

# **Introduction to Causal Inference and Causal Data Science**

---

## Introduction to Target Trials and Target Trial Emulation

---

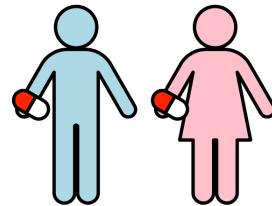
Bas Penning de Vries

# Exercise

- Imagine, we you'd like to learn about the effect of long-term use versus non-use of some drug (statins) on cancer incidence from observational data (about the period Jan 2000 – Dec 2023)
- You start follow-up (counting cancers) in Jan 2001
- Discuss with your neighbour how you might proceed
- When does someone qualify as a long-term user?
- Would you include individuals who started statin use before Jan 2001?

# Why trials?

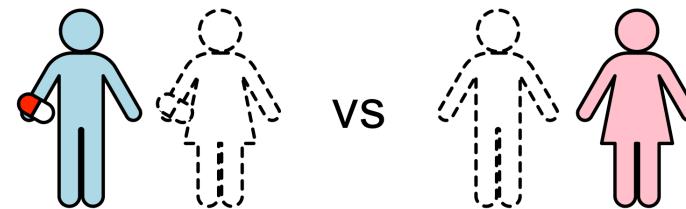
**Causal inference** is about speculating what would happen if ...



A **causal effect** is a contrast between the answers to what-if questions

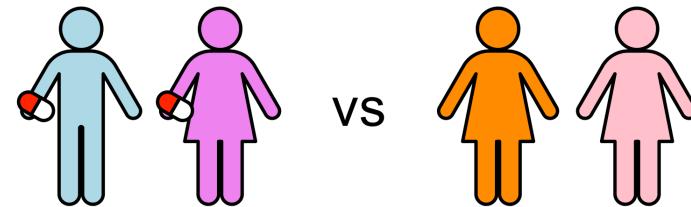
# Fundamental obstacle

Impossible to observe the consequences of  $\geq 2$  mutually exclusive actions (interventions, treatments, etc.)

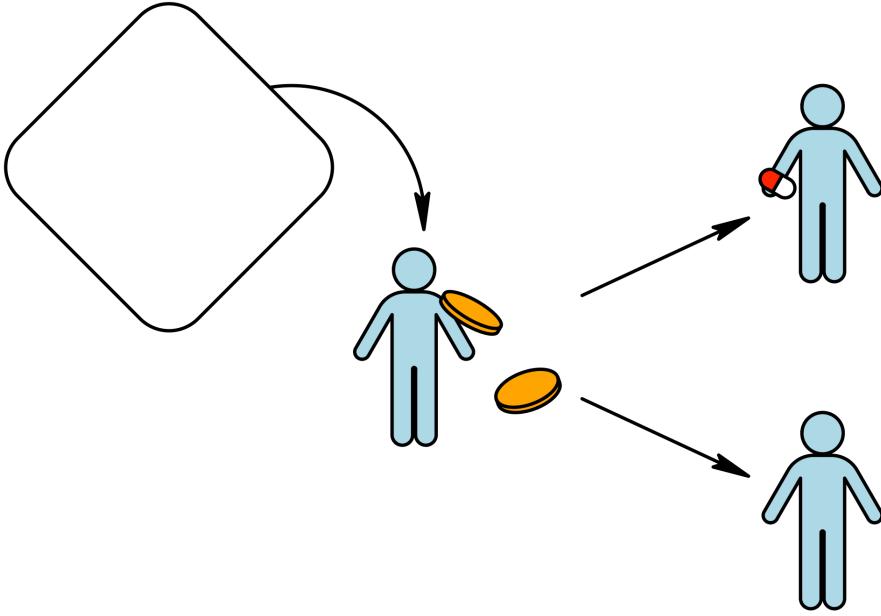


# Solution?

Instead of comparing the same individual between different counterfactual ("what-if") situations, ...

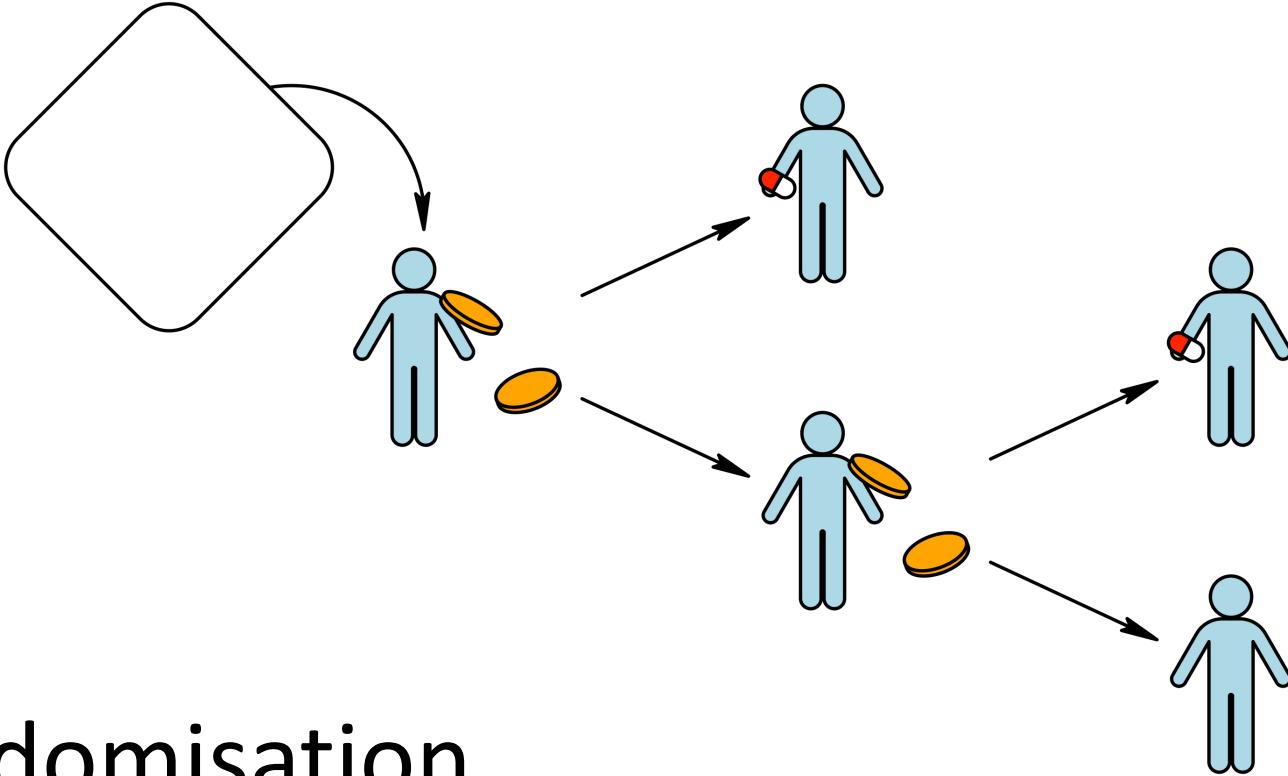


... compare *different individuals* who are *actually treated differently*



## Randomisation

Instead of subjecting exact copies of the same individual to different levels of treatment, with randomisation you get *differently treated individuals* whose characteristics – other than treatment and its consequences – are *identical in distribution*



## Randomisation

- Powerful (conceptual) tool
- Can accommodate all sorts of interventions  
(single or multiple time-point interventions, static or dynamic)

# Why not do trials?

- Expensive
- Unethical
- Impractical
- Untimely
- ...

# Target trial

A *hypothetical trial* that – if implemented – would readily allow us to answer our what-if question

- To help *communicate causal estimand* (because identification is “straightforward”)
- To *facilitate appraisal* of actual research designs (and *avoid methodological problems* with your study)

# Target trial emulation

*Explicit* attempt to address deviations from a target trial, given the (observational) study data at hand

Step 1. *Specify* target trial

Step 2. *Emulate* it!

# Step 1: specify target trial

What do you need to know to implement (and replicate) it?

Target trial	
Eligibility criteria	...
Treatment strategies	...
Outcome	...
Time zero and follow-up	...
Causal contrasts	...
Data analysis	...

# Example: identification from target trials should be straightforward

$$\begin{aligned} \text{ATE} &= E[Y^{A=1}] - E[Y^{A=0}] && \text{(the causal estimand)} \\ &= E[Y^{A=1} | A = 1] - E[Y^{A=0} | A = 0] && \text{(randomisation} \Rightarrow \\ &&& \text{exchangeability, i.e., } A \text{ indep. of } Y^{A=a} \text{ for } a = 0, 1; \\ &&& \text{conditionals are defined only under positivity,} \\ &&& \text{which too is controlled by design)} \\ &= E[Y | A = 1] - E[Y | A = 0] && \text{(consistency, i.e., } Y^{A=a} = Y \text{ if } a = A) \end{aligned}$$

# Target trial emulation vs “silly” questions

Formulating a target trials helps to communicate the causal estimand and helps to avoid asking “silly” questions (about ill-defined or irrelevant interventions)

- Eligibility defined by post-baseline events
- Causal effect of (a reduction/increase in) BMI?
- “Does water kill?” (Hernán, Ann Epidemiol., 2016;26(10):674–680)

# Treatment-variation (ir)relevance and well-definedness

- There may be many variations on an intervention and their impact on the outcome of interest need not be the same
- Interventions are sufficiently **well-defined** if there is no ambiguity about the variation or all possible variations equally affect the outcome variables of interest (i.e., there is **treatment-variation irrelevance**)
- Prerequisite of consistency

*Be explicit and precise (enough)!*

## Step 2: emulate target trial

Compare and address departures from target trial (analytically)

	<b>Target trial</b>	<b>Emulation study</b>
Eligibility criteria	...	...
Treatment strategies	...	...
Outcome	...	...
Time zero and follow-up	...	...
Causal contrasts	...	...
Data analysis	...	...

# Example: do statins prevent cancer?



## Avoidable flaws in observational analyses: an application to statins and cancer

Barbra A. Dickerman<sup>1\*</sup>, Xabier García-Albéniz<sup>1,2</sup>, Roger W. Logan<sup>1</sup>, Spiros Denaxas<sup>3,4,5</sup> and  
Miguel A. Hernán<sup>1,6,7</sup>

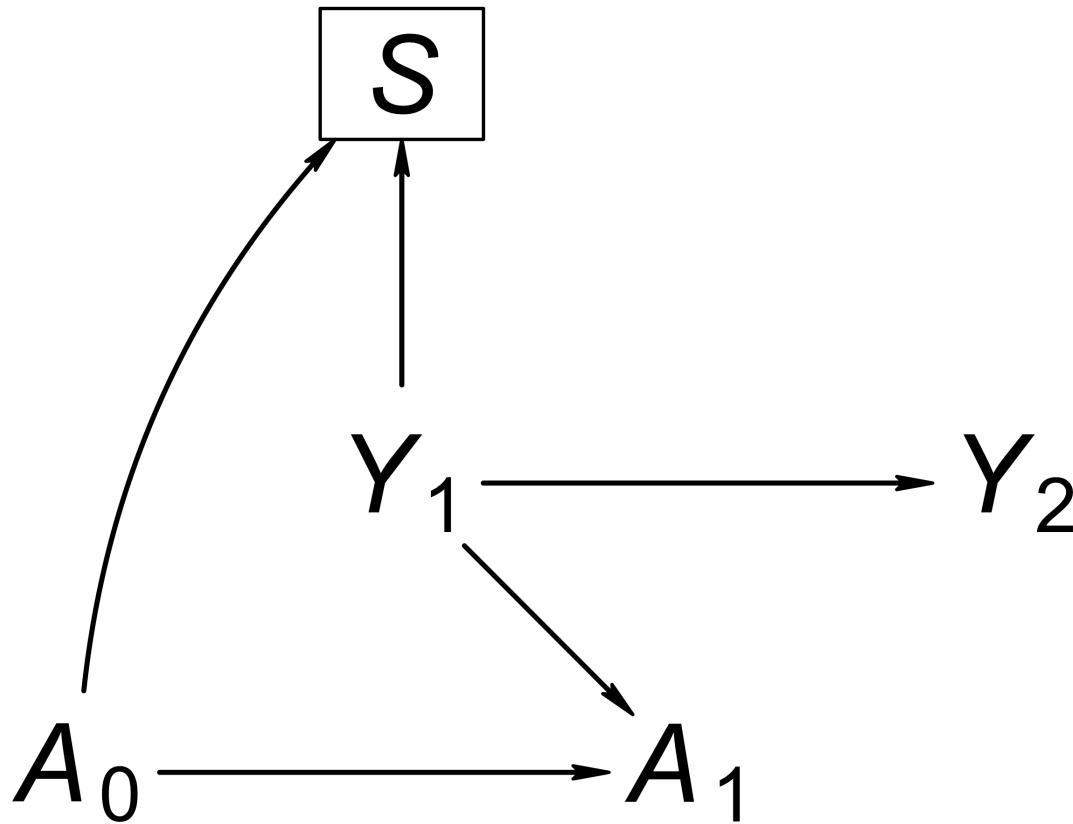
**Table 2—Effect of Statin Therapy and Its Duration on the Odds for Lung Cancer**

Variables	Cases, No.	Controls, No.	Crude OR	95% CI	Adjusted OR*	95% CI*	p Value for Adjusted OR
Overall	7,280	476,453					
Not exposed to statins	5,286	314,785					
Exposed to statins before lung cancer diagnosis (statin use > 0 yr)	1,994	161,668	0.73	0.70–0.77	0.55	0.52–0.59	< 0.01
Duration of statin use, yr							
0–0.5	446	10,259	2.59	2.34–2.86	2.32	2.05–2.63	< 0.01
0.5–1.0	214	15,564	0.82	0.71–0.94	0.75	0.63–0.89	< 0.01
1.0–2.0	416	30,590	0.81	0.73–0.90	0.70	0.61–0.79	< 0.01
2.0–4.0	649	55,516	0.70	0.64–0.76	0.49	0.44–0.55	< 0.01
> 4.0	269	49,739	0.32	0.28–0.36	0.23	0.20–0.26	< 0.01

\*Adjusted for effects of age, race, sex, BMI, smoking, alcohol use, and diabetes.

# Immortal time bias

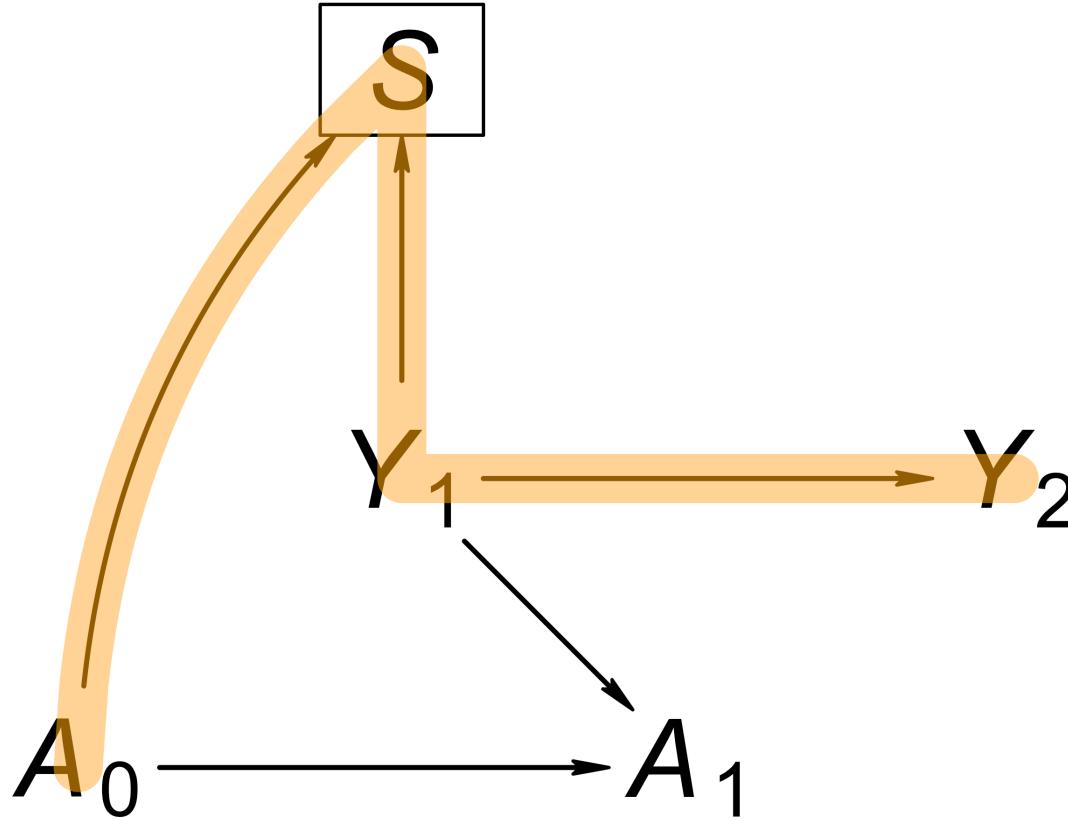
- **Immortal time** – a period of follow-up during which death or the study outcome cannot occur *by design*
- Arises from using postbaseline information to define (1) inclusion/eligibility/selection (selection) or (2) the exposure/contrast – *against trial principles!*
- May result in bias depending on *how it is handled!*
- Key to depicting this in a *DAG* is to include time-specific instances of variables



## Immortal time bias by selection

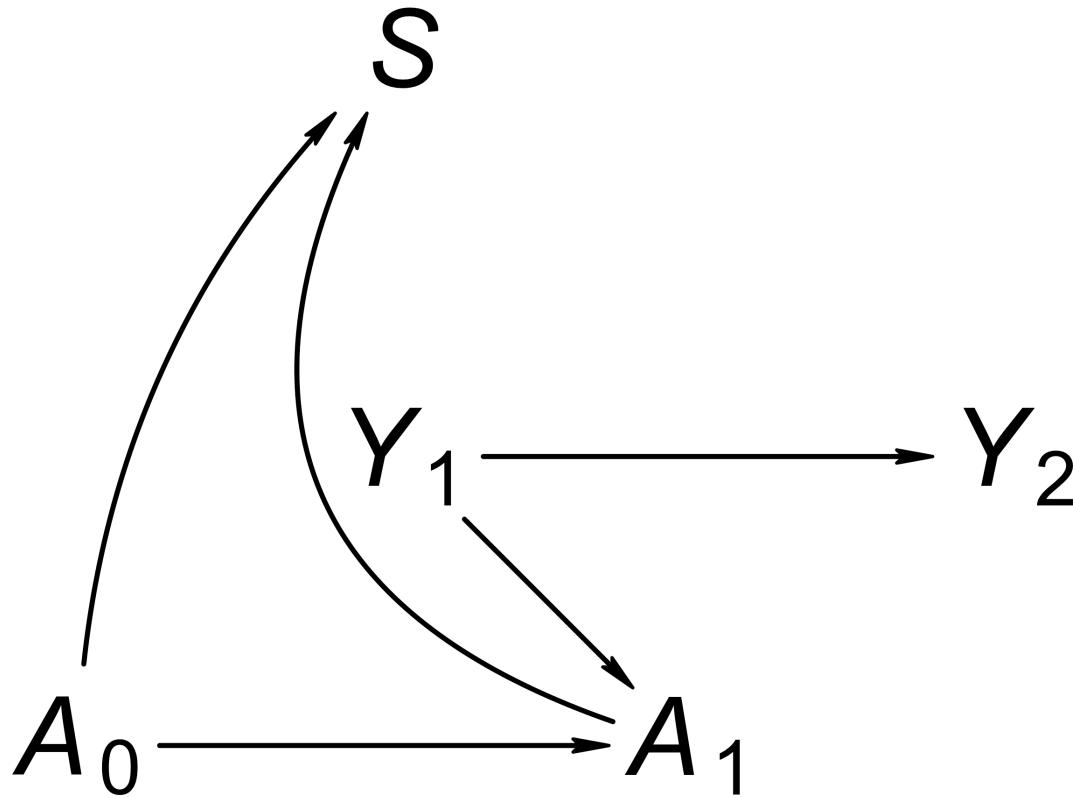
Are groups defined by  $A_0$  exchangeable relative to outcome  $Y_2$  conditional on  $S=1$ ? *Hint: use the backdoor criterion!*





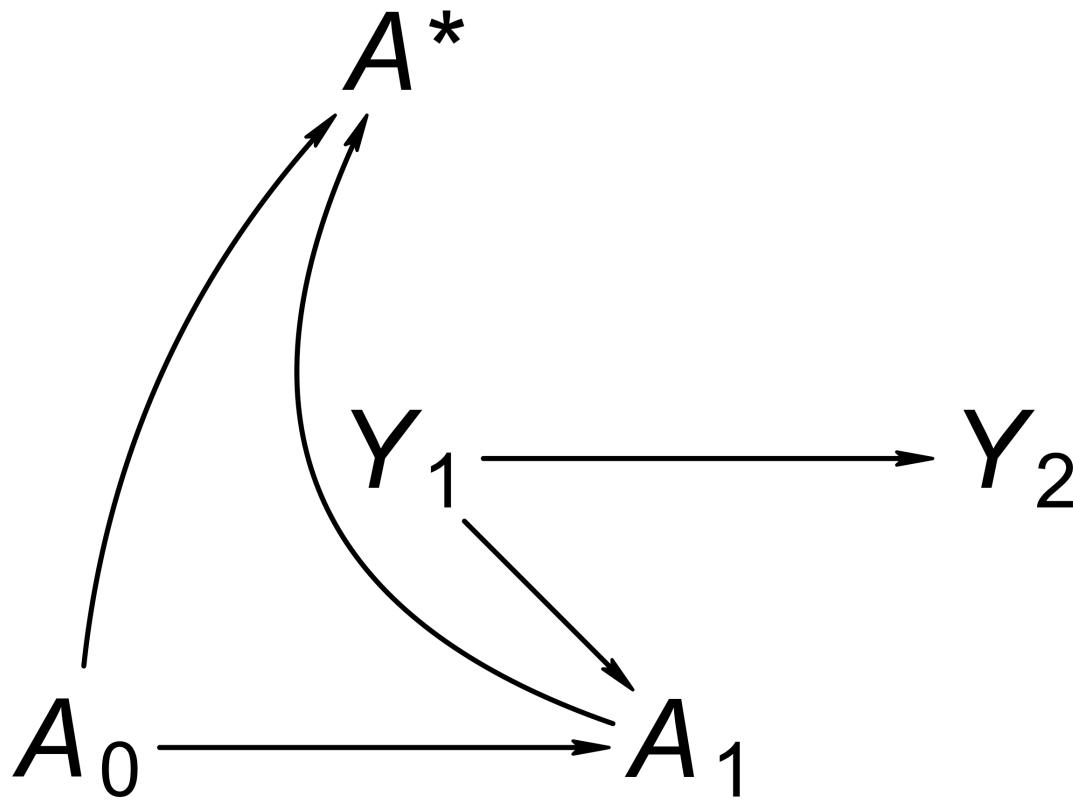
## Immortal time bias by selection

- $S$  is a descendant of  $A_0$  – violation of backdoor criterion!
- $A_0$  and  $Y_2$  are marginally independent (d-separated) but not necessarily conditional on  $S=1$ !



## Immortal time bias by selection (2)

- $S$  is a descendant of  $A_0$  – violation of backdoor criterion!
- $A_0$  and  $Y_2$  are marginally independent (d-separated) but not necessarily conditional on  $S=1$ !



## Immortal time bias by misclassification

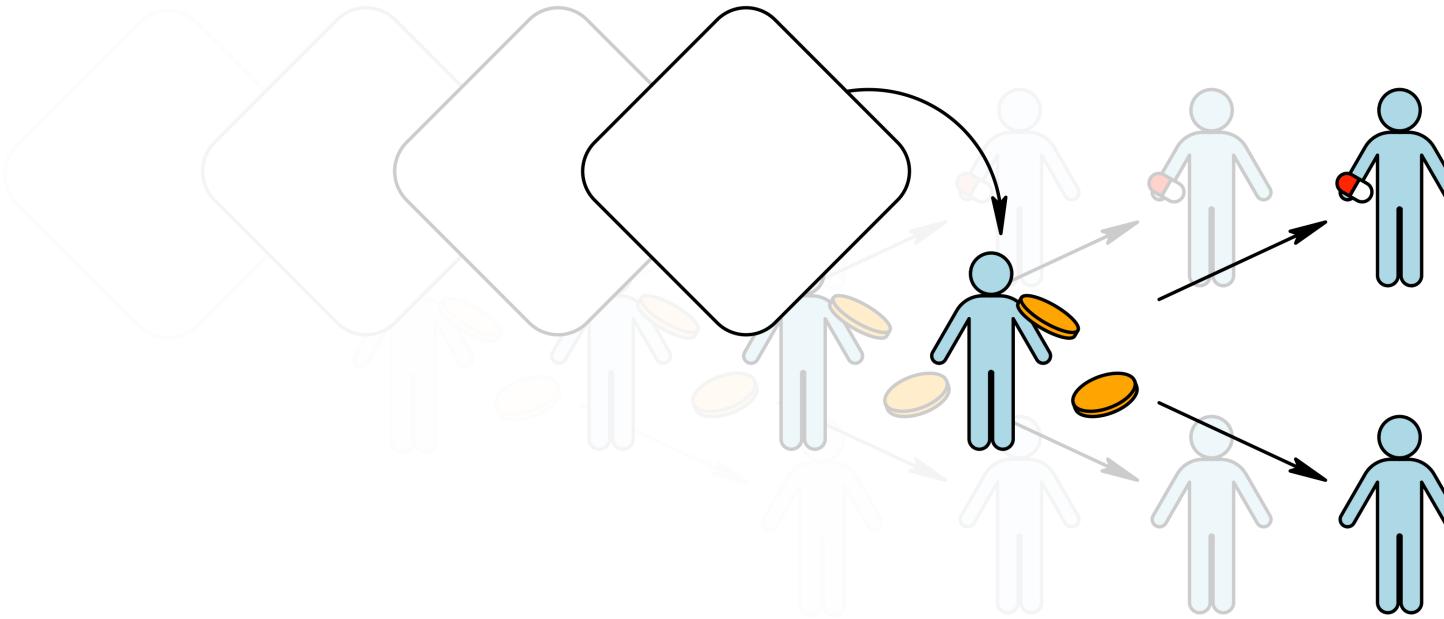
- Eg: compare surgery with wait-time vs no surgery
- Immortal time bias can arise when we include everyone but make the wrong contrast (surgery actually received vs not)

# Including prevalent users to study effects incident use

- Would you consider *initiating* a treatment regime now (at baseline) for a patient who is already on treatment (*prevalent user*)?
- Prevalent users are not part of the target population!
- Inclusion might result in (selection) bias!

# Too few incident users at any given time?

- In studies on the effect of statin use and cancer incidence, there are *few* incident users at specified time baseline  $t_0$  (or in short period starting at  $t_0$ )
- Consider trials that are identical except for their baseline time
- To gain efficiency, could emulate *multiple* such trials and analyse simultaneously (possibly according to flexible modelling assumptions to reflect heterogeneity across trials)
- NB: because individuals can be eligible for randomisation in multiple trials, need to respect *clustering* in estimating standard errors and constructing confidence intervals!



# Sequential trial emulation

# Statin-cancer example revisited

- Dickerman et al. (Nat Med, 2019;25(10):1601-1606):
  - When applying trial principles to analyse observational data (emulating a trial), they found effect estimates close to null
  - When reanalysing the observational data using the same approach as in earlier analyses, they found effect estimates similar to those found in earlier observational studies
- Discrepancies between trials and observational studies are often attributable largely to sources of bias other than residual confounding!

# Addressing departures from randomisation

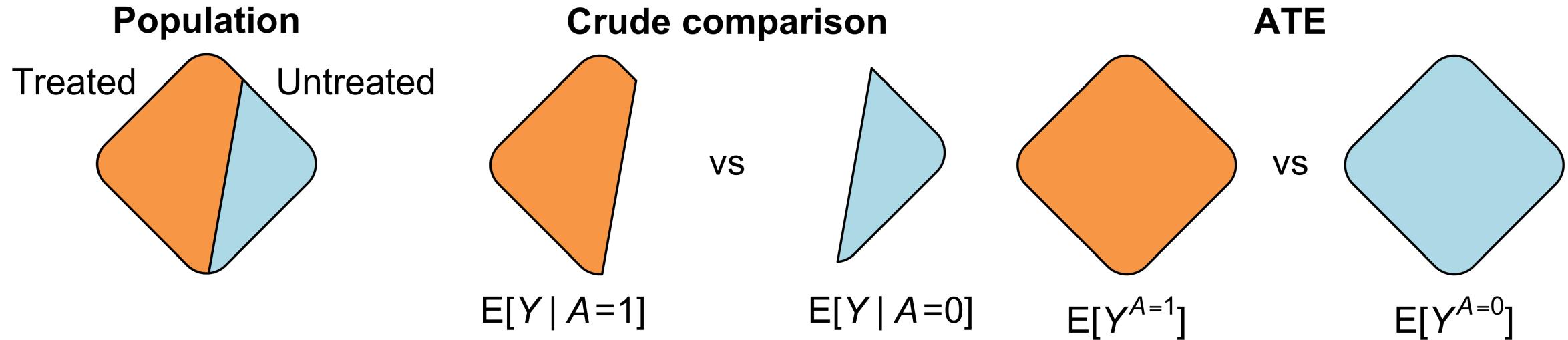
Any method may be used for confounding control

- Restriction
- Matching
- Regression adjustment
- G-computation
- Inverse probability weighting (IPW)
- ...

Choice should be influenced in part by estimand

- G-methods (IPW and g-computation) target quantities typically estimated in trials





## IPW for binary treatments

- Typical estimand is *average treatment effect* (ATE)
- Key idea: reweight the treated and untreated subpopulations so that they look (in some respects) like the entire population

# Goal of IPW for binary treatments

Create a **pseudopopulation** by weighting the original population such that

distribution of  $(L, Y^{A=a})$  in the (original) population

=

distribution of  $(L, Y^{A=a})$  in the pseudopopulation  
among both the treated and among the untreated

This means that, in the pseudopopulation:

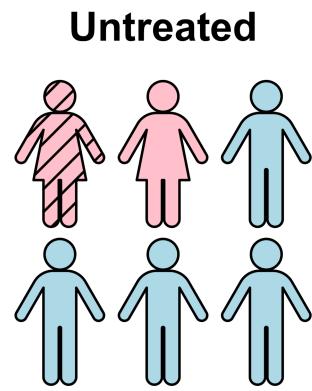
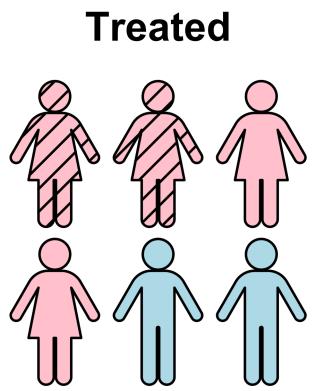
- $L$  and  $Y^{A=a}$  independent of  $A$  (exchangeability, as in RCT)
- Causal effect = crude association, under consistency

# IPW for binary treatments: how?

The pseudopopulation with this property can be made under exchangeability and positivity given  $L$ , by weighting

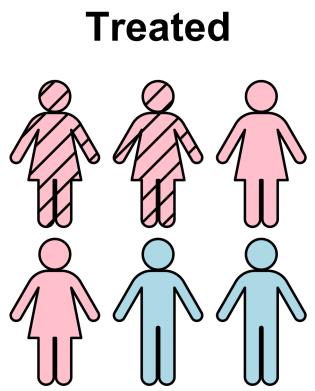
- treated subjects with  $\frac{1}{\text{ps}(L)}$
- untreated subjects with  $\frac{1}{1 - \text{ps}(L)}$

where  $\text{ps}(L)$  is the propensity score  $\Pr(A = 1 \mid L)$

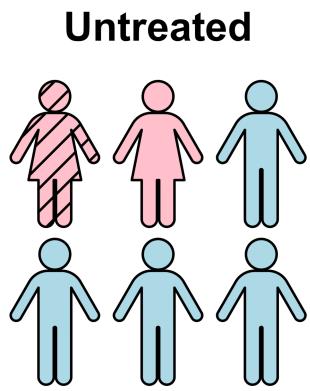
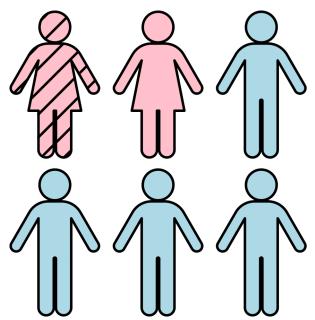


Sex	Treated	Propensity for treatment	Weight
Female	No	?	?
Female	Yes	?	?
Male	No	?	?
Male	Yes	?	?

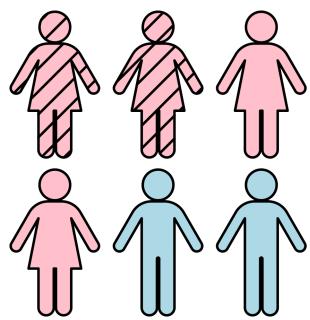
IPW: example



+



+



Sex	Treated	Propensity for treatment	Weight
Female	No	4/6	3.0
Female	Yes	4/6	1.5
Male	No	2/6	1.5
Male	Yes	2/6	3.0

IPW: example

# IPW: extensions

- For inference about average effect among the treated (ATT)
- To accommodate more than two treatment levels
- Point interventions with more than two levels or interventions on time-varying variables
- But with many possible treatment regimes to consider: there may be a need to make modelling assumptions regarding the distribution of the outcome and the treatment(s) in the pseudopopulation (marginal structural modelling)
- ‘Stabilisation’ of weights recommended (to reduce their variability)
- To address selective censoring or account for missing data

# G-methods

- Inverse probability weighting
- G-computation
- G-estimation

# G-computation for binary treatments

- Alternative to IPW that comes under same (non-parametric) identifiability conditions, for the same estimand
- Relies on outcome modelling rather than treatment modelling
- Essentially “standardisation” of conditional quantities

Identification of marginal causal mean

$$\begin{aligned} E[Y^{A=a}] &= E\{E[Y^{A=a} \mid L]\} && \text{(Law of iterated expectations)} \\ &= E\{E[Y^{A=a} \mid A=a, L]\} && \text{(Conditional exchangeability} \\ &&& \quad + \text{positivity)} \\ &= E\{E[Y \mid A=a, L]\} && \text{(Consistency)} \end{aligned}$$

# The gist of g-estimation (for binary treatments)

- Postulate model for contrast between treatments within levels of covariates; eg,  $E[Y^{A=a} | A=a, L] - E[Y^{A=0} | A=a, L] = \beta a$
- Express  $E[Y^{A=0} | L]$  in terms of factuals under consistency;  
 $E[Y^{A=0} | A=a, L] = E[Y - \beta A | A=a, L]$
- *Search for model parameters that are compatible with conditional exchangeability;*  $E[Y^{A=0} | A, L] = E[Y^{A=0} | L]$ , so search for  $\hat{\beta}$  that renders  $A$  independent of  $E[Y - \hat{\beta} A | A, L]$  given  $L$



---

## Education Corner

# An introduction to g methods

**Ashley I Naimi<sup>1</sup>\*, Stephen R Cole<sup>2</sup> and Edward H Kennedy<sup>3</sup>**

<sup>1</sup>Department of Epidemiology, University of Pittsburgh, <sup>2</sup>Department of Epidemiology, University of North Carolina at Chapel Hill and <sup>3</sup>Department of Statistics, Carnegie Mellon University

\*Corresponding author. Department of Epidemiology University of Pittsburgh 130 DeSoto Street 503 Parran Hall Pittsburgh, PA 15261 ashley.naimi@pitt.edu

Accepted 17 October 2016

### Abstract

Robins' generalized methods (g methods) provide consistent estimates of contrasts (e.g. differences, ratios) of potential outcomes under a less restrictive set of identification conditions than do standard regression methods (e.g. linear, logistic, Cox regression). Uptake of g methods by epidemiologists has been hampered by limitations in understanding both conceptual and technical details. We present a simple worked example that illustrates basic concepts, while minimizing technical complications.

# Extentions of g-methods to time-varying treatment settings

- IPW, g-computation and g-estimation suitable to estimate causal effects of time-varying treatment effects
- ... more later in this course!

## Statin-cancer example revisited (2)

Previous studies implicitly compared long-term statin users versus non-users – doesn'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?

More on estimating “per-protocol” effects later ...

# Summary

Target trial emulation = *explicit* attempt to address deviations from a target trial, given the (observational) study data at hand

Step 1. *Specify* target trial

Step 2. *Emulate* it!