[Y|\ A=a, M=m]$ if the pseudopopulation is obtained by weighting everyone with *W*, defined as
$1 / \Pr(M = 1 \mid C,A)$ if $M = 1$ and as $W = 1 / \Pr(M = 0 \mid C,A)$ if $M = 0$.

Identification of natural direct effects

- Point (i.e., exact) identification problematic as identification results rely on cross-world assumptions!
- But partial identification (without cross-world assumptions) may still be informative

  (Robins, J.M. and Richardson, T.S., 2010. Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: Finding the determinants of disorders and their cures*, *84*, pp.103-158)
- Relevance questionable/debatable



Instrumental variable analysis

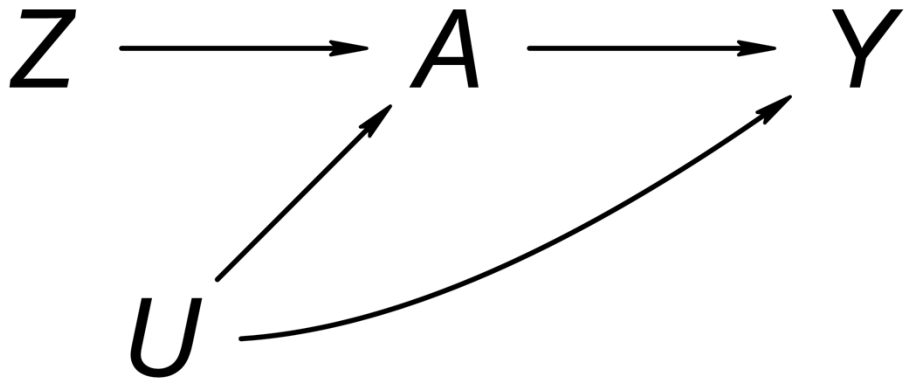
… when we have non–exchangeable treatment groups (i.e., uncontrolled confounding)

Standard identifiability assumptions

For inference about the effect of A on Y from observational data, we typically assume

- **Conditional exchangeability**: A and Y^a are independent given L
- **Consistency**: $Y^a = Y$ if $A = a$
- **Positivity**: $\Pr(A = a \mid L) > 0$

What if we're not comfortable that we've found such L? It may then be worth considering instrumental variable analysis, which exchanges our standard assumptions for a different set of identifiability assumptions



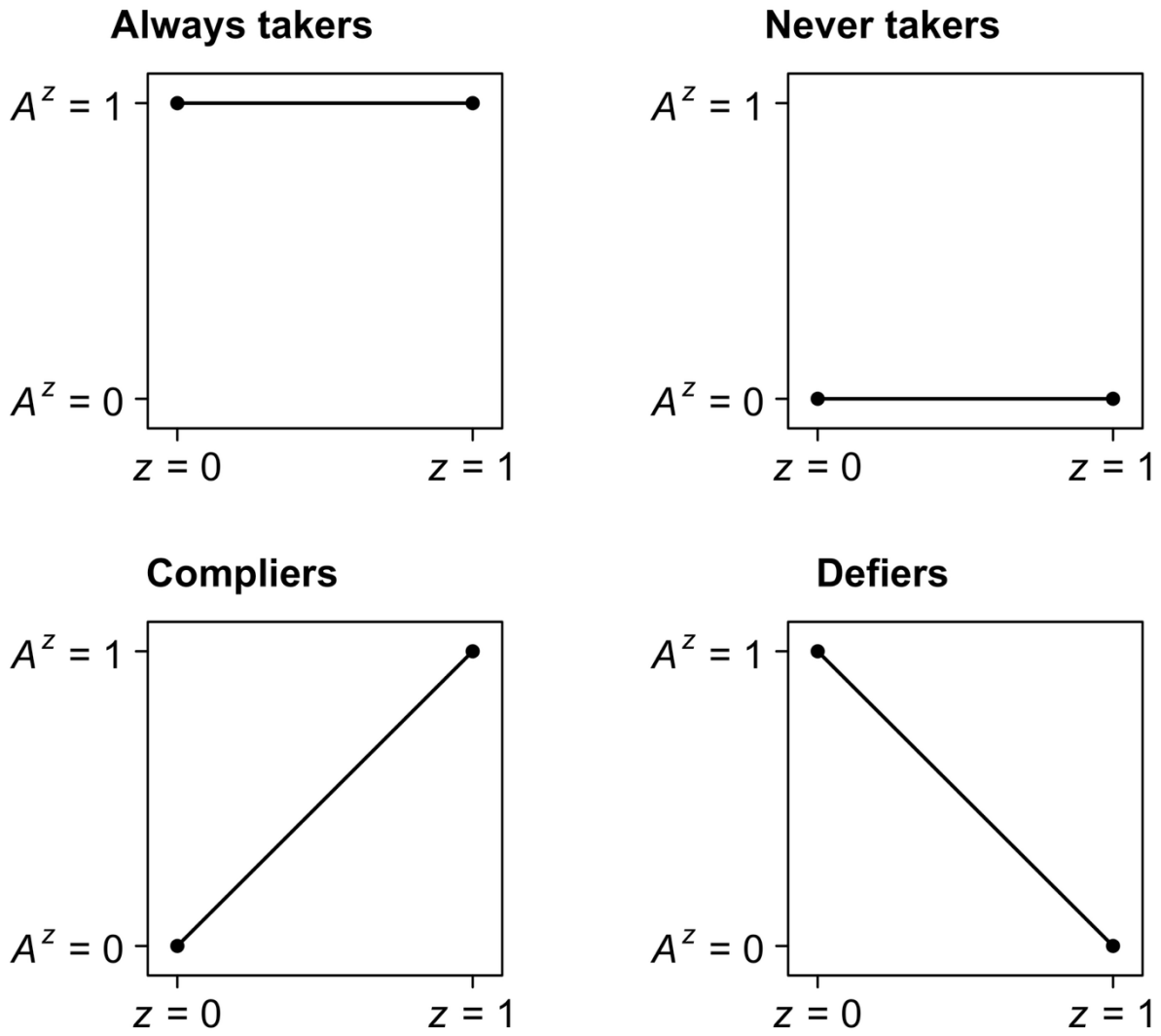
Motivating example:
RCT with non-compliance

Inference about effect of received treated

- Intention-to-treat analysis (misclassification due to non-compliance)
- Per-protocol analysis (≠ per-protocol effect; unmeasured confounding)
- As-treated analysis (≠ treatment effect; unmeasured confounding)

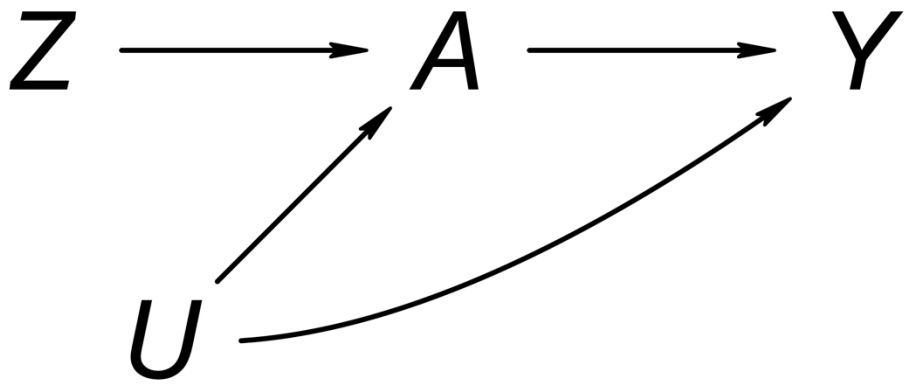
Notation

- Allocated treatment: Z
- Received treatment: A
- Received treatment had Z been z: A^z
- Outcome: Y
- Outcome had Z and A been z and a: $Y^{z,a}$
- Outcome had A been a: $Y^{Z,a}$



Four subgroups

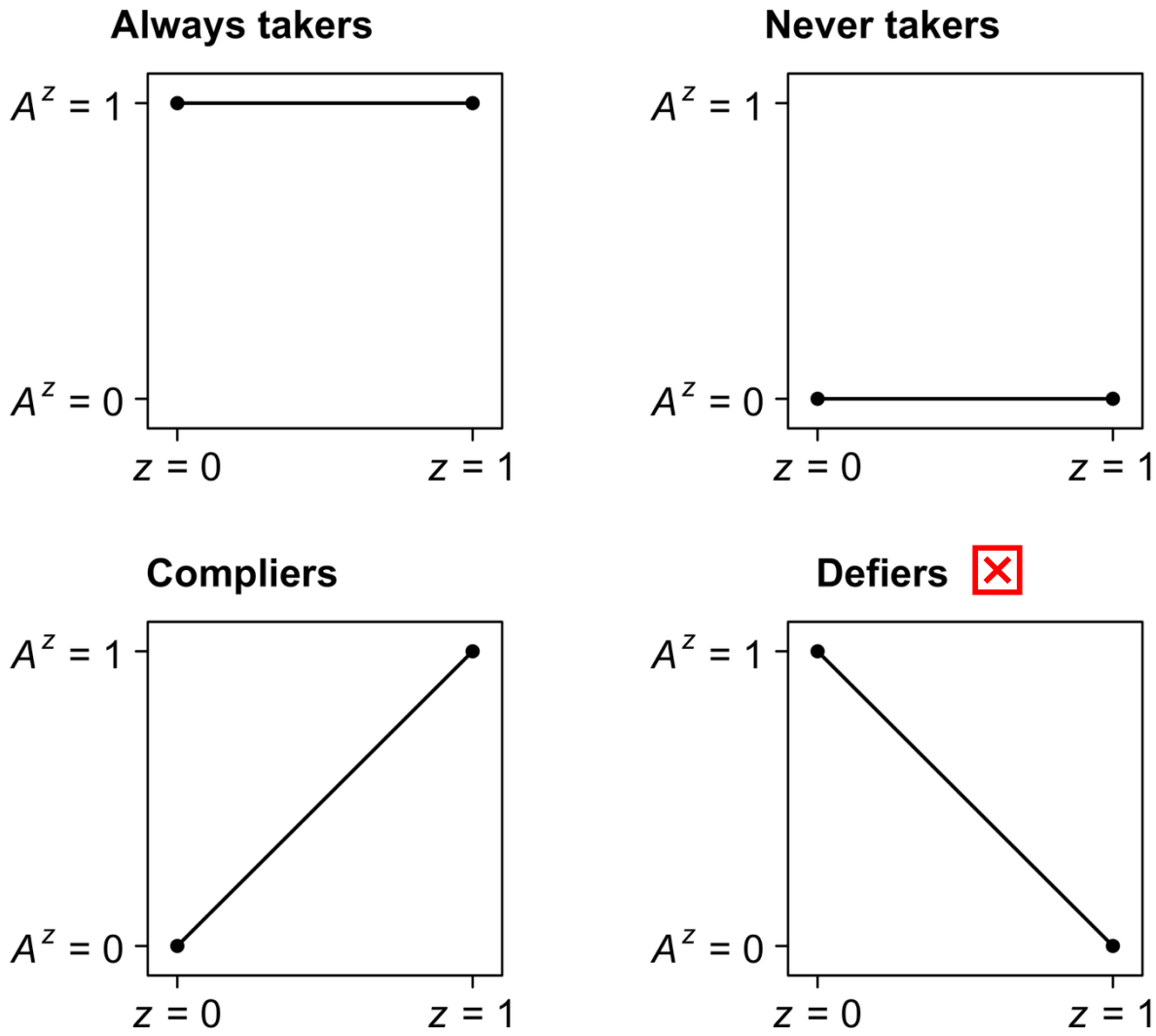
Based on A^1 and A^0, we can distinguish between four subgroups



Instrumental variables

Z is an instrumental variable (IV) for treatment *A* and outcome *Y* if

- Cov(*Z*,*A*) ≠ 0 (**relevance**)
- *Y*(*z*,*a*) constant across levels of *z* (no direct effect of IV on *Y*; **exclusion restriction**)
- **IV exchangeability**: e.g., no causes shared by *Z* and *Y* or by *Z* and *A*



Fourth assumption

Monotonicity: there are no defiers

IV analysis under monotonicity

Suppose instrument Z and treatment A are binary
and, in addition to consistency and positivity, the following assumptions hold:

- Relevance
- Exclusion restriction
- Exchangeability: Y^{z,A^z} and A^z are both independent of Z, for all z
- Monotonicity: there are no defiers

Then, the mean treatment effect among compliers, $E[Y^{Z,1} - Y^{Z,0} | \text{Compliers}]$, equals $\text{Cov}(Y,Z) / \text{Cov}(A,Z)$ or, equivalently,

$$\frac{E[Y \mid Z = 1] - E[Y \mid Z = 0]}{E[A \mid Z = 1] - E[A \mid Z = 0]}$$

Checking assumptions

- Relevance *can* be empirically tested
- Exclusion criterion *cannot* generally be empirically tested
- Exchangeability *cannot* generally be empirically tested
  - But balance can be tested with respect to measured variables
- Monotonicity *cannot* generally be empirically tested

✗

Z	A	$A(1-Z)$	**Probability**
0	0	0	0.00
0	0	1	0.25
0	1	0	0.00
0	1	1	0.25
1	0	0	0.00
1	0	1	0.25
1	1	0	0.00
1	1	1	0.25

✓

Z	A	$A(1-Z)$	**Probability**
0	0	0	0.00
0	0	1	0.25
0	1	0	0.00
0	1	1	0.25
1	0	0	0.25
1	0	1	0.00
1	1	0	0.00
1	1	1	0.25

Z	A	**Probability**
0	0	0.25
0	1	0.25
1	0	0.25
1	1	0.25

Instrumental Variables

Application and Limitations

Edwin P. Martens,† Wiebe R. Pestman,† Anthonius de Boer,* Svetlana V. Belitser,* and Olaf H. Klungel**

Abstract: To correct for confounding, the method of instrumental variables (IV) has been proposed. Its use in medical literature is still rather limited because of unfamiliarity or inapplicability. By introducing the method in a nontechnical way, we show that IV in a linear model is quite easy to understand and easy to apply once an appropriate instrumental variable has been identified. We also point out some limitations of the IV estimator when the instrumental variable is only weakly correlated with the exposure. The IV estimator will be imprecise (large standard error), biased when sample size is small, and biased in large samples when one of the assumptions is only slightly violated. For these reasons, it is advised to use an IV that is strongly correlated with exposure. However, we further show that under the assumptions required for the validity of the method, this correlation between IV and exposure is limited. Its maximum is low when confounding is strong, such as in case of confounding by indication. Finally, we show that in a study in which strong confounding is to be expected and an IV has been used that is moderately or strongly related to exposure, it is likely that the assumptions of IV are violated, resulting in a biased effect estimate. We conclude that instrumental variables can be useful in case of moderate confounding but are less useful when strong confounding exists, because strong instruments cannot be found and assumptions will be easily violated.

(*Epidemiology* 2006;17: 260–267)