



#LEEDSCAUSALSCHOOL

Leeds Spring School in Causal Inference with Observational Data

GEORGIA



@GEORGIATOMOVA

MARK



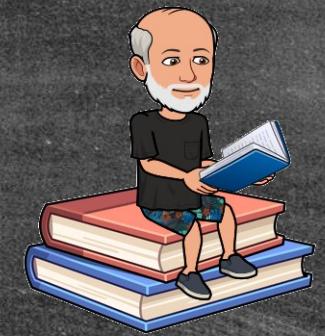
@STATSMETHODS

PETER



@PWGTENNANT

ROB



@WAYNEROBERTLONG

09:00-09:30 REGISTRATION

09:30-10:00 WELCOME

10:00-10:30 LECTURE 1.1

10:30-11:00 DELEGATE INTRO

11:00-11:30 TEA & COFFEE

11:30-12:00 DELEGATE INTRO

12:00-12:45 LECTURE 1.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:15 LECTURE 1.3

15:15-15:30 Q&A

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 1.4

17:00-17:45 ACTIVITY 1-A

17:30-18:00 Q&A

1.1 - THE NEED FOR A CAUSAL FRAMEWORK

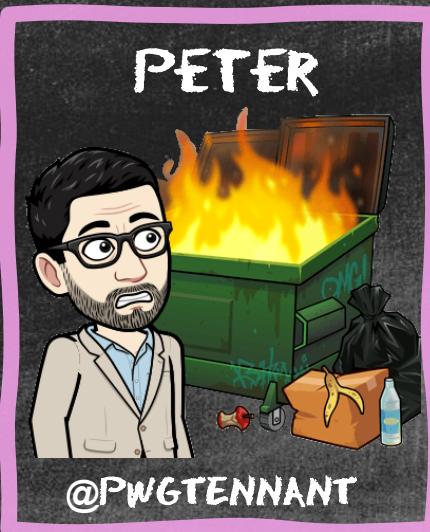
GEORGIA



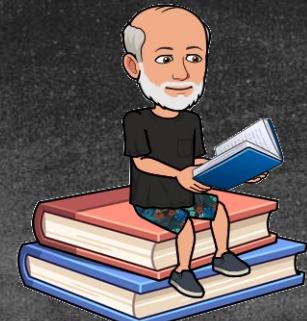
MARK



PETER



ROB



@GEORGIATOMOVA

@STATSMETHODS

@PWTENNANT

@WAYNEROBERTLONG

PROGRAMMED FOR CAUSAL INFERENCE

We learn about **cause-and-effect** from the moment we're born

And we learn to **associate** objects and actions with consequences

Passively: by observation & experience



Actively: by trial & interaction



PROGRAMMED FOR CAUSAL INFERENCE

In time, we each build a **mental model of the world**

This becomes so sophisticated that we can ask (**counterfactual**) questions about
'what might have been?' or 'how things could be different?'

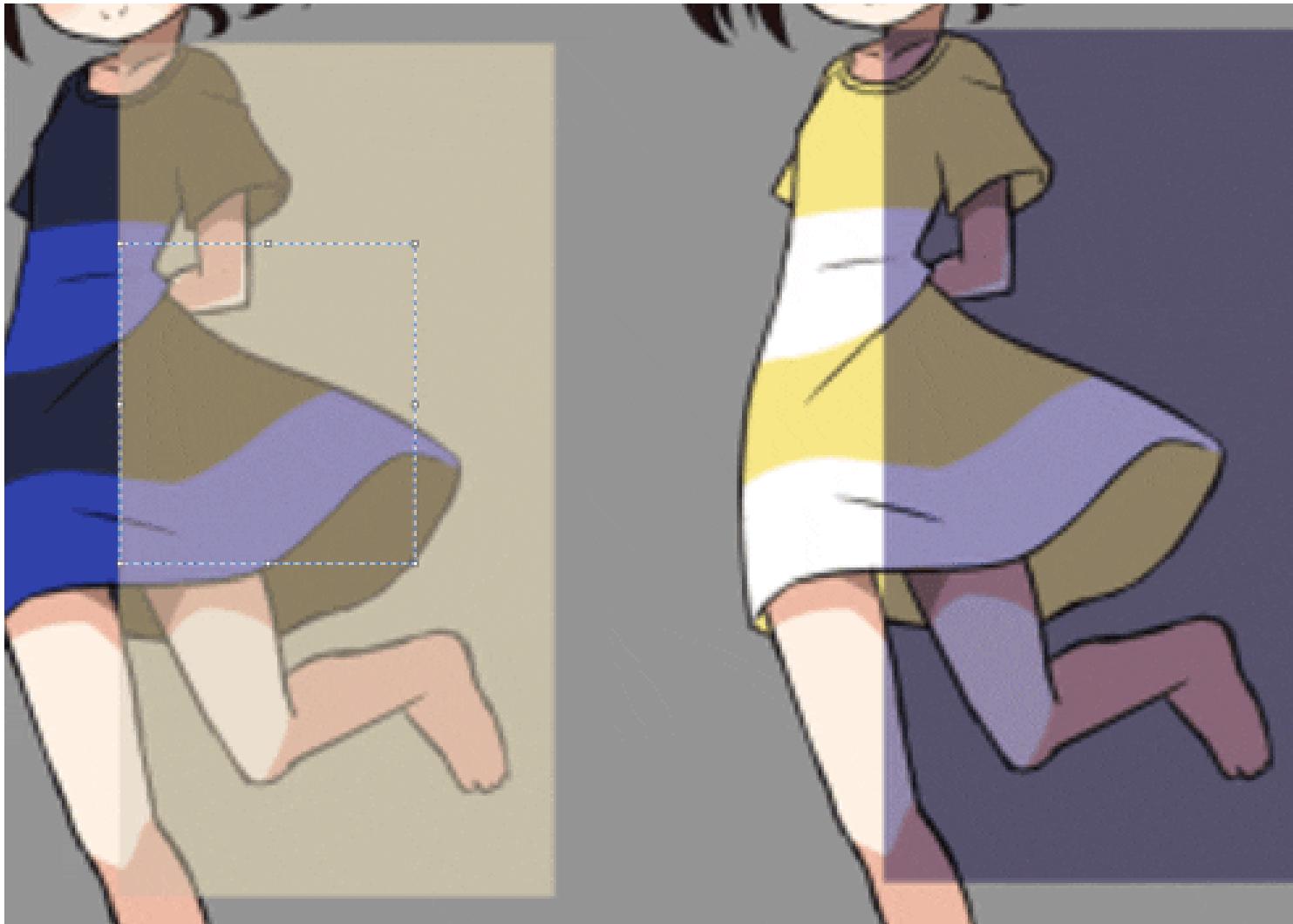
That imagination has helped us to change our world beyond recognition



EASILY FOOLED



EASILY FOOLED



CAUSAL MISINFERENCE

- We are **over-tuned** to make causal inferences
- Our intuition struggles to distinguish **correlation** from **causation**

**NOT A
HEAVY
SEAGULL!**



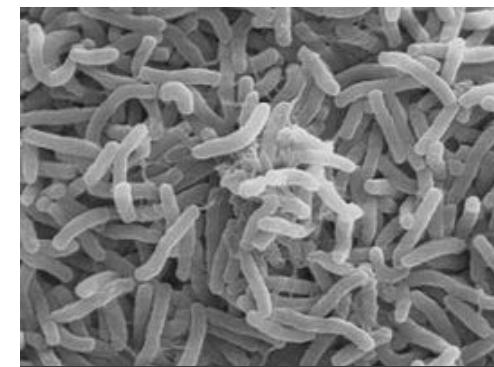
CAUSAL MISINFERENCE

- This is particularly true when faced with situations that don't fit our (**deterministic**) childhood model of cause-and-effect...



HEALTHY

+



VIBRIO CHOLERAE



UNHEALTHY

✓ Deterministic

CAUSAL MISINFERENCE

- This is particularly true when faced with situations that don't fit our (**deterministic**) childhood model of cause-and-effect...



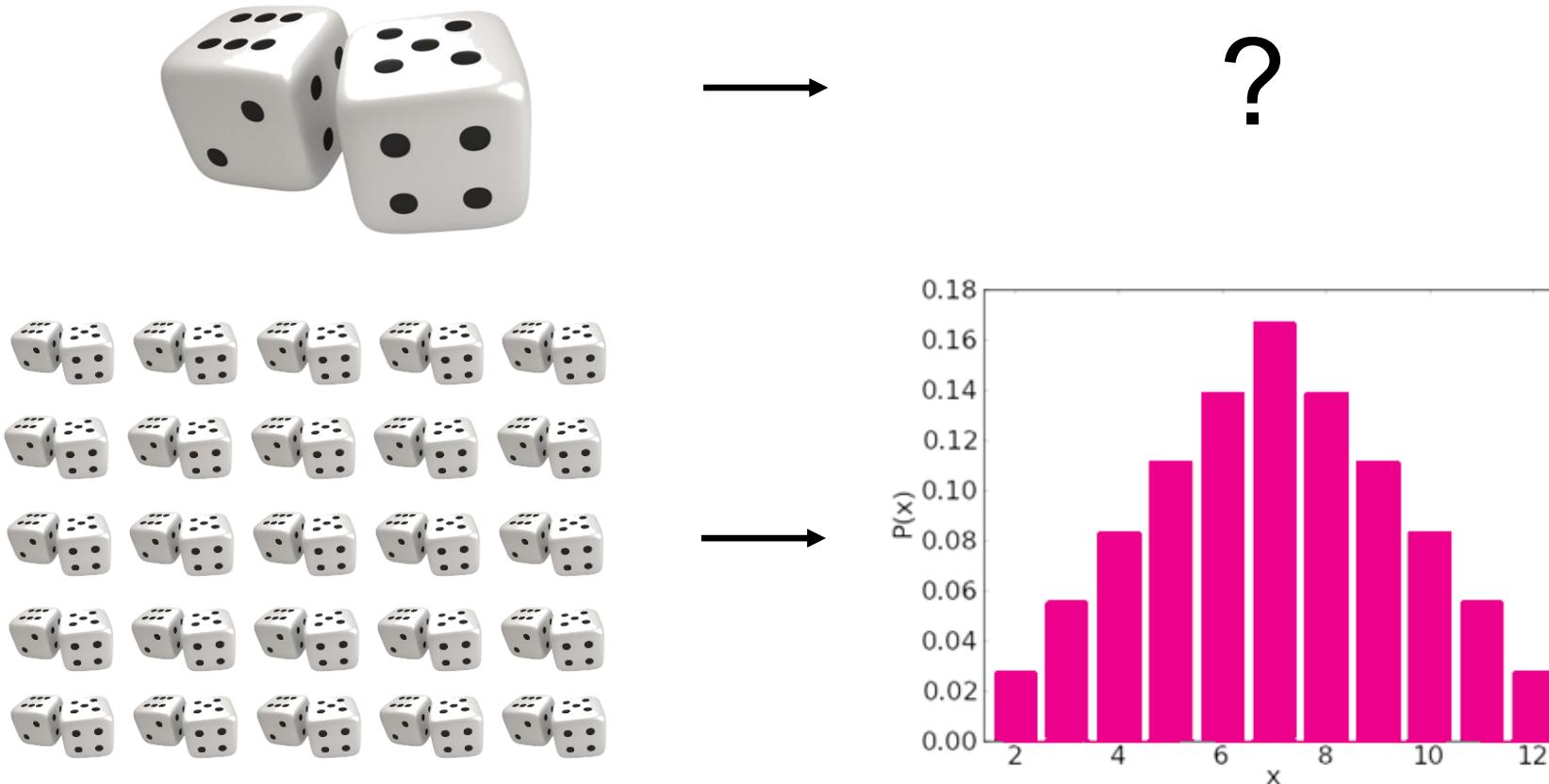
+



✖ Probabilistic

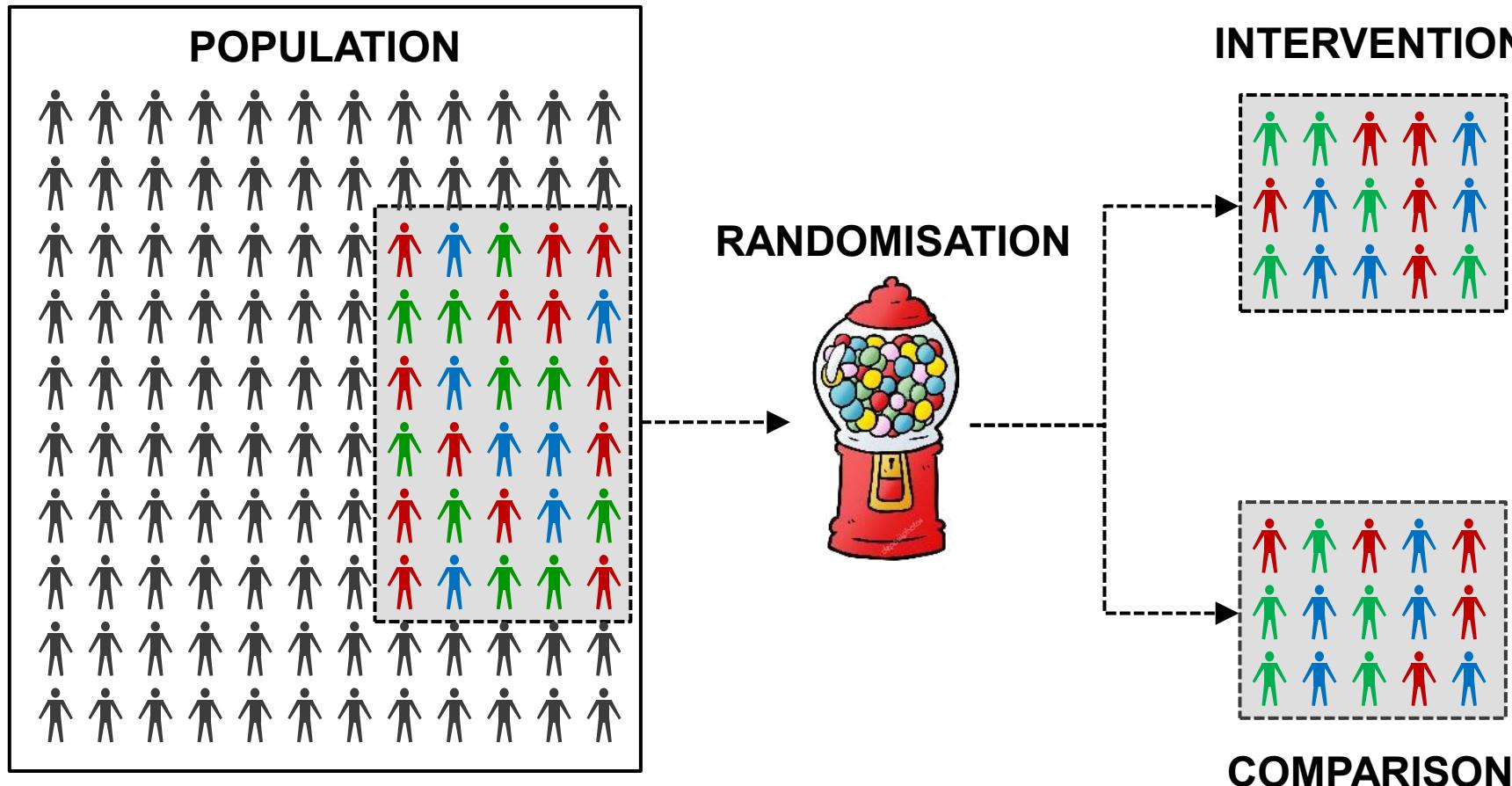
PROBABILISTIC REASONING

- **Statistics & probability** have helped with this problem
- What's unknowable for an individual... can be **predictable** for a group!



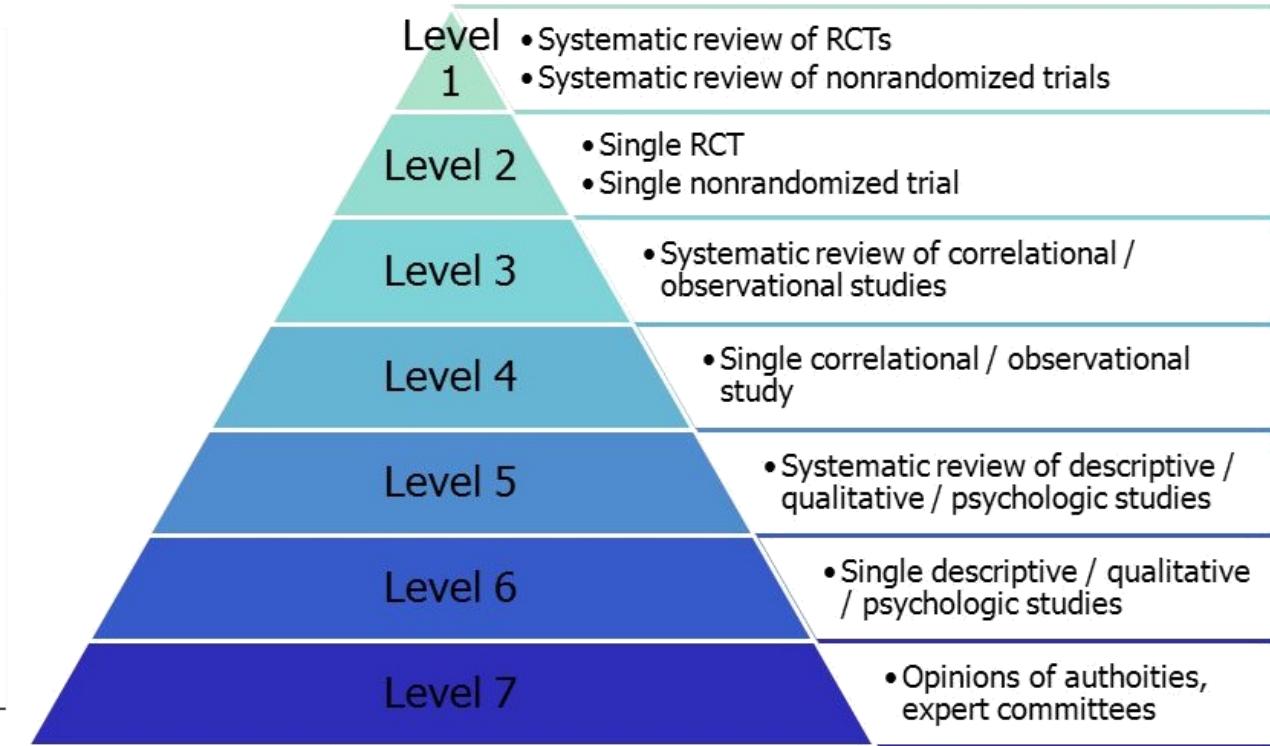
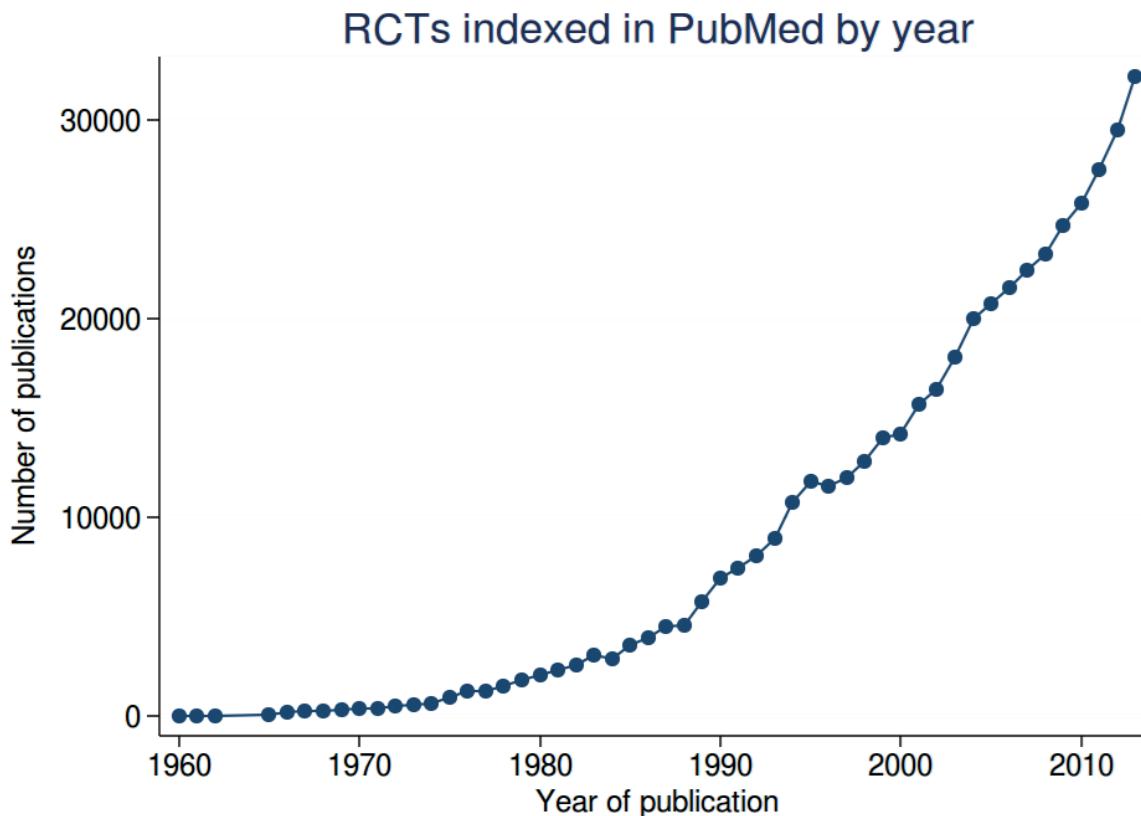
RANDOMISATION

- With **randomisation**, this provides a potent way to **estimate causal effects** that we'd have little/no hope using **intuition** alone



RANDOMISATION

- In health and medical research, **randomised controlled trials** are embraced with huge enthusiasm and canonised beyond all other forms of evidence



DIFFERENT DATA



- Without randomisation & a control group, analysts face a range of nasty biases:

- **Experimental data and observational data** may look very similar
- But understanding our data means understanding how it was generated



Confounding bias



Collider bias



Selection bias

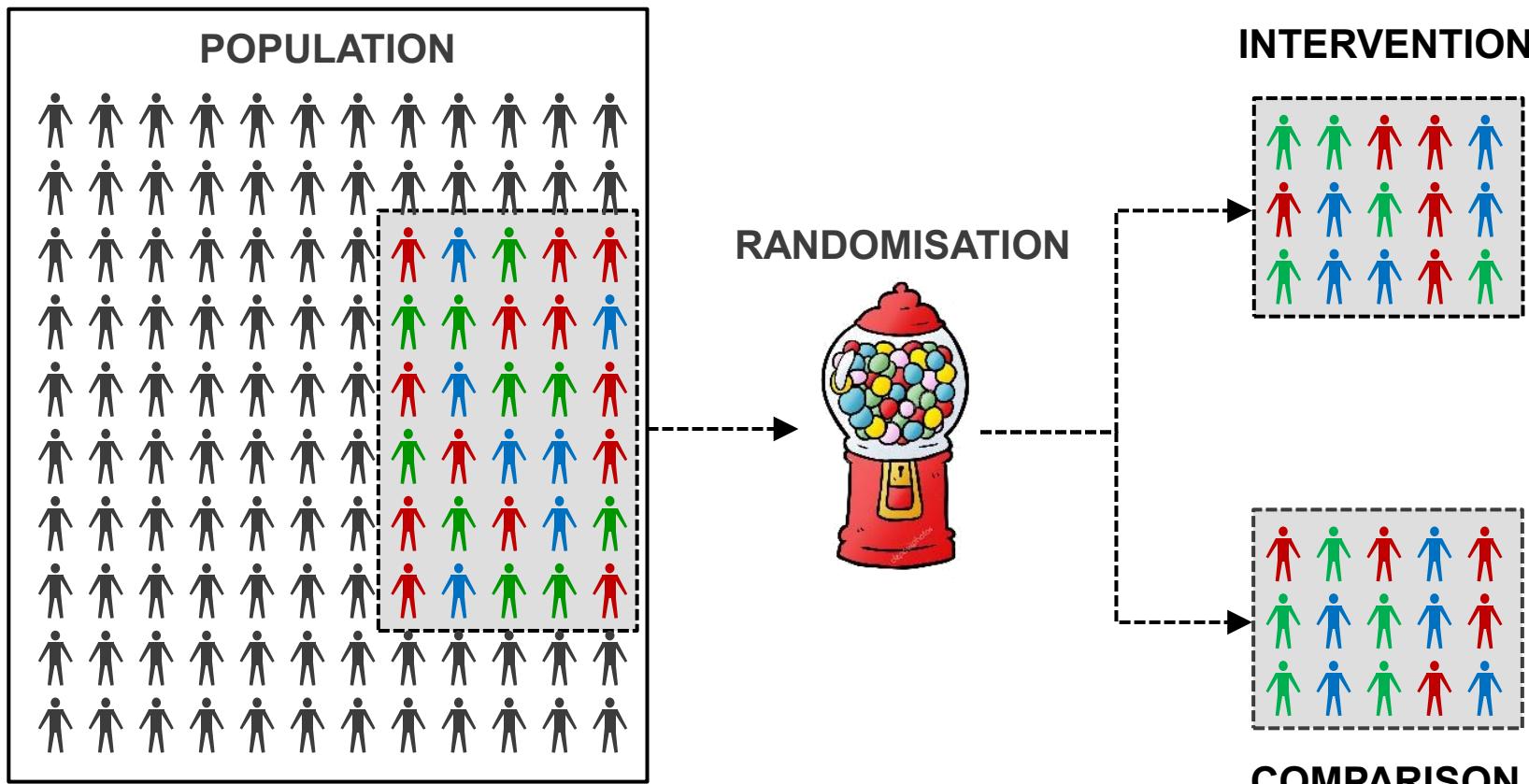


Inferential bias



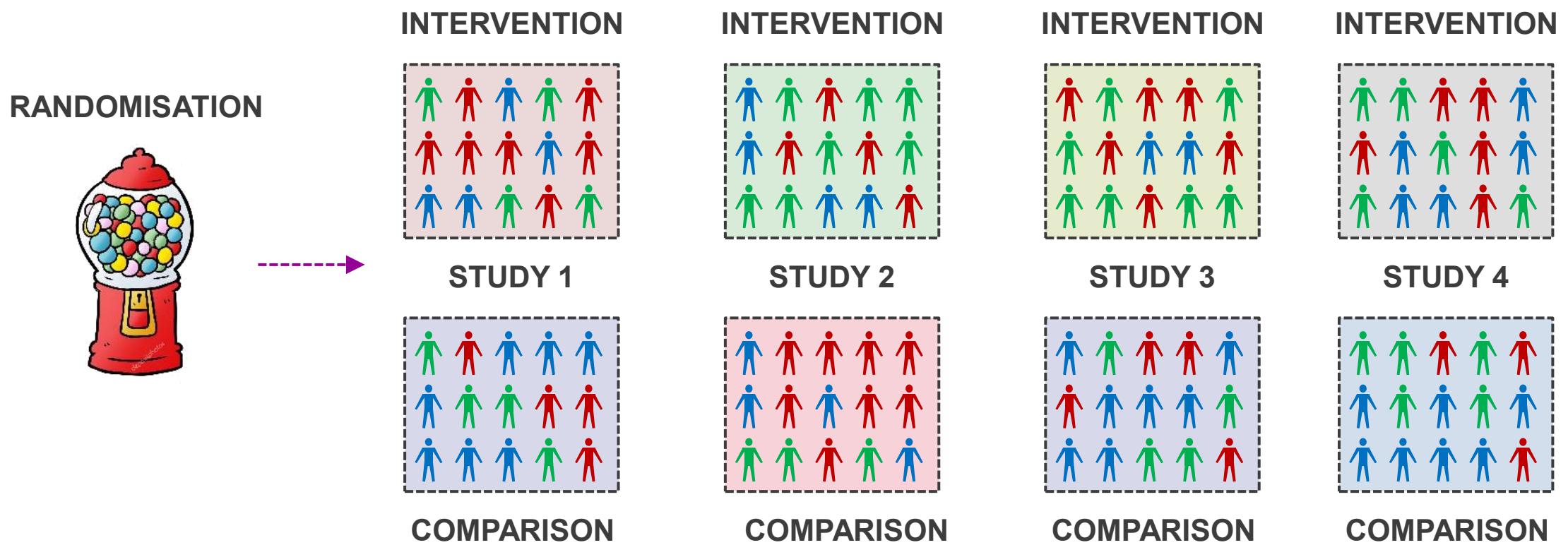
ANALYSING EXPERIMENT DATA

In **randomisation controlled experiments** we use **machinery of chance** to ensure a **fair balance** of people are assigned to different groups



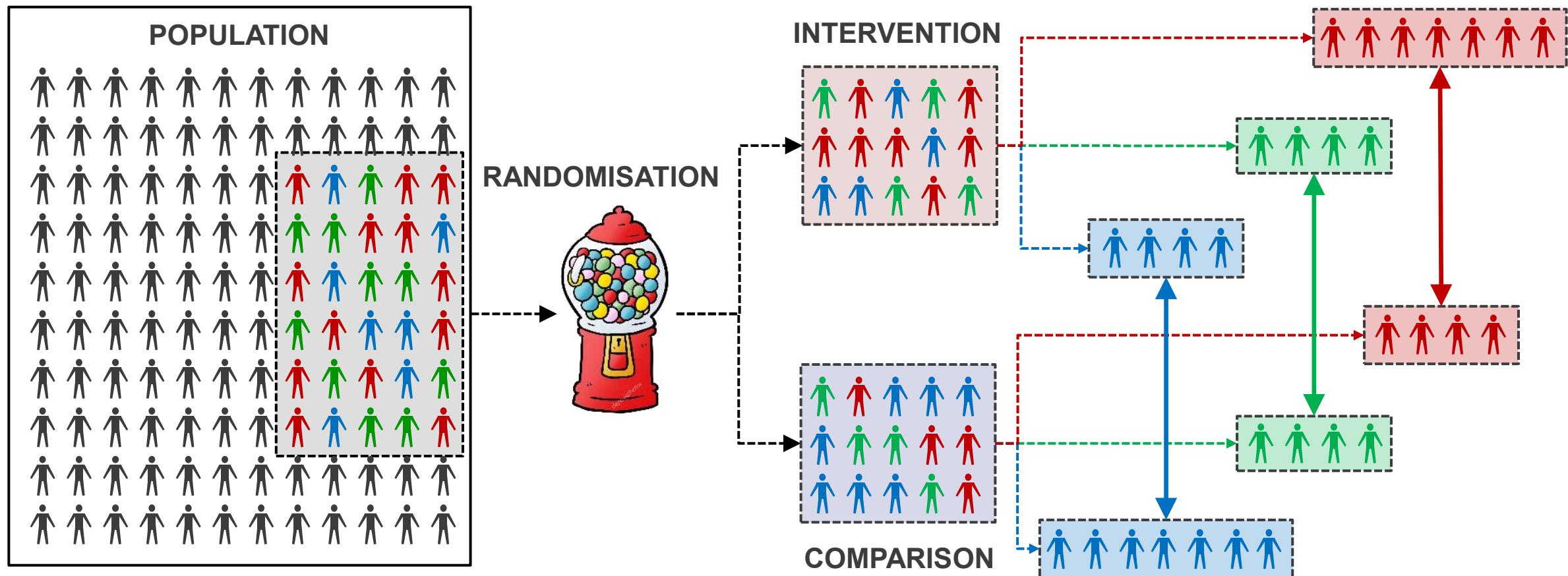
ANALYSING EXPERIMENT DATA

- Reliance on chance means groups aren't always comparable. They're '**imbalanced**' due to **random sampling variable**



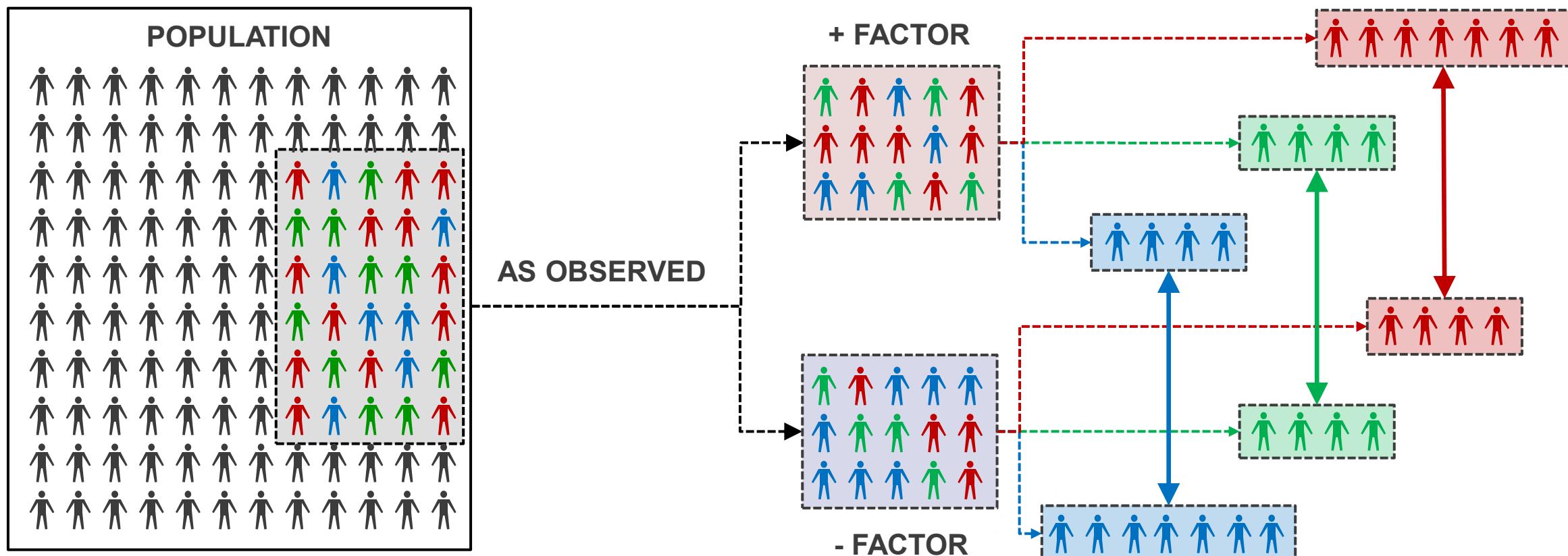
ANALYSING EXPERIMENT DATA

- Can reduce heterogeneity by making comparisons conditional on random differences (or ‘**imbalances**’), by e.g. **subgroup analysis** or **covariate adjustment**



ANALYSING OBSERVATIONAL DATA

- Seems logical to use same approach to analyse **non-experimental data**, and isolate the '**independent association**' of individual '**(risk) factors**' above (non-random) differences in '**case mix**'?



CAUSAL MISINFERENCE

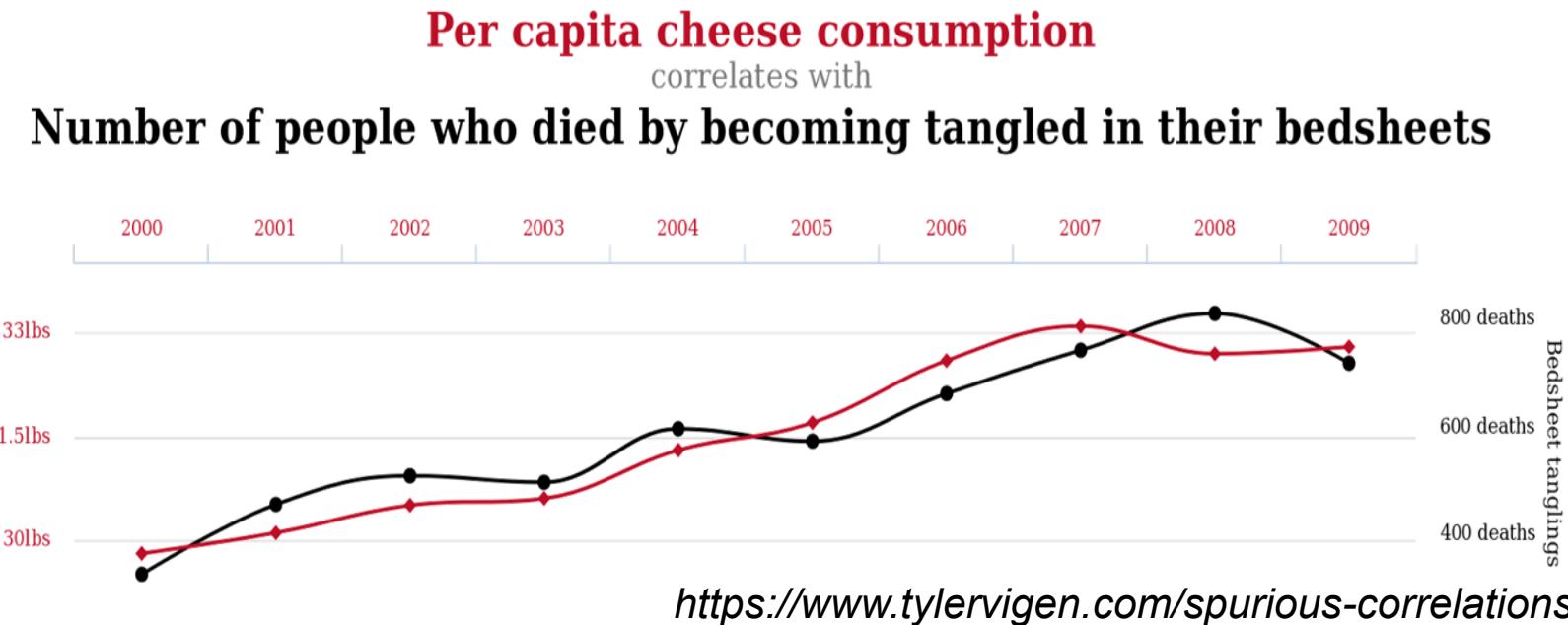
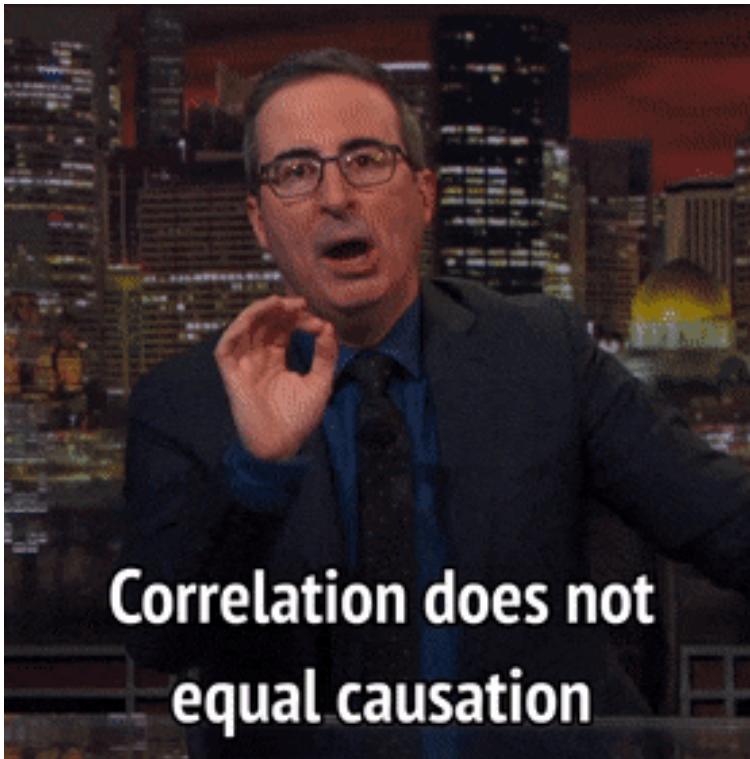
We have poor record of causal inference in observational data

12 RCTs studied **52 claims** from observational data

- **ZERO replicated and 5 found opposite !**

<i>ID no.</i>	<i>Pos.</i>	<i>Neg.</i>	<i>No. of claims</i>	<i>Treatment(s)</i>	<i>Reference</i>
1	0	1	3	Vit E, beta-carotene	NEJM 1994; 330 : 1029–1035
2	0	3	4	Hormone Replacement Ther.	JAMA 2003; 289 : 2651–2662, 2663–2672, 2673–2684
3	0	1	2	Vit E, beta-carotene	JNCI 2005; 97 : 481–488
4	0	0	3	Vit E	JAMA 2005; 293 : 1338–1347
5	0	0	3	Low Fat	JAMA. 2006; 295 : 655–666
6	0	0	3	Vit D, Calcium	NEJM 2006; 354 : 669–683
7	0	0	2	Folic acid, Vit B6, B12	NEJM 2006; 354 : 2764–2772
8	0	0	2	Low Fat	JAMA 2007; 298 : 289–298
9	0	0	12	Vit C, Vit E, beta-carotene	Arch Intern Med 2007; 167 : 1610–1618
10	0	0	12	Vit C, Vit E	JAMA 2008; 300 : 2123–2133
11	0	0	3	Vit E, Selenium	JAMA 2009; 301 : 39–51
12	0	0	3	HRT + Vitamins	JAMA 2002; 288 : 2431–2440
Totals	0	5	52		<i>Young & Karr 2011 Significance, 116:120, 2011</i>

CORRELATION ≠ CAUSATION



CORRELATION ≠ CAUSATION

Discouraged from using causal language when interpreting observational data



Causal language (including use of terms such as effect and efficacy) should be used only for randomized clinical trials. For all other study designs (including meta-analyses of randomized clinical trials), methods and results should be described in terms of association or correlation and should avoid cause-and-effect wording.

Observational studies are algorithmically rated as '**low quality**'

Observational study →	Low
-----------------------	-----

Guyatt G, Oxman AD, Akl EA, et al. *J Clin Epidemiol* 2011;64:383–94

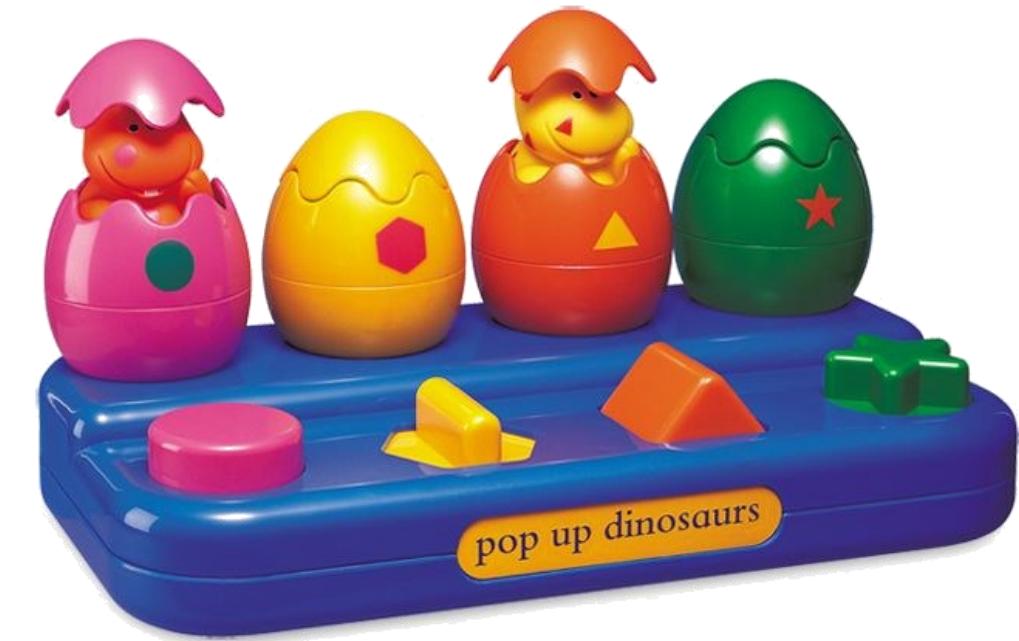
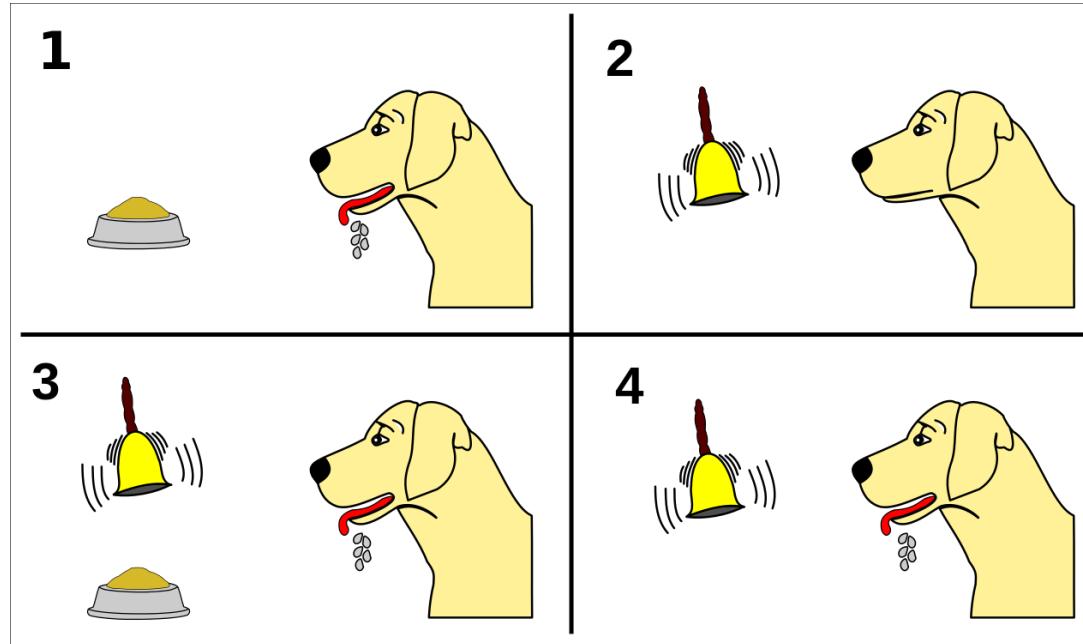
Scientists rely on **euphemisms**



“Risk factor” **“Correlation”**
“Link” **“Association”** **“Predictor”**

PAVLOVIAN INFERENCE

But: We are **programmed** to infer causality



And: **Spurious correlations** have very little **scientific** and **practical value**

People tend to infer causality regardless (**Pavlovian inference**)

SCHRÖDINGER'S CAUSAL INFERENCE

- **Often:** Draw causal inferences while claiming that causal inferences cannot be made (**Schrödinger's causal inference**)

5. Discussion

Among 14-year olds living in the UK, we found an association between social media use and depressive symptoms and that this was stronger for girls than for boys. The magnitude of these associations reduced when potential explanatory factors were taken into account, suggesting that experiences of online harassment, poorer sleep quantity and quality, self-esteem and body image largely explain observed associations. There was no evidence of differences for girls and boys in hypothesised pathways between social media use and depressive symptoms. **Findings are based largely on cross sectional data and thus causality cannot be inferred.**

Our findings add weight to the growing evidence base on the potential pitfalls associated with lengthy time spent engaging on social media. These findings are highly relevant to current policy development on guidelines for the safe use of social media and calls on industry to more tightly regulate hours of social media use for young people [[10], [11]]. Clinical, educational and family settings are all potential points of contact whereby young people could be encouraged to reflect not only on their social media use but also other aspects of their lives including online experiences and their sleep patterns. For instance, in the home setting all family members could reflect on patterns of use and have in place limits for time online, curfews for use and the overnight removal of mobile

Kelly Y et al. Social media use and adolescent mental health: findings from the UK Millennium Cohort Study. EClinicalMedicine 2019.

SCHRÖDINGER'S CAUSAL INFERENCE

Cold
Spring
Harbor
Laboratory

BMJ Yale

Causal and Associational Language in Observational Health Research: A systematic evaluation

✉ Noah A. Haber, ✉ Sarah E. Wieten, ✉ Julia M. Rohrer, ✉ Onyebuchi A. Arah, ✉ Peter W.G. Tennant, ✉ Elizabeth A. Stuart, ✉ Eleanor J. Murray, ✉ Sophie Pilleron, ✉ Sze Tung Lam, ✉ Emily Riederer, ✉ Sarah Jane Howcutt, ✉ Alison E. Simmons, ✉ Clémence Leyrat, ✉ Philipp Schoenegger, ✉ Anna Boaman, ✉ Mi-Suk Kang Dufour, ✉ Ashley L. O'Donoghue, ✉ Rebekah Baglini, ✉ Stefanie Do, ✉ Mari De La Rosa Takashima, ✉ Thomas Rhys Evans, ✉ Daloha Rodriguez-Molina, ✉ Taym M. Alsalti, ✉ Daniel J. Dunleavy, ✉ Gideon Meyerowitz-Katz, ✉ Alberto Antonietti, ✉ Jose A. Calvache, ✉ Mark J. Kelson, ✉ Meg G. Salvia, ✉ Camila Olarte Parra, ✉ Saman Khalatbari-Soltani, ✉ Taylor McLinden, ✉ Arthur Chatton, ✉ Jessie Seiler, ✉ Andreea Steriu, ✉ Talal S. Alshihayb, ✉ Sarah E. Twardowski, ✉ Julia Dabrowskaj, ✉ Eric Au, ✉ Rachel A. Hoopsick, ✉ Shashank Suresh, ✉ Nicholas Judd, ✉ Sebastián Peña, ✉ Cathrine Axfors, ✉ Palwasha Khan, ✉ Ariadne E. Rivera Aguirre, ✉ Nnaemeka U. Odo, ✉ Ian Schmid, ✉ Matthew P. Fox

Haber et al. Causal and Associational Language in Observatioanl Health Research: A Systematic Evaluation.
MedRxiv 2021.

- Although few studies... declared an interest in estimating causal effects, the majority used language that... implied causality
- Although many studies used disclaimers warning readers against making causal inferences, an... interest in causality was apparent
- “Schrödinger’s causal inference”... is common in the observational health literature”

THE CONSEQUENCES

thebmj

BMJ 2015;351:h4596 doi: 10.1136/bmj.h4596 (Published 5 September 2015)

Page 1 of 6



CrossMark
Data updated

ANALYSIS

Increased mortality associated with weekend hospital admission: a case for expanded seven day services?

Nick Freemantle and colleagues discuss the findings of their updated analysis of weekend admissions and the implications for service design

Nick Freemantle *professor of clinical epidemiology and biostatistics*^{1,2}, Daniel Ray *professor of health informatics*^{2,3,4}, David McNulty *medical statistician*^{2,3}, David Rosser *medical director*⁵, Simon Bennett *director, clinical policy and professional standards*⁶, Bruce E Keogh *national medical director*⁶, Domenico Pagano *professor, cardiac surgery*^{2,7}

¹Department of Primary Care and Population Health, University College London, UK; ²Quality and Outcomes Research Unit, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; ³Department of Informatics, University Hospitals Birmingham NHS Foundation Trust; ⁴Farr Institute of Health Informatics Research, University College London; ⁵University Hospitals Birmingham NHS Foundation Trust; ⁶Medical Directorate, NHS England, London, UK; ⁷Department of Cardiothoracic Surgery, University Hospitals Birmingham NHS Foundation Trust

Freemantle et al 2015 BMJ 5;351

THE CONSEQUENCES



Health policy Michael White's political briefing

Quiet hospitals kill, but mindless union bashing can give us a nasty injury

Jeremy Hunt has every right to question NHS consultants' working practices but there are questions they should be allowed to ask him too



Hospital equipment is being underused. Photograph: Justin Paget/Justin Paget/Corbis

THE CONSEQUENCES



home > UK > society law scotland wales northern ireland education medi all

NHS

Doctors urge inquiry into Jeremy Hunt's NHS 'weekend effect' claims

Letter signed by doctors and scientists including Stephen Hawking accuses health secretary of misrepresenting evidence



Health secretary Jeremy Hunt has said 'there are 11,000 excess deaths because we do not staff our hospitals properly at weekends'. Photograph: Peter Byrne/PA

THE CONSEQUENCES

the guardian

home > politics election UK election world sport football opinion cult all

Health policy

Two deaths possibly linked to 'Hunt effect', study suggests

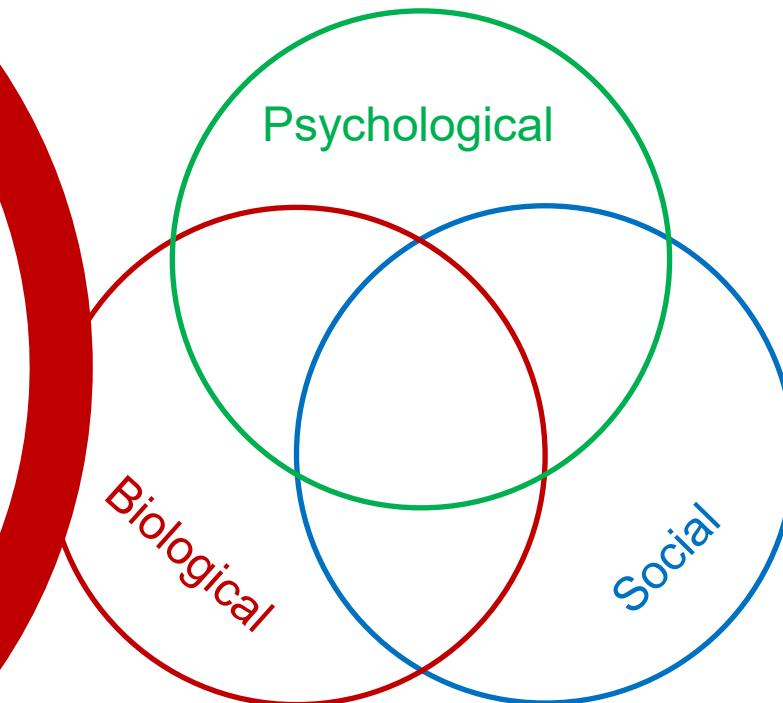
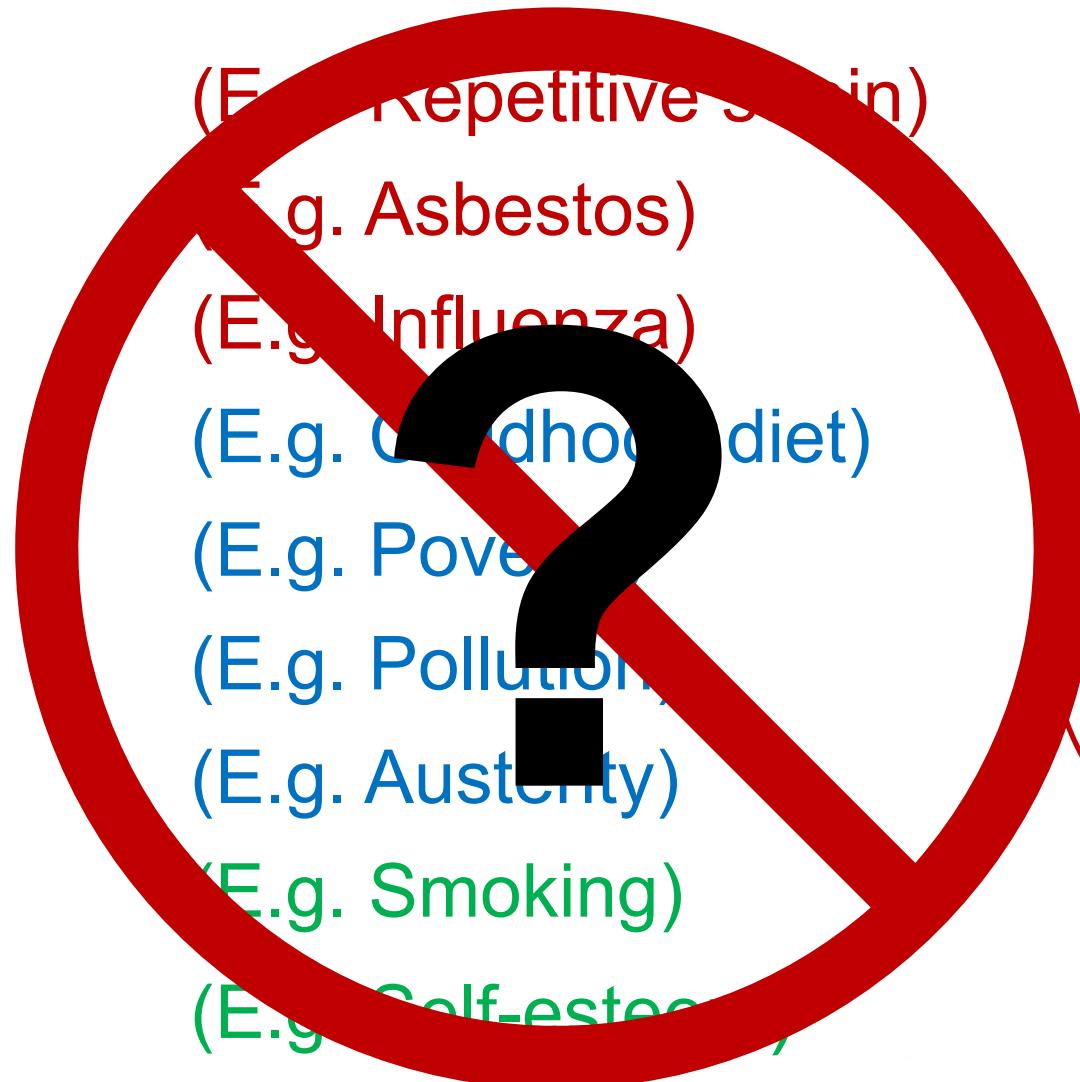
Research suggests some patients may be avoiding going to hospital at the weekend because of health secretary's statements about inadequate NHS staffing levels



Almost a third of the 40 patients studied suffered an increased chance of dying as a result of waiting to seek help. Photograph: Chris Radburn/PA

STOP DOING OBSERVATIONAL RESEARCH?

Physical
Chemical
Biological
Cultural
Economic
Environmental
Political
Behavioural
Personal



DO BETTER OBSERVATIONAL RESEARCH!

Accept & admit our causal ambitions

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

This commentary argues that using the term "causal" is necessary to improve the quality of observational research.

Miguel A. Hernán, MD, DrPH



See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiole Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

You know the story:

Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

Many journal editors request

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating "causal" as a dirty word that respectable investigators do not

glass of red wine per day alcohol drinking. For s disregard measurement random variability—th pose the 0.8 comes fro large population so tha confidence interval aro

The proscription against the C-word is harmful to science because causal inference is a core task of science, regardless of whether the study is randomized or nonrandomized. Without being able to make explicit references to causal effects, the goals of many observational studies can only be expressed in a round-about way. The resulting ambiguity impedes a frank discussion about methodology because the methods used to estimate causal effects are not the same as those used to estimate associations.

American Journal of Public Health - May 2018

DO BETTER OBSERVATIONAL RESEARCH!

We need to upgrade our **tools** and our **epistemology**



Hume



"How can we infer causal
relations from observations?"

Pearl



You



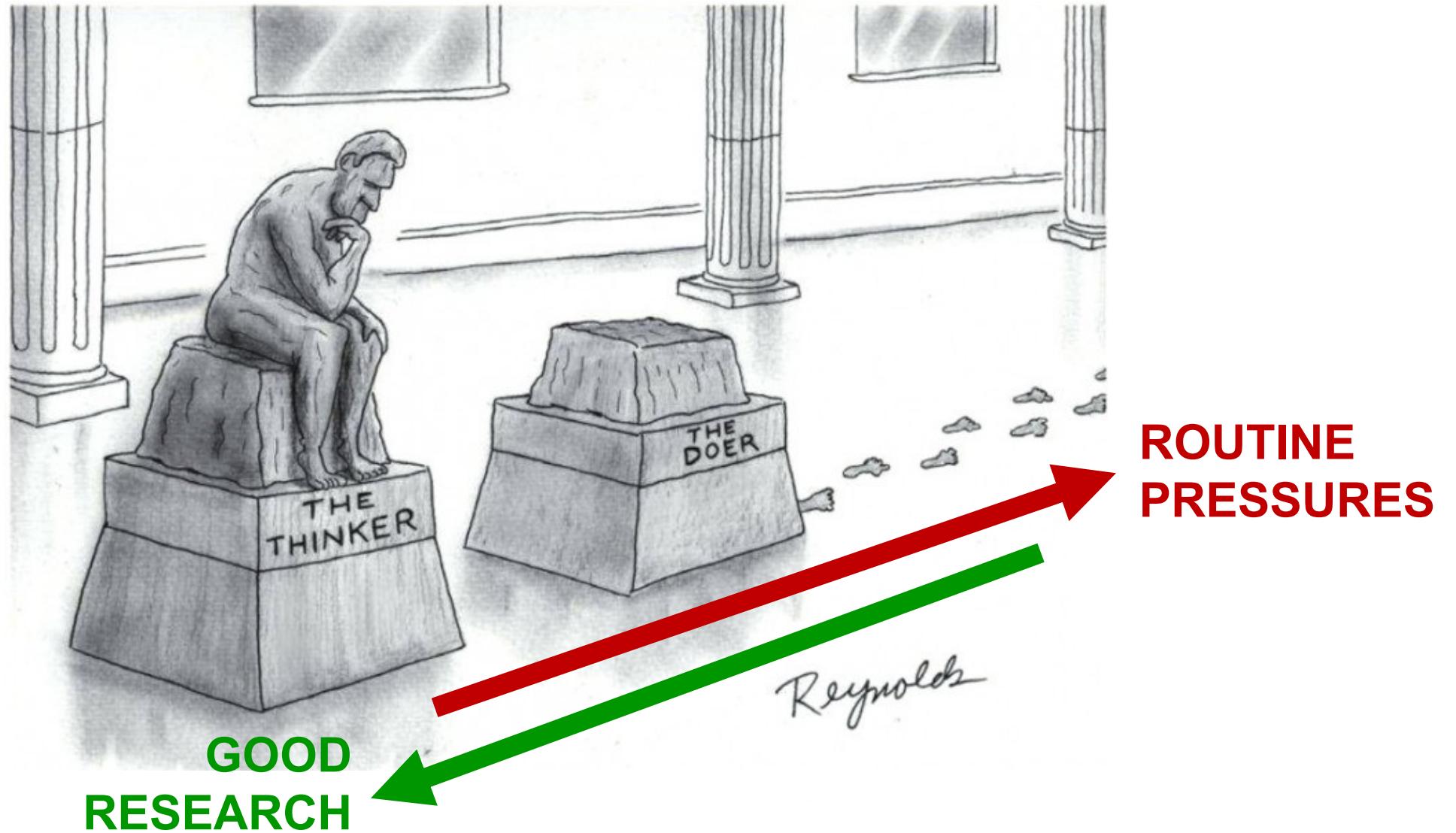
Causal
Philosophy

$$p(\mathbf{X} = \mathbf{x}) = \prod_i p(X_i = x_i | \text{PA}_i = \text{pa}_i)$$

Causal
Statistics

Causal Data
Science

MAKING TIME TO THINK



NOTE OF CAUTION



You can't put the
(causal inference)
genie back in the
bottle!

09:00-09:30 REGISTRATION

09:30-10:00 WELCOME

10:00-10:30 LECTURE 1.1

10:30-11:00 DELEGATE INTRO

11:00-11:30 TEA & COFFEE

11:30-12:00 DELEGATE INTRO

12:00-12:45 LECTURE 1.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:15 LECTURE 1.3

15:15-15:30 Q&A

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 1.4

17:00-17:45 ACTIVITY 1-A

17:30-18:00 Q&A

I.2 - PREDICTION VS CAUSAL INFERENCE

GEORGIA



@GEORGIATOMOVA

MARK



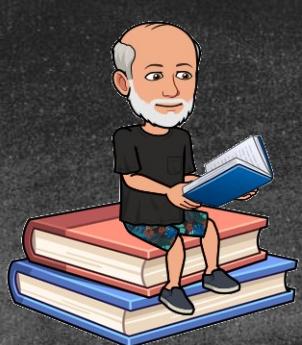
@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

LEARNING OBJECTIVES

- Recognise **description**, **prediction**, and **causal inference** as the three tasks of data science, each requiring different methods & philosophies
- Understand the different aims & priorities of prediction and causal inference
- Explain why machine learning is currently unsuitable for causal inference

THE THREE TASKS OF DATA SCIENCE

Science is about building **knowledge** and **understanding**

Data science is about '**gaining insights**' and '**extracting meaning**' from data

Most data science activity can be divided into **three scientific tasks**, each with different methods and philosophies



Hernán et al 2019. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks, CHANCE, 32:1, 42-49, DOI: 10.1080/09332480.2019.1579578



Description (& visualisation)

- Focussed on **summarising, describing, &/or visualising** features
- **Data driven** - involves simple calculations & unsupervised learning

Questions

- What happened?
 - Who was affected?
 - What was occurrence of Y in people with X?
-
- *What is the risk of death from COVID-19 among bald men?*



Prediction (AKA classification and regression)

- Focussed on **pattern recognition** and **forecasting**
- **Data driven** – involves statistical modelling and supervised learning

Questions

- What **will** happen?
- Who **will** be affected?
- Are people with **X** are **more likely** to have **Y**?
- *Are bald men more likely to die from COVID-19?*



Causal inference (AKA counterfactual prediction)

- Focussed on **understanding**
- **NOT** data driven – involves fusion of external knowledge with statistical modelling and supervised learning

Questions

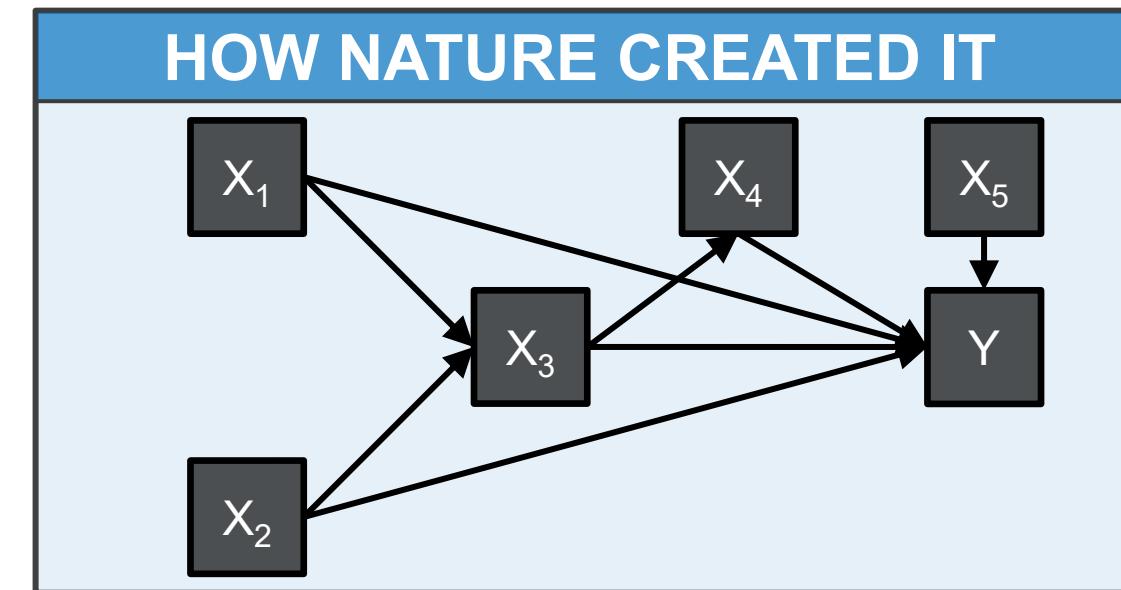
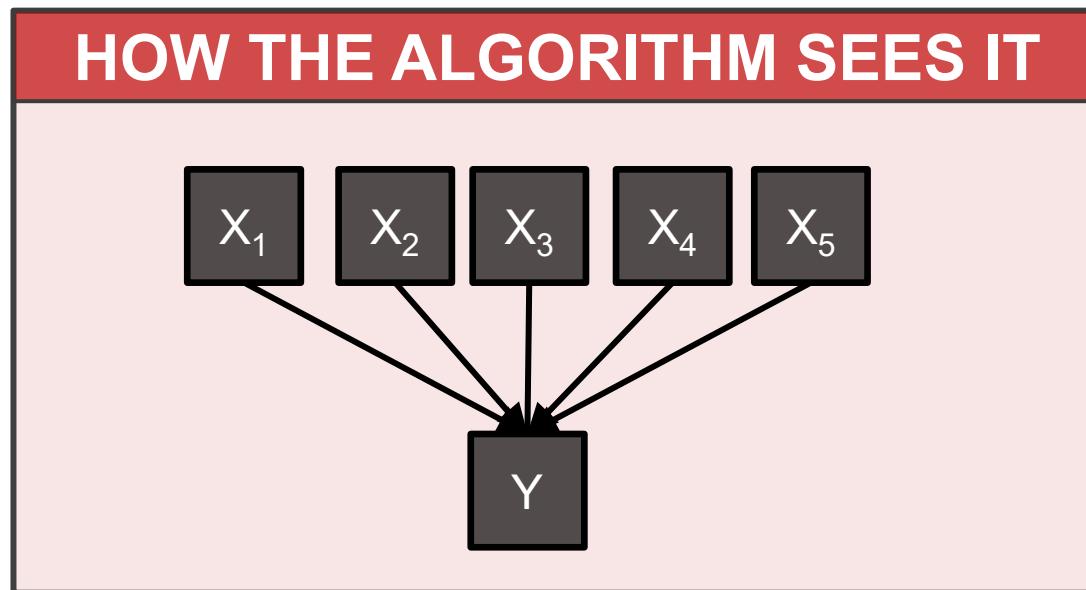
- What will happen **if**...?
 - **Why** were they affected?
 - If we **changed** X, how would it **change** Y?
-
- *If a bald man buys a wig, does this reduce his risk of death from COVID-19?*

WHAT THE MACHINE CANNOT LEARN

Data-driven algorithms are excellent at **finding patterns** in complex data, and are therefore well suited to **prediction**

Causal inference requires **identifying** and **estimating counterfactuals**, which *cannot* be learnt from the data alone

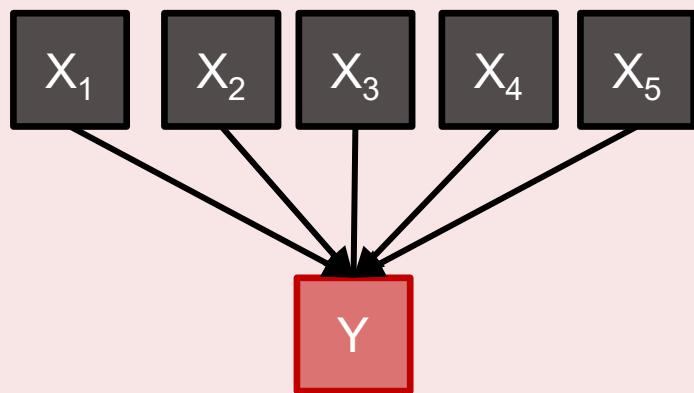
We must therefore provide **external knowledge** of (or control) the '**data generating process**'; the story behind how the data came into being



PREDICTIVE VS CAUSAL MODELLING

PREDICTIVE MODEL

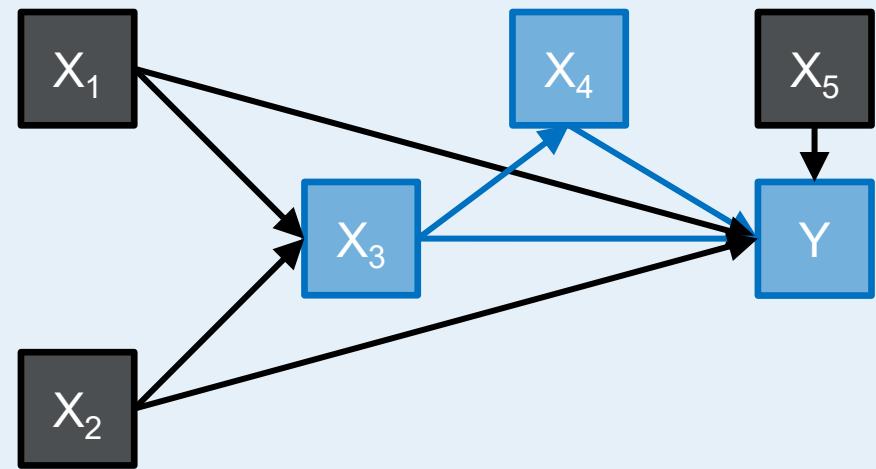
Outcome/label-focused



X_i is _____ Y
'correlated with'
'a predictor of'
'associated with'

CAUSAL MODEL

Effect-focused



The _____ of X_2 on Y is...
'total causal effect'
'direct causal effect'

PREDICTIVE VS CAUSAL MODELLING

PREDICTIVE MODEL

- Aim: **Predict values of outcome**
- Maximise: **Variance 'explained' (R^2)**
- Covariate selection focused on:
 - Balancing **precision** & **parsimony**
 - **Availability** of variables
 - Maximising: **Joint information**
- Coefficients: **Uninterpretable**
- Automation: **Favoured**

CAUSAL MODEL

- Aim: **Estimate a causal effect**
- Maximise: **Accuracy of estimate**
- Covariate selection focused on:
 - External **knowledge** & **judgement**
 - **Role** of variables
 - Minimizing: **confounding** & **selection bias**
- Coefficients: **Interpretable**
- Automation: **Not possible**

COMMON MISTAKES

Prediction is interested in the **joint predictive power** of all covariates, yet people often:

- report '**associations**' for individual covariates
- interpret coefficients as '**independent effects**'
- compare and consider '**power**' or '**contribution**' of individual covariates

Nuance applies:

- causes *can* make good predictors
- they are probably the most **transportable** predictors
- but the best predictors often just *share* important common causes (i.e. confounders)

GENERALISED LINEAR MODELS (GLMS)

$$f(E(Y)) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_n X_n$$

Prediction:

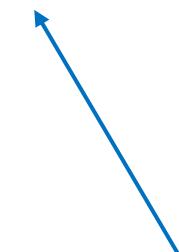
Concerned with estimating the likely value (or risk) of an outcome/event, given information from one or more observed factors

GENERALISED LINEAR MODELS (GLMS)

$$f(E(Y)) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_n X_n$$

Prediction:

Concerned with estimating the likely value (or risk) of an outcome/event, given information from one or more observed factors



Causal inference:

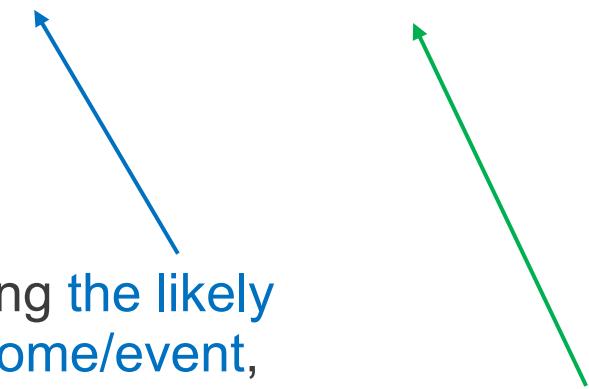
Concerned with estimating the likely change in the value (or risk) of an outcome/event that is due to (potentially hypothetical) change in a *particular factor*

GENERALISED LINEAR MODELS (GLMS)

$$f(E(Y)) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_n X_n$$

Prediction:

Concerned with estimating the likely value (or risk) of an outcome/event, given information from one or more observed factors

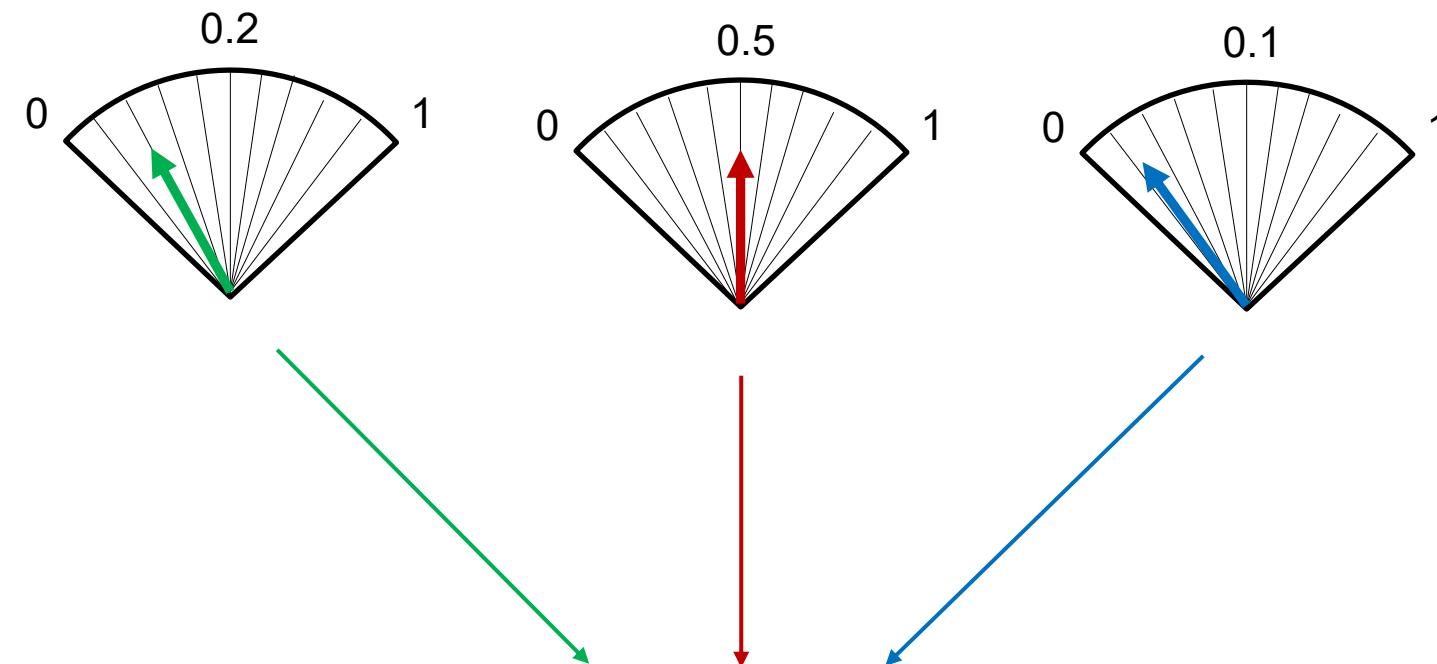


Causal inference:

Concerned with estimating the likely change in the value (or risk) of an outcome/event that is due to (potentially hypothetical) change in a *particular factor*

BUILDING PREDICTION MODELS

Prediction models are typically built by combining the independent predictive power of **multiple predictors**



$$P(\text{Outcome}) = \text{Green} + \text{Red} + \text{Blue}$$

BUILDING PREDICTION MODELS

With modern computers & statistical software, we've become quite good at this!

Framingham Heart Disease Risk Score ★

Age years

Sex Female Male

Smoker No Yes

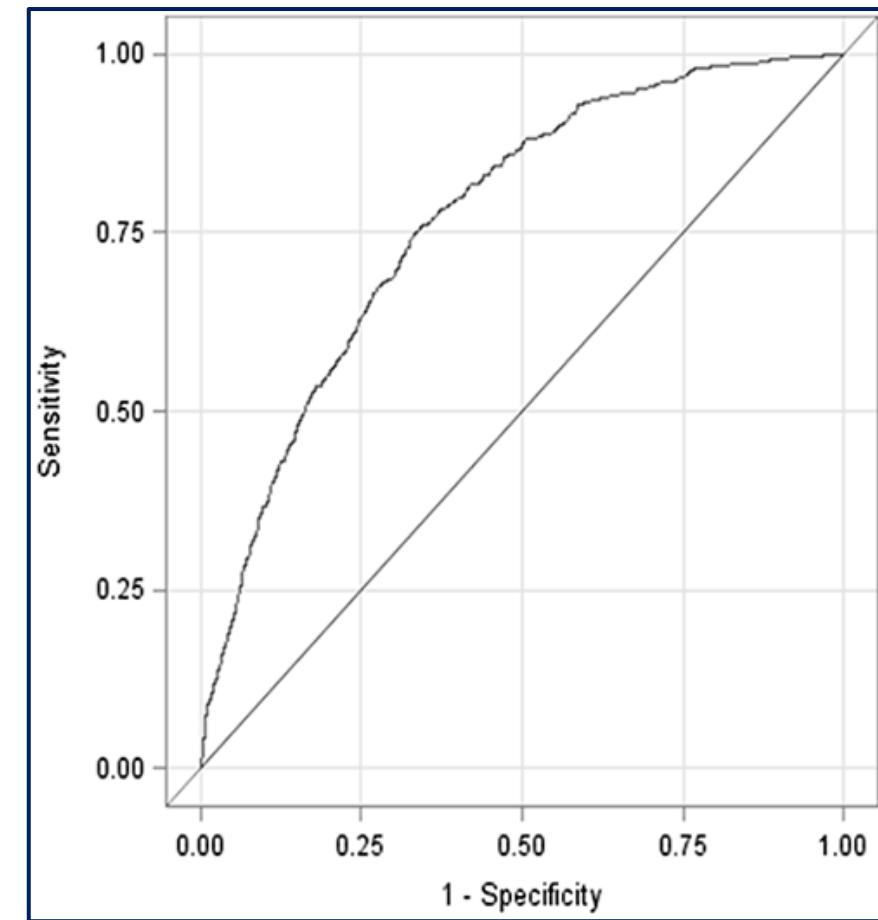
Total cholesterol Norm: 3.9 - 5.2 mmol/L

HDL cholesterol Norm: 1 - 2.1 mmol/L

Systolic BP Norm: 100 - 120 mm Hg

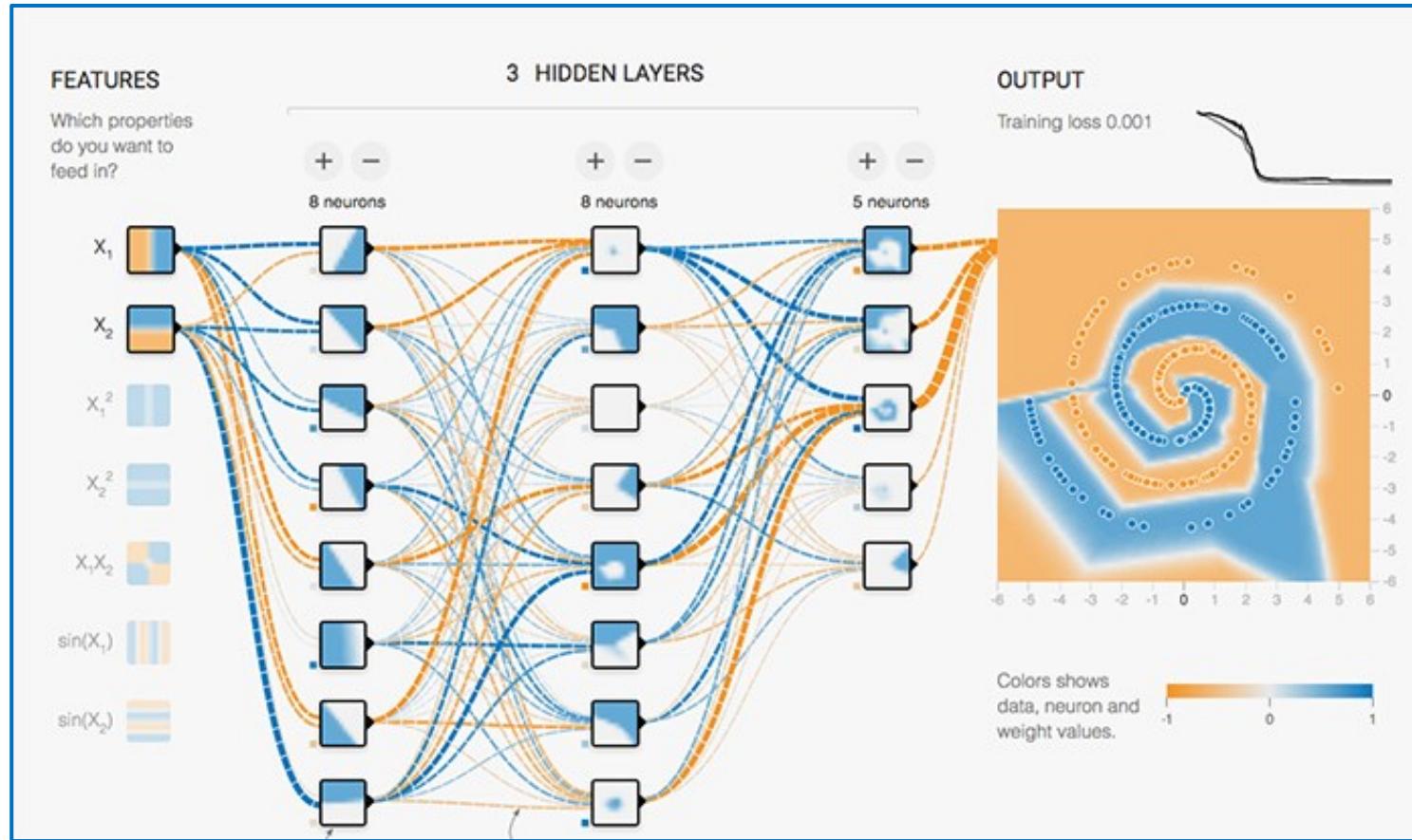
Blood pressure being treated with medicines No Yes

Result:
Please fill out required fields.



BUILDING PREDICTION MODELS

Artificial intelligence & machine learning promises to make us better, finding predictive data patterns we might never have thought to look for



Good models make us better at predicting **prognosis** & identifying those who might benefit from proactive treatment

BUT they can't identify the best targets for **intervention** and **prevention**

CONFLATION OF PREDICTION AND CAUSATION

Utilisation of **same modelling frameworks**

- e.g. (generalised) *linear models*

Language of '**risk factors**'

- *Conflates correlational relationships with causal ones*

'**Taboo**' of causality in mainstream / conventional statistics

- *Lack of training in statistical methods for causal inference*

AJPH PUBLIC HEALTH OF CONSEQUENCE

AJPH AMERICAN JOURNAL OF PUBLIC HEALTH

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Miguel A. Hernán, MD, DrPH

FOCUS

Is
fo
And
you
clar

F
who
The
risk"
not p
sema
prec
desig

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

You know the story:
Dear author: Your observational study cannot prove causation. Please replace all references to causal effects by references to associations.

See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

This commentary argues that using the term "causal" is necessary to improve the quality of observational research.

Specifically, being explicit about the causal objective of a study reduces ambiguity in the scientific question, errors in the data analysis, and excesses in the interpretation of the results. (*Am J Public Health*. 2018;108: 616–619. doi:10.2105/AJPH.2018.304337)

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are made.

We need to stop treating "causal" as a dirty word that respectable investigators do not say in public or put in print. It is true that observational studies cannot definitely prove causation, but this statement misses the point, as discussed in this commentary.

Many journal editors request authors to avoid causal language, and many observational researchers, trained in a scientific environment that frowns upon causality claims, spontaneously refrain from mentioning the C-word ("causal") in their work. As a result, "causal effect" and terms with similar meaning ("impact," "benefit," etc.) are routinely avoided in scientific publications that describe nonrandomized studies. Instead, we see terms like "association" and others that convey a similar meaning ("correlation," "pattern," etc.), or the calculatedly ambiguous "link."

The proscription against the C-word is harmful to science because causal inference is a core task of science.

glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and random variability—that is, suppose the 0.8 comes from a very large population so that the 95% confidence interval around it is tiny.

The risk ratio of 0.8 is a measure of the association between wine intake and heart disease. Strictly speaking, it means that drinkers of one glass of wine have, on average, a 20% lower risk of heart disease than individuals who do not drink. The risk ratio of 0.8 does not imply that drinking a glass of wine every day lowers the risk of heart disease by 20%. It is possible that the kind of people who drink a glass of wine per day would have a lower risk of heart disease even if they didn't drink wine because, for example, they have high enough incomes to buy, besides wine, nutritious food and to take

**OF COURSE
"ASSOCIATION IS NOT
CAUSATION"**

Suppose we want to know whether daily drinking of a glass of wine affects the 10-year risk of coronary heart disease. Because there are no randomized trials of long-term alcohol drinking, we analyze observational data by comparing the risk of heart disease



CONFLATION OF PREDICTION AND CAUSATION

Support The Guardian
Available for everyone, funded by readers

Contribute → Subscribe →

Sign in

The
Guardian

News | Opinion | Sport | Culture | Lifestyle

Risk of obesity can be accurately predicted in babies, study finds

Factors including birth weight could determine likelihood of child becoming obese aged 10



▲ In the Netherlands, children are assessed for being overweight at the age of two. Photograph: Image Source/Getty Images

Sarah Boseley Health editor

Thu 2 May 2019 00.01 BST

CONFLATION OF PREDICTION AND CAUSATION

RESEARCH



OPEN ACCESS

the
bmj

Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study

Carlos A Celis-Morales,¹ Donald M Lyall,² Paul Welsh,¹ Jana Anderson,² Lewis Steell,¹ Yibing Guo,¹ Reno Maldonado,¹ Daniel F Mackay,² Jill P Pell,² Naveed Sattar,¹ Jason M R Gill¹

OBJECTIVE

To investigate the association between active commuting and incident cardiovascular disease (CVD), cancer, and all cause mortality.

CONCLUSIONS

Cycle commuting was associated with a lower risk of CVD, cancer, and all cause mortality. Walking commuting was associated with a lower risk of CVD independent of major measured confounding factors. Initiatives to encourage and support active commuting could reduce risk of death and the burden of important chronic conditions.

CONFLATION OF PREDICTION AND CAUSATION

Causal interpretation of a prediction model can do actual **HARM** ...

- implied causal interpretation for ‘**understanding**’ of what leads to **Covid-19 related death**
- one massive ‘**association**’ study
⇒ i.e. prediction!



Article

Factors associated with COVID-19-related death using OpenSAFELY

<https://doi.org/10.1038/s41586-020-2521-4>

Received: 15 May 2020

Accepted: 1 July 2020

Published online: 8 July 2020

Check for updates

Elizabeth J. Williamson^{1,6}, Alex J. Walker^{2,6}, Krishnan Bhaskaran^{1,6}, Seb Bacon^{2,6}, Chris Bates^{3,6}, Caroline E. Morton², Helen J. Curtis², Amir Mehrkar², David Evans², Peter Inglesby², Jonathan Cockburn², Helen I. McDonald^{1,4}, Brian MacKenna², Laurie Tomlinson¹, Ian J. Douglas¹, Christopher T. Rentsch¹, Rohini Mathur¹, Angel Y. S. Wong¹, Richard Grieve¹, David Harrison⁵, Harriet Forbes¹, Anna Schultze¹, Richard Croker², John Parry³, Frank Hester³, Sam Harper³, Rafael Perera², Stephen J. W. Evans¹, Liam Smeeth^{1,4,7} & Ben Goldacre^{2,7,8}

Coronavirus disease 2019 (COVID-19) has rapidly affected mortality worldwide¹. There is unprecedented urgency to understand who is most at risk of severe outcomes, and this requires new approaches for the timely analysis of large datasets. Working on behalf of NHS England, we created OpenSAFELY—a secure health analytics platform that covers 40% of all patients in England and holds patient data within the existing data centre of a major vendor of primary care electronic health records. Here we used OpenSAFELY to examine factors associated with COVID-19-related death. Primary care records of 17,278,392 adults were pseudonymously linked to 10,926 COVID-19-related deaths. COVID-19-related death was associated with: being male (hazard ratio (HR) 1.59 (95% confidence interval 1.53–1.65)); greater age and deprivation (both with a strong gradient); diabetes; severe asthma; and various other medical conditions. Compared with people of white ethnicity, Black and South Asian people were at higher risk, even after adjustment for other factors (HR 1.48 (1.29–1.69) and 1.45 (1.32–1.58), respectively). We have quantified a range of clinical factors associated with COVID-19-related death in one of the largest cohort studies on this topic so far. More patient records are rapidly being added to OpenSAFELY, we will update and extend our results regularly.

CONFLATION OF PREDICTION AND CAUSATION

Incorrect causal interpretation of ‘**clinical factors**’ misused in France
... but thrown out by courts (due to a causal inference expert)



Article

Factors associated with COVID-19-related death using OpenSAFELY

<https://doi.org/10.1038/s41586-020-2521-4>

Received: 15 May 2020

Accepted: 1 July 2020

Published online: 8 July 2020

Check for updates

Elizabeth J. Williamson^{1,6}, Alex J. Walker^{2,6}, Krishnan Bhaskaran^{1,6}, Seb Bacon^{2,6}, Chris Bates^{3,6}, Caroline E. Morton², Helen J. Curtis², Amir Mehrkar², David Evans², Peter Inglesby², Jonathan Cockburn², Helen I. McDonald^{1,4}, Brian MacKenna², Laurie Tomlinson¹, Ian J. Douglas¹, Christopher T. Rentsch¹, Rohini Mathur¹, Angel Y. S. Wong¹, Richard Grieve¹, David Harrison⁵, Harriet Forbes¹, Anna Schultze¹, Richard Croker², John Parry³, Frank Hester³, Sam Harper³, Rafael Perera², Stephen J. W. Evans¹, Liam Smeeth^{1,4,7} & Ben Goldacre^{2,7,8}

Coronavirus disease 2019 (COVID-19) has rapidly affected mortality worldwide¹. There is unprecedented urgency to understand who is most at risk of severe outcomes, and this requires new approaches for the timely analysis of large datasets. Working on behalf of NHS England, we created OpenSAFELY—a secure health analytics platform that covers 40% of all patients in England and holds patient data within the existing data centre of a major vendor of primary care electronic health records. Here we used OpenSAFELY to examine factors associated with COVID-19-related death. Primary care records of 17,278,392 adults were pseudonymously linked to 10,926 COVID-19-related deaths. COVID-19-related death was associated with: being male (hazard ratio (HR) 1.59 (95% confidence interval 1.53–1.65)); greater age and deprivation (both with a strong gradient); diabetes; severe asthma; and various other medical conditions. Compared with people of white ethnicity, Black and South Asian people were at higher risk, even after adjustment for other factors (HR 1.48 (1.29–1.69) and 1.45 (1.32–1.58), respectively). We have quantified a range of clinical factors associated with COVID-19-related death in one of the largest cohort studies on this topic so far. More patient records are rapidly being added to OpenSAFELY, we will update and extend our results regularly.

RECOMMENDED READING

- A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. Hernán MA, et al. *CHANCE*, 2019, <https://doi.org/10.1080/09332480.2019.1579578>
- To explain or predict? Shmueli, G. *Statistical Science*, 2010, <https://doi.org/10.1214/10-STS330>
- Reflection on modern methods: generalized linear models for prognosis and intervention—theory, practice and implications for machine learning. KF Arnold, et al. *International Journal of Epidemiology*, 2020, <https://doi.org/10.1093/ije/dyaa049>
- Williamson EJ, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 2020, <https://doi.org/10.1038/s41586-020-2521-4>
- Ellie Murray Twitter thread: <https://twitter.com/EpiEllie/status/1258607277357006849>
- French court reporting: <https://www.affiches-parisiennes.com/suspension-des-nouveaux-criteres-de-vulnerabilite-au-covid-19-ouvrant-droit-au-chomage-partiel-11123.html>

SUMMARY

- **Description, prediction, and causal inference** are three distinct tasks of data analysis, and each requires different methods and philosophies
- Prediction and causation are commonly conflated
- To unlock causal meaning, we must recognise and embrace causal inference as a distinct goal requiring distinct methods
- AI and ML promise a revolution in predictive modelling, but this improved '**curve fitting**' remains fundamentally unsuitable for causal inference
- Without massive education around these issues, much avoidable **harm** will continue to happen through ignorance

I.3 - COUNTERFACTUALS AND POTENTIAL OUTCOMES

GEORGIA



@GEORGIATOMOVA

MARK



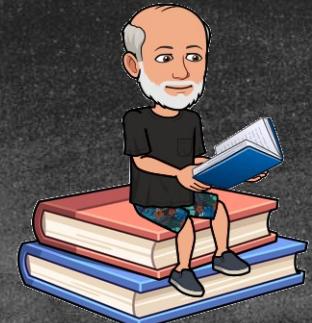
@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 1

09:00-09:30 REGISTRATION

09:30-10:00 WELCOME

10:00-10:30 LECTURE I.1

10:30-11:00 DELEGATE INTRO

11:00-11:30 TEA & COFFEE

11:30-12:00 DELEGATE INTRO

12:00-12:45 LECTURE I.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:15 LECTURE I.3

15:15-15:30 Q&A

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE I.4

17:00-17:45 ACTIVITY I-A

17:30-18:00 Q&A

LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- Describe the core principles of **counterfactual reasoning**
- Use and interpret the notation of the **Potential Outcome Framework**
- Explain the two stages of estimating causal effects:
 - ✓ **identifying** your **estimand**
 - ✓ **estimating** your **estimand** with an **estimator** (!!)
- Explain why **interval estimation** is encouraged, and **significance testing** is discouraged when doing causal inference
- Define the four **identifiability conditions**

THE CAUSAL REVOLUTION

What tools can help us with causal inference?



Hume



"How can we infer causal relations from observations?"

Pearl



You

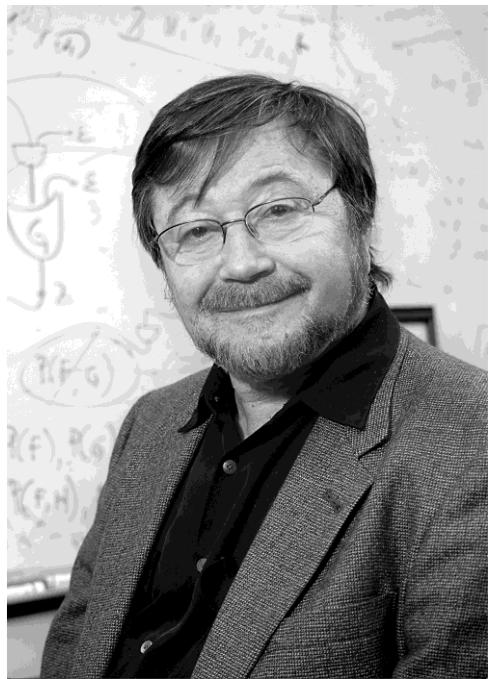


Causal
Philosophy

Causal
Statistics

Causal Data
Science

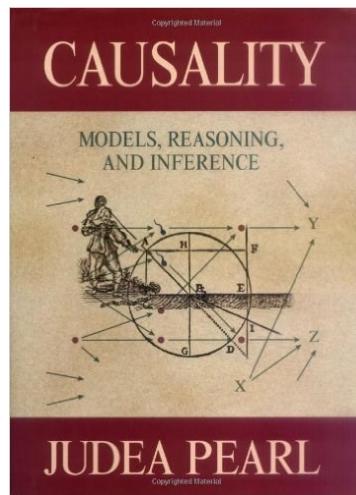
STRUCTURAL CAUSAL MODEL



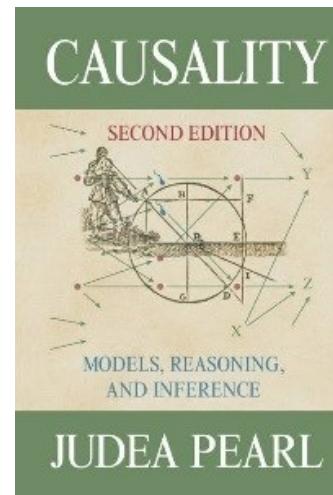
Judea Pearl, Computer scientist, philosopher, Turing Prize winner

Judea Pearl's '**Structural Causal Model**' is a formal framework for considering causal effects that draws together three mathematical and/or philosophical tools:

- **Probability theory**
- **Counterfactual reasoning**
- **Graphical model theory**



2000



2009

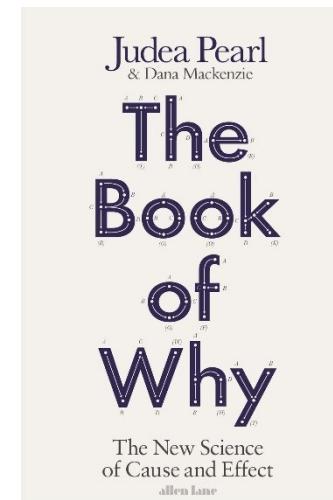


CAUSAL INFERENCE
IN STATISTICS
A Primer

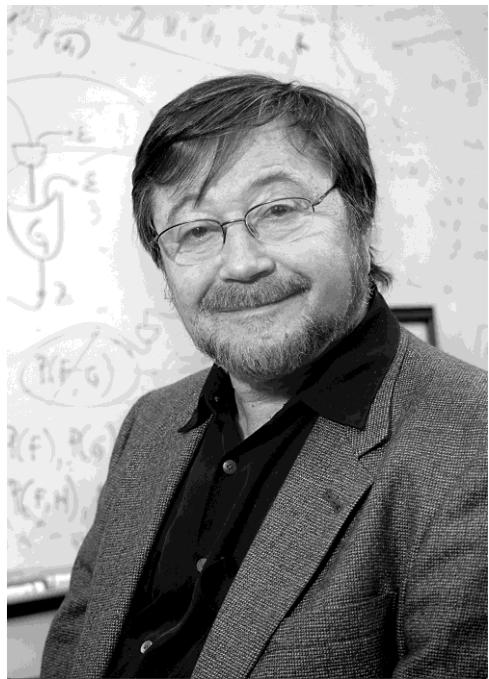
Judea Pearl
Madelyn Glymour
Nicholas P. Jewell

WILEY

2016



2018



Judea Pearl, Computer scientist, philosopher, Turing Prize winner

Causal Diagrams for Epidemiologic Research

Sander Greenland,¹ Judea Pearl,² and James M. Robins³

Causal diagrams have a long history of informal use and, more recently, have undergone formal development for applications in expert systems and robotics. We provide an introduction to these developments and their use in epidemiologic research. Causal diagrams can provide a starting point for identifying variables that must be measured and controlled to obtain unconfounded effect estimates. They also provide a method for

critical evaluation of traditional epidemiologic criteria for confounding. In particular, they reveal certain heretofore unnoticed shortcomings of those criteria when used in considering multiple potential confounders. We show how to modify the traditional criteria to correct those shortcomings.
(Epidemiology 1999;10:37–48)

Keywords: bias, causation, confounding, epidemiologic methods, graphical methods, observational studies.

Greenland et al 1999 *Epidemiology*; 10: 37–48

“Epidemiology has been a pioneer in accepting the DAG-counterfactuals symbiosis as a ruling paradigm — way ahead of mainstream statistics and its other satellites... adopting new tools... at dazzling speed”

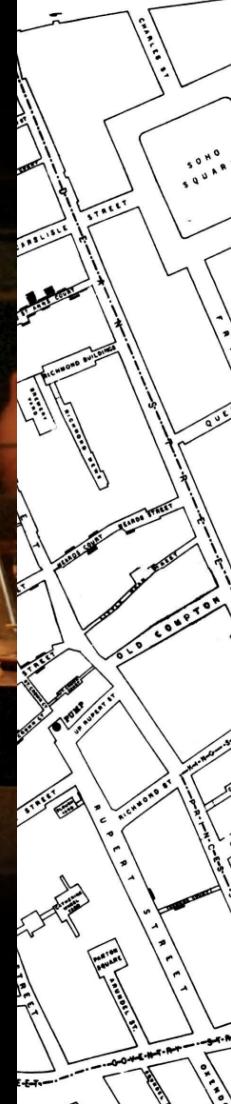
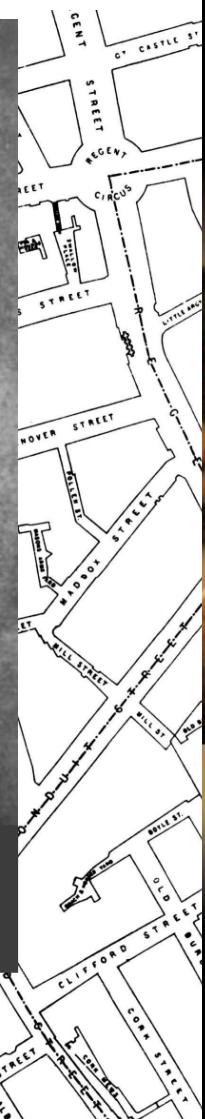
POTENTIAL OUTCOME FRAMEWORK

- The **Structural Causal Model** is built on (**Rubin's Potential Outcome Framework**, which provides a formal language for **counterfactual reasoning**)
- The Potential Outcome Framework continues to be widely used among 'causal' epidemiologists, particularly leading (American) advocates of '**causal inference methods**'
- It can quickly become tedious (and technical), but familiarity with the core features is necessary to understand later concepts, and inoculates against embarrassment!
- Judea Pearl's Structural Causal Model uses its own slightly different notation (including the '**do**' operator); it's debatable whether this adds any function

POTENTIAL OUTCOMES



John Snow, often described
as the 'first Epidemiologist'



The 'Broad Street' Pump

POTENTIAL OUTCOMES

1854

- 31 Aug** Cholera outbreak in Soho
- 5 Sep** Snow maps where deaths are occurring
- 6 Sep** 83% died drank from Broad Street pump
- 7 Sep** Snow meets parish guardians to argue for pump closure
- 8 Sep** Snow removes handle > Cholera outbreak ends!



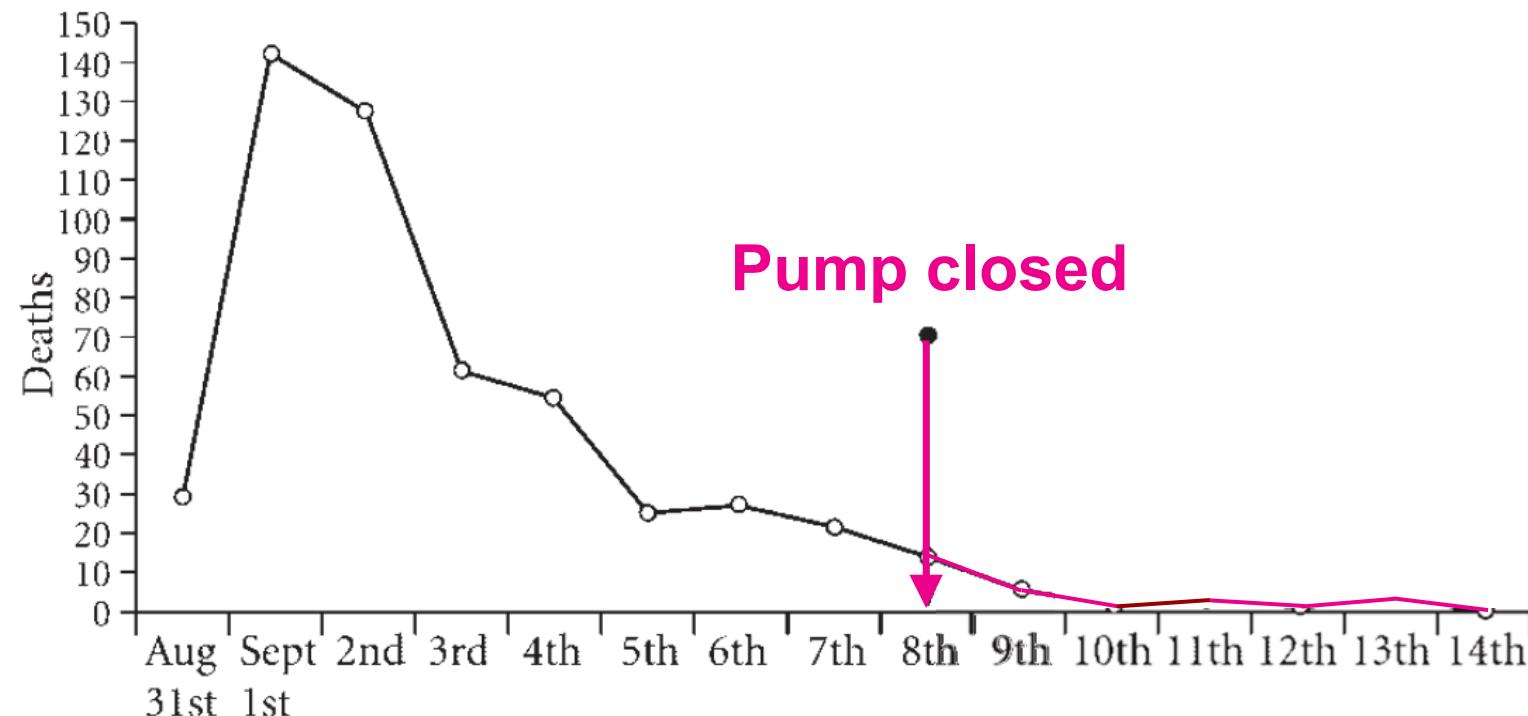
POTENTIAL OUTCOMES



JOHN SNOW, LEGEND,
DATA SCIENTIST

Snow observed **outcome (Y)** of what happened when the **exposure (X)** was '**closed**' (**factual**)

This is written $Y|X=\text{closed} = Y_{X=0} = Y_0$



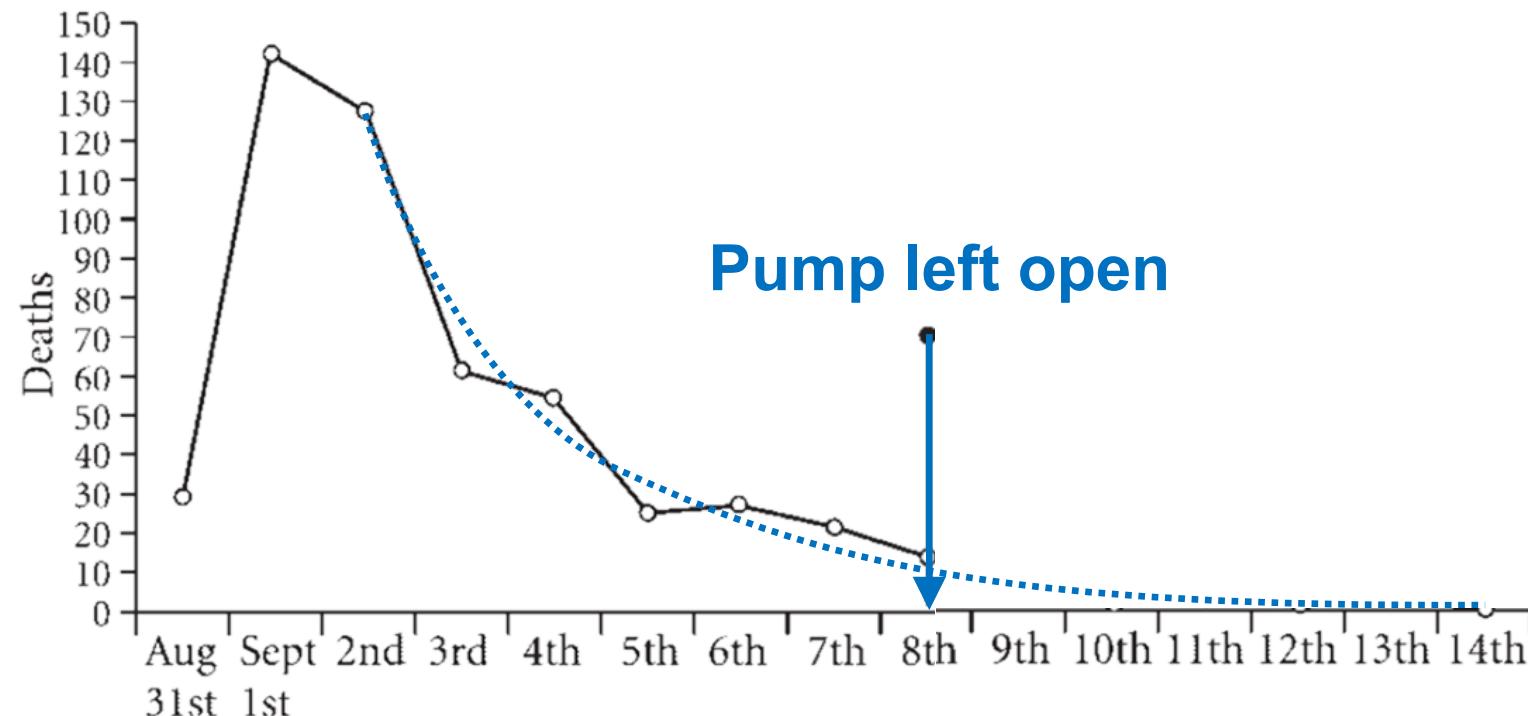
POTENTIAL OUTCOMES



JOHN SNOW, LEGEND,
DATA SCIENTIST

We don't know the **potential outcome (Y)** that would have happened if the **exposure (X)** had been – *counter to fact* - left 'open' (**counterfactual**)

This is written $Y|X=open = Y_{x=1} = Y_1$



POTENTIAL OUTCOMES

To identify the causal effect of removing the pump handle on subsequent mortality, we would need to know (and compare):

Y_0

vs

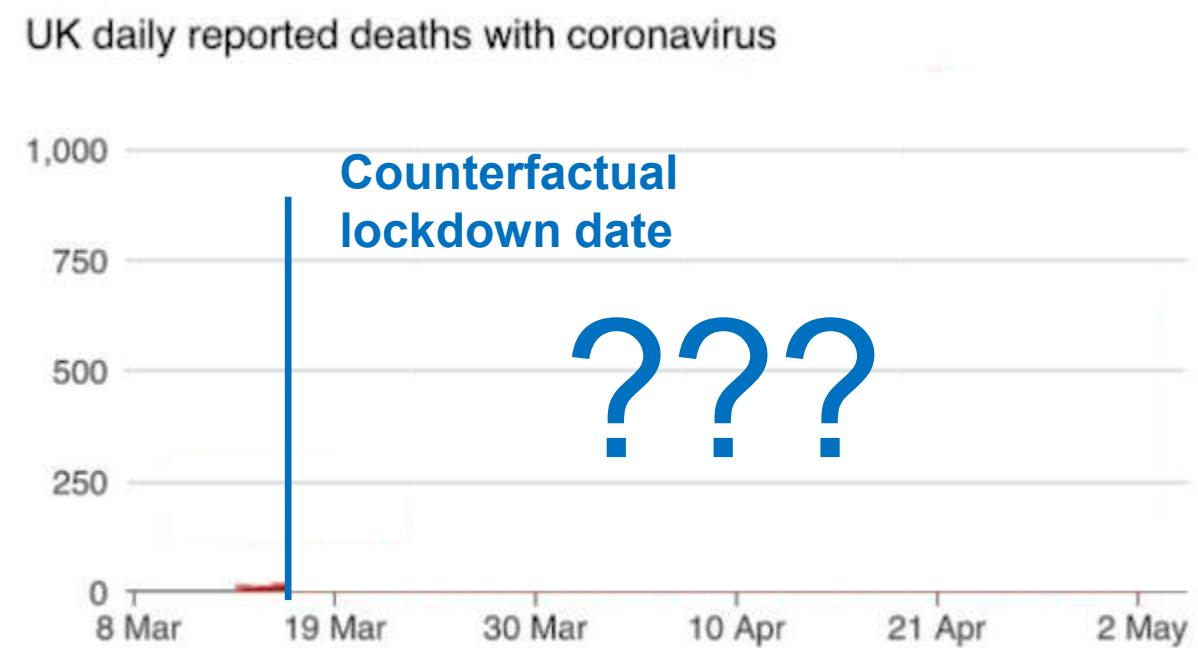
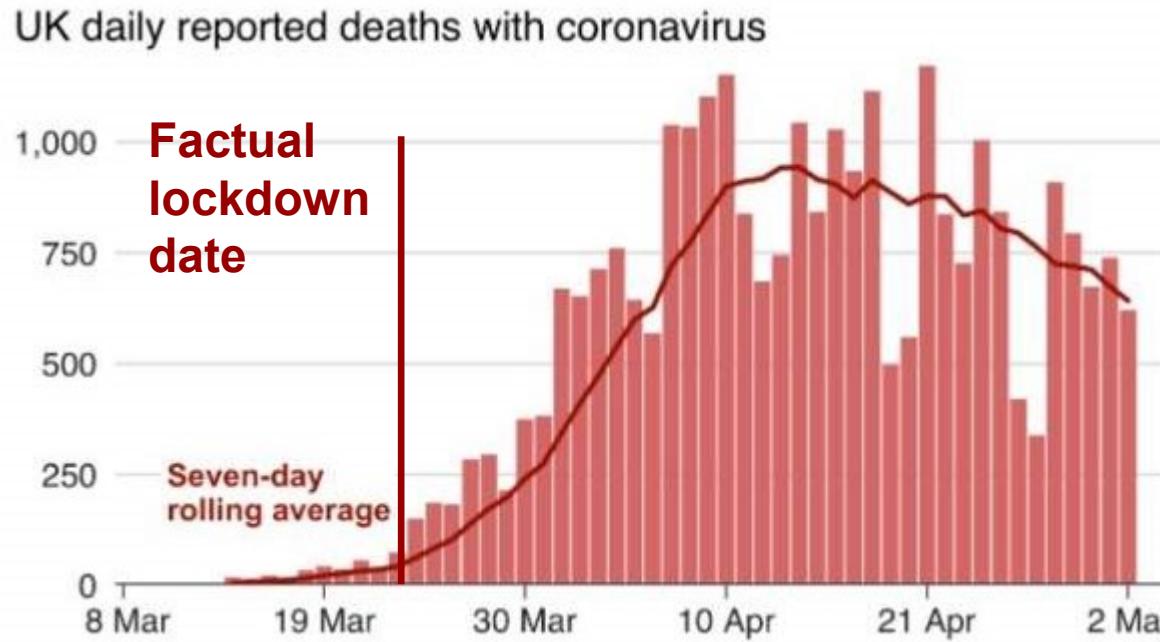
Y_1

The number of deaths (Y) when the pump was closed ($X=0$)

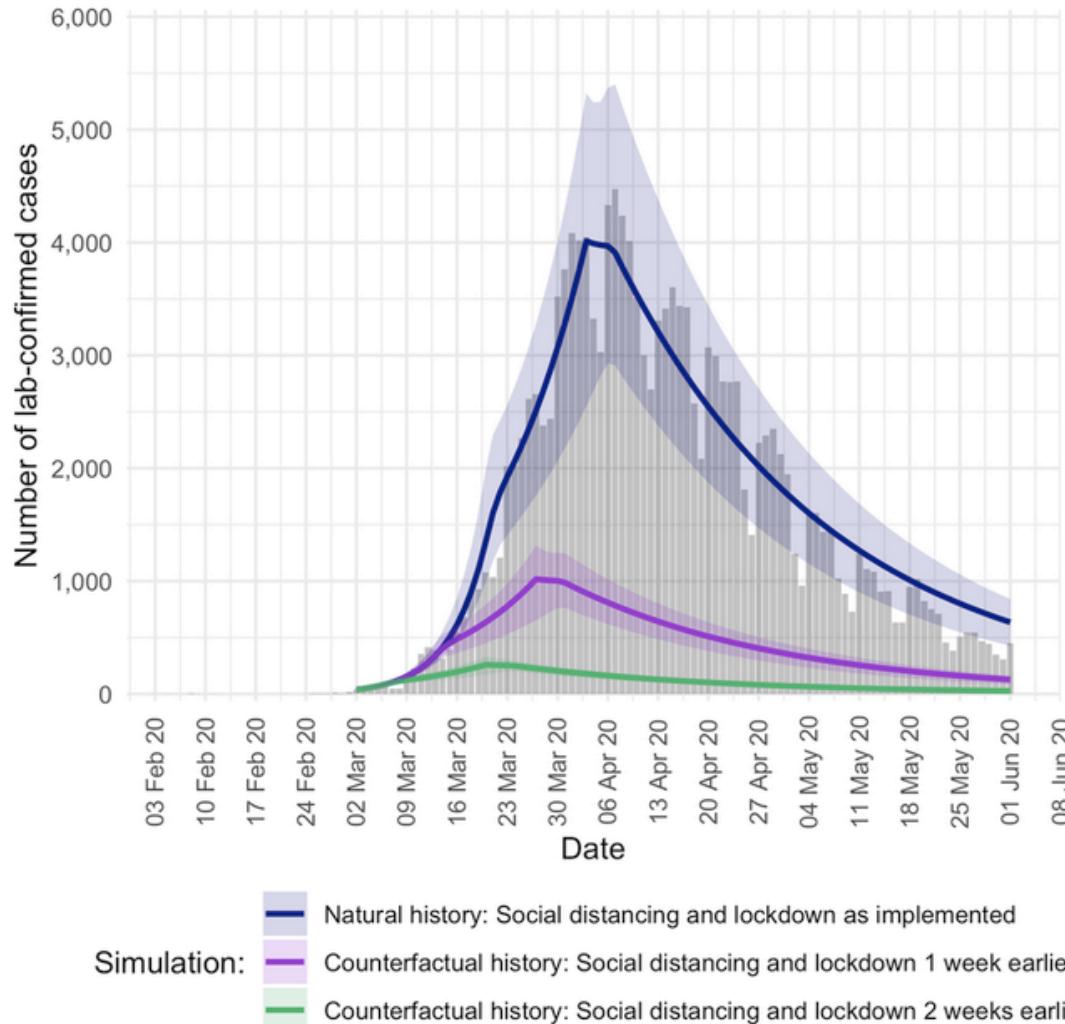
The number of deaths (Y) when the pump was open ($X=1$)

FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

- We can never know the **potential outcome** for a **counterfactual exposure**!
- For each ‘unit of analysis’ we can only observe one potential outcome
- This is known as the **fundamental problem of causal inference**



FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE



Arnold et al 2022. Estimating the effects of lockdown timing on COVID-19 cases and deaths in England: A counterfactual modelling study



Thomas House (``▽``)☕
@TAH_Sci

...

Replies to @james_e_b_ and @PWGTennant

This kind of "study" is just people dressing up their prejudices as science. I'm embarrassed on behalf of the people who put their names to it.



Philippe Lemoine
@phl43

...

This study doesn't confirm jackshit and is just another example of the kind of ridiculous methods that pass as counterfactual analysis in the field of epidemiology. Here is a short thread explaining what they did and why it doesn't show anything. 1/n

ESTIMATING POTENTIAL OUTCOMES

Instead we must *estimate* the **potential outcome** for the **counterfactual exposure** from **exchangeable units of analysis**

What would have
happened if teenage
Peter had eaten
MAGIC BEANS?



Know:
 $Y_0(Peter)$

Want to know:
 $Y_1(Peter)$

ESTIMATING POTENTIAL OUTCOMES

Instead we must *estimate* the **potential outcome** for the **counterfactual exposure** from **exchangeable units of analysis**

What would have happened if teenage **Peter** had eaten **MAGIC BEANS?**

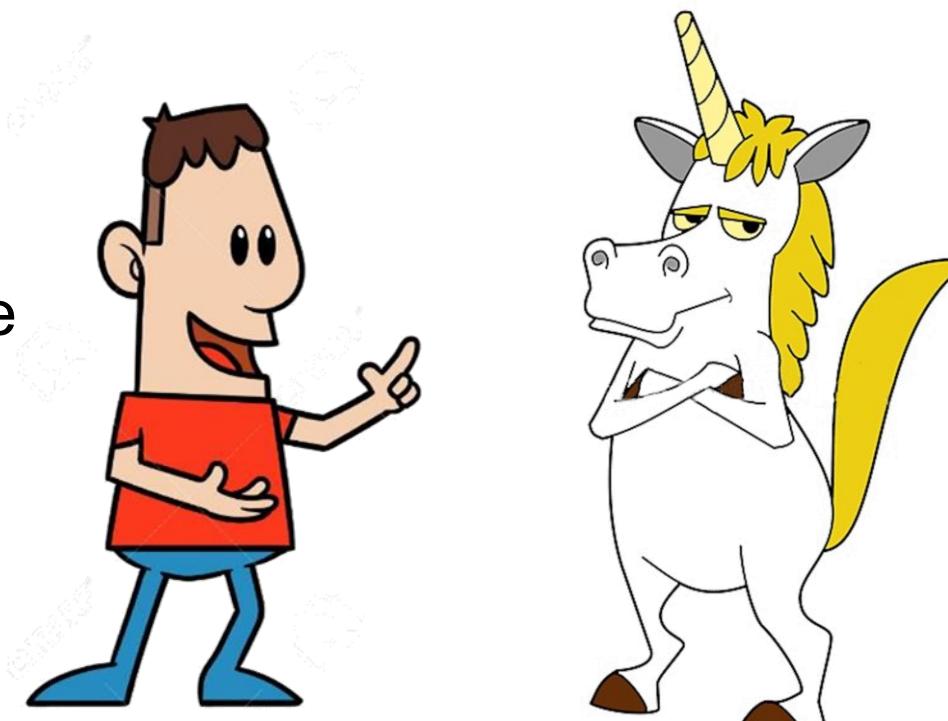


ESTIMATING POTENTIAL OUTCOMES

Instead we must *estimate* the **potential outcome** for the **counterfactual exposure** from **exchangeable units of analysis**

What would have happened if teenage **Peter** had eaten **MAGIC BEANS?**

(*What would have been Peter's expectation, E?*)



$$E[Y_1(\text{Peter})]$$

≈

$$E[Y_1(\text{Twin})]$$

POOR COUNTERFACTUALS

Problem: Units (e.g. people) are very different - even the same units can respond differently at different times

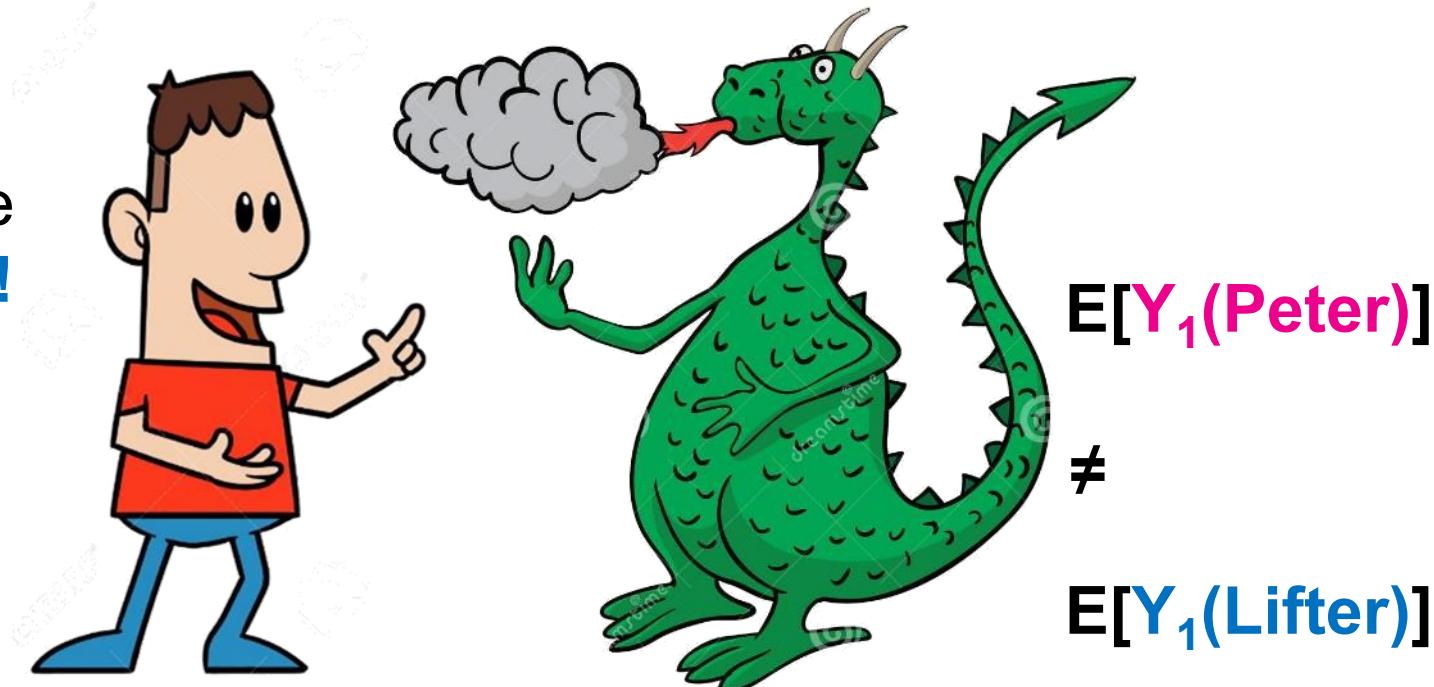


Not **exchangeable** units of analysis!

POOR COUNTERFACTUALS

Problem: Units (e.g. people) are very different - even the same units can respond differently at different times

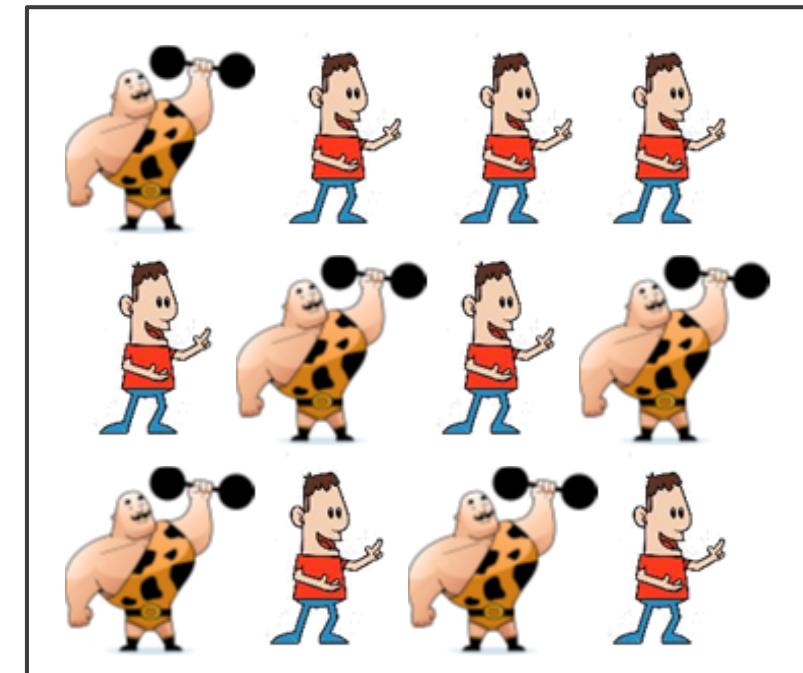
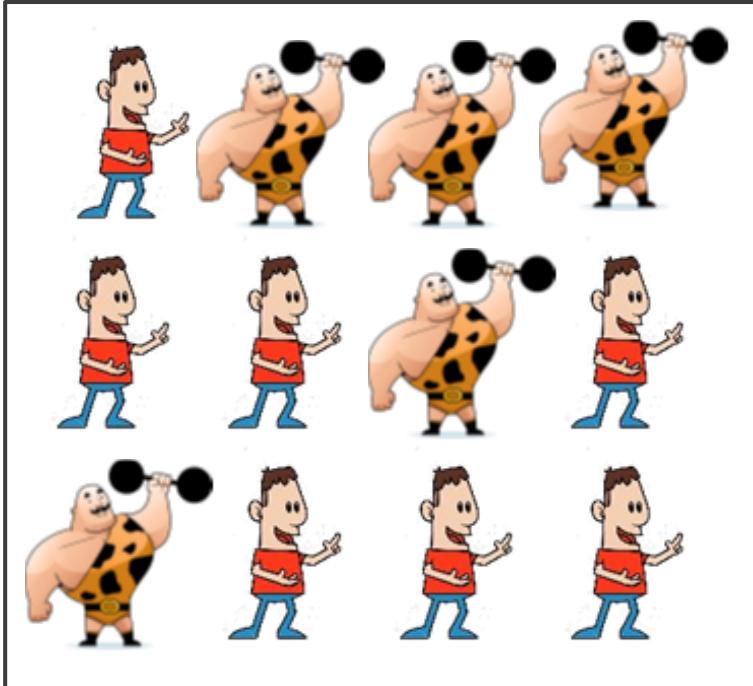
Poor estimate of the
potential outcome!



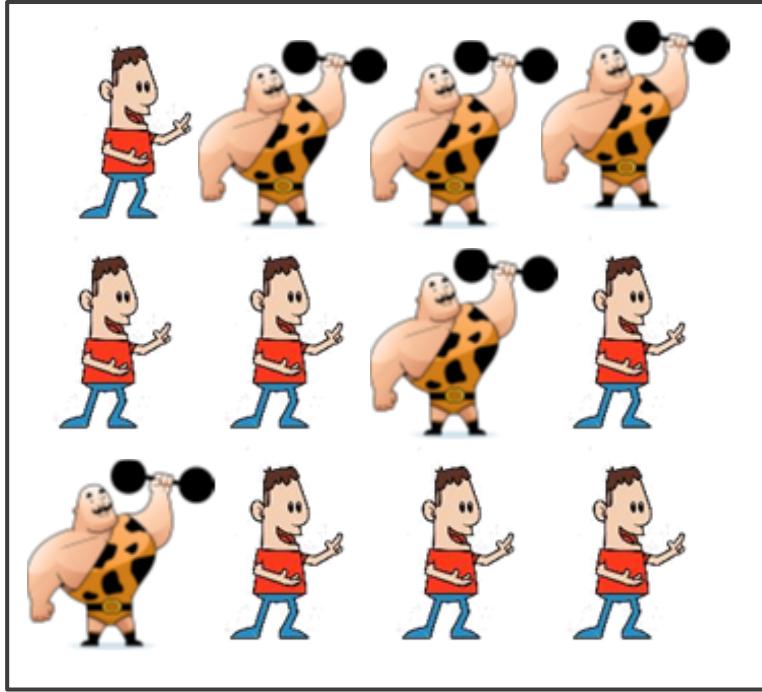
Not **exchangeable** units of analysis!

AVERAGE CAUSAL EFFECTS

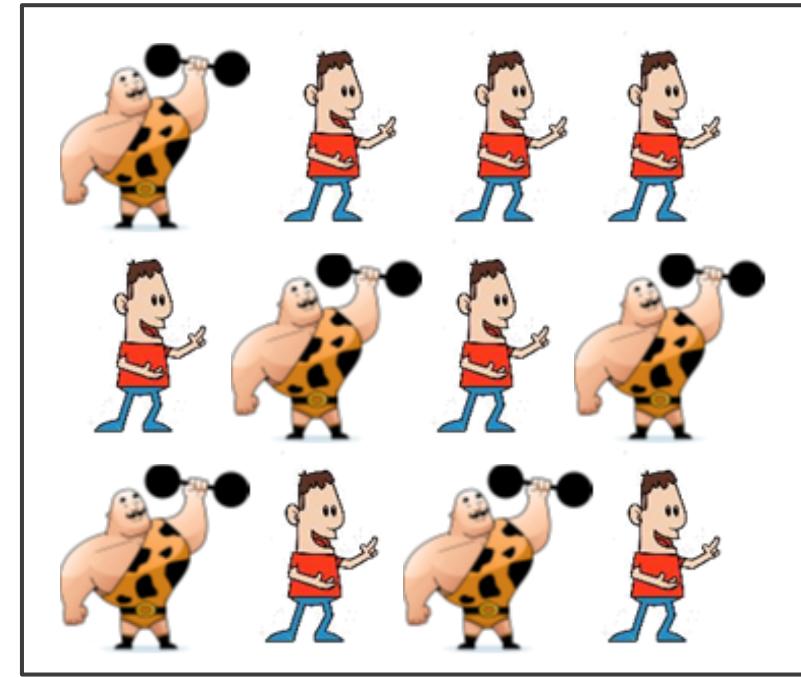
- We therefore have to **identify** (sub)groups of units that were **exchangeable at the time of exposure**
- We can then estimate the **average causal effect** (AKA **average treatment effect, ATE**) by comparing the outcomes between these (sub)groups



AVERAGE CAUSAL EFFECTS



TOOK MAGIC BEANS



AVERAGE CAUSAL EFFECTS

$E[Y_0(\text{Originals})]$

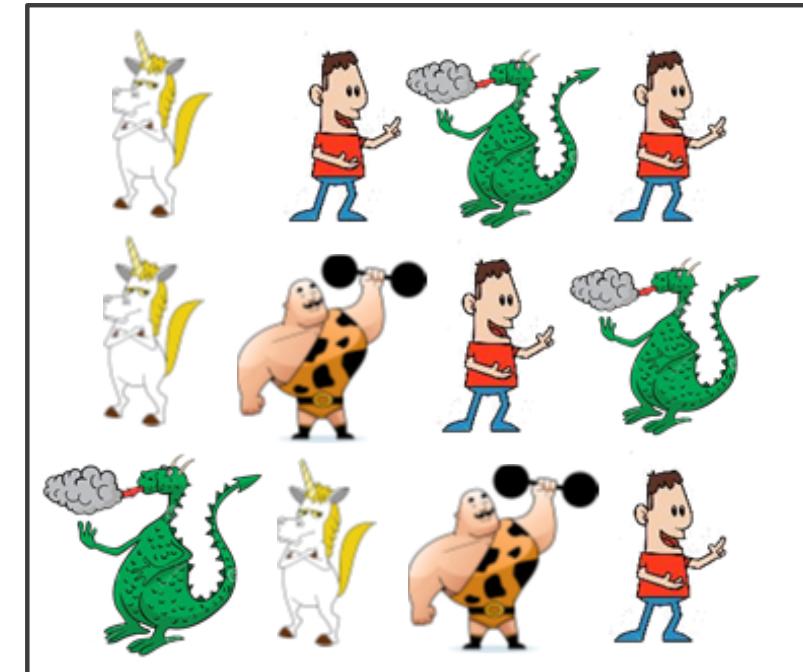
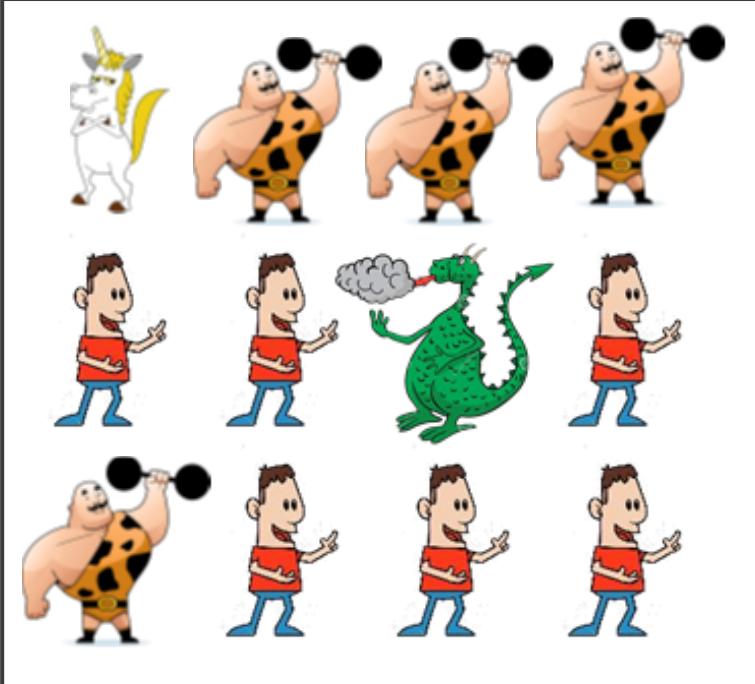
\approx

$E[Y_0(\text{Clones})]$

$E[Y_1(\text{Originals})]$

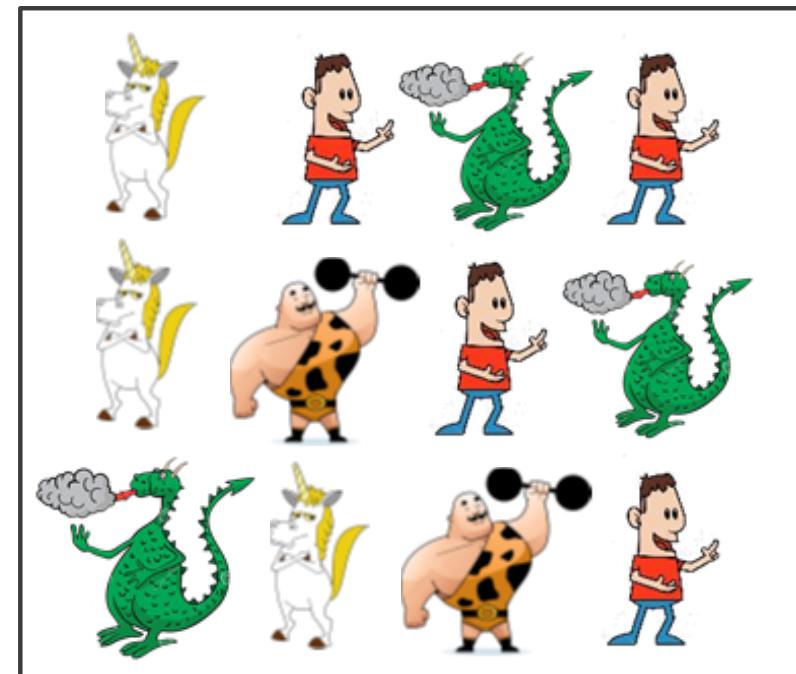
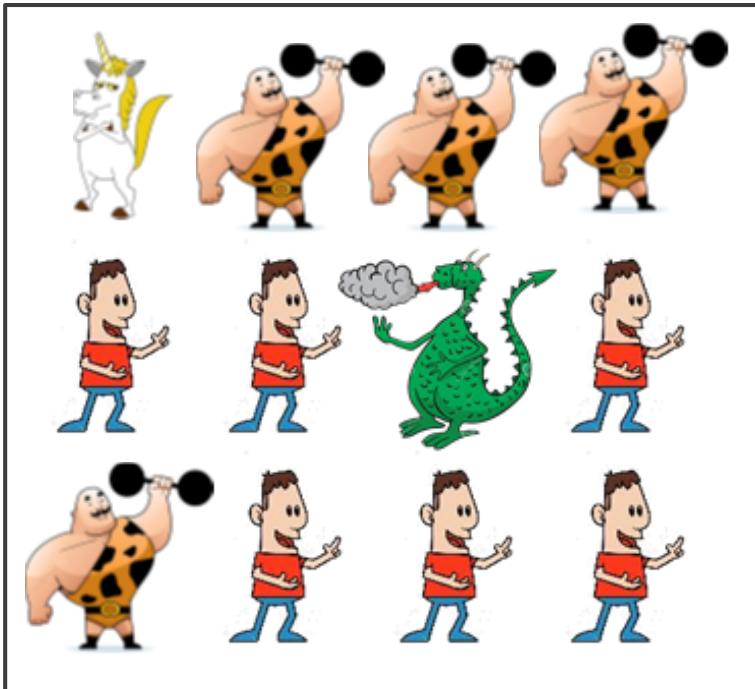
\approx

$E[Y_1(\text{Clones})]$



IDENTIFYING THE ESTIMAND

- But... determine the '**average causal effect**' of **MAGIC BEANS**, we must first **identify** our **estimand**
- The **estimand** is the quantity of interest, the thing we *wish* to estimate
- **Identification** is the first step of estimating a causal effect



IDENTIFYING THE ESTIMAND

- This may sound trivial but surprisingly few data scientists stop to think what they really want before they start modelling!
- The **estimand** is ‘**identified**’ from the appropriate ‘**counterfactual contrast**’
- Suppose we want to know the **average causal/treatment effect (ATE)** of magic beans ($X=$ Beans) on the **probability (P)** of turning into a dragon ($Y=$ Dragon)
 - The estimand can therefore be obtained by contrasting:

$P(Y_{\text{Beans}}=\text{Dragon})$ vs $P(Y_{\text{NoBeans}}=\text{Dragon})$

- Whether you chose to divide or subtract these depends on whether you want to obtain absolute or relative effect estimates

ESTIMATING OUR ESTIMAND

We turn our **estimand** into our **estimate** by applying an **estimator** (!!!)

ESTIMAND



E.g. The true difference in Y
due to exposure

ESTIMATOR

Method

1. Preheat your oven to 190°C /170°F / Gas Mark 5. Grease and line the base of 2 cake tins, one 8 inch/20cm and one 6 inch/15cm
2. Cream together the butter and caster sugar until light and fluffy.
3. Add the eggs one at a time with a spoonful of flour and blend in well.
4. Sift in the flour and baking powder and gently fold in. Finally add the milk and mix until you have a smooth batter.
5. Pour 1/3 of the batter into the small tin and 2/3 into the large tin.
6. Bake on the same shelf in the preheated oven, the smaller tin at the front.
7. Check the smaller cake after 20 minutes. When it is cooked remove from the oven, leaving the larger one still baking. The large cake should be done by 30 minutes.
8. Leave the cakes for 5 minutes in the tins, then turn out onto a rack to cool completely.
9. To make the icing, beat together the butter and icing sugar, add the vanilla and then the milk. Whisk the icing hard using an electric stand mixer if you can. Whisk it for 5 minutes and it will become really pale and light.

E.g. Your regression model

ESTIMATE



E.g. the estimated difference
in Y from model coefficient

ESTIMATING OUR ESTIMAND

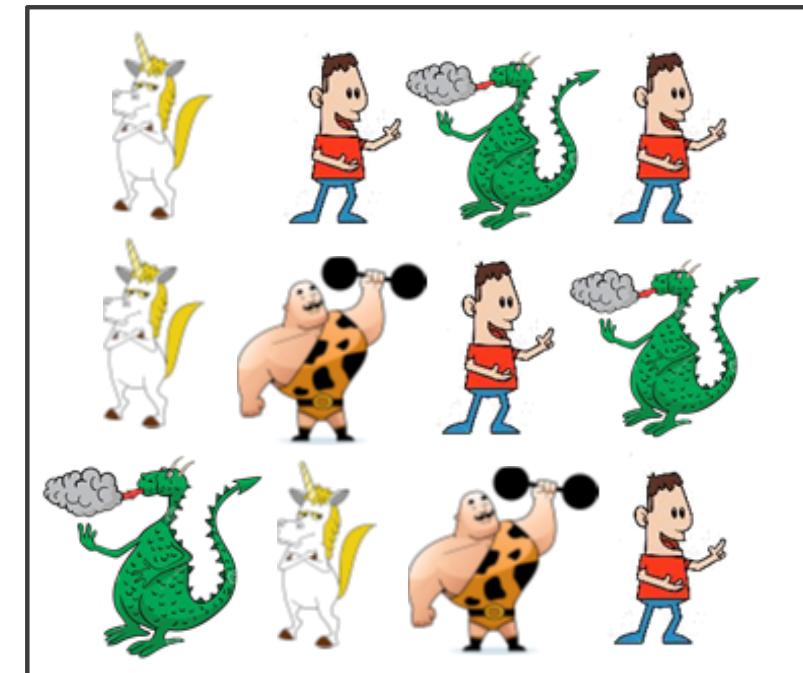
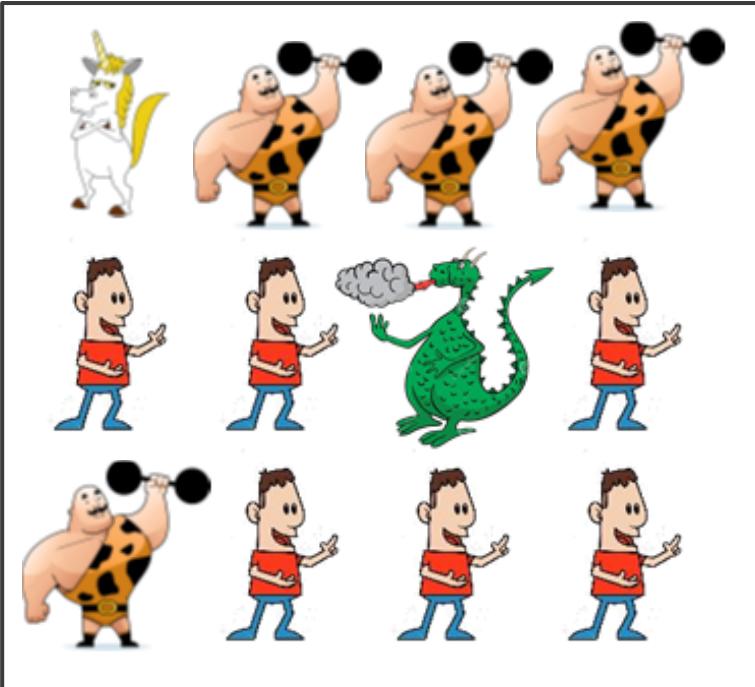
Estimated causal effect of **MAGIC BEANS** on $p(\text{Dragon})$ is:

$$P(Y_{\text{NoBeans}} = \text{Dragon}) = 1/12$$

$$P(Y_{\text{Beans}} = \text{Dragon}) = 3/12$$

$$\text{Causal Risk Ratio} = (3/12)/(1/12) = 3$$

$$\text{Causal Risk Difference} = (3/12 - 1/12) = 17\%$$



THE PHILOSOPHY OF ESTIMATION



- Causal inference encourages **estimation** not **testing**
- **Testing** is a binary approach – we ask whether a ‘significant’ effect is observed
 - e.g. **null-hypothesis significance testing** asks: are the observed data consistent with the null distribution
 - testing encourages bad practices (e.g. **p-hacking**)
- **Interval estimation** is a continuous approach – we seek the most accurate estimate and measure of uncertainty
 - estimation is less glamorous; it places trust in the accumulation of collective knowledge
 - estimation encourages good practices (e.g. **quantitative bias analysis**)

PROBLEMS WITH TESTING

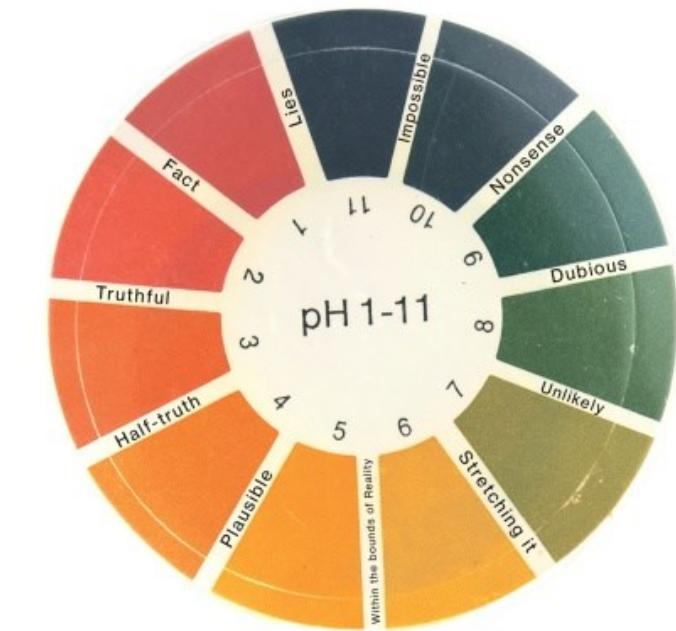
Together with perverse incentives and the flawed publication system, null-hypothesis significance testing can be dangerous and misleading

- Honest researchers may inadvertently torture their data to produce ‘significance results’, at the cost of accuracy
- Fluke findings are often presented as pre-planned investigations (**‘hypothesizing after the results are known’** or **HARKing**), a type of **Texas sharpshooter fallacy**
- Less honest researchers can use a range of **p-hacking** and **fishing** approaches (including selective reporting) to build an apparent evidence that doesn’t exist

‘Required reading for everyone’ ADAM RUTHERFORD

SCIENCE FICTIONS

STUART RITCHIE



Exposing Fraud, Bias,
Negligence and Hype in Science

PROBLEMS WITH TESTING

 [REDACTED]
Director, Payer Decision Sciences at
GSK

[View profile](#)
[Message](#)

Industry Insurance
Region Durham, North Carolina
Skills **Business Analytics, Segmentation, Databases**, and 45 other skills
Groups **Global Analytics Network, FiercePharma: A Network of Pharmaceutical Professionals, Pivotal: Data Science Group**, and 30 other groups

[REDACTED] 6d
ese comments read a bit like they are advising to use (see Simmons, Nelson, & Simonsohn, 2011). Trying to lower p hacking and considered unethical.

[REDACTED] 5d
to see cause and effect out come out of your research and if will be the problem

[REDACTED] 3d
n in mind ...of changing insignificant into significant....when most like defiling the very sanctity of scientific approach...

[REDACTED] 5h
Depending on your research question and design, sometimes I have found it useful to use decision trees to identify a sub-population within my sample that demonstrate a statistically significant and meaningful relationship between key variables.

For example, imagine I'm testing the efficacy of my "migraine reduction program", and find it does not significantly reduce the intensity, duration, or treatment cost of migraines in my sample. If I have additional details about my sample (demographics, comorbidities, lifestyle, etc.), I can use a decision tree to see if any subpopulation defined by those additional details DID show the desired effect of my intervention. This finding should serve as the a priori hypothesis tested in a new sample before any efficacy claims are made.

[Like](#) [REDACTED] 1h

AN END TO TESTING

American Statistical Association 2016
The American Statistician, 70: 129-133

ASA Statement on Statistical Significance and P-Values

1. Introduction

Increased quantification of scientific research and a proliferation of large, complex datasets in recent years have expanded the scope of applications of statistical methods. This has created new avenues for scientific progress, but it also brings concerns about conclusions drawn from research data. The validity of scientific conclusions, including their reproducibility, depends on more than the statistical methods themselves. Appropriately chosen techniques, properly conducted analyses and correct interpretation of statistical results also play a key role in ensuring that conclusions are sound and that uncertainty surrounding them is represented properly.

Underpinning many published scientific conclusions is the concept of "statistical significance," typically assessed with an index called the *p*-value. While the *p*-value can be a useful statistical measure, it is commonly misused and misinterpreted. This has led to some scientific journals discouraging the use of *p*-values, and some scientists and statisticians recommending their abandonment, with some arguments essentially unchanged since *p*-values were first introduced.

In this context, the American Statistical Association (ASA) believes that the scientific community could benefit from a formal statement clarifying several widely agreed upon principles underlying the proper use and interpretation of the *p*-value. The issues touched on here affect not only research, but research funding, journal practices, career advancement, scientific education, public policy, journalism, and law. This statement does not seek to resolve all the issues relating to sound statistical practice, nor to settle foundational controversies. Rather, the statement articulates in nontechnical terms a few select principles that could improve the conduct or interpretation of quantitative science, according to widespread consensus in the statistical community.

a proposed model for the data. The most common context is a model, constructed under a set of assumptions, together with a so-called "null hypothesis." Often the null hypothesis postulates the absence of an effect, such as no difference between two groups, or the absence of a relationship between a factor and an outcome. The smaller the *p*-value, the greater the statistical incompatibility of the data with the null hypothesis, if the underlying assumptions used to calculate the *p*-value hold. This incompatibility can be interpreted as casting doubt on or providing evidence against the null hypothesis or the underlying assumptions.

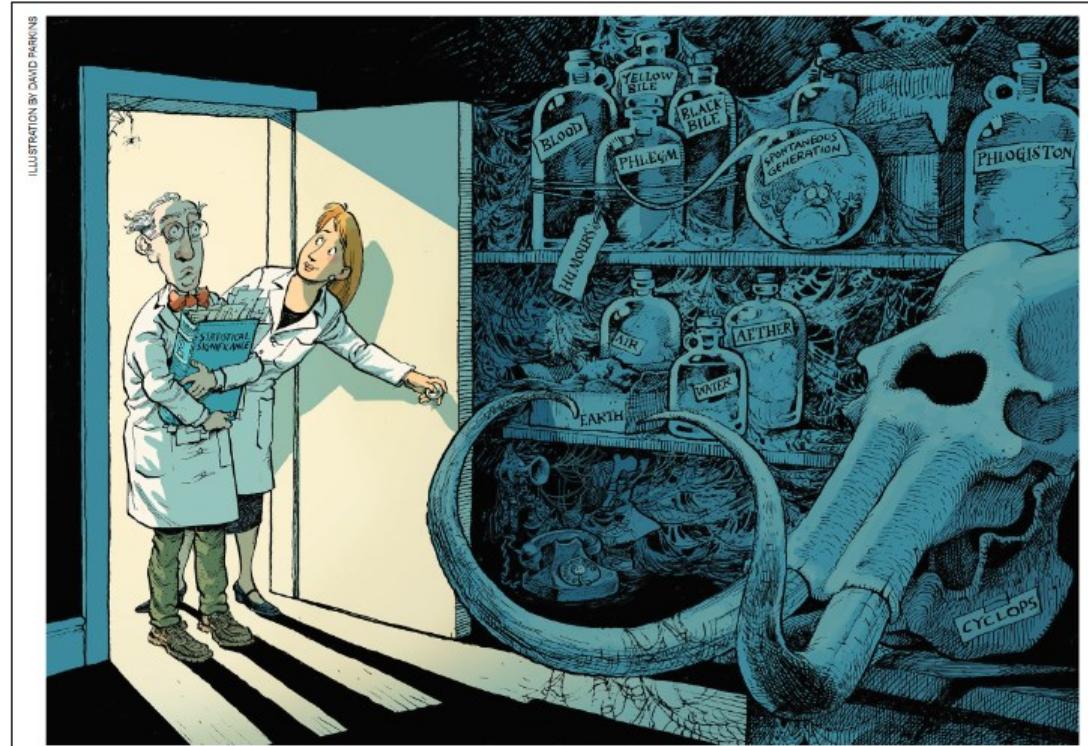
2. *P*-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

Researchers often wish to turn a *p*-value into a statement about the truth of a null hypothesis, or about the probability that random chance produced the observed data. The *p*-value is neither. It is a statement about data in relation to a specified hypothetical explanation, and is not a statement about the explanation itself.

3. Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold.

Practices that reduce data analysis or scientific inference to mechanical "bright-line" rules (such as " $p < 0.05$ ") for justifying scientific claims or conclusions can lead to erroneous beliefs and poor decision making. A conclusion does not immediately become "true" on one side of the divide and "false" on the other. Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of

Amrhein et al 2019 *Nature*, 567: 305-307



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

FREQUENTIST CONFIDENCE INTERVALS

- **Interval estimation** is about more than just calculating the **(frequentist) confidence interval!**
- (Frequentist) confidence intervals describe the degree of estimate uncertainty **ONLY due to random sampling variation...**
 - assuming the model is correct
 - assuming no bias occurs!



Ken Rothman
@ken_rothman

...

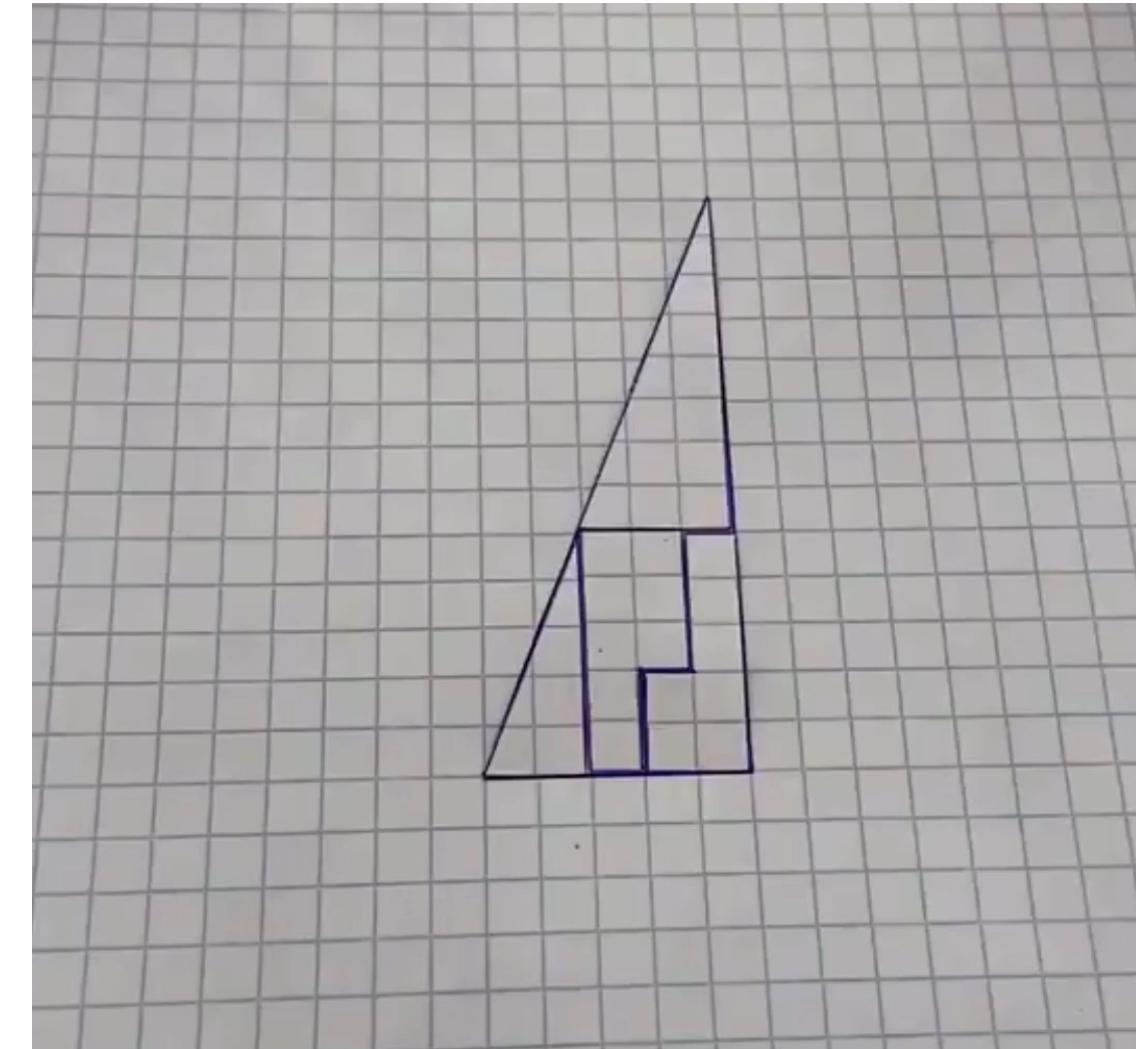
My toughest exam question: as study size increases,
the prob that a 95%CI includes the truth a) increases b)
decreases c) remains the same.

4:55 AM · Mar 24, 2015 · Twitter Web Client

DISCUSSING ERROR AND BIAS

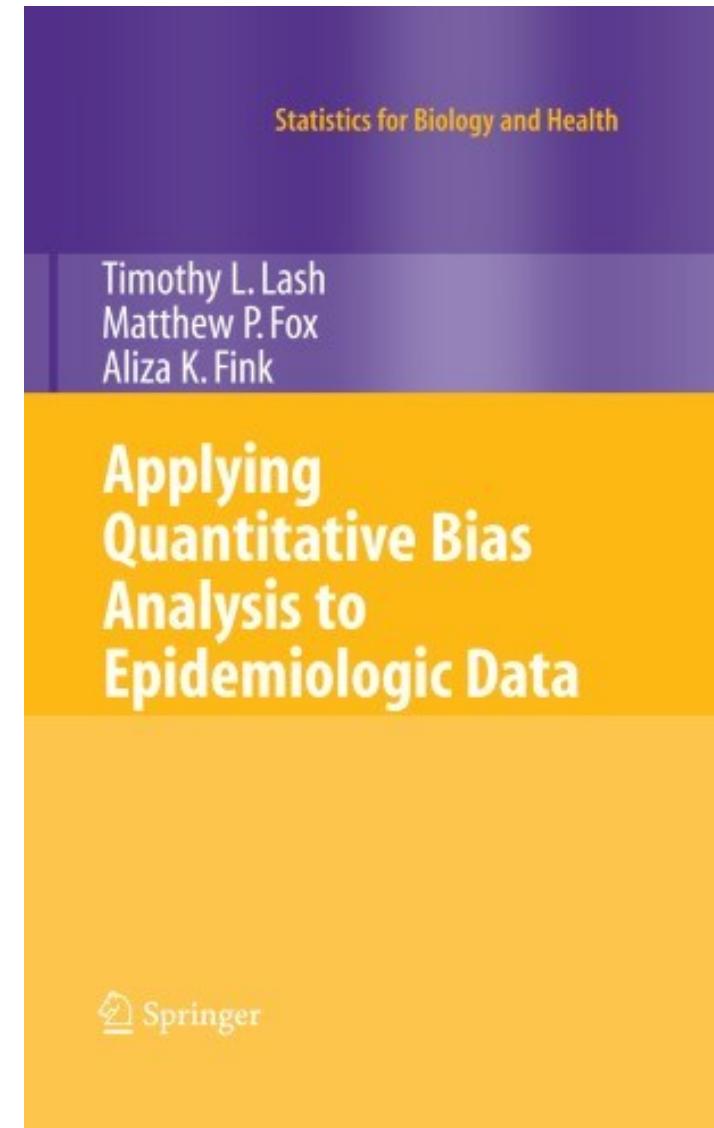
Most good observational studies:

1. Conduct study using appropriate methods, accounting for measured errors and biases
2. Calculate point estimate & confidence intervals
3. Interpret based on practical ‘significance’ of point estimate and degree of uncertainty
4. May conduct some sensitivity analyses
5. Most residual sources of error and bias are simply discussed qualitatively



QUANTITATIVE BIAS ANALYSIS

- Rather than adding our *qualitative* judgements to the discussion, we should seek *quantitative* assessments
- **Quantitative bias analysis** is a form of sensitivity analysis focused on estimating error and bias due to unmodelled or imperfectly measured issues
- External information (known as '**bias parameters**', usually from validation studies or the literature) are used to estimate:
 - magnitude of bias
 - direction of bias
 - uncertainty added by bias



THE IDENTIFIABILITY ASSUMPTIONS

According to Rubin, accurately **identifying** and estimating a causal effect **estimand** requires three conditions (known as the **identifiability assumptions**):

- **Exchangeability**
- **Positivity**
- **(Causal) consistency**

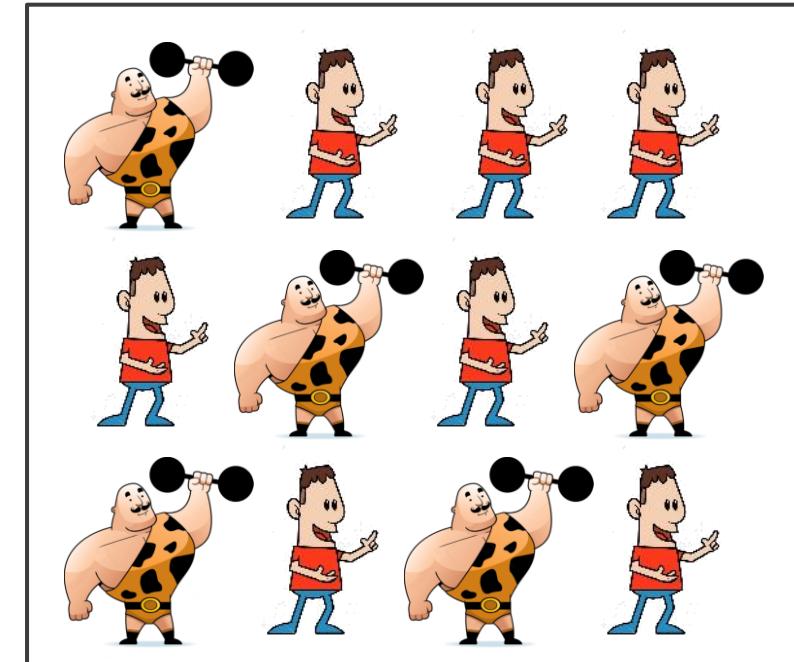
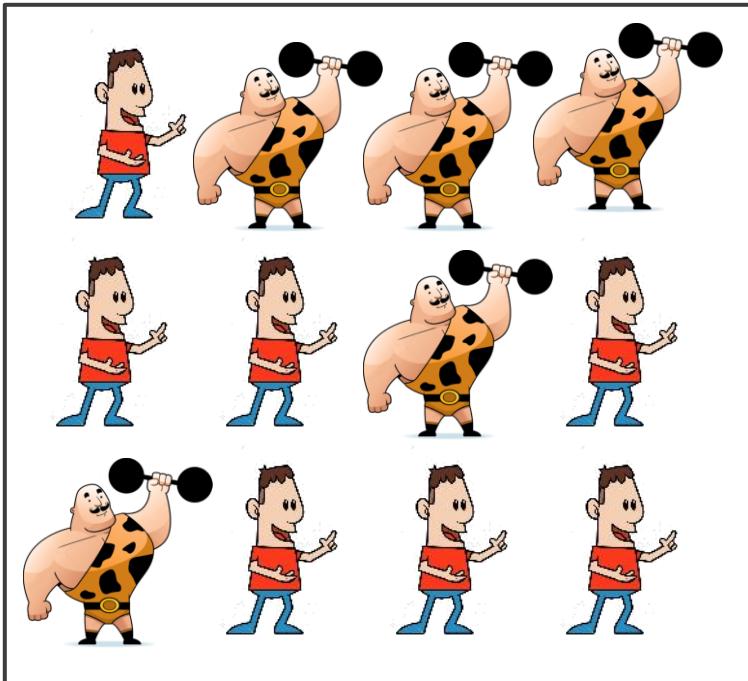
Consistency is sometimes also combined with another condition (**no interference**) and termed the **Stable Unit Treatment Values Assumption (SUTVA)**



EXCHANGEABILITY

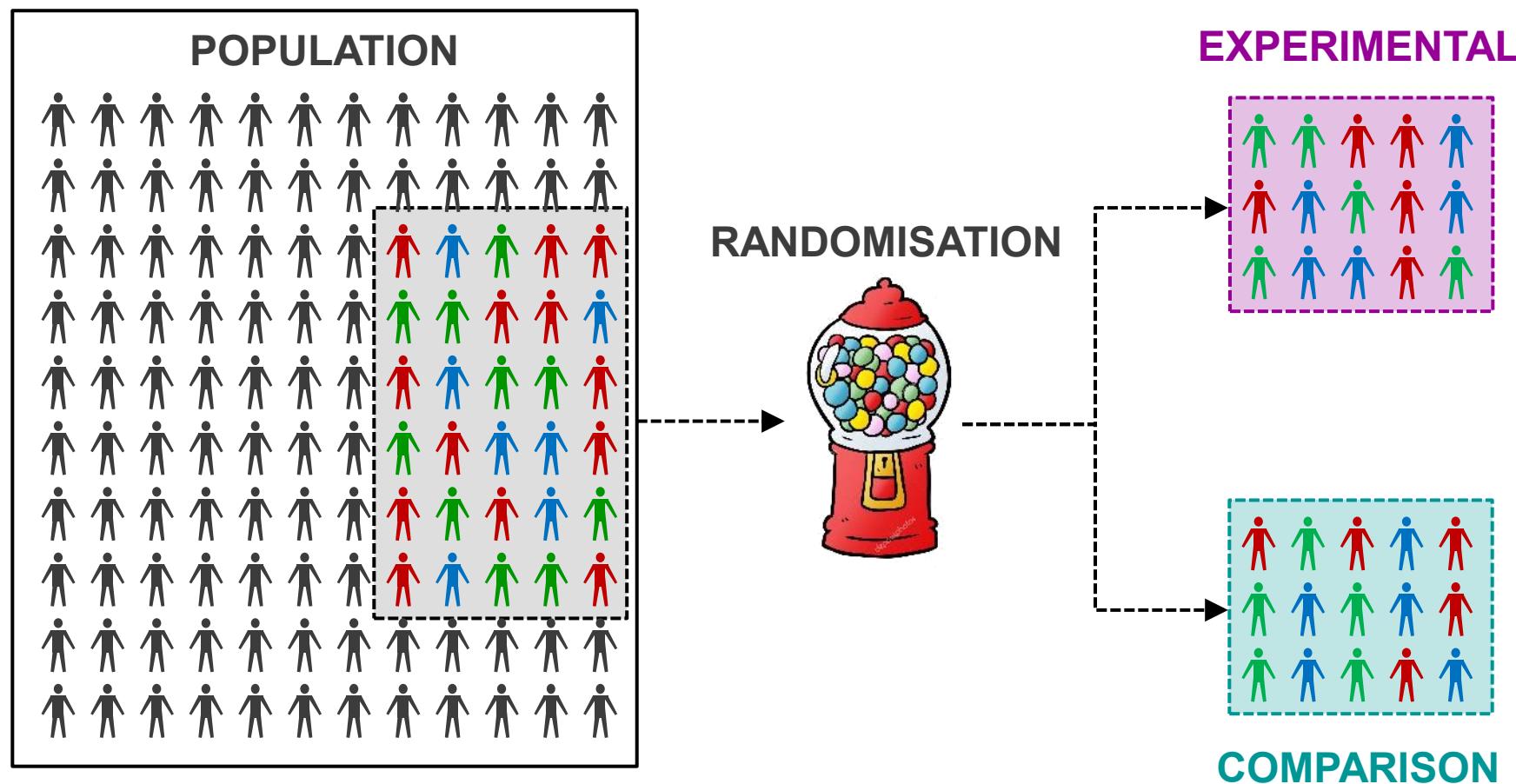
- **Exchangeability (AKA ignorability):**

- The probability of being assigned to a particular value of the exposure ($X=x$) should be independent of the probability of the outcome, $P(Y=y)$
- I.e. the units of analysis should have similar risks of the outcome



UNCONDITIONAL EXCHANGEABILITY

- The easiest way to achieve **exchangeability** is through **randomisation**
- If exposure is assigned **at random** it cannot be related to the probability of the outcome

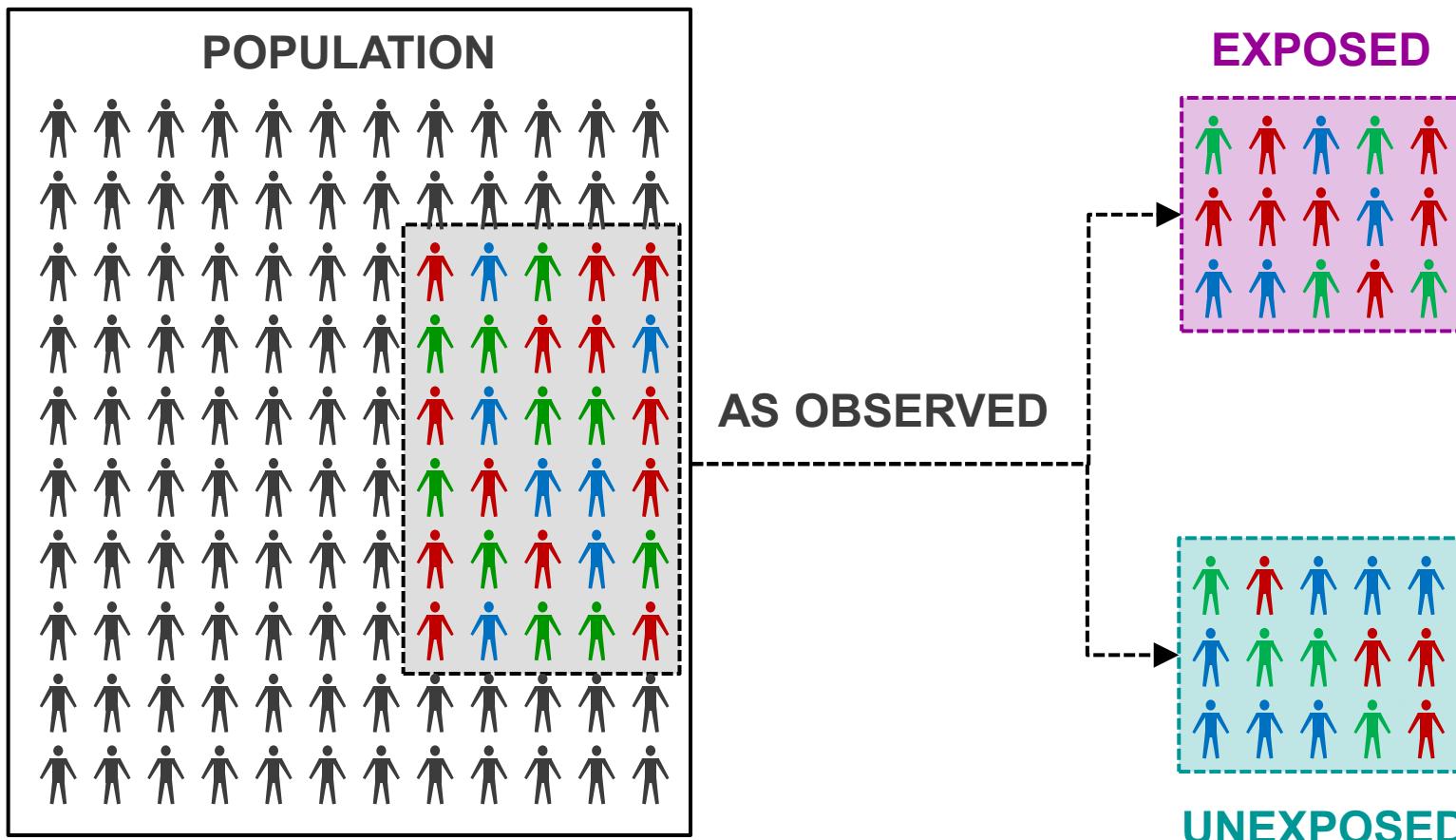


Units of analysis are **unconditionally exchangeable**, because exposure was assigned *independently* of the probability of the outcome

CONDITIONAL EXCHANGEABILITY

Without randomisation we have to aim for **conditional exchangeability**

- This does **NOT** mean **conditioning** on all differences / imbalances!!!
- We condition on variables that cause *both* the exposure & outcome (**confounders**)



Units of analysis are *not* unconditionally exchangeable, because exposure assignment is related to probability of outcome!

Conditional exchangeability occurs when exposure assignment is independent of outcome probability, *conditional on* C_1, C_2, C_3 etc

POSITIVITY

- **Positivity (AKA the experimental treatment assignment assumption):**
 - All levels of the exposure must be *possible* in all participants
 - All levels of the exposure must be *observed* for all subgroups being studied (i.e. for all combinations of all variables being conditioned)
- A full or partial **positivity violation** occurs when certain values are not observed (or are extremely rare) in certain subgroups
- A **random positivity violation** occurs when certain values are not observed due to **data sparsity**, but these values are still *possible*
 - here, we may be able to borrow information from our other observations
- A **structural positivity violation** occurs when certain values are not observed because they are *not possible*
 - here, we must restrict our inferences to those about whom we have information

EXAMPLE: POSITIVITY VIOLATION

We recently sent you a letter about measuring Roxanne's height and weight in school as part of the National Child Measurement Programme. The measurements have now been completed.

Knowing if your child's weight is within the healthy range for their age, sex and height can help you make informed choices about their lifestyle.

Roxanne's results	
Height (cm)	110.4
Weight (kg)	23.60
Date of measurement	06 January 2017

These results suggest that your child is very overweight for their age, sex and height. Being very overweight can lead to health problems for your child, such as high blood pressure, early signs of type 2 diabetes and low self-confidence. But you and your child can make simple changes to be more active and eat more healthily. As a first step, please call us on the number above to find out how you can benefit from free local support. You can also:

- Take a look at the tips in the enclosed leaflet
- Go online for practical advice at: www.nhs.uk/change4life and www.nhs.uk/ncmp4

- Causal effect of child '**fat letters**' (exposure) on adult weight (outcome)
- Letters sent to parent of 'obese' children

- The level of exposure (letter vs no letter) is therefore determined by the 'current weight', breaking the assumption that all levels of exposure must be possible

NO INTERFERENCE

- **No interference (AKA no spillover effects, no dependent happenings)**
 - First part of the **Stable Unit Treatment Value Assumption (SUTVA)**
 - One unit's exposure must not affect the potential outcomes of other units
- Interference commonly occurs where units are networked or clustered, such that effects can spread between participants over time
- **Example:** Vaccination trials
 - An unvaccinated individual has their risk of the virus lowered by other people getting vaccinated

(CAUSAL) CONSISTENCY

- **Causal consistency (AKA treatment variation irrelevance)**

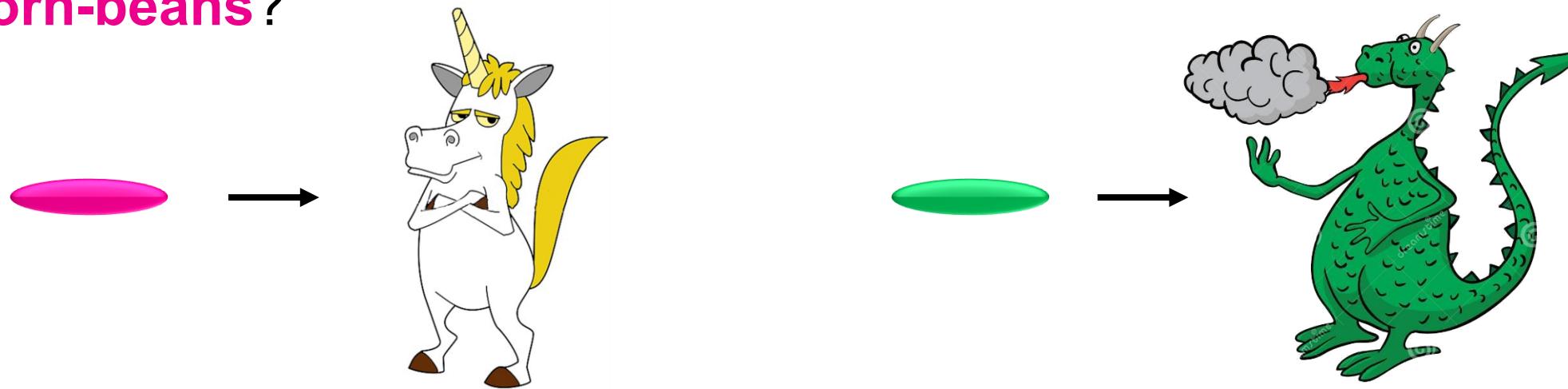
- The effect of the exposure must be the same whether observed (i.e. in the factual universe) as if given by intervention (i.e. in the counterfactual universe)
- Second part of the **Stable Unit Treatment Value Assumption (SUTVA)**
- Rehkopf et al: “*there are no two “flavors” or versions of treatment*”

Rehkopf et al 2016. The Consistency Assumption for Causal Inference in Social Epidemiology: When a Rose is Not a Rose. *Curr Epidemiol Rep.* 2016 Mar; 3(1): 63–71.

- Consistency may therefore be ‘violated’ if the *same* values of the exposure can have different effects in different units or contexts
- This is likely to occur when the exposure is poorly defined or multidimensional

EXAMPLE: CONSISTENCY VIOLATION

- To be consistent, the causal effect of 'eats **MAGIC BEAN**' needs to be the same for all people and in all environments
- But what if there were two types of **MAGIC BEAN**, **dragon-beans** and **unicorn-beans**?



- 'Eats **MAGIC BEAN**' would be an inconsistent exposure because $P(Y_1)$ would depend on whether the participant ate the dragon-beans or unicorn-beans

WHAT IS CONSISTENT ENOUGH?

Q1) Does water kill?

What, drinking it or swimming in it?

From: Hernan Ann Epidemiol. 2016 Oct; 26(10): 674–680.

Q2) Does drinking water kill?

We mean fresh water, not salty water.

Q3) Does drinking fresh water kill?

How much water? 1000 litres per day will kill you.

Q4) Does drinking a swig of fresh water kill?

What is the source of the water? Tap, fountain, directly from the river...

Q5) Does drinking a swig of water from the Broad Street pump kill?

Over which period?

Q6) Does drinking a swig of water from the Broad Street pump between 31 Aug and 10 Sep kill?

OK, compared with what? With drinking 3 litres of beer?

Q7) Does drinking a swig of water from the Broad Street pump between 31 Aug and 10 Sep kill

compared with drinking all your water from other pumps?

What about other factors that may affect the causal effect of interest?

Q8) Does drinking a swig of water from the Broad Street pump between 31 Aug and 10 Sep and not initiating a rehydration treatment if diarrhoea starts kill ...

(CAUSAL) CONSISTENCY

Lessons:

- Think carefully about what your exposure *means*!
- Try to define it as clearly as reasonably possible
- It may help to consider what '**well-defined intervention**' you would use to study your exposure of interest using a randomised controlled trial
- If your exposure cannot be (conceivably) manipulated, there is a pragmatic argument for considering alternative (e.g. downstream) exposures

For complex / multi-dimensional exposures:

- Your causal effect will be an (imperfect) average of many effects
- It may not **transport** as well to other settings or populations

RECOMMENDED READING

Books

- Ritchie S. Science fictions: Exposing fraud, bias, negligence and hype in science. Random House; 2020 Jul 16.
- Westreich D. Epidemiology by design: a causal approach to the health sciences. Oxford University Press; 2019.
- Morgan SL, Winship C. Counterfactuals and causal inference. Cambridge University Press; 2015.

Papers

- Rubin DB. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*. 2005 Mar 1;100(469):322-31.
- Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*. 2016 Apr;31(4):337-50.
- Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Current epidemiology reports*. 2016 Mar;3(1):63-71.
- Westreich D, Cole SR. Invited commentary: positivity in practice. *American journal of epidemiology*. 2010 Mar 15;171(6):674-7.

SUMMARY

- Counterfactual reasoning helps with considering whether we are comparing like-with-like
- This process is formalised by the **potential outcomes framework**
- In this framework, we estimate causal effects by:
 - 1) **Identifying** our **estimand** of interest from its **counterfactual contrast**
 - 2) **Estimating** that quantity from our data using a statistical model (**estimator**)
- The **validity** of this estimate depends (initially) on four conditions:
 - **Exchangeability**: do the units of analysis have similar outcome probabilities?
 - **Positivity**: were all units of analysis eligible for all values of exposure?
 - **No interference**: does the exposure have any overspill effects on other participants?
 - **Consistency**: does the exposure have a consistent effect on the outcome?

09:00-09:30 REGISTRATION

09:30-10:00 WELCOME

10:00-10:30 LECTURE 1.1

10:30-11:00 DELEGATE INTRO

11:00-11:30 TEA & COFFEE

11:30-12:00 DELEGATE INTRO

12:00-12:45 LECTURE 1.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:15 LECTURE 1.3

15:15-15:30 Q&A

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 1.4

17:00-17:45 ACTIVITY 1-A

17:30-18:00 Q&A

1.4 - CAUSAL DAGS AND COVARIATE ROLES

GEORGIA



@GEORGIATOMOVA

MARK



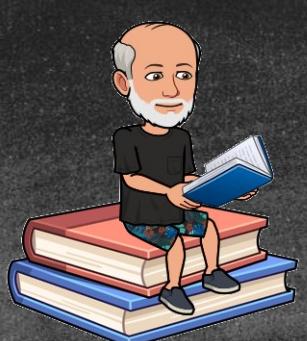
@STATSMETHODS

PETER



@PWTENNANT

ROB



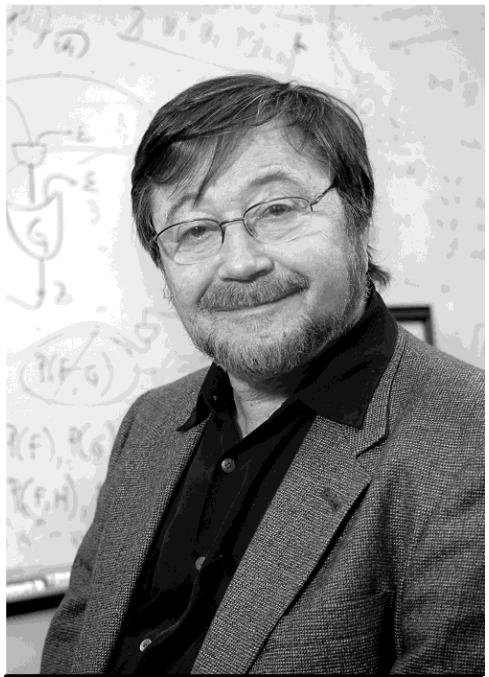
@WAYNEROBERTLONG

LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- Define the core features and language of causal **directed acyclic graphs**
- Recognise and describe **confounders**, **mediators**, and **competing exposures**
- Appreciate and describe the value of DAGs for distinguishing between these, and for identifying appropriate **adjustment sets**
- Distinguish between **unobserved confounding** and **residual confounding**
- Appreciate and explain the **Table 2 Fallacy** and its implications for interpretability

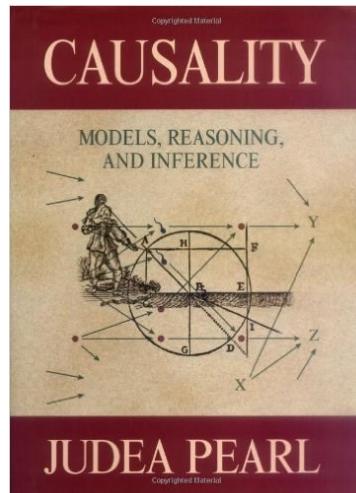
CAUSAL INFERENCE METHODS



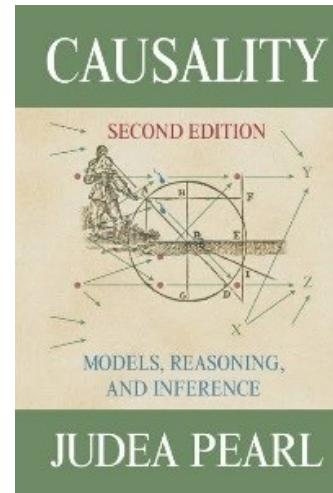
Judea Pearl, Computer scientist, philosopher, Turing Prize winner

Contemporary causal inference methods – like **Judea Pearl's 'Structural Causal Model'** – provide a formal mathematical and philosophical framework for estimating causal effects:

- **Probability theory**
- **Counterfactual reasoning**
- **Graphical model theory**



2000



2009

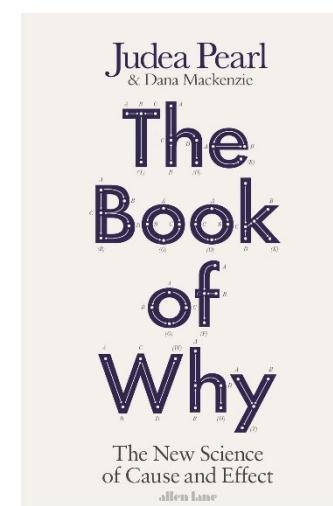


WILEY

CAUSAL INFERENCE
IN STATISTICS

A Primer

Judea Pearl
Madelyn Glymour
Nicholas P. Jewell



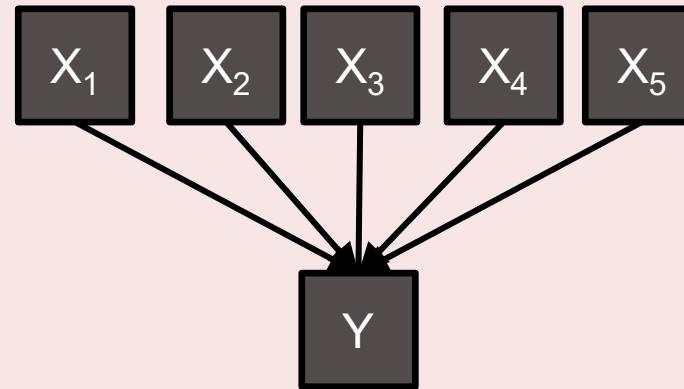
2018

#LEEDSCAUSALSCHOOL

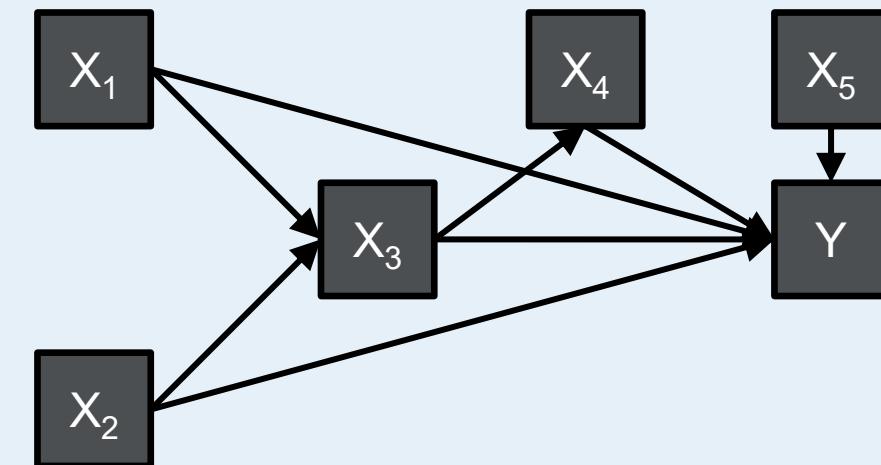
CAUSAL DIAGRAMS

- **Causal diagrams** – such as **directed acyclic graphs** – encode many of the ideas of e.g. the **potential outcomes framework** into diagrammatic form
- The first benefit is identifying which variables are **confounders** and need conditioning to obtain **conditional exchangeability**
- We must do this with outside knowledge; the statistical software cannot!

HOW THE SOFTWARE SEES IT

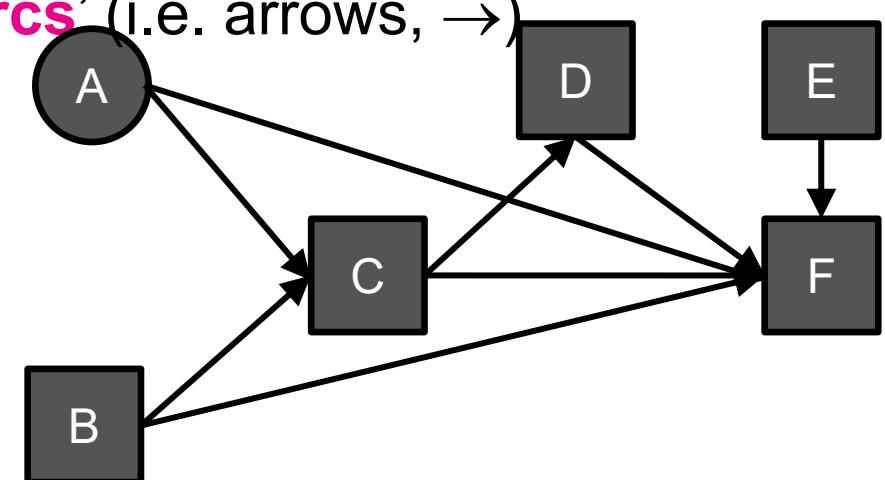


HOW NATURE CREATED IT



CAUSAL DIRECTED ACYCLIC GRAPHS (DAGS)

- **Causal diagrams** – such as **directed acyclic graphs (DAGs)** – help us to identify our **estimand** by encoding our theory of the **data generating mechanism**
- DAGs are **non-parametric** graphical representations of (hypothesised) causal relationships between variables, in which:
 - Variables represented as ‘**nodes**’ (e.g. A-F, below)
 - Causal relationships represented as directed ‘**arcs**’ (i.e. arrows, →)
 - There are no circular paths (hence ‘**acyclic**’)



ARCS AND CAUSATION

DAGs are *directed* because causality is *directed* (over time); a cause cannot occur *after* a consequence



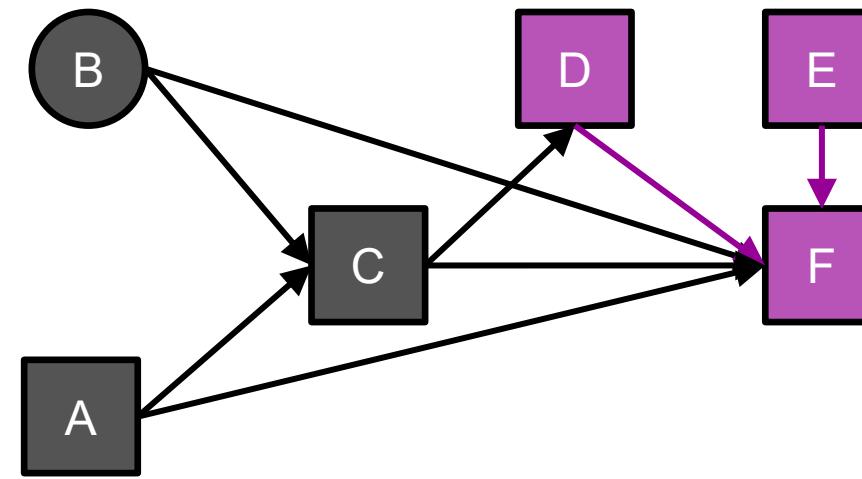
When we draw an **arc** between **X** and **Y**, we state that we believe:

- *Changing X modifies the probability of Y (probabilistic reasoning)*
- *If X had been different, Y would have been different (counterfactual reasoning)*
- **Y** ‘listens’ to **X**
- *If we could wiggle X, it would wiggle Y*

PATHS

A **path** exists between two variables if they are **connected by one or more arcs** (regardless of the arrow directions)

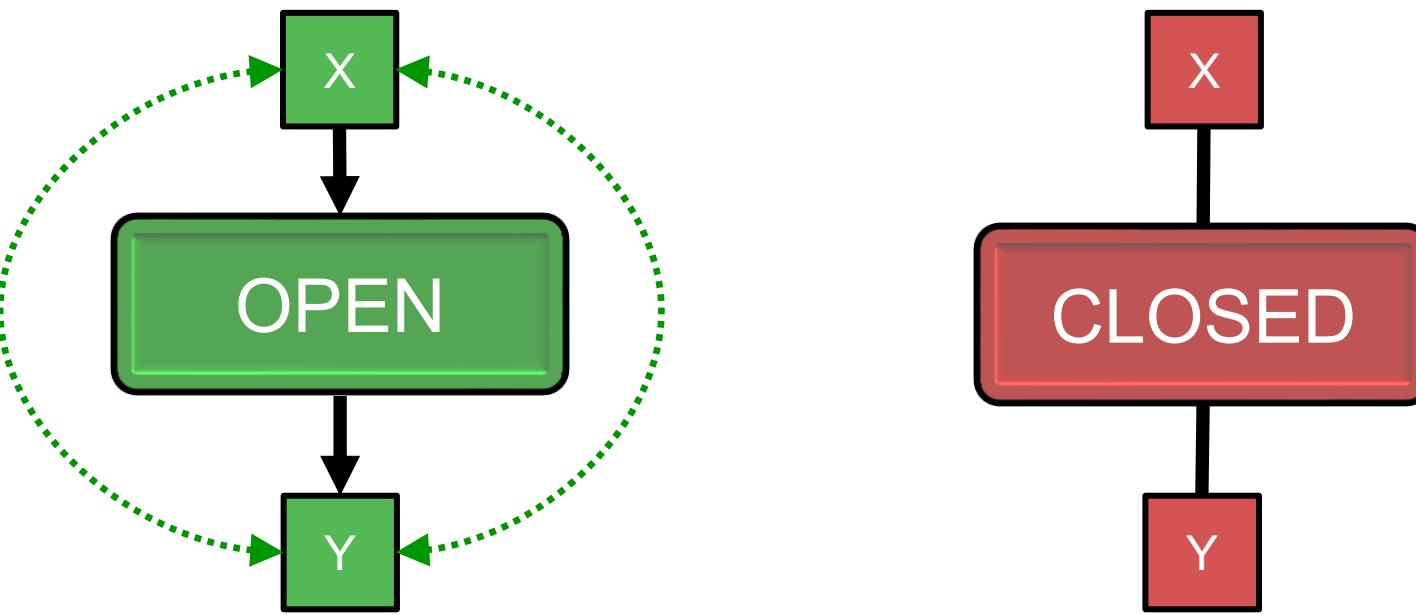
- e.g. $D \rightarrow F \leftarrow E$



OPEN AND CLOSED PATHS

A path may be **open** or **closed**

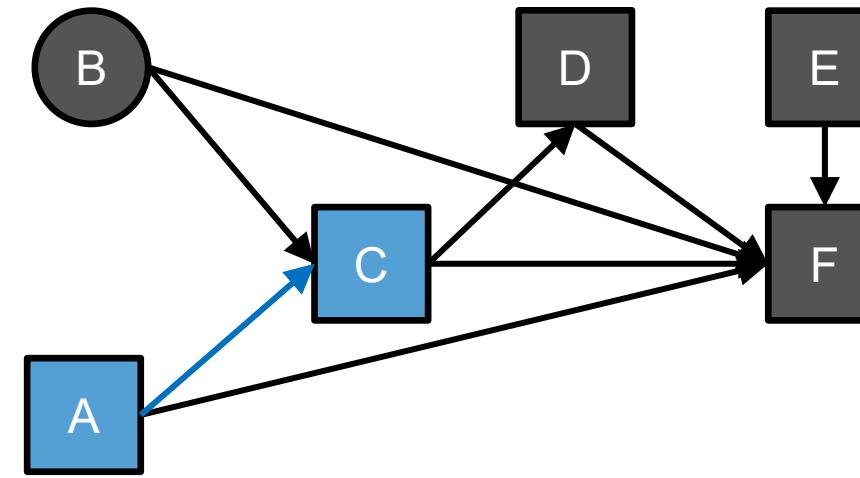
- **Open paths** transmits associations (AKA **correlations** or **dependencies**)
- **Closed paths** do not



CAUSAL PATHS

A **causal path** (AKA **directed path**) is one where all arrows run in the same direction

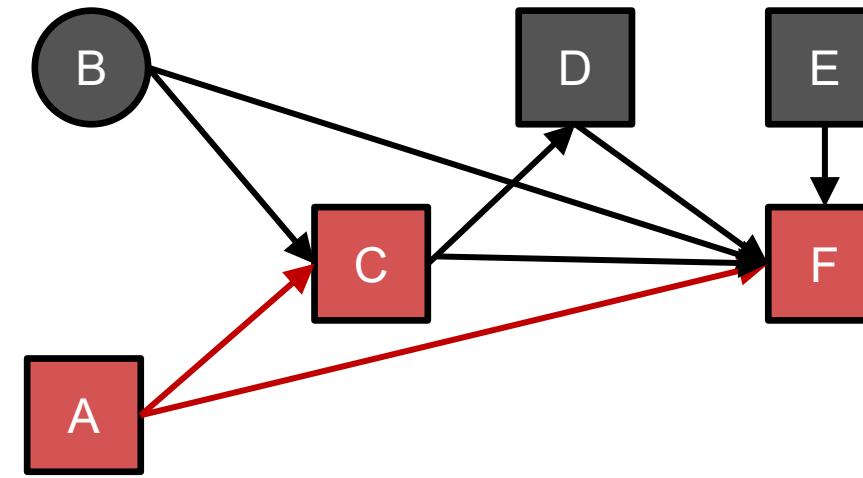
- e.g. $\mathbf{A} \rightarrow \mathbf{C}$



CONFOUNDING PATH

A **confounding path** (AKA **backdoor path**) is one where the arrows do not flow in the same direction – initially backwards, then forwards

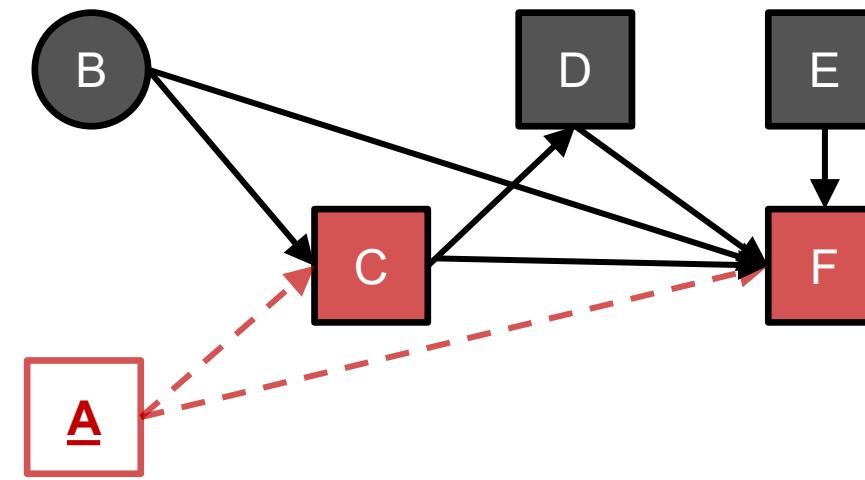
- e.g. **C ← A → F**



Without conditioning, **confounding paths** are **open** and will transmit dependencies between the variables caused by the confounder (e.g. A)

CONFOUNDING PATH

A **confounding path** can be **closed** by conditioning on the confounding node(s) – or a proxy thereof



Conditioning on A closes $C \leftarrow A \rightarrow F$

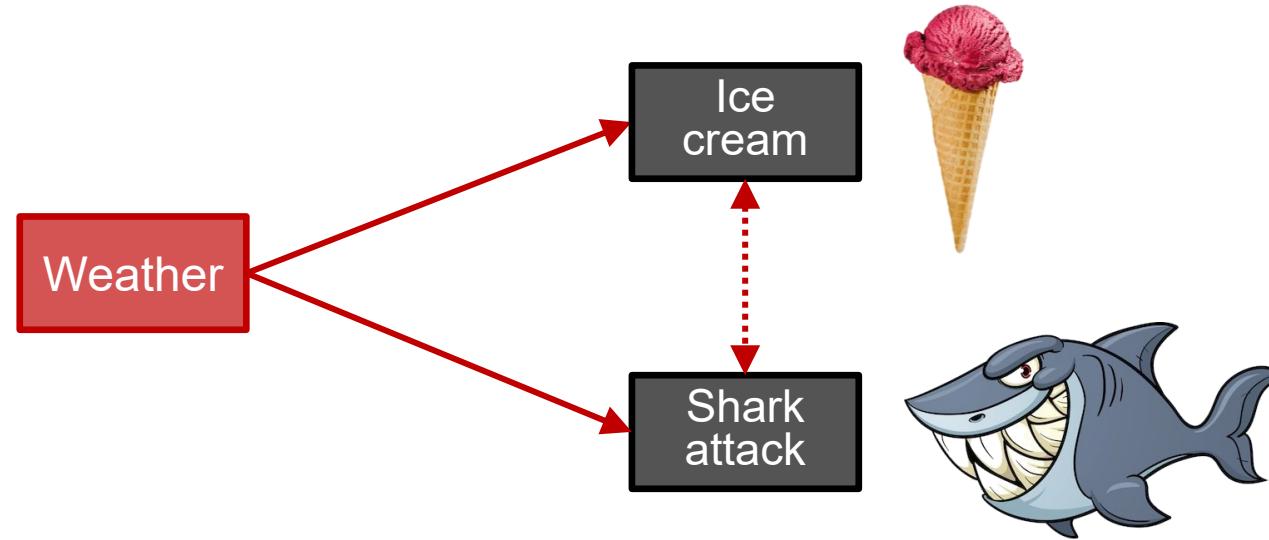
CONDITIONING (AKA ADJUSTING / CONTROLLING)

- **Conditioning** is the process of estimating a statistic (e.g. coefficient in a regression model) at **fixed levels of one or more other variables**
- **Restriction**
 - Estimating the effect in a sample with similar values of one or more other variables
 - e.g. Non-smokers only
- **Stratification**
 - Estimating the effect in strata with similar values of one or more other variables
 - E.g. Non-smokers, Ex-smokers, Current-smokers
- **Covariate adjustment**
 - Estimating the effect while controlling for values of one or more other variables
 - E.g. including smoking as a covariate in a regression model

EXAMPLE: CONFOUNDING BIAS

Example: The ‘Ice-Cream Hazard’

- **Ice cream consumption** is associated with a higher incidence of **shark attack**

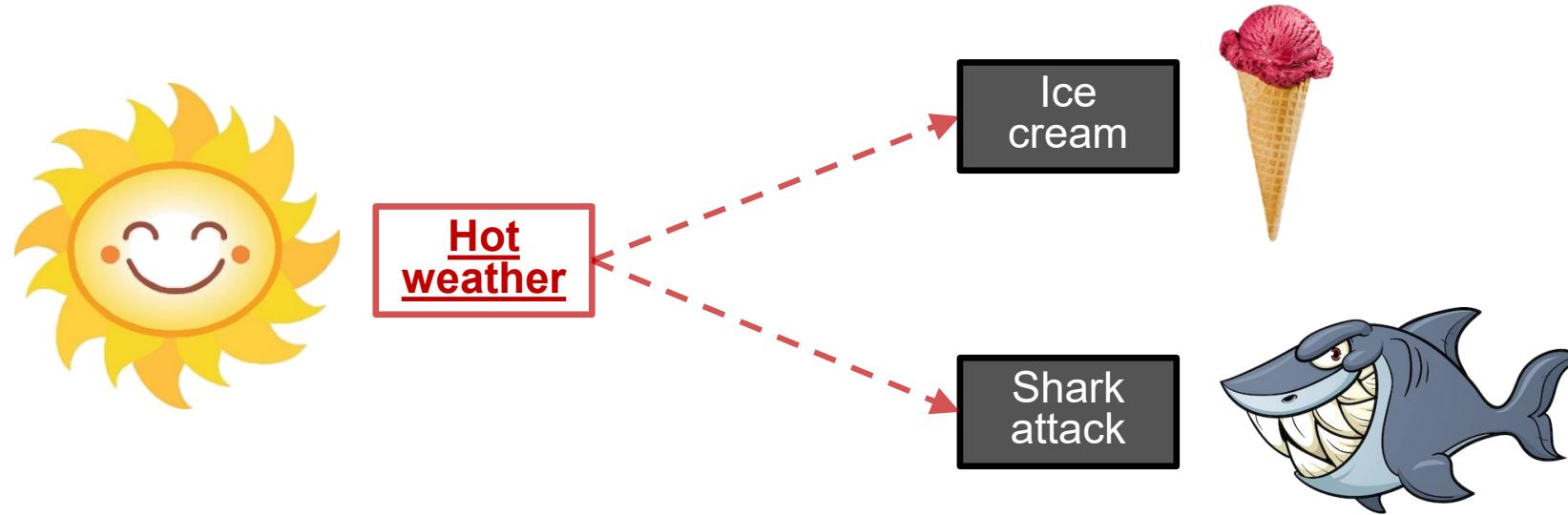


- Happens because **weather** causes **ice-cream consumption** and risk of **shark attack**
- **Weather** is a **confounder**, creating an **unconditional dependency** through:
 $\text{Ice-cream} \leftarrow \text{Weather} \rightarrow \text{shark attack}$

EXAMPLE: CONFOUNDING BIAS

Example: The ‘Ice-Cream Hazard’

- Association disappears if you examine days with similar weather:



- By conditioning on the **confounder** we close the path:
 $\text{Ice-cream} \leftarrow \text{Weather} \rightarrow \text{shark attack}$

ESTIMATING CAUSAL EFFECTS

DAGs help us to chose an appropriate **estimator** (i.e. help build our model)

They are useful for identifying and avoiding various common errors

BUT: They are **NOT** a substitute for thinking (quite the opposite!)

BUT: They cannot help you to *prove* whether an ‘effect’ is causal

- ✓ **Remember:** We are trying to produce the best possible *estimate*

DON'T SAY

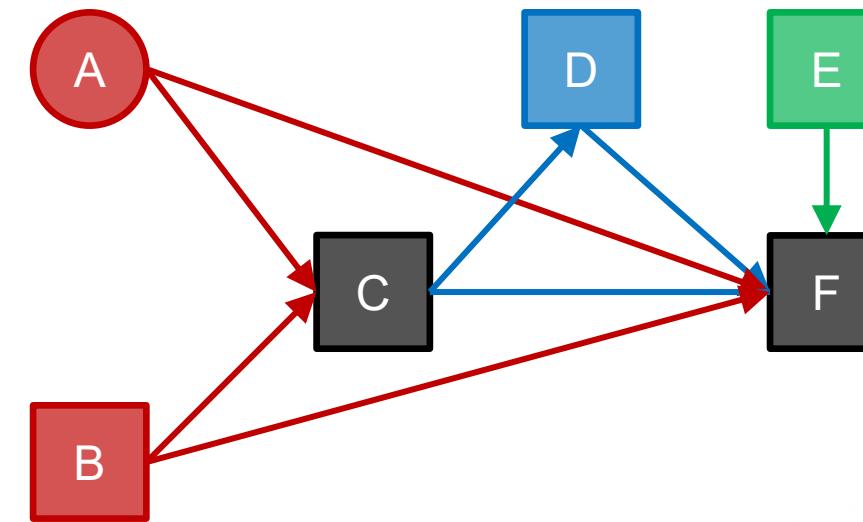
*“We found a **significant** causal effect of X on Y ($p<0.01$)”*

DO SAY

*“The estimated **total causal effect** of X on Y was RR=2.0 (95% CI: 1.5-3.0)”*

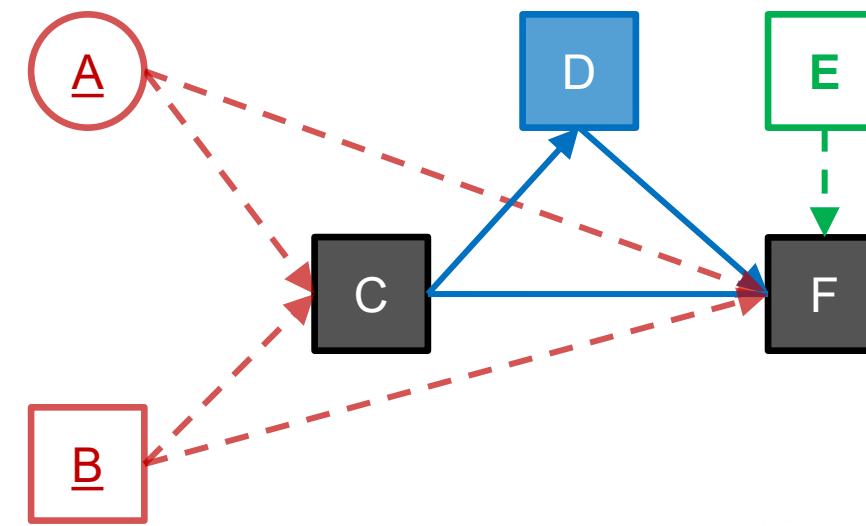
CONDITIONAL EXCHANGEABILITY

- To estimate the **total causal effect** of **C** on **F** (the ‘**focal relationship**’):
 - Want all **causal paths** to be **open**
 - Want all **confounding paths** to be **closed**



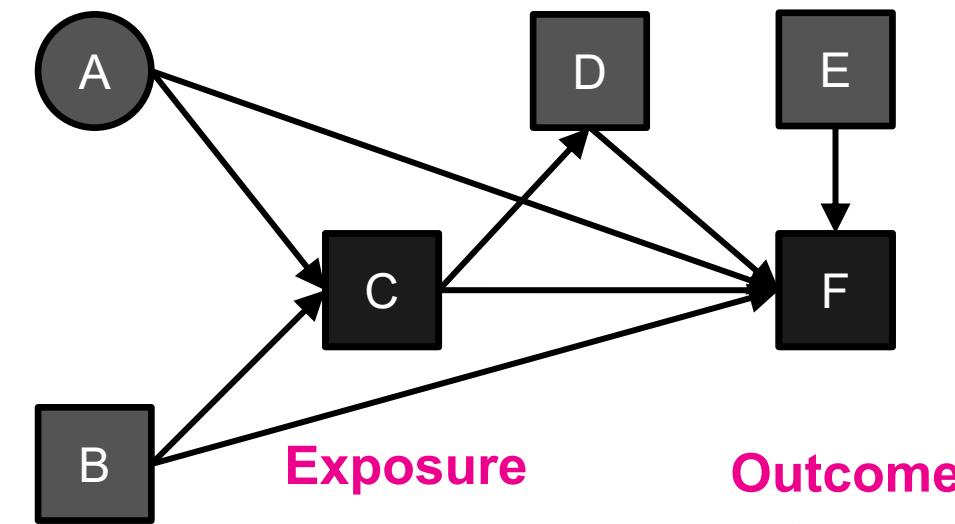
CONDITIONAL EXCHANGEABILITY

- To estimate the **total causal effect** of **C** on **F** (the ‘**focal relationship**’):
 - Want all **causal paths** to be **open**
 - Want all **confounding paths** to be **closed**
 - This means identifying and conditioning on all **confounders**



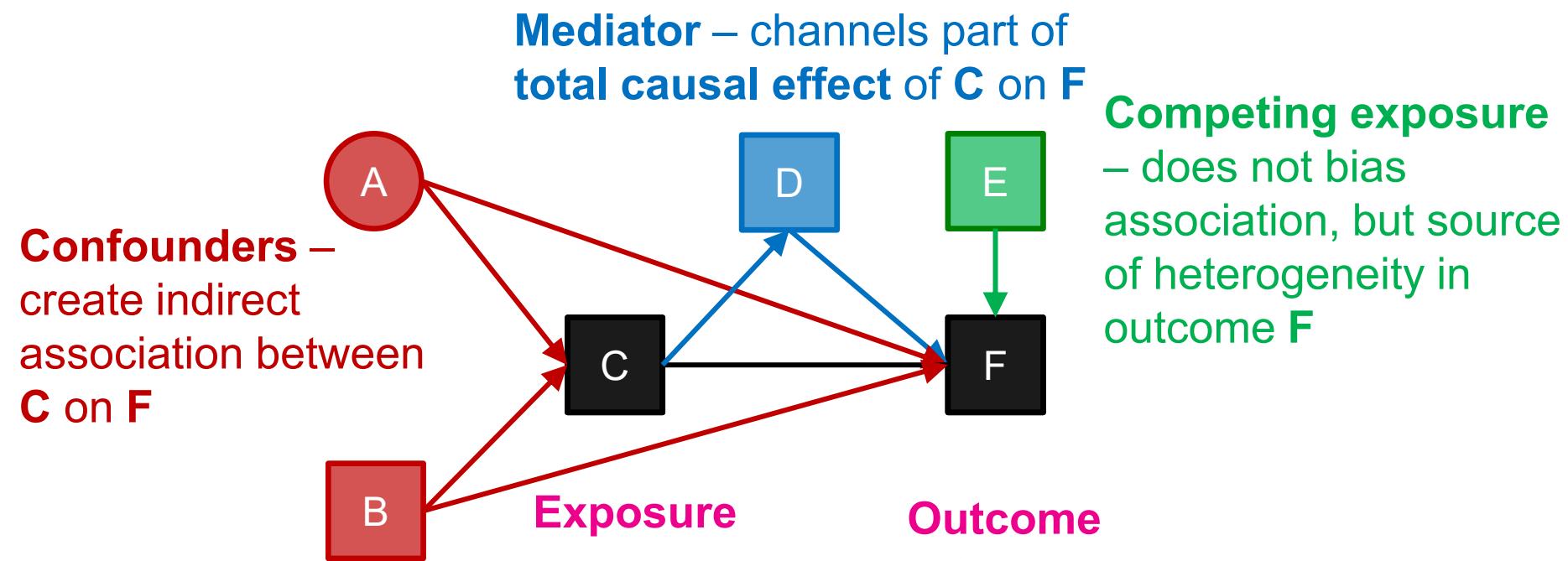
CONTEXTUAL VARIABLES

- The role of each variable - and which require **conditioning** - is defined by their relationship to your **focal relationship** of interest
- E.g. If considering **total causal effect** of C on F



CONTEXTUAL VARIABLES

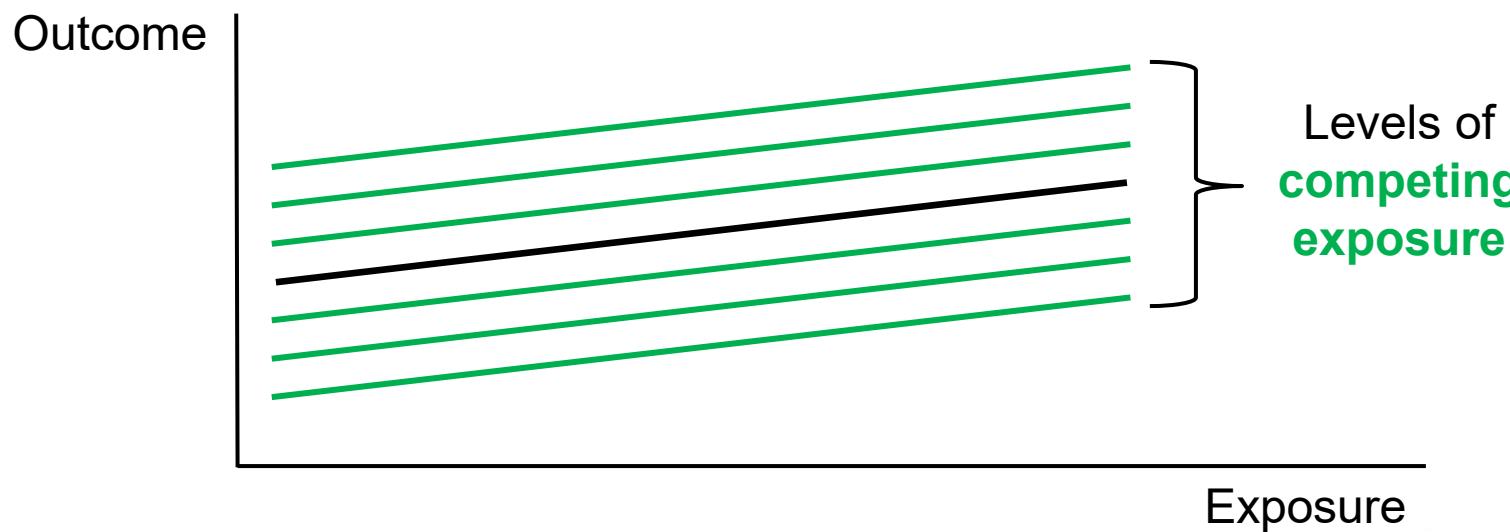
- The role of each variable - and which require **conditioning** - is defined by their relationship to your **focal relationship** of interest
- E.g. If considering **total causal effect** of C on F



COMPETING EXPOSURES

■ Competing exposures

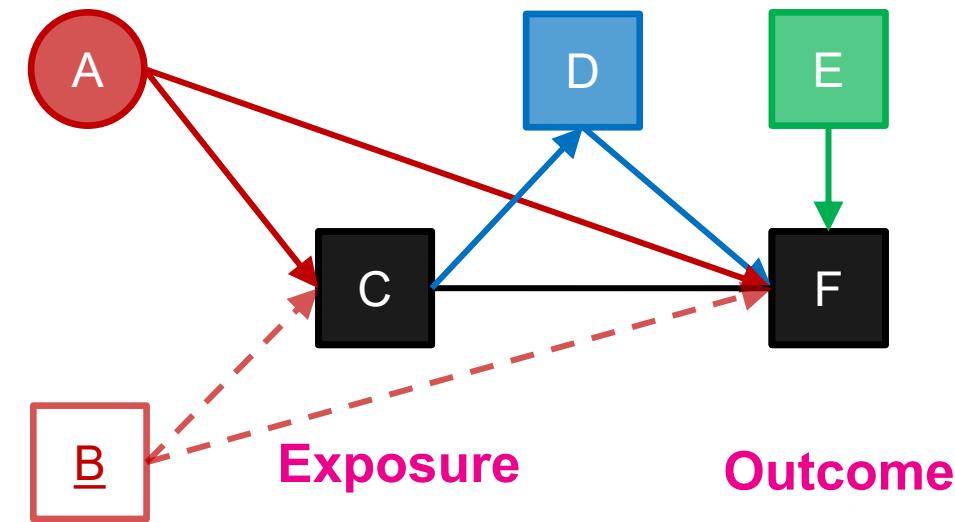
- Causes outcome (or is a proxy cause)
- Does **NOT** cause exposure and is **NOT** caused by exposure
- Does **NOT** bias effect of exposure on outcome; adds **heterogeneity** (increasing error > reducing precision)



UNOBSERVED CONFOUNDING

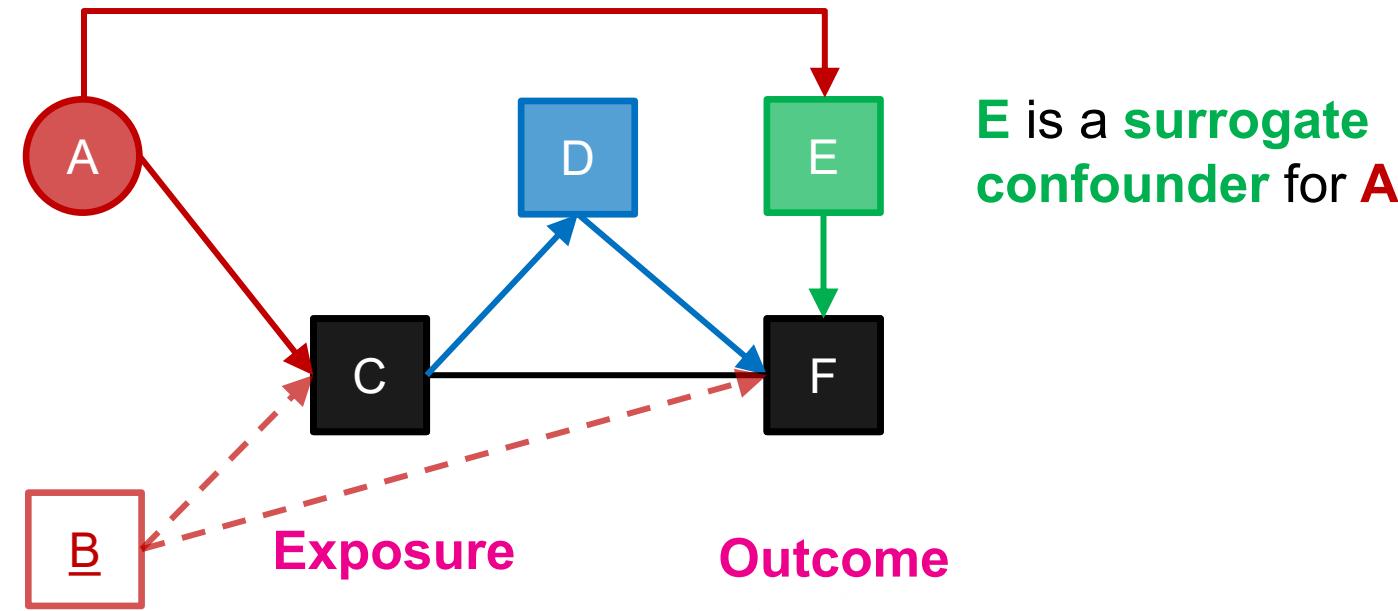
- **Unobserved confounding** is created by **confounders** that we haven't measured (and cannot condition on)
 - If we do measure it, but don't condition on it, it's **uncontrolled confounding**
- Unobserved variables are sometimes shown as **circles** or **ellipses**

A is unobserved so
cannot be conditioned,
creating unobserved
confounding



SURROGATE CONFOUNDERS

- Suppose **A** actually acts largely through **E**
- Now **E** is a **surrogate (proxy) confounder** for **A**



SURROGATE CONFOUNDERS

Sometimes the confounder can't be measured directly:

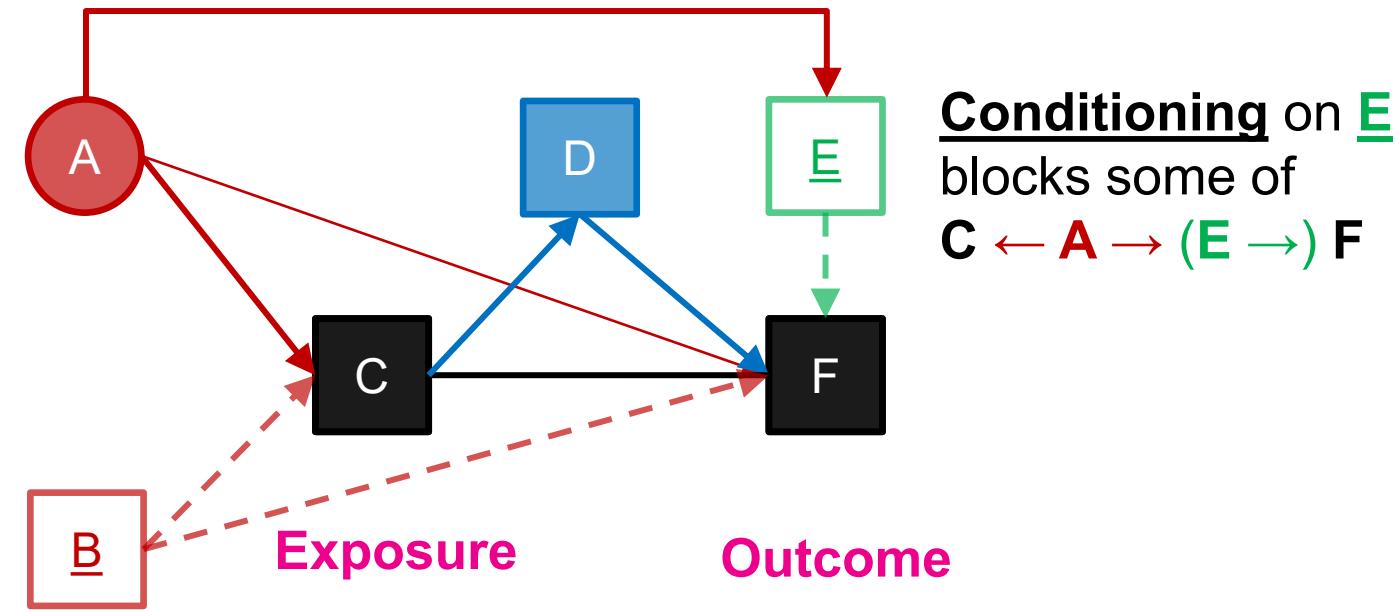
- e.g. % body fat - not obtainable directly, can be estimated indirectly

Sometimes concepts can't be directly measured:

- e.g. socio-economic background is a latent concept
- we can construct a proxy through statistical means

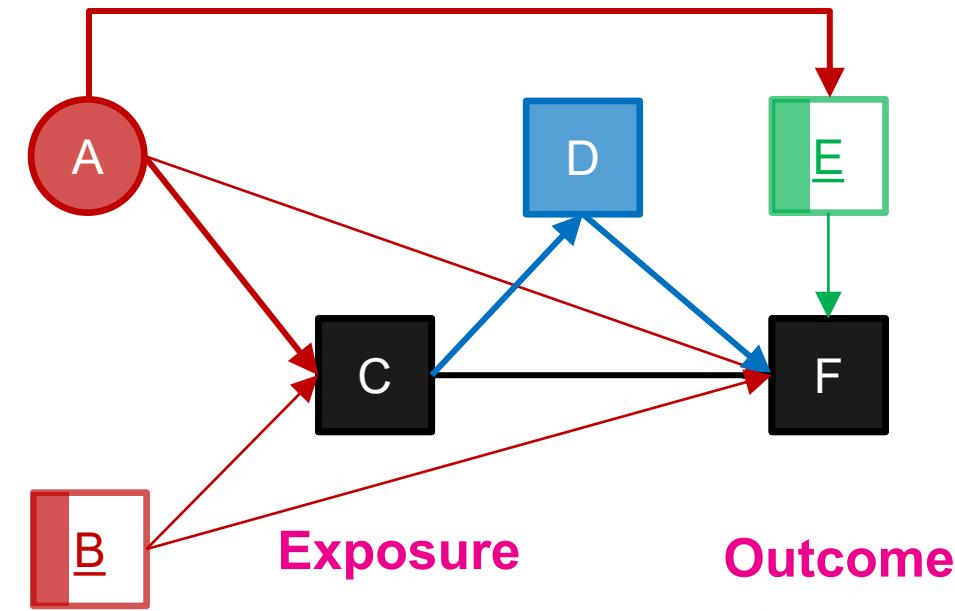
RESIDUAL CONFOUNDING

- Conditioning on **E** would block some of the confounding from **A**
- But some **residual confounding** would almost certainly remain



RESIDUAL CONFOUNDING

- In fact, there is *always* **residual confounding** after conditioning, because you can *never* measure a concept/variable perfectly



ESTIMATING TOTAL CAUSAL EFFECTS

DO

- Condition on **confounders** to block confounding paths

DO NOT

- Condition on **mediators** as this would block true causal paths (and far worse...)

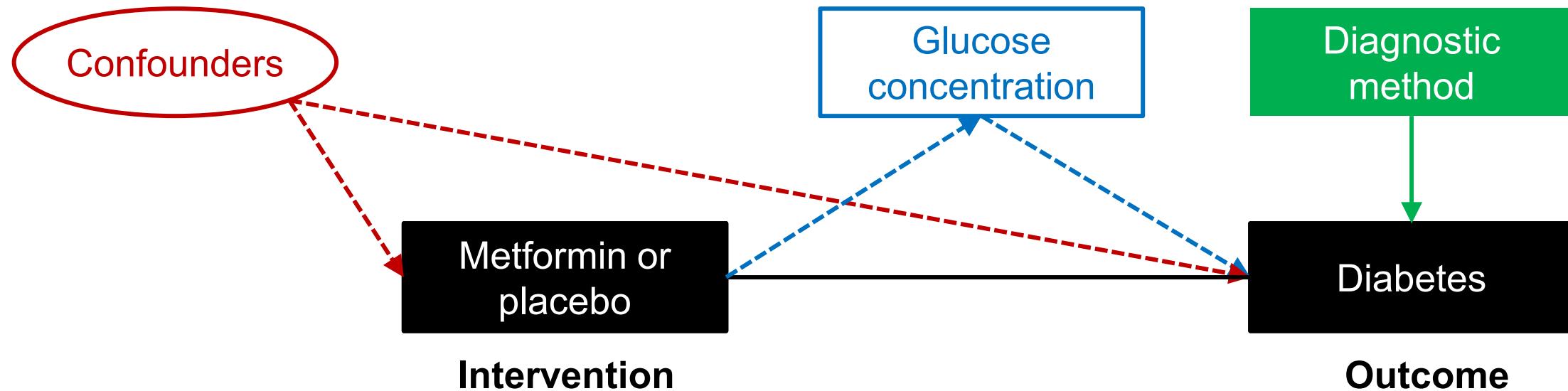
OPTIONALLY

- Condition on **competing exposures** to improve the precision of your estimates

CONDITIONING ON MEDIATORS

Conditioning on mediators is **RISKY** as we will discover when we learn about '**collider bias**'. Even ignoring **collider bias**, there remains a question of what the resulting coefficient(s) *mean*...

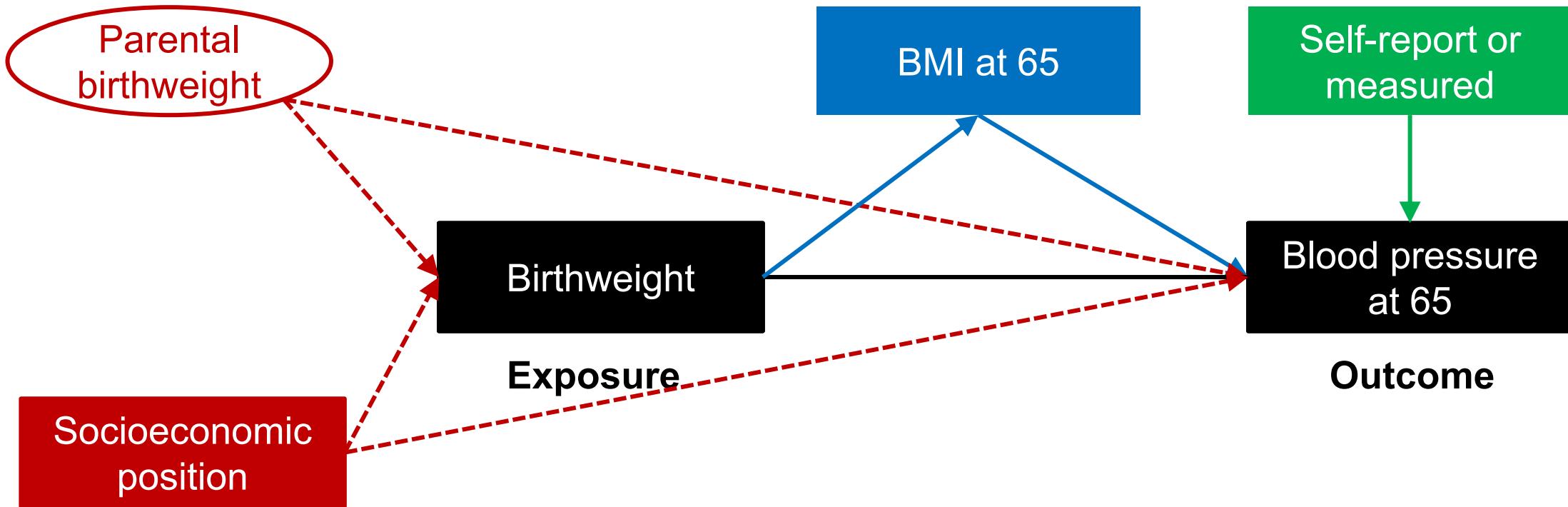
Example: RCT of metformin vs placebo on diabetes diagnosis at 6-months



What would 'causal effect' of metformin on diabetes *mean* independent of its effect on glucose concentration?! This is the mechanism!

CONDITIONING ON MEDIATORS

Example: Cohort study of birthweight and (adult) blood pressure



What does causal effect of birthweight on adult blood pressure, independent of any effect on adult BMI *mean*?



Journal of Human Hypertension (2006) 20, 646–657
© 2006 Nature Publishing Group All rights reserved 0950-9240/06 \$30.00
www.nature.com/jhh

ORIGINAL ARTICLE

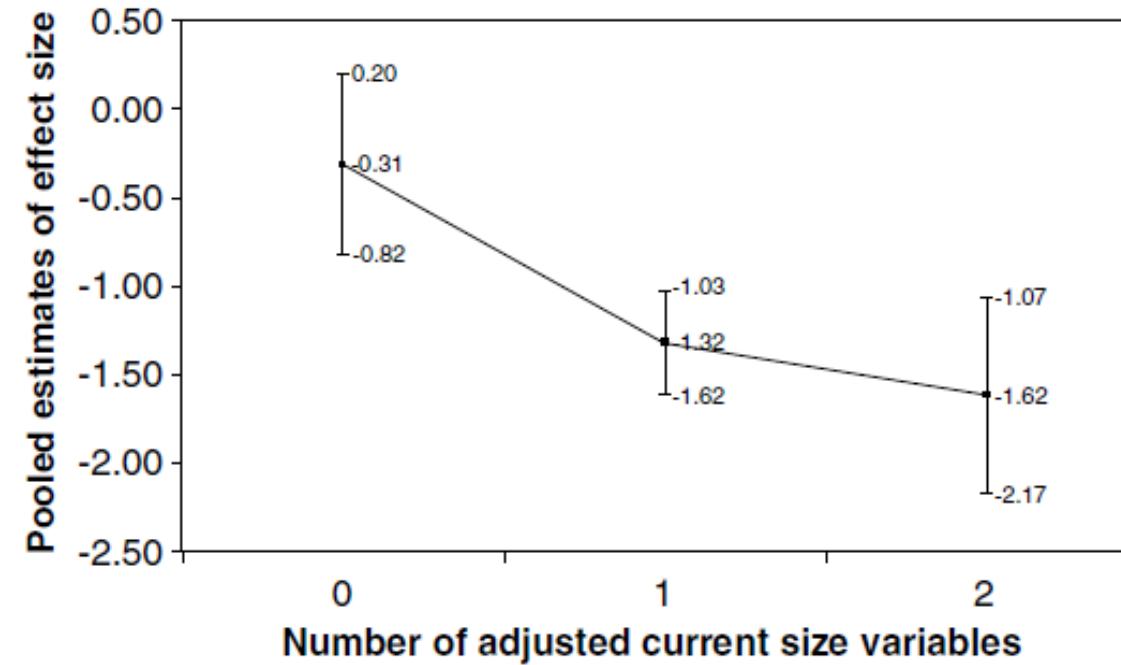
What is the effect of adjusting for more than one measure of current body size on the relation between birthweight and blood pressure?

Y-K Tu^{1,2}, MS Gilthorpe¹ and GTH Ellison³

¹Biostatistics Unit, Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, UK; ²Leeds Dental Institute, University of Leeds, Leeds, UK and ³St George's, University of London, Cranmer Terrace, London, UK

CONDITIONING ON MEDIATORS

How important is birthweight on adult blood pressure?



It seems to depend on how many times you adjust for adult BMI?

Results like this suggest **collider bias**...

ESTIMATING TOTAL CAUSAL EFFECTS

DO

- Condition on **confounders** to block confounding paths

DO NOT

- Condition on **mediators** as this would block true causal paths (and far worse...)

OPTIONALLY

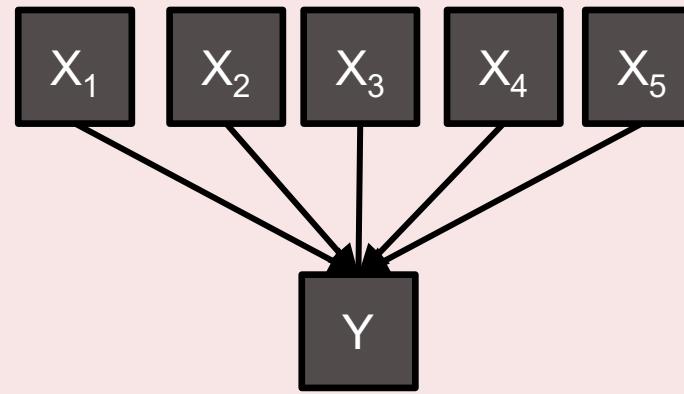
- Condition on **competing exposures** to improve the precision of your estimates

The choice of covariates is determined from your DAG **AND** the causal relationship of interest → different models are needed for different exposures!

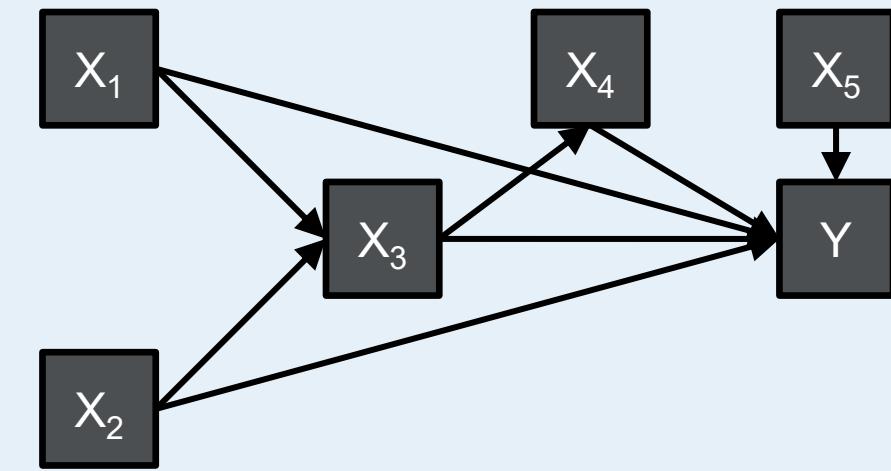
ESTIMATING TOTAL CAUSAL EFFECTS

- **Recall:** The statistical software does not understand this context

HOW THE SOFTWARE SEES IT



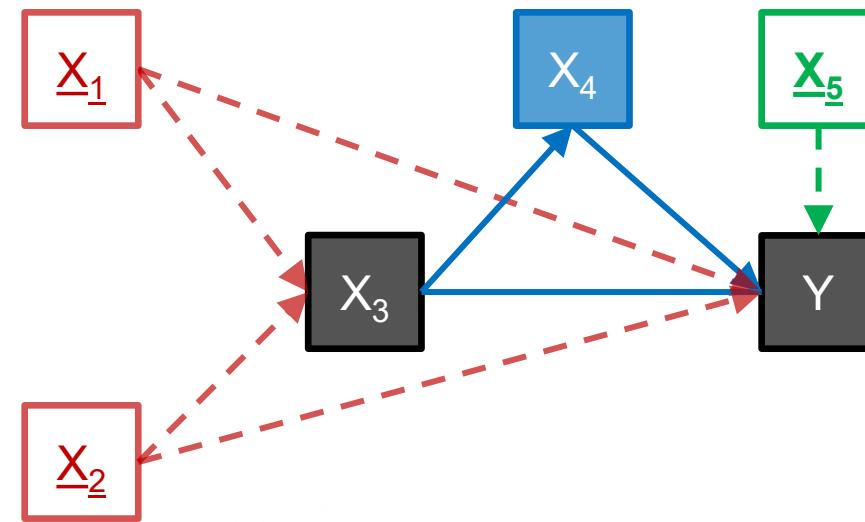
HOW NATURE CREATED IT



ESTIMATING TOTAL CAUSAL EFFECTS

- After drawing your DAG, you can use tools such as *Dagitty.net* to identify '**sufficient adjustment sets**' of variables for conditioning that satisfy these rules
- For **Total causal effect** of X_3 on Y :
 - Model should include **confounders** (X_1 , X_2) and **competing exposures** (X_5)

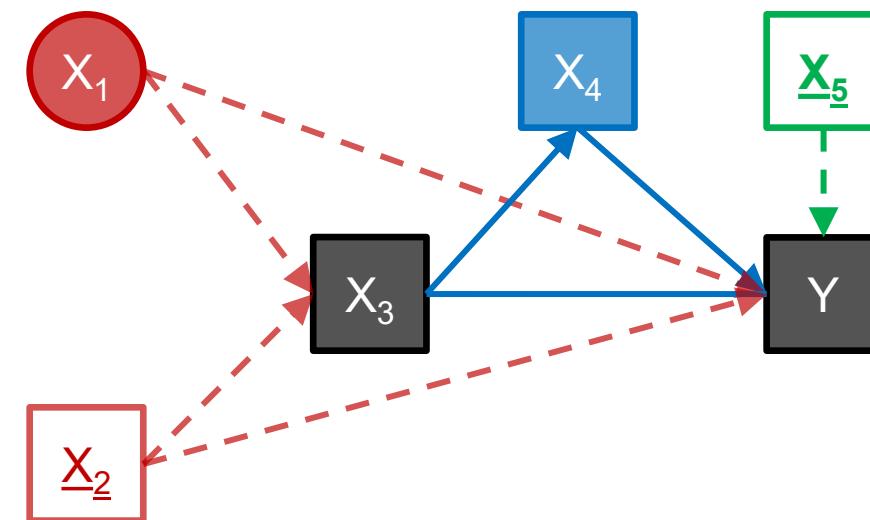
$$Y \sim X_3 + X_1 + X_2 + X_5$$



ESTIMATING TOTAL CAUSAL EFFECTS

- After drawing your DAG, you can use tools such as *Dagitty.net* to identify '**sufficient adjustment sets**' of variables for conditioning that satisfy these rules
- For **Total causal effect** of X_3 on Y :
 - Model should include **confounders** (X_1, X_2) and **competing exposures** (X_5)

$$Y \sim X_3 + \quad + X_2 + X_5$$



Where you have unobserved confounding,
you could conduct **quantitative bias analysis** simulations to estimate the impact

ESTIMATING TOTAL CAUSAL EFFECTS

- In our model of the **total causal effect** of X_3 on Y :
 - It would be wrong to *interpret* coefficients for other covariates (X_1, X_2, X_5), because they would require different adjustment sets!

$$Y \sim X_3 + X_1 + X_2 + X_5$$

E.g coefficient on X_1 is **NOT** total causal effect of X_1 on Y , due to conditioning on X_3

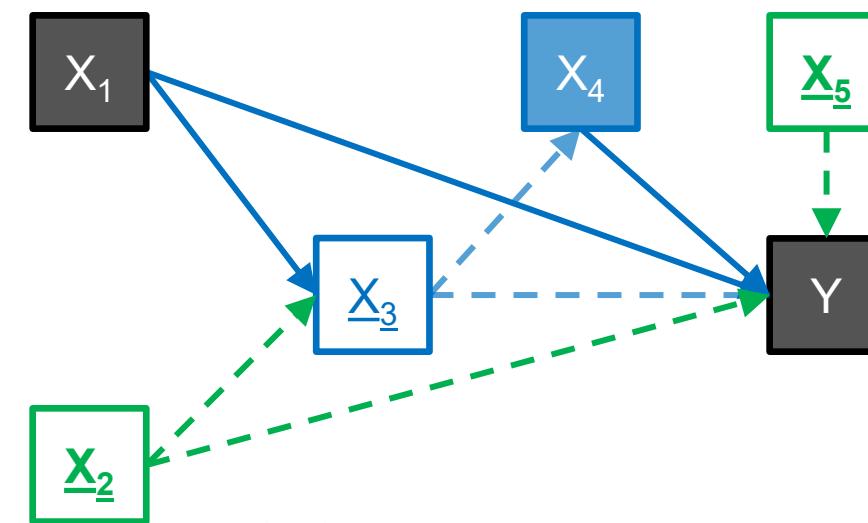


TABLE 2 FALLACY

- The tradition of including all ‘**predictors**’ of our outcome/label (**Y**) in a **single model**, and *interpreting* the coefficients/weights (X_1, X_2, X_3, X_4, X_5) as has been dubbed the ‘**Table 2 Fallacy**’



American Journal of Epidemiology

© The Author 2013. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 177, No. 4

DOI: 10.1093/aje/kws412

Advance Access publication:
January 30, 2013

Commentary

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients

Daniel Westreich* and Sander Greenland

* Correspondence to Dr. Daniel Westreich, Department of Obstetrics and Gynecology, Duke Global Health Institute, Duke University, DUMC 3967, Durham, NC 27710 (e-mail: daniel.westreich@duke.edu).

Initially submitted January 13, 2012; accepted for publication October 11, 2012.

Source: Westreich & Greenland 2013 - *Am J Epidemiol*

INTERPRETABILITY

- This has very important implications on **interpretability**
- Recently there has been a lot of focus on making models ‘transparent’ which is expected to make them more ‘interpretable’...
- But what is the actual **meaning** of the model coefficients/weights?
- Just because we have obtained model coefficients, does not mean that we can interpret them as effects on the outcome! This is no different to the **Table 2 Fallacy!**

SUMMARY

- Causal diagrams like DAGs, help us to identify our **assumptions** about how variables are causally related
- They force us to make those assumptions explicit, which is a huge advance in **transparency** compared with ‘black box’ and post-hoc approaches
- However, we must remember that we can only **interpret** the effect of the exposure
- DAGs conveniently encode counterfactual and probabilistic theories of causation
 - this is managed so cleverly, that we need not learn or understand them!
 - Although the rules (and terminology) are initially daunting, with time and familiarity they will become second nature!

RECOMMENDED READING

Books

- Morgan SL, Winship C. *Counterfactuals and causal inference*. Cambridge University Press; 2015.

Papers

- Digitale, J.C., Martin, J.N. and Glymour, M.M., 2022. Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142, pp.264-267.
- Lipsky, A.M. and Greenland, S., 2022. Causal Directed Acyclic Graphs. *JAMA*, 327(11), pp.1083-1084.
- Shrier, I. and Platt, R.W., 2008. Reducing bias through directed acyclic graphs. *BMC medical research methodology*, 8(1), pp.1-15.

SUMMARY

DAGs are a simple tool, they do not:

- indicate whether an effect is harmful, protective, or meaningful
- indicate if effect modification (interaction) is present
- indicate if a cause is sufficient or necessary (non-linearity)

DAGs are semi-parametric but linear regression models are parametric

- many different DAGs will be compatible with your data structure
- several data structures may be compatible with your DAG
- It is impossible to ‘prove’ which is correct

The accuracy of an estimated causal effect is conditional on the accuracy of the corresponding DAG

- this may sound like a limitation, but it was ever thus!

09:30-10:15 LECTURE 2.1

10:15-11:00 ACTIVITY 2-A

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 2.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:00 ACTIVITY 2-B

15:00-15:30 LECTURE 2.3

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 2.4

17:00-17:45 ACTIVITY 2-C

17:45-18:00 Q&A

2.1 - THE TARGET TRIAL FRAMEWORK

GEORGIA



MARK



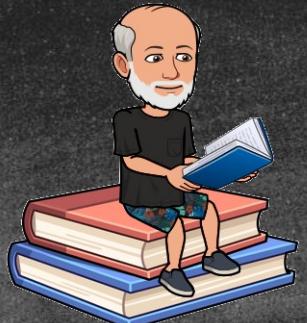
@GEORGIATOMOVA

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

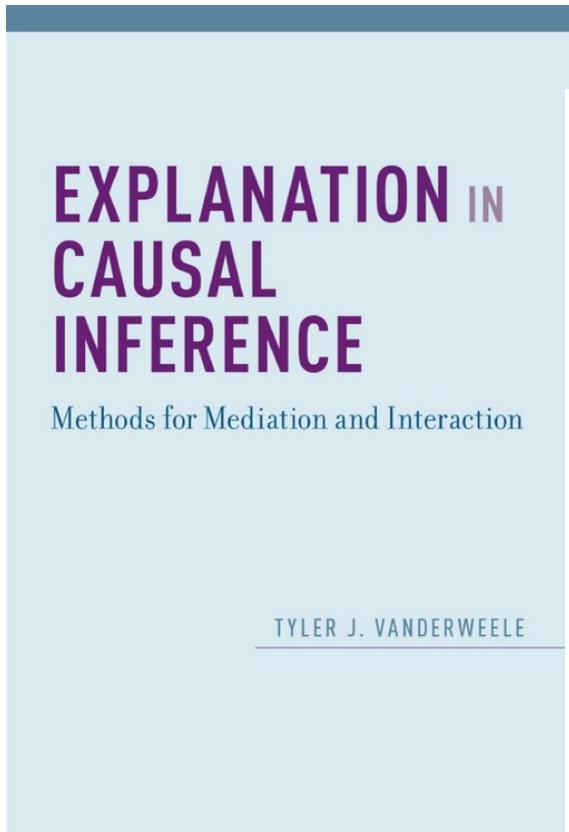
LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- Describe the broad aims of the **target trial framework**
- Identify the **seven components** of the target trial framework and discuss the main features of each
- Describe **healthy user bias**
- Describe **immortal time bias** and discuss strategies to avoid it

CAUSAL INFERENCE METHODS

Causal inference methods can be rather daunting to the uninitiated ...



2.5. BINARY MEDIATORS

A similar approach, allowing for exposure–mediator interaction, also works with binary mediators. Suppose that the mediator is binary and the outcome is continuous and that the following models fit the observed data:

$$\begin{aligned}\mathbb{E}[Y|a,m,c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta'_4 c \\ \text{logit}\{P(M = 1|a,c)\} &= \beta_0 + \beta_1 a + \beta'_2 c\end{aligned}$$

If the covariates C satisfy no-confounding assumptions (A2.1)–(A2.4) above, then average controlled direct effect and the average natural direct and indirect effects on the outcome difference scale are given by

$$\begin{aligned}CDE(m) &= (\theta_1 + \theta_3 m)(a - a^*) \\ NDE &= \{\theta_1(a - a^*)\} + \{\theta_3(a - a^*)\} \frac{\exp[\beta_0 + \beta_1 a^* + \beta'_2 c]}{1 + \exp[\beta_0 + \beta_1 a^* + \beta'_2 c]} \\ NIE &= (\theta_2 + \theta_3 a) \left\{ \frac{\exp[\beta_0 + \beta_1 a + \beta'_2 c]}{1 + \exp[\beta_0 + \beta_1 a + \beta'_2 c]} - \frac{\exp[\beta_0 + \beta_1 a^* + \beta'_2 c]}{1 + \exp[\beta_0 + \beta_1 a^* + \beta'_2 c]} \right\}\end{aligned}$$

The expressions for the direct and indirect effects are once again simply combinations of the coefficients from the two regressions above. These effects were derived in Valeri and VanderWeele (2013), and standard errors are also given there and in the Appendix. The macros described below will implement this approach and estimate standard errors and confidence intervals automatically.

Similarly, suppose that both the mediator and the outcome are binary. Suppose that the outcome is rare and that the following models are fit to the observed data:

$$\begin{aligned}\text{logit}\{P(Y = 1|a,m,c)\} &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta'_4 c \\ \text{logit}\{P(M = 1|a,c)\} &= \beta_0 + \beta_1 a + \beta'_2 c\end{aligned}$$

If the covariates C satisfied assumptions (A2.1)–(A2.4) above, then the conditional controlled direct effect and natural direct and indirect effects on the odds ratio scale would be given by

$$\begin{aligned}OR^{CDE}(m) &= (\theta_1 + \theta_3 m)(a - a^*) \\ OR^{NDE} &= \frac{\exp(\theta_1 a)\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta'_2 c)\}}{\exp(\theta_1 a^*)\{1 + \exp(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \beta'_2 c)\}} \\ OR^{NIE} &= \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta'_2 c)\}\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta'_2 c)\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta'_2 c)\}\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta'_2 c)\}}\end{aligned}$$

These expressions apply also if the outcome is not rare and log-linear rather than logistic models are fit to the data; the expressions are then for direct and indirect effect risk ratios rather than odds ratios. The mediator does not need to be rare for these expression to apply. Once again, derivations and standard errors for these are given in Valeri and VanderWeele (2013), but the macros described below will implement this approach and give estimates and confidence intervals automatically.

THE TARGET TRIAL FRAMEWORK



American Journal of Epidemiology

© The Author 2016. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 183, No. 8

DOI: 10.1093/aje/kwz254

Advance Access publication:

March 18, 2016

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115
(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 using an appropriately designed and conducted randomized experiment. When we cannot conduct a randomized experiment, we analyze observational data. Causal inference from large observational databases (big data) can be viewed as an attempt to emulate a randomized experiment—the target experiment or target trial—that would answer the question of interest. When the goal is to guide decisions among several strategies, causal analyses of observational data need to be evaluated with respect to how well they emulate a particular target trial. We outline a framework for comparative effectiveness research using big data that makes the target trial explicit. This framework channels counterfactual theory for comparing the effects of sustained treatment strategies, organizes analytic approaches, provides a structured process for the criticism of observational studies, and helps avoid common methodologic pitfalls.

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

To make it easier to conduct robust analyses without counterfactual language & graphical models, Hernan & Robins devised the '**Target Trial Framework**',

Not a new method, but an aid to thinking ("organising principle")

THE TARGET TRIAL

The target trial is a *hypothetical* trial that you would conduct to estimate your estimand of interest

Not a real trial, but Hernan and Robins encourage the target trial should be as plausible as possible

The target trial framework encourages you to think of a '**well defined intervention**', i.e. something you could introduce in practice

- e.g. drugs, therapies, protocols, and policies
- not vague concepts (i.e. 'socioeconomic position')



TWO PROTOCOLS

Protocol Component	Description	Notes
Eligibility criteria	How the patient population is recruited into the trial.	All inclusion and exclusion criteria are based on characteristics ascertained exclusively at baseline.
Treatment strategies	Each of the clinical interventions that are to be compared.	The description needs to include the initial treatment as well as protocol-approved reasons for discontinuation or switching.
Treatment assignment	How participants will be assigned to each treatment strategy at baseline.	The assignment is randomized, possibly conditional on baseline prognostic factors. Patients will be aware of the treatment strategy to which they were assigned.
Start and end of follow-up	Define when the follow-up period starts and ends for each participant.	For each eligible individual, follow-up starts at baseline (the time of treatment assignment) and ends at death, outcome, loss to follow-up, or administrative end of follow-up.
Outcomes	Outcomes of interest and how to ascertain them.	If possible, include negative controls, i.e., outcomes that are known to be unaffected by the studied treatments.
Causal contrast	What comparative effects of the treatment strategies will be estimated.	The intention-to-treat effect (the comparative effect of being assigned to the treatment strategies at baseline) or per-protocol effect (the comparative effect of receiving the treatment as specified in the protocol).
Statistical analysis	How to estimate the intention-to-treat effect or per-protocol effect via intention-to-treat and per-protocol analyses that appropriately adjust for pre- and post-baseline prognostic factors associated with adherence and loss to follow-up.	Investigators should specify and measure the covariates potentially related to treatment choice, adherence, and outcomes at baseline and during the follow-up. Other variables that may need to be specified include those that define key sub-groups.

Create two protocols:

- Target Trial
- Emulation Study

From: National Academies of Sciences, Engineering, and Medicine. An Approach to Evaluate the Effects of Concomitant Prescribing of Opioids and Benzodiazepines on Veteran Deaths and Suicides. Washington (DC): National Academies Press (US); 2019 Sep 24. 2, Specifying the Target Trial. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK547516/>

ELIGIBILITY CRITERIA

Clearly explain who would be eligible for recruitment into the target trial & emulation study, outlining all inclusion and exclusion criteria

- e.g. “*Postmenopausal women within 5 years of menopause between the years 2005 and 2010 and with no history of cancer and no use of hormone therapy in the past 2 years.*”



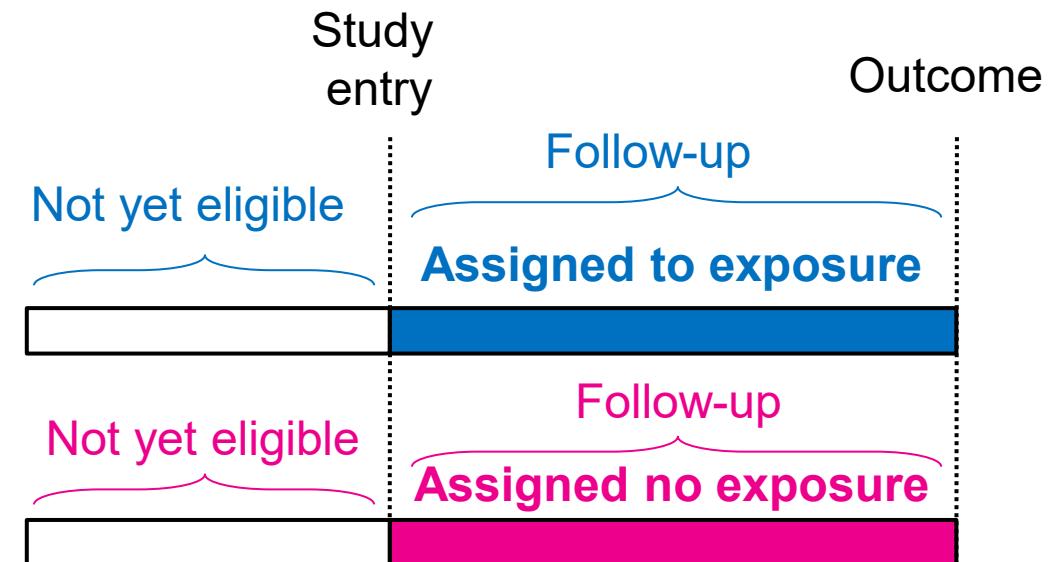
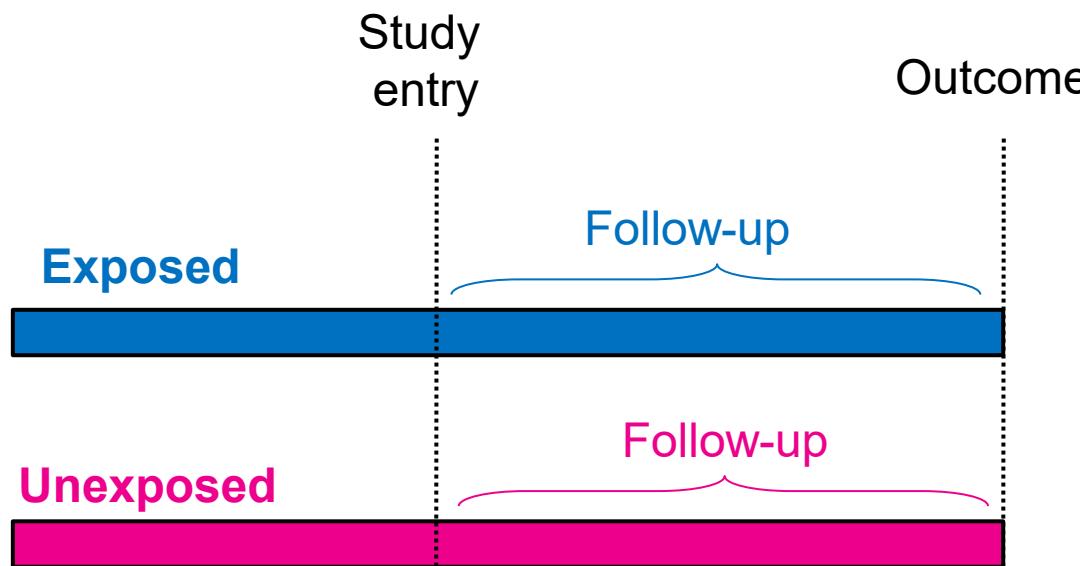
- This might seem very obvious
- But requires that you set criteria based *only* on what is known at baseline!
- This avoids '**conditioning on the future**'
- Cannot define sample as people with 'complete data' or 'who attended >3 appointments' etc.

TREATMENT STRATEGIES

Outline all treatment strategies (i.e. all intervention and control groups)

Your target trial will typically be a '**pragmatic trial**', comparing an intervention to 'usual care' (a placebo doesn't make sense in an observational study)

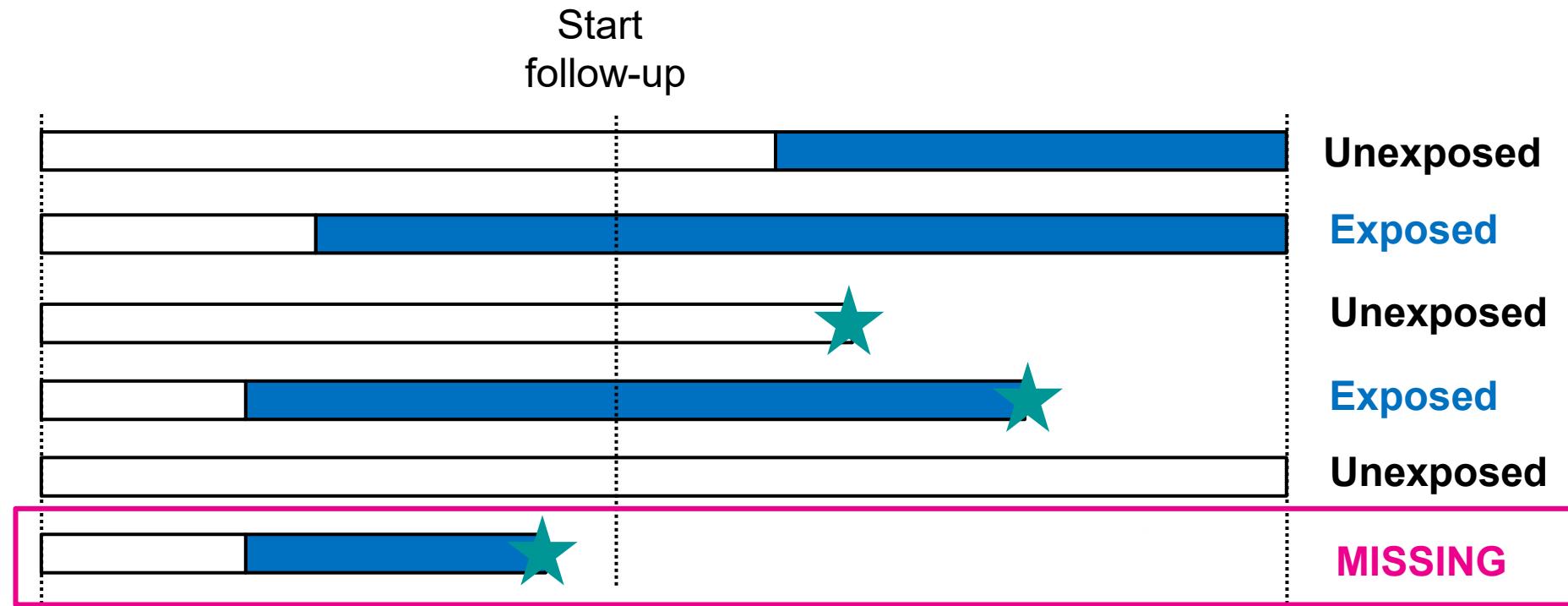
Important: You cannot simply compare 'exposed' to 'unexposed', as this is not how a trial would work; you must *assign* the treatment to all participants



HEALTHY USER BIAS

When you simply compare exposed to unexposed, you risk '**healthy user bias**' (AKA '**prevalent user bias**')

To be a current user, you must be healthy enough to continue using the treatment and survive until the start of the study follow-up begins



EXAMPLE: HEALTHY USER BIAS

RESEARCH

OPEN ACCESS

Check for updates

The art of life and death: 14 year follow-up analyses of associations between arts engagement and mortality in the English Longitudinal Study of Ageing

Daisy Fancourt,¹ Andrew Steptoe¹

¹Department of Behavioural Science and Health, University College London, London WC1E 7HB, UK

Correspondence to: D Fancourt
d.fancourt@ucl.ac.uk
(or @Daisy_Fancourt on Twitter;
ORCID 0000-0002-6952-334X)

Cite this as: *BMJ* 2019;367:l6377
<http://dx.doi.org/10.1136/bmj.l6377>

Accepted: 24 September 2019

ABSTRACT

OBJECTIVE

To explore associations between different frequencies of arts engagement and mortality over a 14 year follow-up period.

DESIGN

Prospective cohort study.

PARTICIPANTS

English Longitudinal Study of Ageing cohort of 6710 community dwelling adults aged 50 years and older (53.6% women, average age 65.9 years, standard deviation 9.4) who provided baseline data in 2004–05.

INTERVENTION

Self reported receptive arts engagement (going to museums, art galleries, exhibitions, the theatre, concerts, or the opera).

MEASUREMENT

Mortality measured through data linkage to the National Health Service central register.

RESULTS

People who engaged with receptive arts activities on an infrequent basis (once or twice a year) had a 14% lower risk of dying at any point during the follow-up (809/3042 deaths, hazard ratio 0.86, 95% confidence interval 0.77 to 0.96) compared with those who never engaged (837/1762 deaths). People who engaged with receptive arts activities on a frequent basis (every few months or more) had a 31% lower risk of dying (355/1906 deaths, 0.69, 0.59 to 0.80), independent

of demographic, socioeconomic, health related, behavioural, and social factors. Results were robust to a range of sensitivity analyses with no evidence of moderation by sex, socioeconomic status, or social factors. This study was observational and so causality cannot be assumed.

CONCLUSIONS

Receptive arts engagement could have a protective association with longevity in older adults. This association might be partly explained by differences in cognition, mental health, and physical activity among those who do and do not engage in the arts, but remains even when the model is adjusted for these factors.

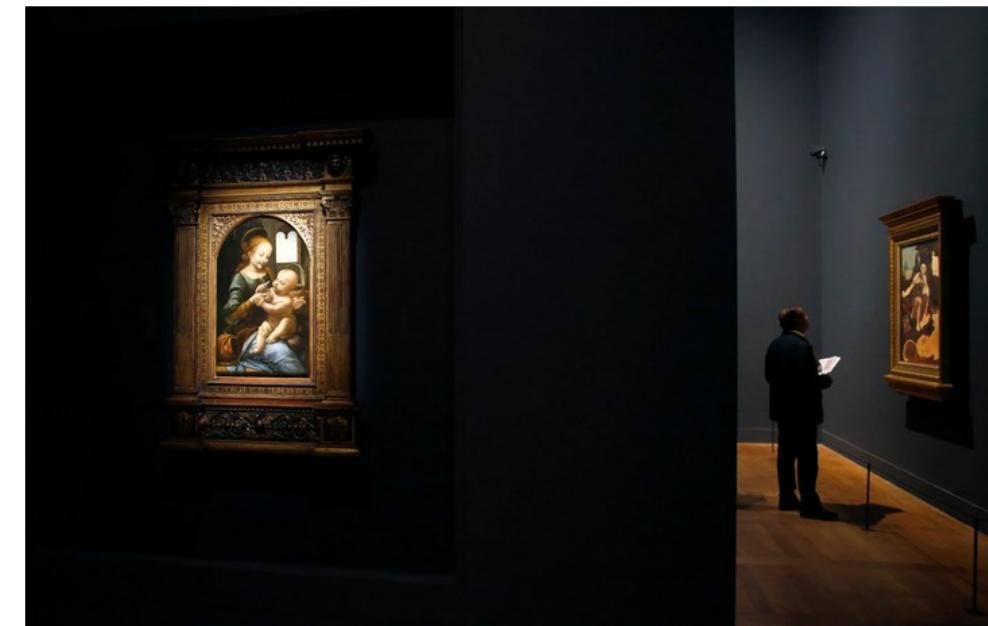
Introduction

Interest in the salutogenic (health promoting) benefits of the arts is increasing. Arts activities are classified as “multimodal” health interventions because they combine multiple psychological, physical, social, and behavioural factors with an intrinsic aesthetic motivation to engage.¹ While previous studies have shown the association between arts engagement and the prevention and treatment of mental and physical health conditions, including depression, dementia, chronic pain, and frailty,^{2,4} whether arts engagement actually confers survival benefits remains unclear. Some research has proposed that the universality of art and the strong emotional responses it induces are indications of its association with evolutionary

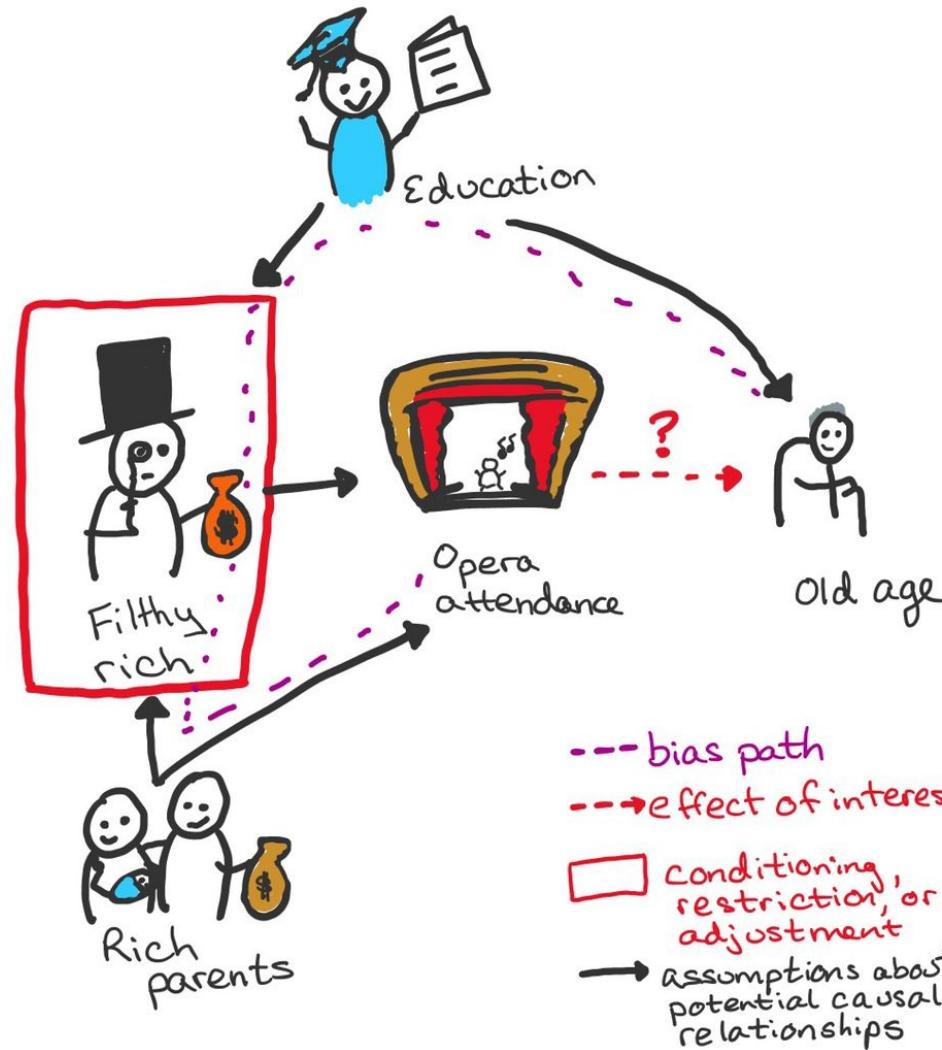
The New York Times

Another Benefit to Going to Museums? You May Live Longer

Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.



EXAMPLE: HEALTHY USER BIAS



Reproduced with permission from @EpiEllie: <https://twitter.com/EpiEllie/status/1433060155734822916>

ASSIGNMENT PROCEDURES

The assignment procedure(s) describes how we will achieve exchangeable units of analysis

In a target trial: this will involve random assignment of the exposure

In an emulation study: this will usually involve conditioning on confounders

Recommended sensitivity analysis:

- examine treatment in reverse – i.e. effect of *discontinuing* treatment in current users, and compare to effect of starting treatment in non-users
- examine a **control exposure** – something with a similar indications, but not expected to have a (strong) causal effect
- examine a **control outcome** – alternatives outcomes where a causal effect is not expected

FOLLOW-UP PERIOD

The **follow-up period** describes the start and end of the time that all participants will be observed

e.g. "*Starts at randomization and ends at diagnosis of breast cancer, death, loss to follow-up, or 5 years after baseline, whichever occurs first*"

Carefully consider the time that follow-up starts to minimise health user bias and immortal time bias (discussed soon...)

Follow-up Plan

Send invitation
1st Reminder
2nd Reminder
1st Call
2nd Call
Send silly cartoon
Beg
Hire goons
Release hounds

OUTCOME

The outcome describes our effort to monitor and identify the outcome

- e.g. "*Breast cancer diagnosed by an oncologist within 5 years of baseline*"

Where possible, we'd like the outcome to be ascertained systematically without knowledge of the treatment status

In most observational studies, this won't be possible, since participant/diagnostician will be aware of the participants exposure status

Therefore, need to aim for highest possible ascertainment; **prefer objective ascertainment, multiple ascertainment, and/or outcome validation**

CAUSAL CONTRASTS OF INTEREST

In a target trial: **intention-to-treat effect** and **per-protocol effect**

Intention-to-treat effect: total causal effect of being assigned to a specific treatment strategy, regardless of whether it is then followed

≈ the effect of randomisation

Per-protocol effect: total causal effect of being assigned to a specific treatment strategy *and* adhering to that strategy for the duration of treatment

≈ the effect of the treatment when used as intended

In the emulation study: may not be possible to estimate the intention-to-treat effect if cannot distinguish between protocol violators and participants assigned to different treatment strategies

ANALYSIS PLAN

The analysis plan is a short summary of the planned approach

Where an intention-to-treat analysis is possible, adjustment will be necessary for baseline confounders

Where a per-protocol analysis is attempted, adjustment will be necessary for baseline confounders and time-varying confounders (using **g-methods**)

In both the target trial and emulation study, appropriate methods (e.g. **inverse probability weighting**) will be needed to account for **attrition** (AKA **drop-out / loss-to-follow-up**)

EXAMPLE: TARGET TRIAL PROTOCOL

The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening

Xabier García-Albéniz^{1,2} · John Hsu^{2,3} · Miguel A. Hernán^{1,4,5}

Component	Target trial	Emulated trial using real world data
Aim	To estimate the effect of screening colonoscopy on the 8-year risk of CRC in U.S. individuals aged 70–74 years	Same
Eligibility	Persons without gastrointestinal symptoms aged 70–74 years with no history of CRC, and continuously enrolled in Medicare for 5 years with no adenoma, inflammatory bowel disease, colectomy, or CRC screening in that period, and who were regular users of preventive services (at least 2 of the following: influenza vaccine, preventive visit, breast or prostate screening, in the 2 years before enrollment)	Same, except CRC history is evaluated in the 5 years before enrollment
Treatment strategies	1. Screening colonoscopy at baseline 2. No screening for CRC at baseline Patients receive usual care after the intervention	Same
Treatment assignment	Patients are randomly assigned to either strategy	Patients are assigned to screening colonoscopy if they receive a screening colonoscopy in the 7 days following eligibility and to no screening otherwise.
Follow-up	Follow-up starts at treatment assignment and ends at CRC diagnosis, at death, at loss to follow-up, 8 years after baseline, or on 31 December 2012, whichever occurs first	Randomization is emulated via adjustment for baseline covariates: sex, race, age, original reason for Medicare entitlement, use of preventive services, U.S. Census Bureau division, combined comorbidity score, calendar month, presence of each CCW condition (Alzheimer's disease, acute myocardial infarction, asthma, atrial fibrillation, cataract, chronic heart failure, chronic kidney disease, endometrial cancer, breast cancer, lung cancer, prostate cancer, chronic obstructive pulmonary disease, depression, diabetes, glaucoma, hip/pelvic fracture, hyperlipidemia, benign prostatic hyperplasia, hypertension, hypothyroidism, ischemic heart disease, osteoporosis, osteoarthritis, stroke)
Outcome	CRC diagnosis within 8 years of baseline	Same
Causal contrast	Intention-to-treat effect, i.e., effect of being assigned to screening colonoscopy versus no screening at baseline. Per-protocol effect, i.e., effect of receiving screening colonoscopy versus no screening at baseline	Observational analog of per-protocol effect
Statistical analysis	Intention-to-treat analysis. Per-protocol analysis: comparison of 8-year CRC risk between groups receiving each treatment strategy with adjustment for baseline covariates (and post-baseline covariates when adjusting for loss to follow-up)	Same as per-protocol analysis

DEFINING TIME ZERO

Key benefit of target trial approach is it forces careful thinking around timing of:

- participant eligibility
- treatment assignment
- follow-up strategy

This helps to avoid:

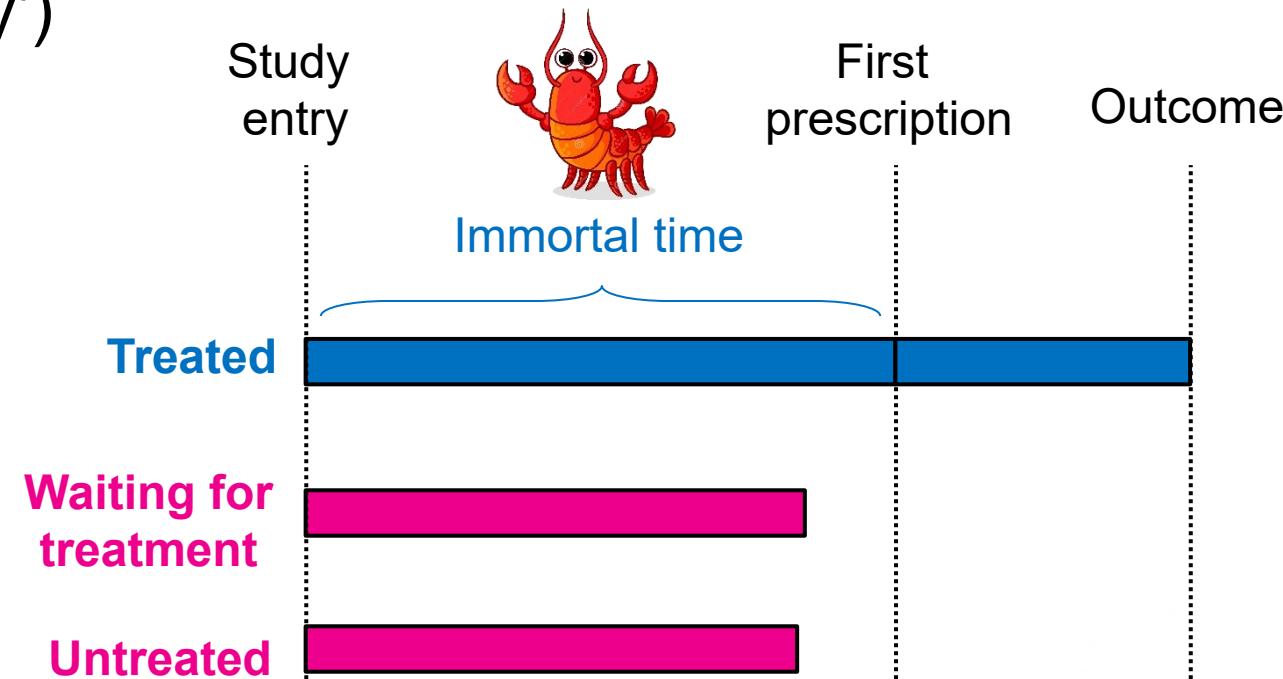
- **healthy user bias**
- **immortal time bias**



IMMORTAL TIME BIAS

Immortal time bias occurs when there is a time (differential) gap between entrance into a study and assignment of one or more treatments

This commonly occurs when the exposure explicitly or implicitly includes time (e.g. '7-days of steroid therapy' or 'attended at least three sessions of chemotherapy')



EXAMPLE 1: IMMORTAL TIME BIAS



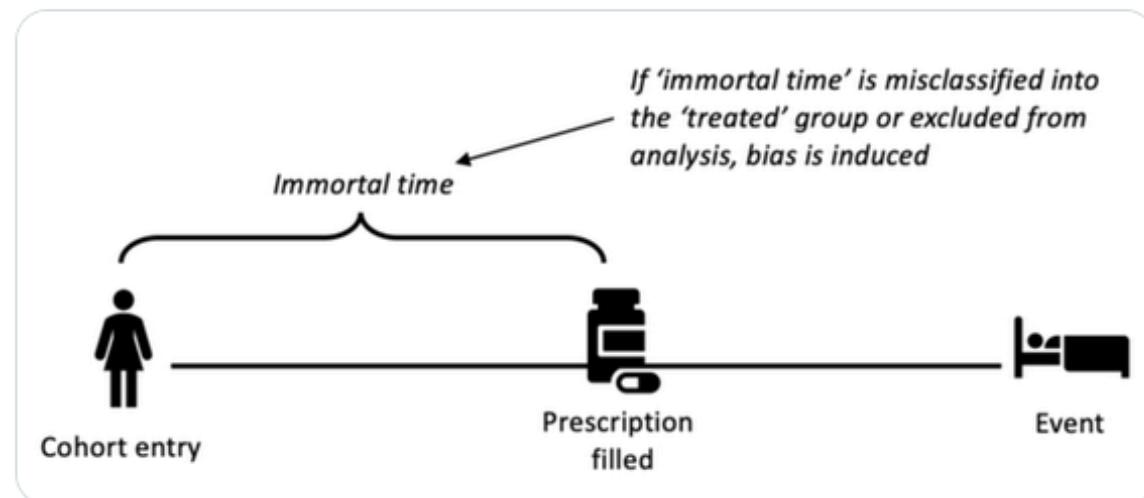
CEBM Catalogue of Bias
@Catalogofbias

📢 BING BONG - NEW BIAS ALERT 📢

The latest entry to the Catalogue is an important one and very relevant to the COVID-19 evidence-base.

IMMORTAL TIME BIAS [catalogofbias.org/biases/immortal...](https://catalogofbias.org/biases/immortal-time-bias)

via [@hopinlee](#) [@dnunan79](#)



...



Ken Rothman
@ken_rothman

...

Replies to [@Catalogofbias](#) [@hopinlee](#) and 11 others

Immortal time bias has been around a long time. See this example from 1793: [twitter.com/ken_rothman/st...](https://twitter.com/ken_rothman/status/143811111111111111)

EXAMPLE 1: IMMORTAL TIME BIAS

For example, Benjamin Rush, a renowned American physician, philosopher, and statesman, claimed in 1793 that he had saved countless patients during the yellow fever epidemic in Philadelphia by bleeding and purging them. No patient died, he said, if he had bled them at least seven times. The fact that some of his patients didn't live long enough to be bled seven times apparently was not part of his calculus, and the only true relief his treatment likely offered was to shorten his patients' suffering by hastening their deaths. Although bloodletting had been popular for millennia, Rush's claims contributed to its persistence until well into the nineteenth century, when it was finally abandoned after researchers put the practice to a test by comparing two groups: patients subjected to bloodletting, and a control group of those who were not.

EXAMPLE 2: IMMORTAL TIME BIAS



DOI: 10.1111/j.1464-5491.2004.01263.x

Statin use in Type 2 diabetes mellitus is associated with a delay in starting insulin

A. Yee*, S. R. Majumdar*§, S. H. Simpson§, F. A. McAlister*§, R. T. Tsuyuki§ and J. A. Johnson‡§

Abstract

Aims It has been suggested that HMG Co-A reductase inhibitors ('statins') may reduce the risk of developing Type 2 diabetes mellitus. This study was designed to evaluate whether use of statins would also delay progression to insulin therapy.

Methods This was a retrospective cohort study using Saskatchewan Health databases to identify subjects newly started on oral antidiabetic agents from 1991 to 1996.

Subjects < 30 years of age or with previous lipid-lowering drug use were excluded. Medications known to influence glycaemic control, co-morbidity, and demographic data were collected. Statin exposure was defined as at least 1 year of use. Primary outcome was starting insulin treatment. Multivariate Cox proportional hazards models were used to examine the association between statin use and starting insulin.

Results The final cohort included 10 996 new users of oral antidiabetic agents, of which 484 (4.4%) used statins. Mean age was 64 years and 55% were male. Mean duration of follow-up was 5.1 years; 11.1% ($n = 1221$) eventually started insulin treatment. Statin users were no less likely than non-users to start insulin treatment eventually (11.6% vs. 11.1%, $P = 0.74$). After multivariate adjustment, however, statin use was associated with a 10-month delay before newly treated diabetic subjects needed to start insulin treatment (adjusted hazard ratio 0.74; 95% confidence interval 0.56, 0.97, $P = 0.028$).

Conclusion The use of statins is associated with a delay in starting insulin treatment in patients with Type 2 diabetes initially treated with oral antidiabetic agents. Whether this relationship exists for patients at high risk of developing diabetes should be examined in a randomized trial.

Diabet. Med. 21, 962–967 (2004)

We conclude that statin use is associated with a delay in starting insulin treatment in patients with Type 2 diabetes mellitus. The clinical significance of our findings remains to be demonstrated. Our results, when taken together with other recent studies, lends further support to the need for a large randomized controlled trial of statin therapy for preventing (or

EXAMPLE 2: IMMORTAL TIME BIAS



DOI: 10.1111/j.1464-5491.2004.01263.x

Statin use in Type 2 diabetes mellitus is associated with a delay in starting insulin

A. Yee*, S. R. Majumdar*§, S. H. Simpson§, F. A. McAlister*§, R. T. Tsuyuki§ and J. A. Johnson‡§

Abstract

Aims It has been suggested that HMG Co-A reductase inhibitors ('statins') may reduce the risk of developing Type 2 diabetes mellitus. This study was designed to evaluate whether use of statins would also delay progression to insulin therapy.

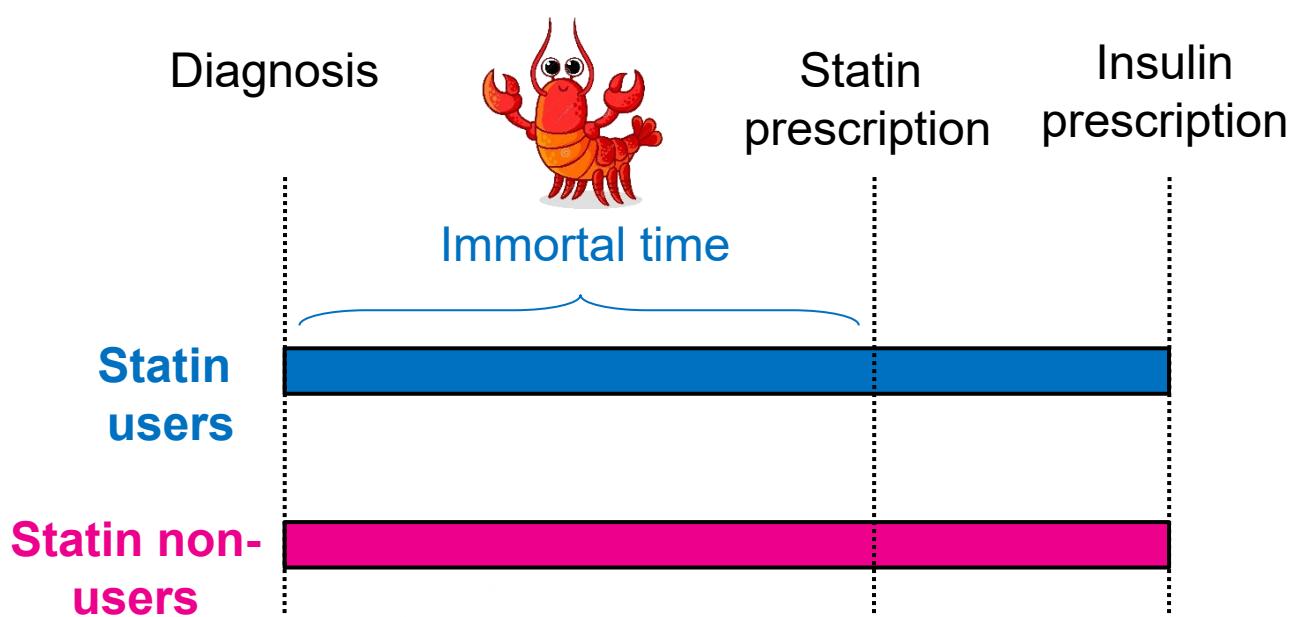
Methods This was a retrospective cohort study using Saskatchewan Health databases to identify subjects newly started on oral antidiabetic agents from 1991 to 1996.

Subjects < 30 years of age or with previous lipid-lowering drug use were excluded. Medications known to influence glycaemic control, co-morbidity, and demographic data were collected. Statin exposure was defined as at least 1 year of use. Primary outcome was starting insulin treatment. Multivariate Cox proportional hazards models were used to examine the association between statin use and starting insulin.

Results The final cohort included 10 996 new users of oral antidiabetic agents, of which 484 (4.4%) used statins. Mean age was 64 years and 55% were male. Mean duration of follow-up was 5.1 years; 11.1% ($n = 1221$) eventually started insulin treatment. Statin users were no less likely than non-users to start insulin treatment eventually (11.6% vs. 11.1%, $P = 0.74$). After multivariate adjustment, however, statin use was associated with a 10-month delay before newly treated diabetic subjects needed to start insulin treatment (adjusted hazard ratio 0.74; 95% confidence interval 0.56, 0.97, $P = 0.028$).

Conclusion The use of statins is associated with a delay in starting insulin treatment in patients with Type 2 diabetes initially treated with oral antidiabetic agents. Whether this relationship exists for patients at high risk of developing diabetes should be examined in a randomized trial.

Diabet. Med. 21, 962–967 (2004)



EXAMPLE 2: IMMORTAL TIME BIAS



covid-19

Research ▾

Education ▾

News & Views ▾

Campaigns ▾

Research Methods & Reporting

Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes

BMJ 2010 ; 340 doi: <https://doi.org/10.1136/bmj.b5087> (Published 12 March 2010)

Cite this as: BMJ 2010;340:b5087

Article	Related content	Metrics	Responses	Peer review
---------	-----------------	---------	-----------	-------------

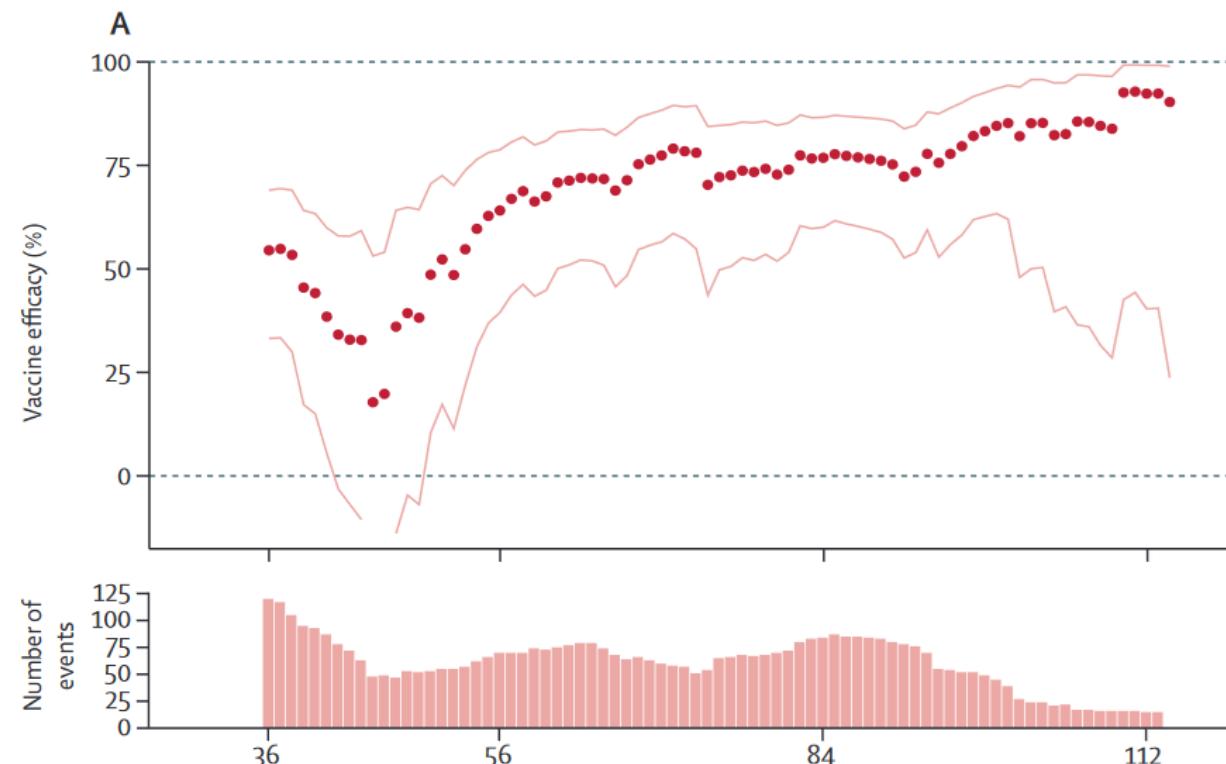
Linda E Lévesque, assistant professor of epidemiology^{1 2 3}, James A Hanley, professor of biostatistics^{1 4}, Abbas Kezouh, biostatistician⁴, Samy Suissa, professor of epidemiology and biostatistics^{1 4}

Source of immortal time*	Immortal and misclassified person time, years (proportion of total statin user person time (n=3221))†	Immortal period corrected by time dependent analysis	Corrected immortal and misclassified person time (years)	Adjusted hazard ratio‡ (95% CI)
All periods	2174 (67.5)	None	0	0.74 (0.58 to 0.95)
Period 1	266 (8.3)	1	266	0.82 (0.64 to 1.05)
Period 2	1376 (42.7)	1 and 2	1642	1.37 (1.07 to 1.76)
Period 3	532 (16.5)	1, 2, and 3	2174	1.97 (1.53 to 2.52)

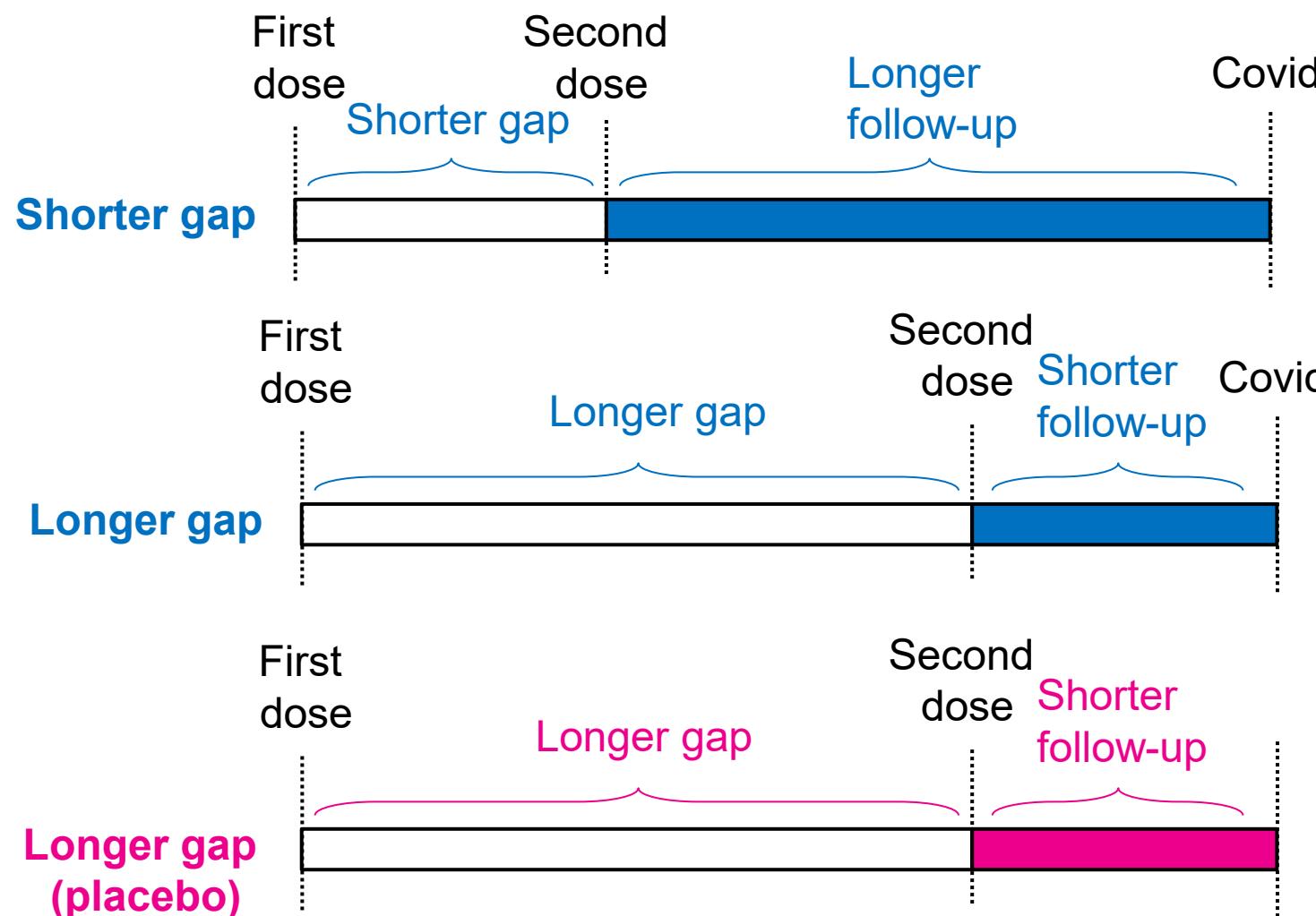
EXAMPLE 3: IMMORTAL TIME BIAS?

Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials

Merryn Voysey*, Sue Ann Costa Clemens*, Shabir A Madhi*, Lily Y Weckx*, Pedro M Folegatti*, Parvinder K Aley, Brian Angus, Vicky L Baillie, Shaun L Barnabas, Qasim E Bhorat, Sagida Bibi, Carmen Briner, Paola Cicconi, Elizabeth A Clutterbuck, Andrea M Collins, Clare L Cutland, Thomas C Darton, Keertan Dheda, Christina Dold, Christopher J A Duncan, Katherine R W Emery, Katie J Ewer, Amy Flaxman, Lee Fairlie, Saul N Faust, Shuo Feng, Daniela M Ferreira, Adam Finn, Eva Galiza, Anna I Goodman, Catherine M Green, Christopher A Green, Melanie Greenland, Catherine Hill, Helen C Hill, Ian Hirsch, Alane Izu, Daniel Jenkin, Carina C D Joe, Simon Kerridge, Anthonet Koen, Gaurav Kwatra, Rajeka Lazarus, Vincenzo Libri, Patrick J Lillie, Natalie G Marchevsky, Richard P Marshall, Ana V A Mendes, Eveline P Milan, Angela M Minassian, Alastair McGregor, Yama F Mujadidi, Anusha Nana, Sherman D Padayachee, Daniel J Phillips, Ana Pittella, Emma Plested, Katrina M Pollock, Maheshi N Ramasamy, Adam J Ritchie, Hannah Robinson, Alexandre V Schwarzbold, Andrew Smith, Rinn Song, Matthew D Snape, Eduardo Sprinz, Rebecca K Sutherland, Emma C Thomson, M Estée Török, Mark Toshner, David P J Turner, Johan Vekemans, Tonya L Villafana, Thomas White, Christopher J Williams, Alexander D Douglas*, Adrian V S Hill*, Teresa Lambe*, Sarah C Gilbert*, Andrew J Pollard*, on behalf of the Oxford COVID Vaccine Trial Group†



EXAMPLE 3: IMMORTAL TIME BIAS?



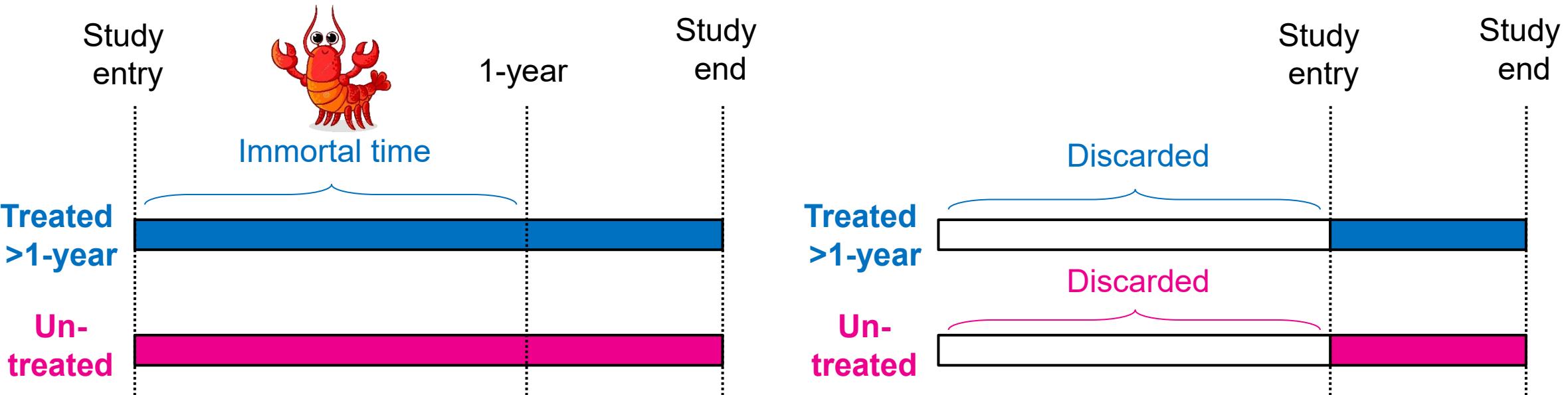
Because this will equally affect the **active** and **control** groups, there is no **overall bias!**

Immortal time bias only occurs when the immortal period is **unbalanced** across exposure groups

SOLUTIONS TO IMMORTAL TIME BIAS

Solution 1: Equal minimum follow-up

All participants must be observed for at least the same period as required for 'treatment'. Discard untreated that die / censored beforehand

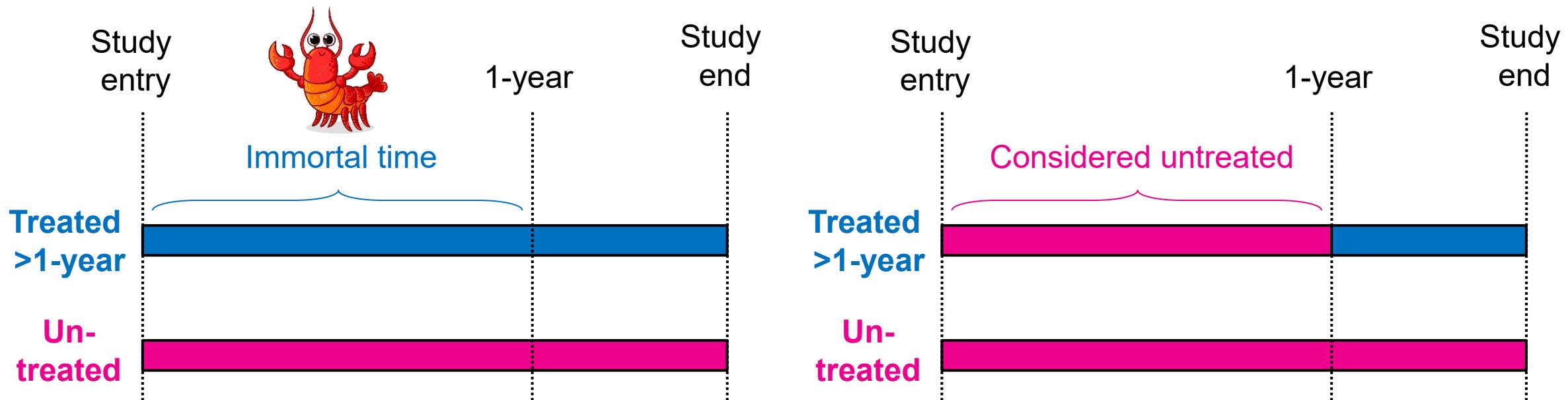


Inefficient – discards useful observations

SOLUTIONS TO IMMORTAL TIME BIAS

Solution 2: Time-dependent analysis

All participants are initially coded as ‘untreated’; the treated flip status after completing the eligibility time



Assumes no effect of partial treatment (seems implausible?)

SOLUTIONS TO IMMORTAL TIME BIAS

Solution 3: Cloning!



RESEARCH METHODS AND REPORTING

How to estimate the effect of treatment duration on survival outcomes using observational data

Miguel A Hernán

Departments of Epidemiology and Biostatistics, Harvard T H Chan School of Public Health, Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts, MA 02115, USA

Correspondence to: M A Hernán
miguel.hernan@post.harvard.edu
Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;360:k182
<http://dx.doi.org/10.1136/bmj.k182>

Accepted: 5 December 2017

When using observational data, quantifying the effect of treatment duration on survival outcomes is not straightforward because only people who live for a long time can receive treatment for a long time. This problem doesn't apply to randomised trials because people are classified based on the treatment duration they are assigned, rather than the treatment duration that they achieve. This approach accepts that dead people do not deviate from their assigned treatment strategy. By transferring this insight to the analysis of observational data, we can follow three steps to estimate the effect of treatment duration from observational data without the bias of naive comparisons between long term and short term users. The first step is cloning people to assign them to multiple treatment strategies. The second step is censoring clones when they deviate from their assigned treatment strategy. The third step is performing inverse probability weighting to adjust for the potential selection bias introduced by censoring. The procedure can be used

to compare any treatment strategies that are sustained over time. Cloning, censoring, and weighting eliminates immortal time bias in the estimates of absolute and relative risk, which helps researchers focus their attention on other biases that may be present in observational analyses and are not so easily eliminated.

Introduction

Quantifying the effect of treatment duration on survival outcomes is not straightforward because only people who survive for a long time can receive a treatment for a long time. Suppose we want to estimate the effect of statins on the mortality of patients with cancer using a healthcare database.¹ A direct comparison of long term users, short term users, and non-users would be biased because long term users have, by definition, survived for a long time. Several methods can be used to tackle this bias, but some do not enable estimation of absolute risks or appropriate adjustment for time varying confounders. To overcome these limitations, I first review an uncontroversial approach to estimating the effect of treatment duration in randomised trials and then explain how to emulate this approach in observational data analyses.

Estimating the effect of treatment duration in a randomised trial with full adherence

Let us consider a simple example that encapsulates some key features of the problem. Table 1 shows data from a trial with perfect adherence and no loss to follow-up, in which 12 people are randomly assigned to one of three treatment strategies: no aspirin ($\text{durA}=0$), one year of aspirin ($\text{durA}=1$), or two years of

ANALYTICAL SOLUTIONS TO IMMORTAL TIME BIAS

The ‘cloning’ approach is useful when you’re interested in a ‘length of..’ exposure (e.g. duration of aspirin treatment)

Step	Goal	Method
1	Assign each individual to every strategy consistent with her data at time zero	Cloning
2	End follow-up when an individual’s data stop being consistent with her assigned strategy	Censoring
3	Adjust for selection bias introduced by the previous step	IP weighting

- Step 1: At time zero, individuals are cloned, with one clone assigned to each possible exposure regime
- Step 2: Clones are censored once the observed data is incompatible with the allocated treatment
- Step 3: Reweight sample appropriately

CLONING

e.g. study of duration of aspirin (durA), either 0 years, >0-1 years, or >1-2 years:

Person/clone	durA	Aspirin at start of first year	Dead at end of first year	Aspirin at start of second year	Dead at end of second year	Inverse probability weight
1	0	No	No	No	No	1
2	0	No	No	No	Yes	1
3	0	No	No	No	Yes	1
4	0	No	Yes	-	Yes	1
5a	1	Yes	No	No	No	2
6a	1	Yes	No	No	Yes	2
7a	1	Yes	No	No	Yes	2
8a	1	Yes	Yes	-	Yes	1
9a	1	Yes	No	Yes	Censored	0
10a	1	Yes	No	Yes	Censored	0
11a	1	Yes	No	Yes	Censored	0
12a	1	Yes	Yes	-	Yes	1
5b	2	Yes	No	No	Censored	0
6b	2	Yes	No	No	Censored	0
7b	2	Yes	No	No	Censored	0
8b	2	Yes	Yes	-	Yes	1
9b	2	Yes	No	Yes	No	2
10b	2	Yes	No	Yes	Yes	2
11b	2	Yes	No	Yes	Yes	2
12b	2	Yes	Yes	-	Yes	1

FURTHER READING

- Hernán, M.A. and Robins, J.M., 2016. Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8), pp.758-764.
- Hernán, M.A., 2018. How to estimate the effect of treatment duration on survival outcomes using observational data. *Bmj*, 360.
- Kucher, S.A., Brophy, J.M., Banack, H.R., Kaufman, J.S. and Samuel, M., 2021. Emulating a Randomised Controlled Trial With Observational Data: An Introduction to the Target Trial Framework. *Canadian Journal of Cardiology*, 37(9), pp.1365-1377.
- Yadav, K. and Lewis, R.J., 2021. Immortal time bias in observational studies. *Jama*, 325(7), pp.686-687.

SUMMARY

The **target trial framework** is a simple aid to planning the analysis of observational data that is designed to help avoid several pitfalls

The approach involves preparing a protocol for an imaginary ‘**target trial**’ that you would use to estimate your causal effect of interest

You then create a second protocol for the ‘**emulation study**’

The target trial framework is useful for avoiding:

- **healthy user bias** – which occurs because a certain degree of health is required to stay in treatment
- **immortal time bias** – which occurs when there is some time mismatch between eligibility, treatment assignment, and time zero



2.2 - DRAWING AND EVALUATING DAGS FOR APPLIED RESEARCH

GEORGIA



MARK



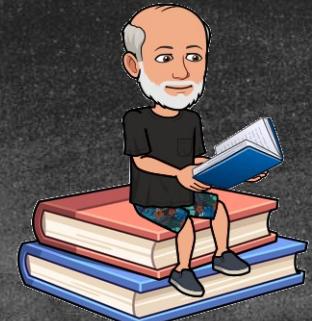
@GEORGIATOMOVA

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 2

09:30-10:15 LECTURE 2.1

10:15-11:00 ACTIVITY 2-A

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 2.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:00 ACTIVITY 2-B

15:00-15:30 LECTURE 2.3

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 2.4

17:00-17:45 ACTIVITY 2-C

17:45-18:00 Q&A

LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- Depict complex causal scenarios in a DAG
- Depict **bidirectional relationships**
- Depict variables with **common correlations** but unclear causality
- Draw DAGs like a pro!

You will also be able to define and explain:

- **Crystallisation**
- **Time-variant** and **time-invariant** variables
- **Temporal precedence**

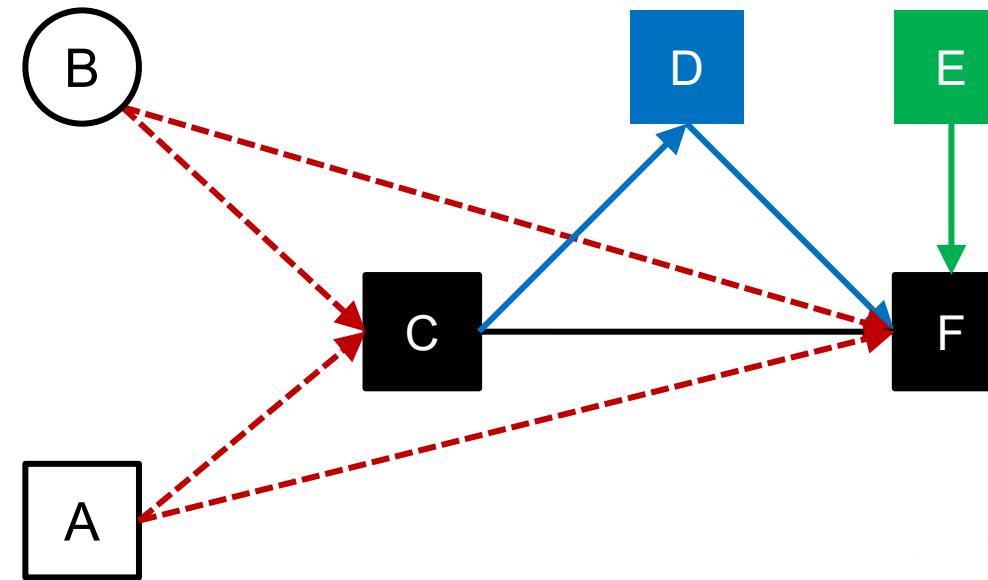
RECAP: DAGS

- Directed acyclic graphs are a common form of '**causal diagram**'
- They are an aid to planning and conducting observational data analysis where we seek to estimate **causal effects**
- **Causal effects** are interesting, because they symbolise the (potentially modifiable) effect of an **exposure** on an **outcome**
- Most quantitative social science is ultimately interested in estimating causal effects

RECAP: DAGS

To estimate the causal effect of **C** on **F**:

- Aim for all **causal paths** to be **open**
- Aim for all **confounded paths** to be **closed**



RECAP:

DO:

- Condition on **confounders** to block **confounding paths**

DO NOT:

- Condition on **mediators** as this would block **true causal paths** (and far worse...)

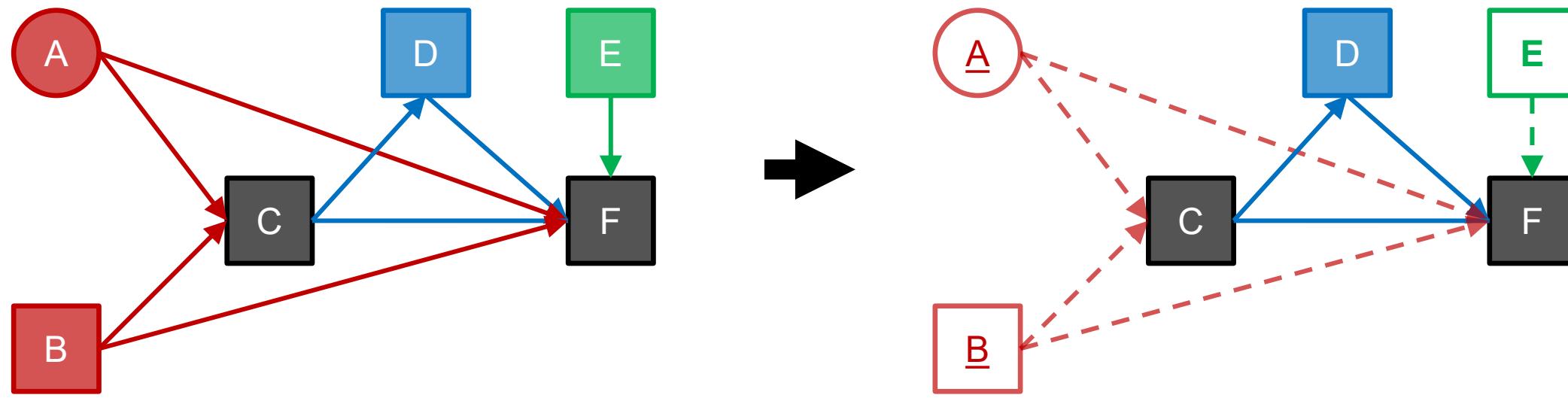
OPTIONALLY:

- Condition on **competing exposures** to improve the precision of your estimates

After drawing your **DAG**, you can use tools such as ***dagitty.net*** to identify adjustments sets that follow these rules

DRAWING DAGS

These rules are very easy to learn and apply to toy examples

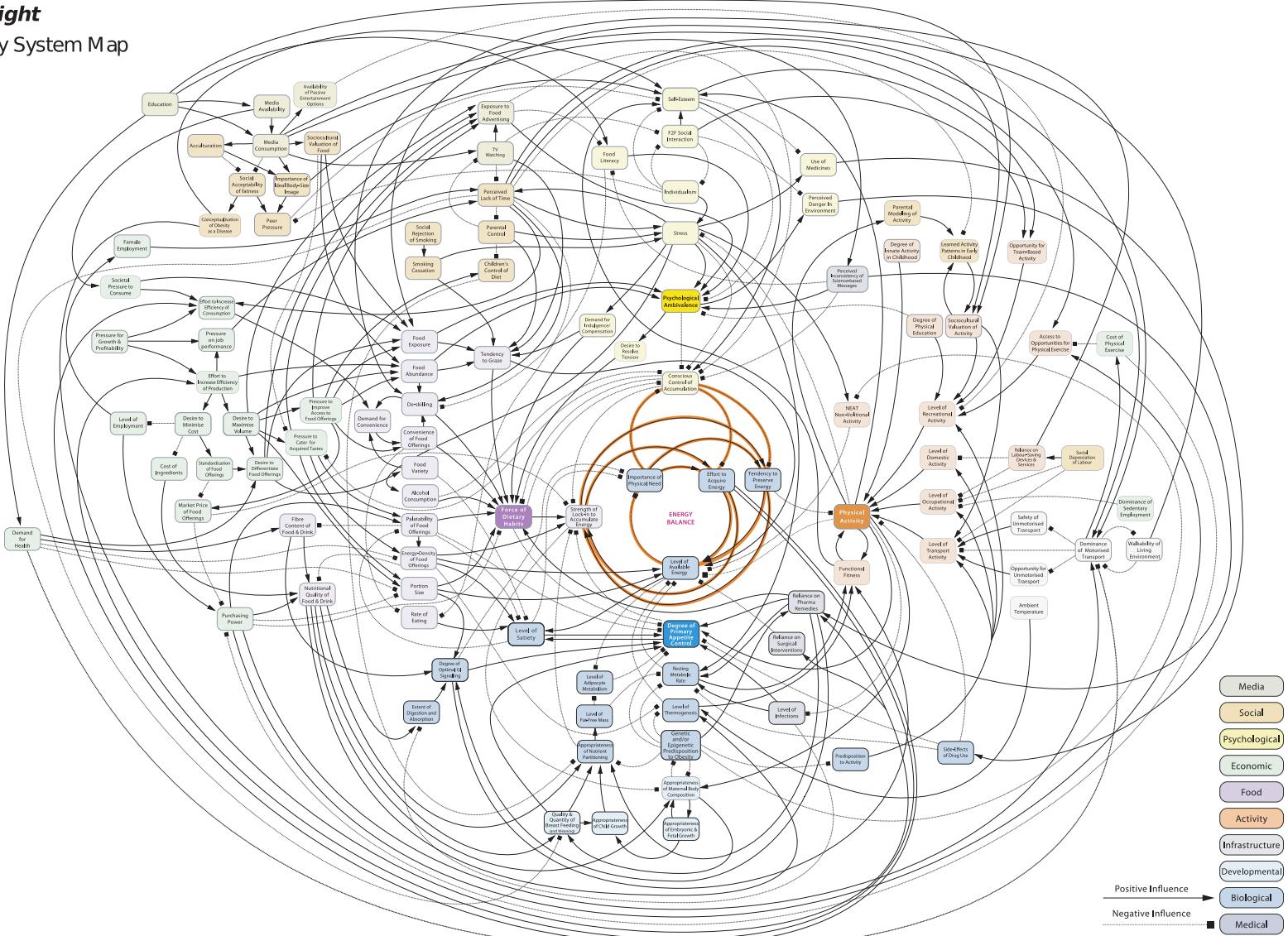


DRAWING DAGS

But the real world can be a little more complex...

Foresight

Obesity System Map



ANALYTIC PERSPECTIVE

Open Access



Situating agent-based modelling in population health research

Eric Silverman¹ , Umberto Gostoli¹, Stefano Picascia¹, Jonatan Alr
Claudio Angione²

Abstract

Today's most troublesome population health challenges are often driven by complex systems which are difficult to model using traditional epidemiological methods. To encourage a wider adoption of agent based modelling (ABM) in taking on these challenges, we argue that for ABM to be most effective for answering questions normally inaccessible to the traditional epidemiologist, it must demonstrate the utility of ABM for population health research, and to clear up persistent misconceptions about its conceptual underpinnings, we offer a detailed presentation of the strengths and limitations of ABM for population health. We summarise why simulations are essential to the study of complex systems, highlight the strengths of ABM for population health, and propose they are well-suited for the study of population health, and could make significant contributions to theory and intervention.

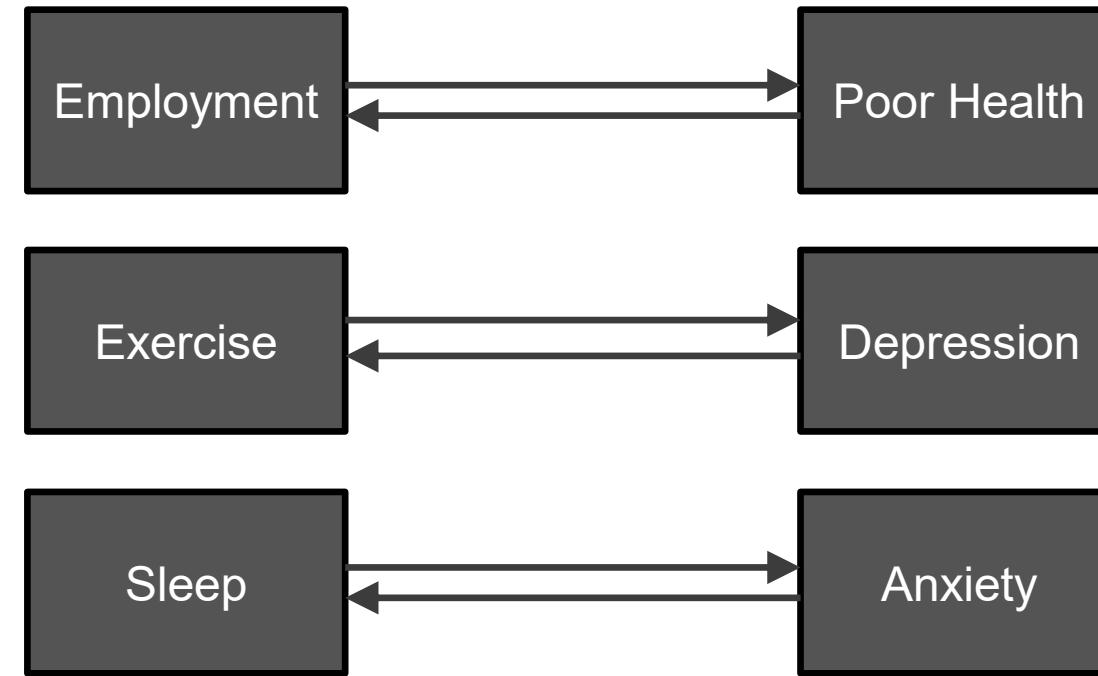
Keywords: Agent-based modelling, Population health, Complexity

However, CIM as applied in epidemiology today, has coalesced around a set of tools with certain limitations when applied to complex systems. For example, [directed acyclic graphs \(DAGs\)](#) are frequently used in CIM, but DAGs are unsuitable for modelling systems containing feedback loops (given they are acyclic), a common feature of complex social systems. Critics of CIM also suggest it takes an overly linear view of the decision-making process leading from evidence to decision-making, and have called for alternative concepts of cause that are not solely based on probabilistic statements about population outcomes in alternative worlds [\[2\]](#). Decisions over causality can be based on pragmatic pluralism [\[3\]](#), or inference to the best explanation [\[2, 4\]](#) (often characterised as a form of abductive reasoning). Decision-makers ultimately must make decisions, even while accepting that their evidence is incomplete or flawed or both, and that the exact causal process underlying the system of interest is still uncertain.

SPECIFYING DAGS

How do we manage the ‘*directed*’ rule, when:

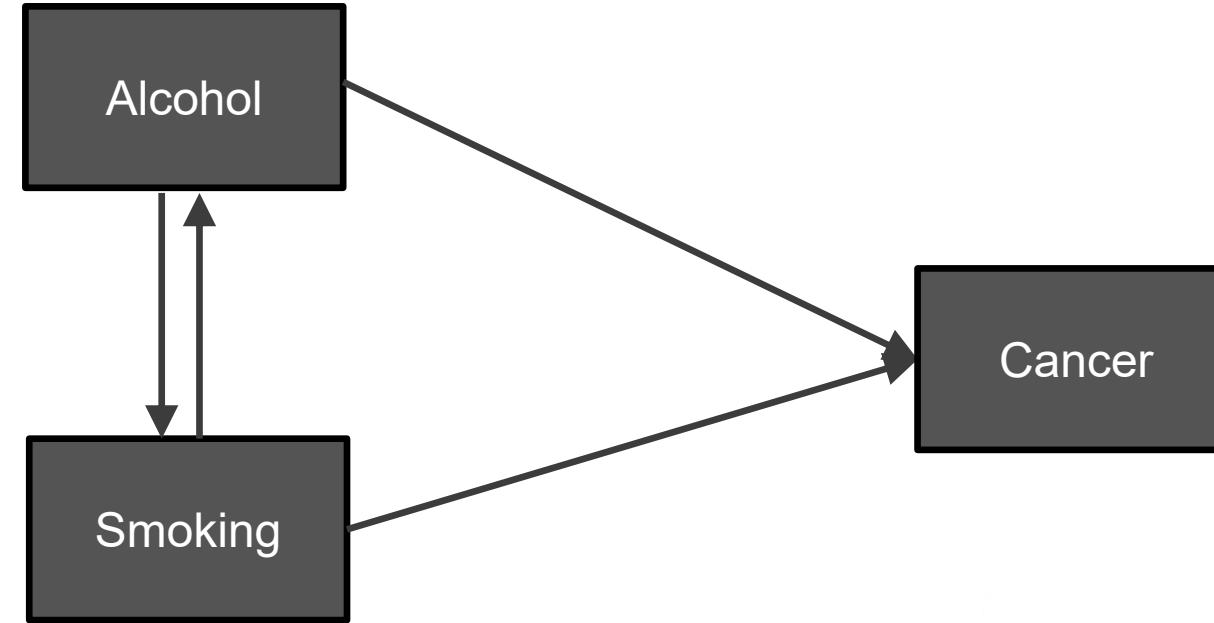
- some associations are bidirectional?



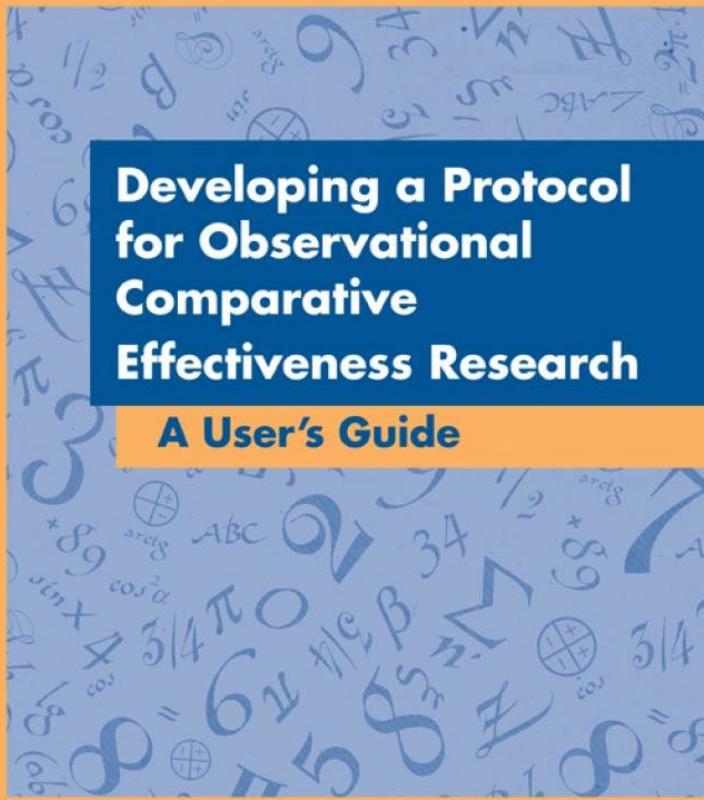
SPECIFYING DAGS

How do we manage the ‘*directed*’ rule, when:

- some associations are bidirectional?
- some variables are completely interrelated with no clear direction?



A NEED FOR GUIDENCE



Sauer & VanderWeele 2013

Guidance	Key Considerations	Check
Develop a simplified DAG to illustrate concerns about bias.	<ul style="list-style-type: none">Use a DAG to illustrate and communicate known sources of bias, such as important well known confounders and causes of selection bias.	<input type="checkbox"/>
Develop complete DAG(s) to identify a minimal set of covariates.	<ul style="list-style-type: none">Construction of DAGs should not be limited to measured variables from available data; they must be constructed independent of available data.The most important aspect of constructing a causal DAG is to include on the DAG any common cause of any other two variables on the DAG.Variables that only causally influence one other variable (exogenous variables) may be included or omitted from the DAG, but common causes must be included for the DAG to be considered causal.Identify a minimal set of covariates that blocks all backdoor paths and does not inadvertently open closed pathways by conditioning on colliders or descendants.	<input type="checkbox"/>

"Developing DAGs is not always easy... A disciplined approach to developing DAGs may be useful for communicating findings and providing rationale for covariate selection"

DRAWING DAGS

There are currently no formal ‘how to’ guides to help with drawing DAGs

Applied scientists have struggled to adapt the theory into real settings

The image shows a screenshot of a research article from the International Journal of Epidemiology. At the top left is the IEA logo. To the right is the journal information: International Journal of Epidemiology, 2020, 1–13, doi: 10.1093/ije/dyaa213, Original Article. Below this is a decorative circular emblem. A horizontal line separates this from the article title. The title is 'Original Article' followed by 'Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations'. Below the title is a list of authors and their institutions. At the bottom is a block of small text detailing the affiliations.

IEA
International Epidemiological Association

International Journal of Epidemiology, 2020, 1–13
doi: 10.1093/ije/dyaa213
Original Article

Original Article

Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations

Peter WG Tennant ^{1,2,3*} Eleanor J Murray,⁴ Kellyn F Arnold ^{1,2},
Laurie Berrie,^{1,5,6} Matthew P Fox,^{4,7} Sarah C Gadd,
Wendy J Harrison ^{1,2} Claire Keeble,¹ Lynsie R Ranker,⁴ Johannes
Textor ⁸, Georgia D Tomova,^{1,2,3} Mark S Gilthorpe^{1,2,3†}
and George TH Ellison^{1,2,9†}

¹Leeds Institute for Data Analytics, University of Leeds, Leeds, UK, ²Faculty of Medicine and Health, University of Leeds, Leeds, UK, ³Alan Turing Institute, British Library, London, UK, ⁴Department of Epidemiology, School of Public Health, Boston University, Boston, MA, USA, ⁵School of Geography, University of Leeds, Leeds, UK, ⁶School of GeoSciences, University of Edinburgh, Edinburgh, UK, ⁷Department of Global Health, Boston University, Boston, MA, USA, ⁸Department of Tumour Immunology, Radboud University Medical Center, Nijmegen, The Netherlands and ⁹Centre for Data Innovation, Faculty of Science and Technology, University of Central Lancashire, Preston, UK

- Some useful recommendations are available in this recent review
- But it isn't really a guidebook...

DRAWING DAGS

We have therefore drafted:

THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. **Develop and state a clear research question**
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



1) DEVELOP AND STATE A CLEAR RESEARCH QUESTION



Sarah B. Andrea, PhD, MPH @SarahBAndrea · Jan 20, 2019

Hi! I'm an epidemiologist. You might know me from my greatest hits including "**what is your research question?**" Hang on, let me draw the DAG" and "Please stop picking your potential confounders based on whether they are statistically significant in the model."

...



Melissa K Sharp, PhD² @sharpmelk · Mar 2, 2020

Hi, I'm an **#epidemiologist**. You might know me from my greatest hits "well, it depends" "but, **what is your research question?**" "But, did they control for..." and the ever present "No, I don't work on skin"

...



Andrew Althouse @ADAlthousePhD · Jan 19, 2019

Hi! I'm a biostatistician. You might know me from my greatest hits including "No, we can't really calculate 10 year survival for something that's only been used for seven years" and "No, you shouldn't say that p=0.08 is 'trending towards' significance" twitter.com/mad_sters/stat...



Helen Ward ✅ @profhelenward · Jan 20, 2019

...

Replies to [@trishgreenhalgh](#)

Hi, I'm an epidemiologist. You may know me from my greatest hits like "no, I don't specialise in skin", "but **what is your research question?**", or "I repeat, correlation does not imply causation".



Maarten van Smeden @MaartenvSmeden · Jan 20, 2019

...

Hi! I'm a statistician. You might know me from my greatest hits including "**let's talk about your research question**", "No, really **let's talk about your research question**" and No. 1 hit: "can't help you without a research question"



Jules Glegg 🏳️‍🌈 @heyjulesfern · Feb 27, 2020

Hi, I'm a senior engineer. You might know me from my greatest hits "but *why* is it null?" "There is greater honor in deleting code than in writing more" And the chart-topping "let's just walk over to them and ask"
[Show this thread](#)



Madeline Sterling, MD, MPH @mad_sters · Jan 18, 2019

Hi! I'm a general internist. You might know me from my greatest hits including "So, tell me how you are? you heard me, I **really care about everything!**" or "Yup, I can (try to) coordinate all of that!" or "Not only will I figure out what you have, but I'll fill out your forms" twitter.com/marklewismd/st...

1) DEVELOP AND STATE A CLEAR RESEARCH QUESTION

For most causal inference research, this means identifying and stating your **focal relationship**:

- What is your **exposure**
- What is your **outcome**
- What is your **estimand**

Examples:

- The **total causal effect** of **fasting plasma glucose (FPG)** at 28 week's gestation on the risk of **stillbirth**
- $P[\text{Stillbirth}=1|\text{FPG}=\text{raised}]-P[\text{Stillbirth}=1|\text{FPG}=\text{normal}]$

This is fundamental to interpreting a DAG because your contextual variables (i.e. '**confounders**', '**mediators**', and '**competing exposures**') only take those roles *with respect to a specific focal relationship*

THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. **Consider and state your context**
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



2) CLEARLY DEFINE YOUR CONTEXT

DAGs are context specific

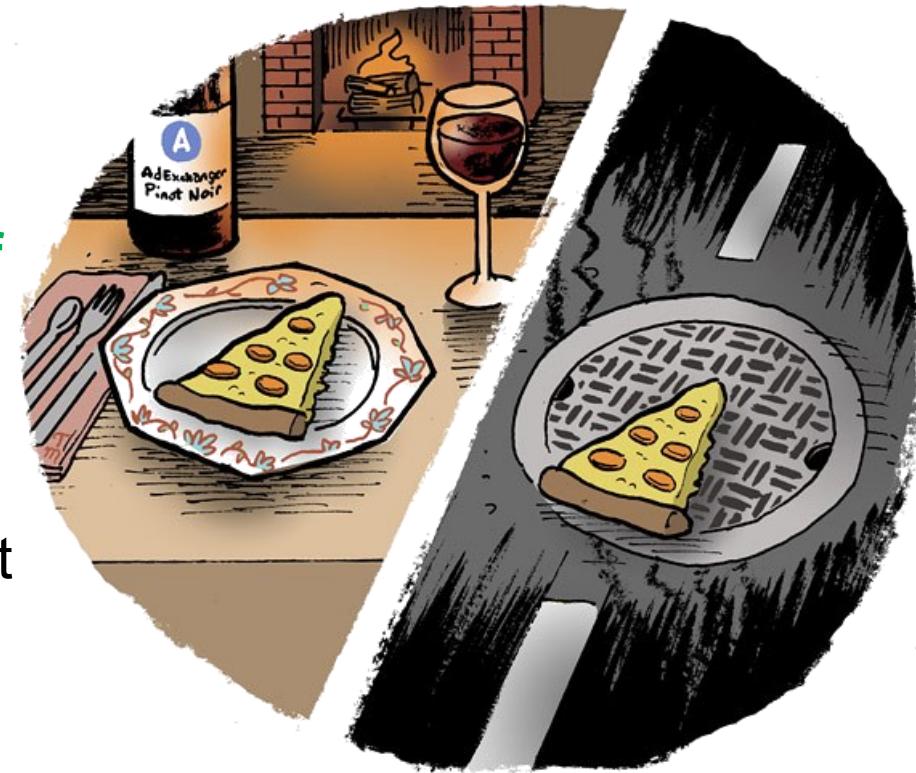
- Your DAG is *your attempt to describe the data-generating process for a specific population*

Population = Who? Where? When?

- e.g. **Women attending maternity care** in the **North of England** during **2012-2015**

Focus on that **context** when drawing your DAG

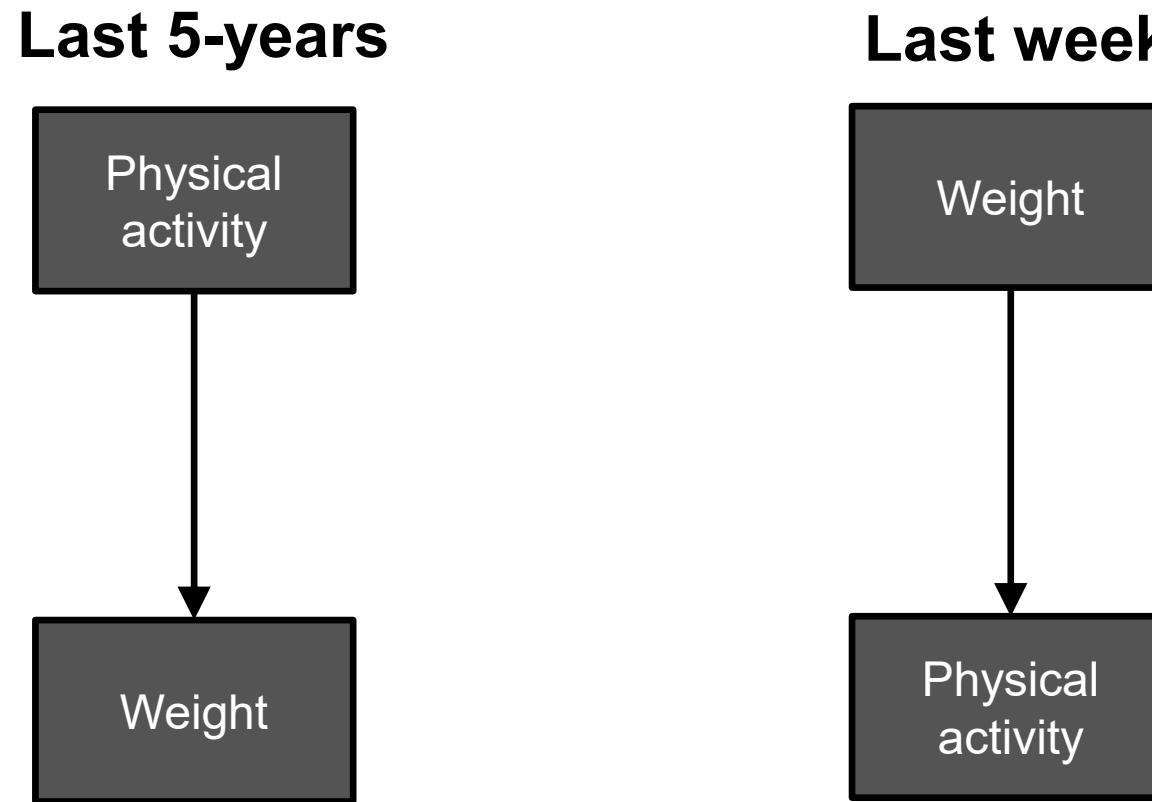
- Doesn't matter how things work in other settings – what matters is the data generation mechanism for your context / dataset!
- Note: A cause in one context may be a consequence in another!



Context Matters

2) CLEARLY DEFINE YOUR CONTEXT

Example: Weight and physical activity in adult British males:



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. **Draw your DAG(s) as early as possible**
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



3) DRAW YOUR DAG AS EARLY AS POSSIBLE

DAGs are an important aid to identifying your assumptions and which variables require conditioning to (robustly) estimate a causal effect

- **Ideally:** draw your DAG before data collection – as part of the study design
- **At least:** draw your DAG before downloading data and conducting analyses

You *can* create updated DAGs from expert feedback or sensitivity analyses

- **But:** do not redraw your DAG to hide missing variables or because you're not getting the results you want !



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. **Get help - don't draw it alone**
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



4) GET HELP – DON’T DRAW IT ALONE

When we draw a **DAG**, we’re trying to use our **knowledge, expertise, and theory** to describe the **data-generating mechanism**

The more accurately our DAG reflects reality, the more accurate our causal estimates will be

Try to use the maximum knowledge and expertise!

- Drawing DAGs should be a **team effort**
- Get input from methods and context experts
- Consider sharing widely for feedback



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. **Include all relevant variables**
6. Draw your DAG(s) in temporal order
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)

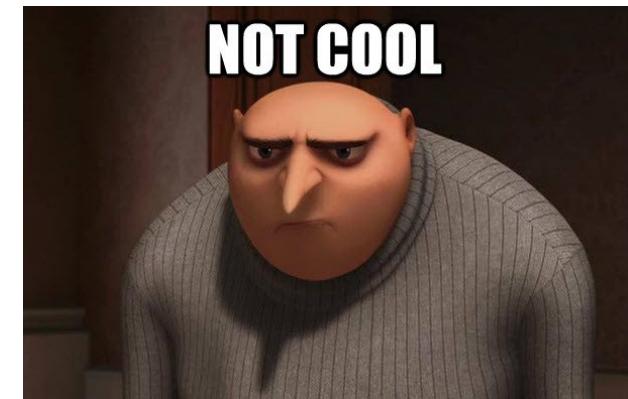


5) INCLUDE ALL RELEVANT VARIABLES

Your DAG should include **all** variables that (potentially) cause the **exposure** and **outcome**

Not just those:

- that are easiest to measure
- that were measured when the cohort was established
- without much missing data
- that give you the result you want (...again, *not cool!*)



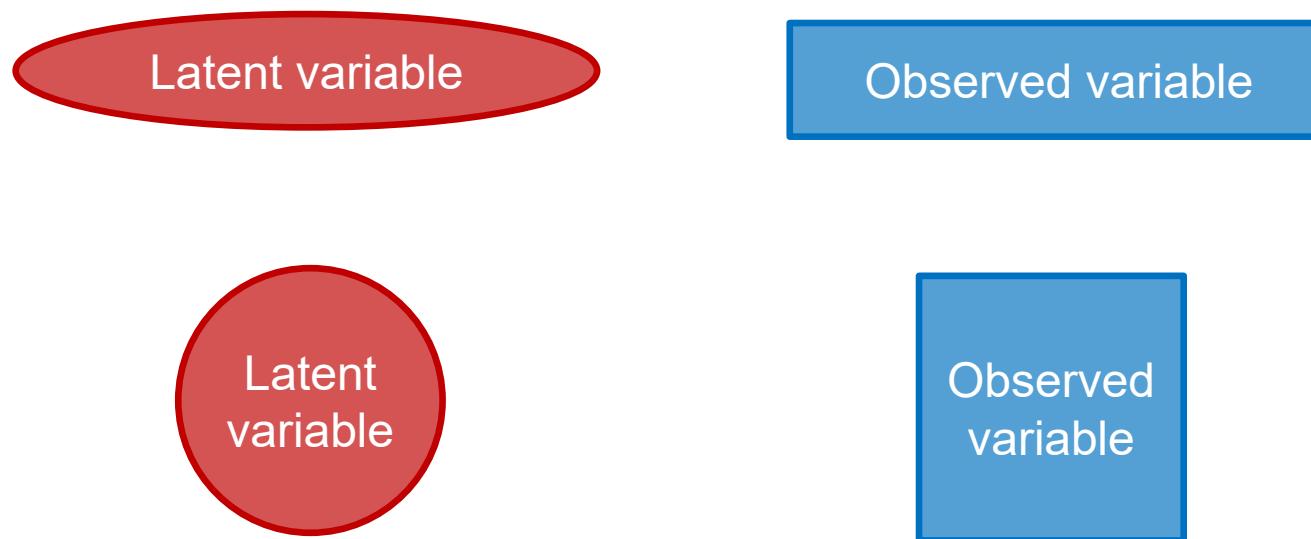
If interested in multiple exposures and/or outcomes:

- include **all** variables that determine **all** the exposures and outcomes of interest
- consider separate DAGs to make sure you're not overlooking some variables

5) INCLUDE ALL RELEVANT VARIABLES

If a variable can't be observed, or isn't available, include it as a **latent variable**

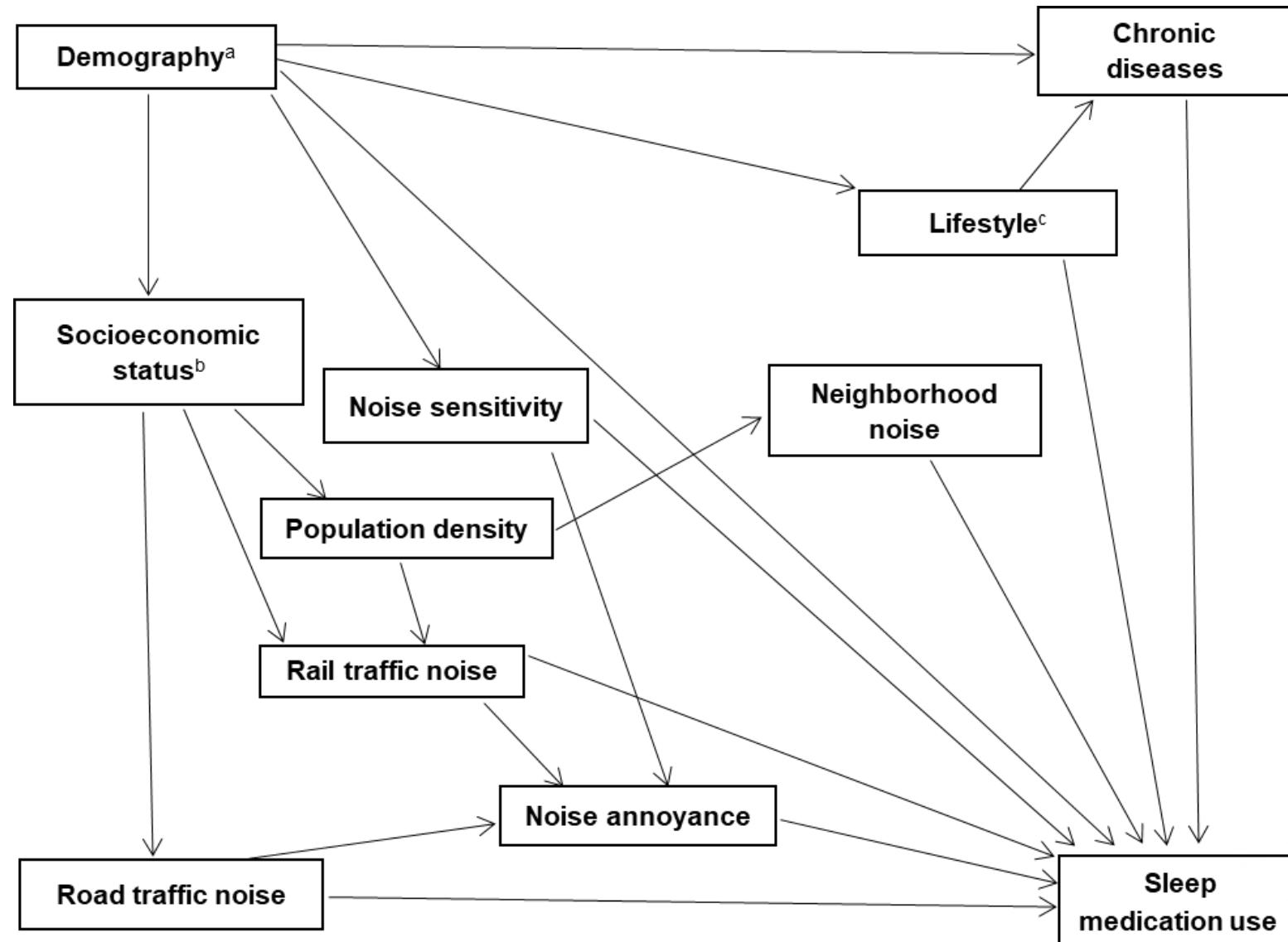
- **Latent variables** are typically depicted in **circles** or **ellipses**
- **Observed variables** as **squares** or **rectangles**



5) INCLUDE ALL RELEVANT VARIABLES

- If the number of variables becomes excessively large, you may wish to consider collapsing similar variables into '**super-nodes**'
- '**Super-nodes**' can take two forms
 1. Similar variables are depicted as a single node with a summary name
 - E.g. *socioeconomic position at birth, highest level of education, and parental income*, are simply depicted as 'socioeconomic confounders'

5) INCLUDE ALL RELEVANT VARIABLES



^a'Demography' includes 'age', 'sex', and 'having children ≤ 5 years'

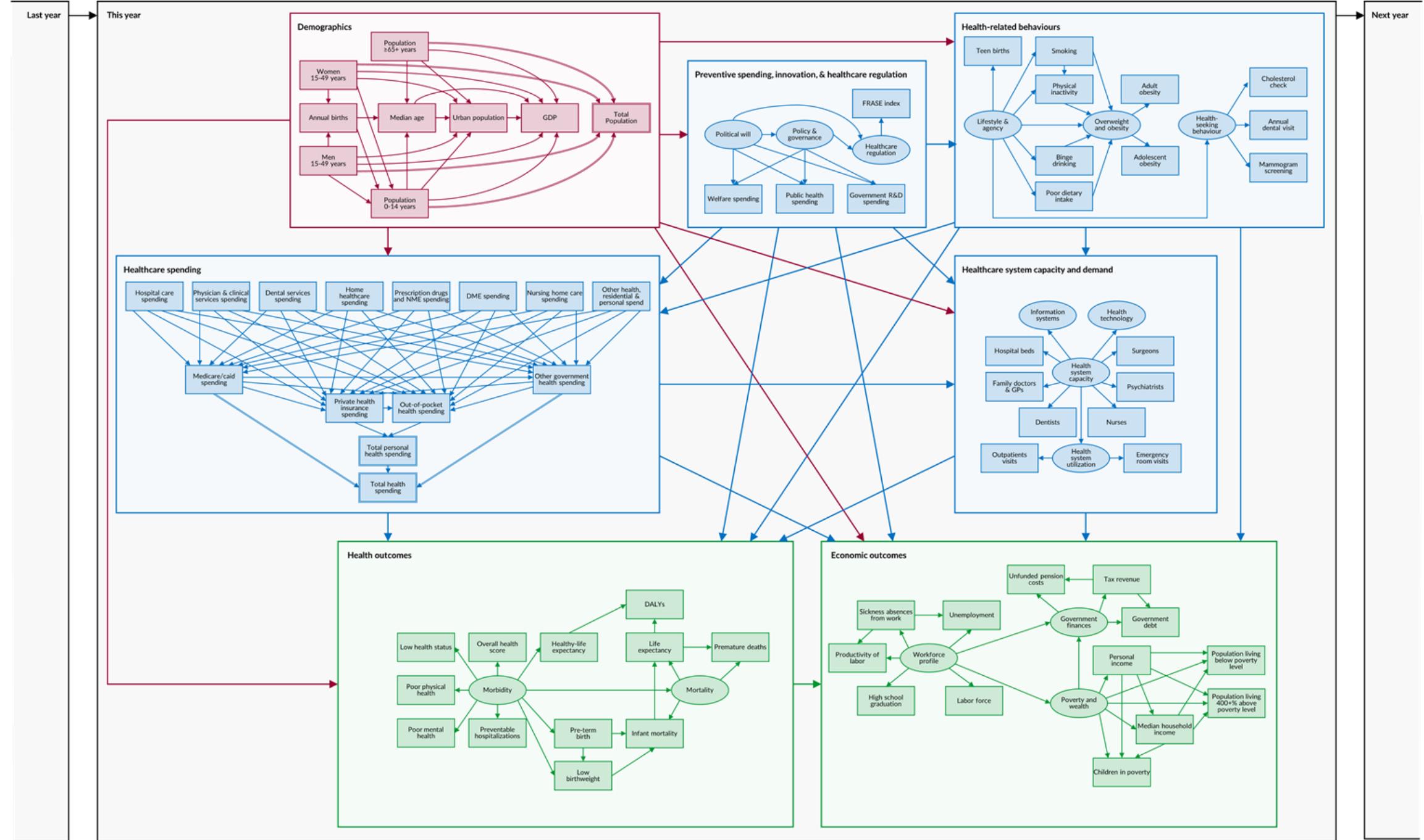
^b'Socioeconomic status' includes 'education' and 'household income'

^c'Lifestyle' includes 'smoking status', 'alcohol use', 'caffeine use', 'physical activity', and 'night shift work'

Source: Evandt J, Oftedal B, Krog NH, et al. Road traffic noise and registry based use of sleep medication. *Environ Health.* 2017;16(1):110. DOI: 10.1186/s12940-017-0330-5

5) INCLUDE ALL RELEVANT VARIABLES

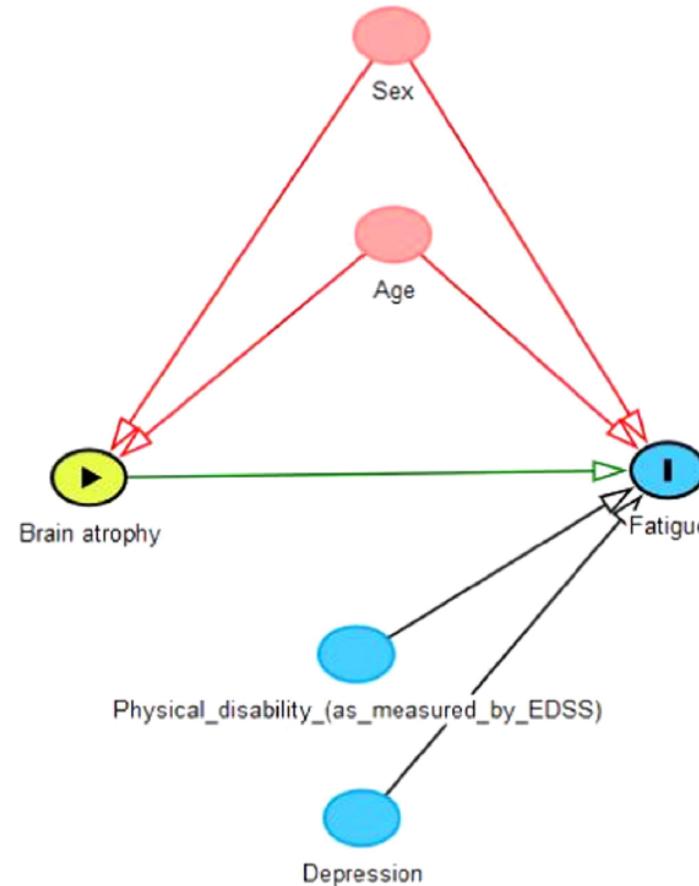
- If the number of variables becomes excessively large, you may wish to consider collapsing similar variables into ‘super-nodes’
- ‘Super-nodes’ can take two forms
 1. Similar variables are depicted as a single node with a summary name
 - E.g. *socioeconomic position at birth, highest level of education, and parental income*, are simply depicted as ‘socioeconomic confounders’
 2. Similar variables are grouped together into a ‘mini-DAG, within a larger box variable that otherwise behaves like a single node



5) INCLUDE ALL RELEVANT VARIABLES

- If the number of variables becomes excessively large, you may wish to consider collapsing similar variables into ‘super-nodes’
- ‘Super-nodes’ can take two forms
 1. Similar variables are depicted as a single node with a summary name
 - E.g. *socioeconomic position at birth, highest level of education, and parental income*, are simply depicted as ‘socioeconomic confounders’
 2. Similar variables are grouped together into a ‘mini-DAG, within a larger box variable that otherwise behaves like a single node
- ‘Super-nodes’ require that constituent variables can be situated together
- The first approach is more appropriate when the interrelationship between constituent nodes is not especially relevant to the focal relationship

DON'T BE LIKE THIS!



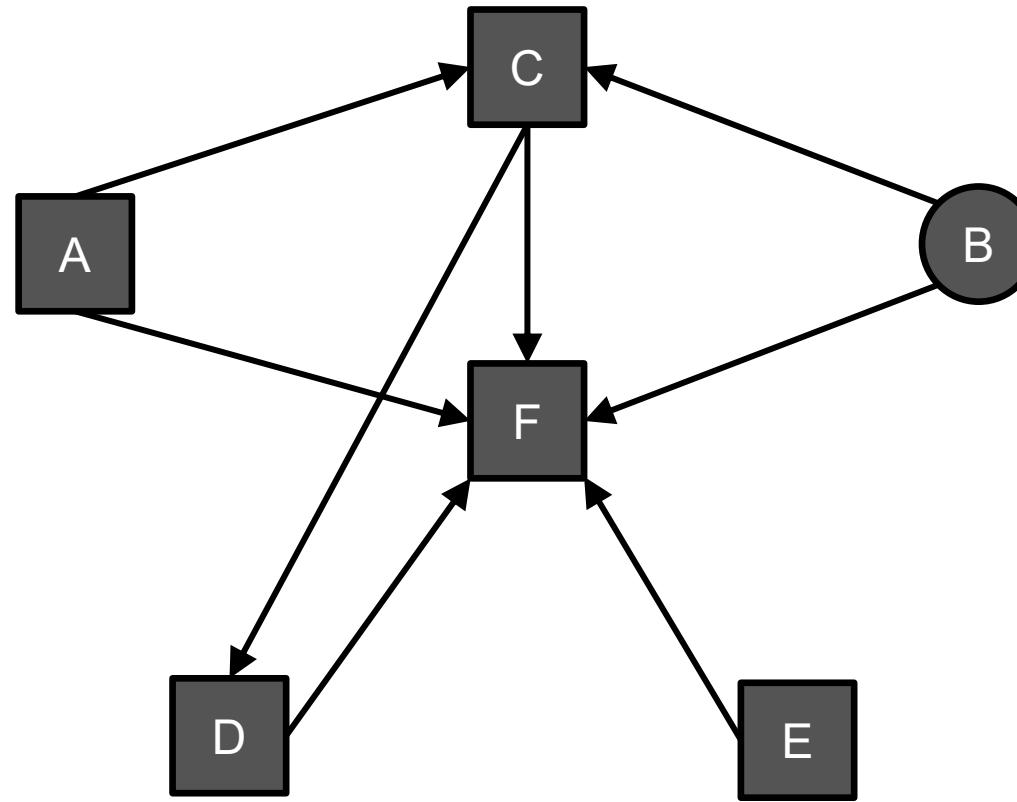
THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. **Draw your DAG(s) in temporal order**
7. Draw forward arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

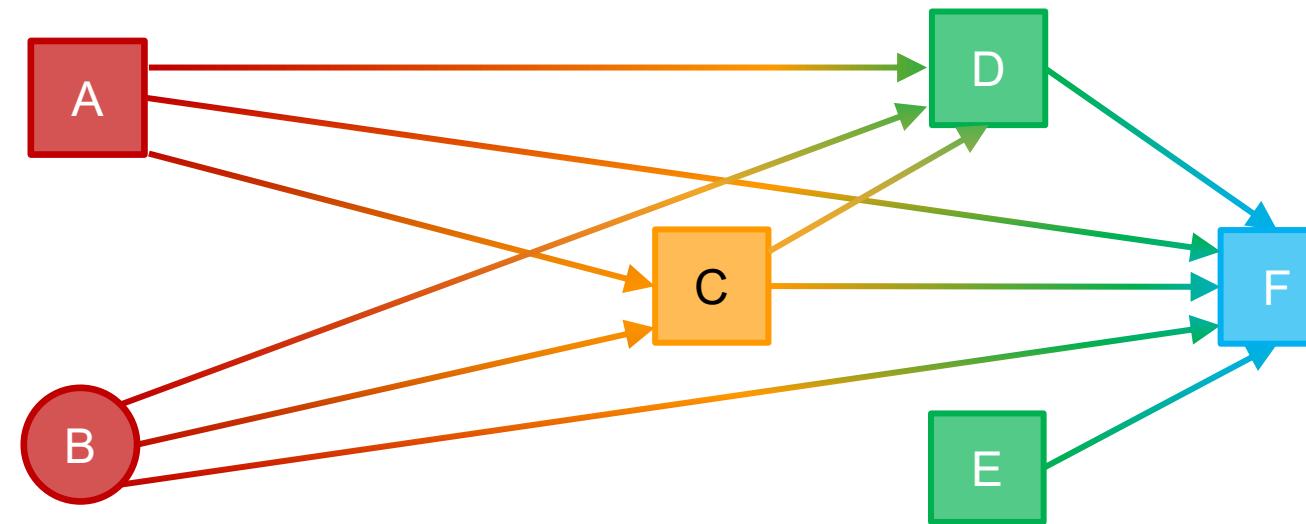
DAGs can be drawn in various attractive ways



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

DAGs can be drawn in various attractive ways

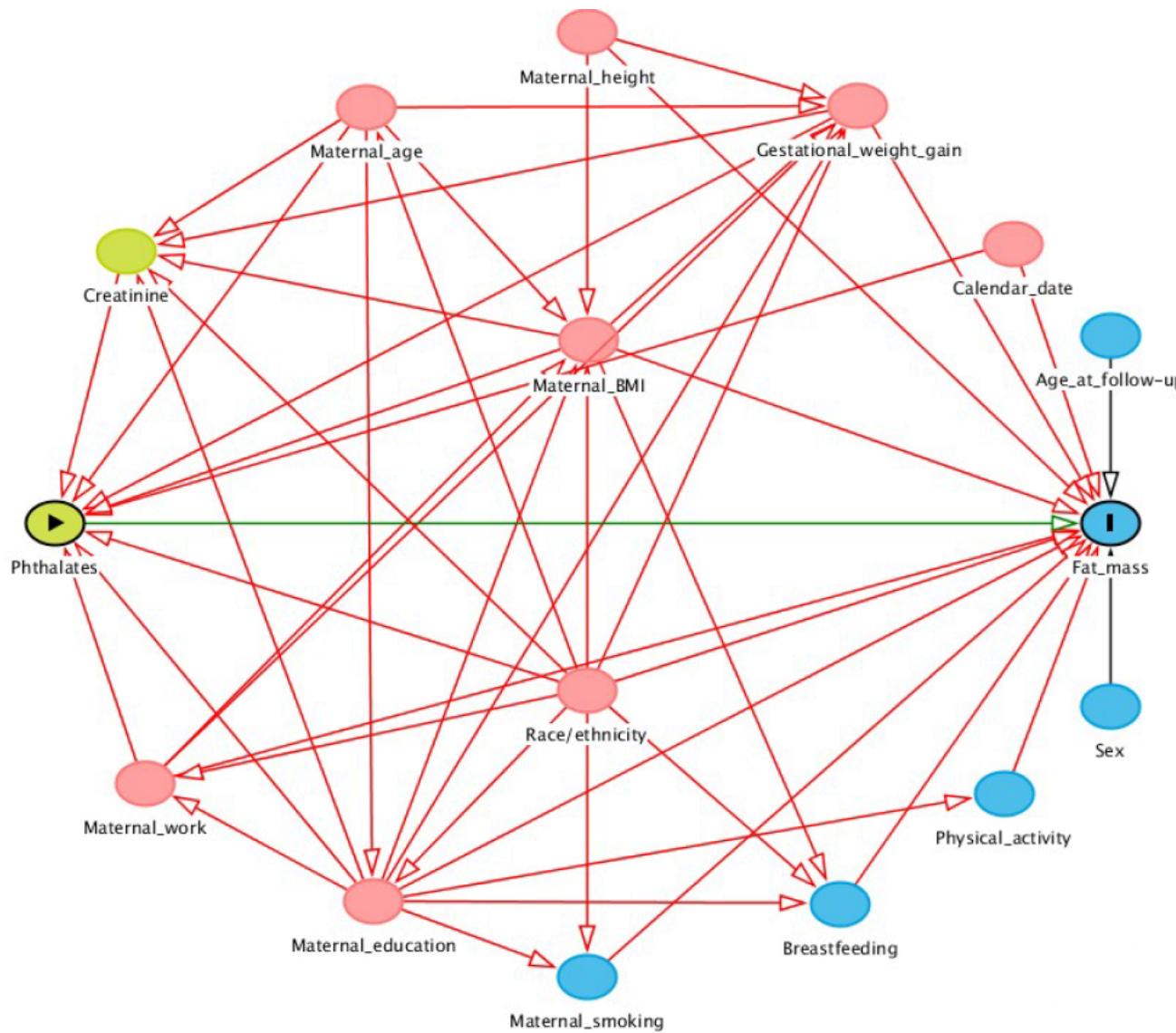
But it is very helpful to arrange them in time order



Nodes depict **events** at specific moments in time and arcs depict **causal processes** occurring in/over time

Drawing DAGs this way helps avoid mistakes and is easier for others to interpret

DON'T BE LIKE THIS!



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

This requires careful **THOUGHT**

- What does each variable **really mean?**
- When did it **crystallise?**
- What about variables that '**feedback**' or occur at similar times?



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

A) What does each variable really *mean*?

Variables are abstract representations of one or more concepts

- Some are relatively simple:
e.g. peak adult height
- Others less so:
e.g. race
- Some are deceptively complex:
e.g. age (includes information on age, period, and cohort effects)

For each variable, you need to think carefully about what it truly ‘means’

- What information does it capture, and when did this come into being

5) ARRANGE THE VARIABLES IN TEMPORAL ORDER

A) What is the nature of each variable?

Variables may be **time-variant**, **time-invariant**, or a mix

Time invariant variables are easiest to place, as they crystallised at a fixed point in time

- e.g. maternal age at birth ⇒ occurred at conception

Time variant variables are much harder, you need to consider the timeframe over which the measured values formed

- e.g. anxiety levels ⇒ occurred over life, and perhaps more recently

Some variables depend on context

- e.g. height ⇒ variant in children, invariant if young adults, variant in older adults

5) ARRANGE THE VARIABLES IN TEMPORAL ORDER

Is biological sex time-invariant?

- What does biological sex mean?
 - ✓ Your genes (**time-invariant**)
 - ✓ Your hormonal exposures (**time-variant**)
 - ✓ Your behaviours (?)

Some variables contain information on several concepts?

- What does ‘socio-economic circumstances’ mean?
 - ✓ Your parents’ jobs?
 - ✓ Your education?
 - ✓ Your income?
 - ✓ Where you live?

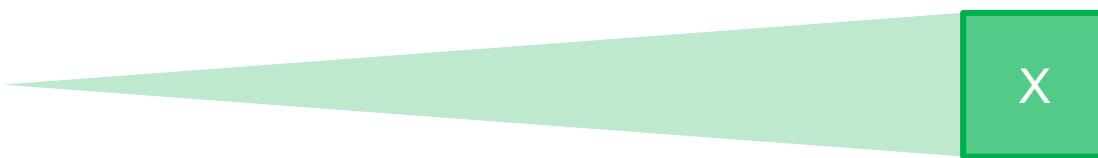
To place these, we often need to be more explicit!

6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

B) When did each variable **crystallise**?

The time something was measured *can* help anchor its position in time ...

... but not always, since a measure might represent insight into an underlying process that may have been occurring over a long time period



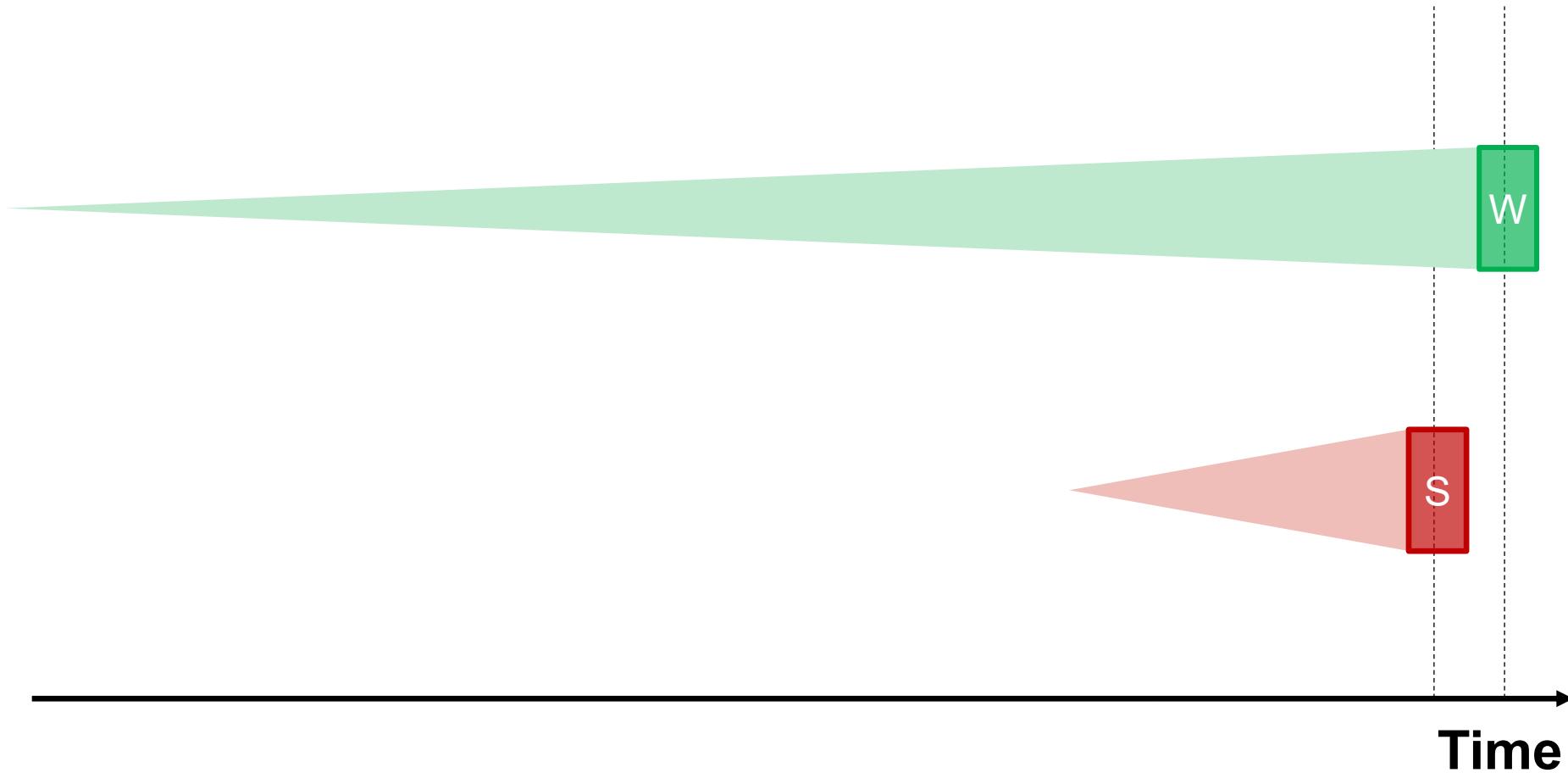
Suppose:

- on waking, we ask you how well you **slept** *last night*
- we then measure your **weight**

Which goes first? **Sleep (last night)** or **weight (this morning)**?

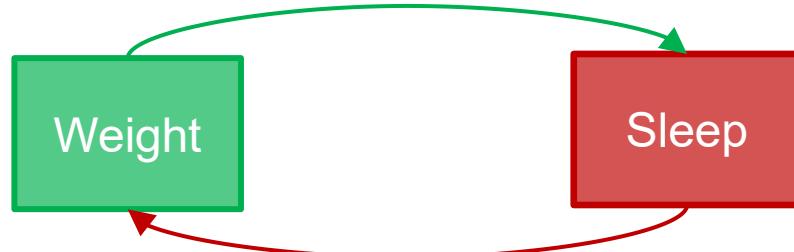
6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

Although **weight** was *measured* after last night's **sleep**, it has mostly occurred beforehand !



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

What if the relationship is **bi-directional**?



Many relationships are bi-direction (and **feedback**) over time

Think carefully about your context – and time window – being measured / studied

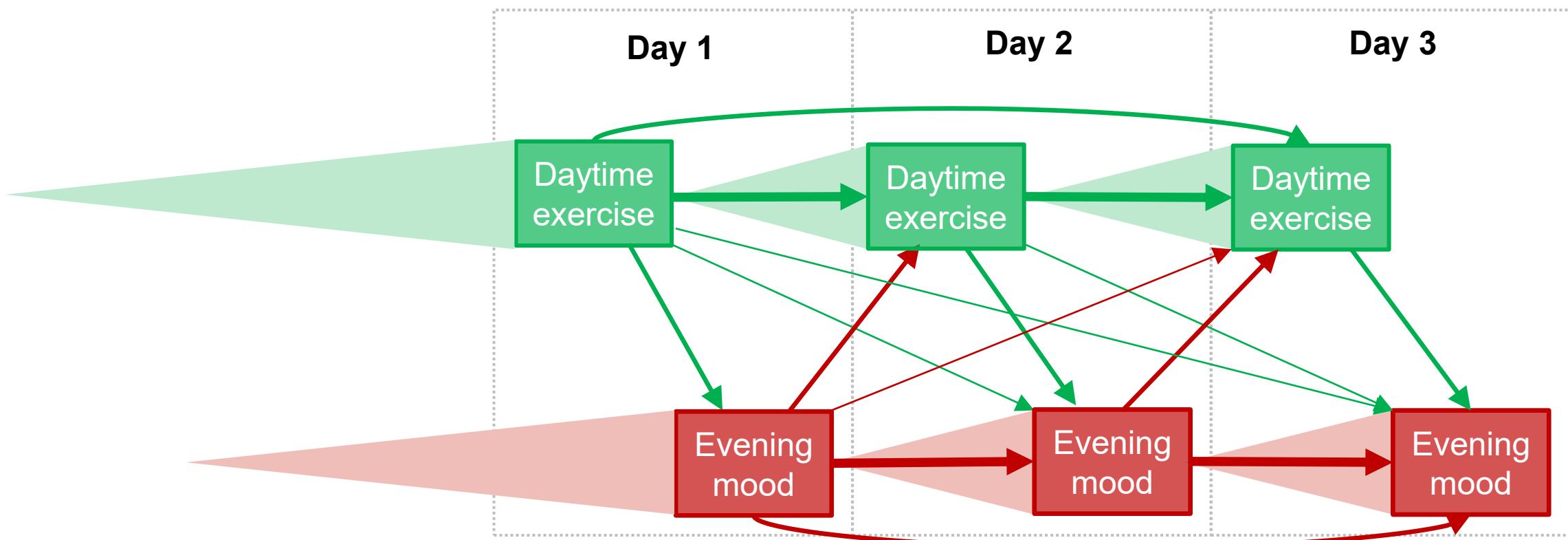
A single measure of weight captures a long autocorrelated process, a single measure of sleep perhaps less so ...

Perhaps **weight** \Rightarrow **sleep** is the **dominant direction of causality**

6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

Where there is strong **feedback** over your time window, you cannot identify a single causal effect and you need more measurements

This is how we turn a '**bi-directional relationship**' into a *directed* graph



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

C) What about interrelated variables that **crystallised** over a similar time?

Without repeated measures, it can be difficult to say which (among several) variables occurred first

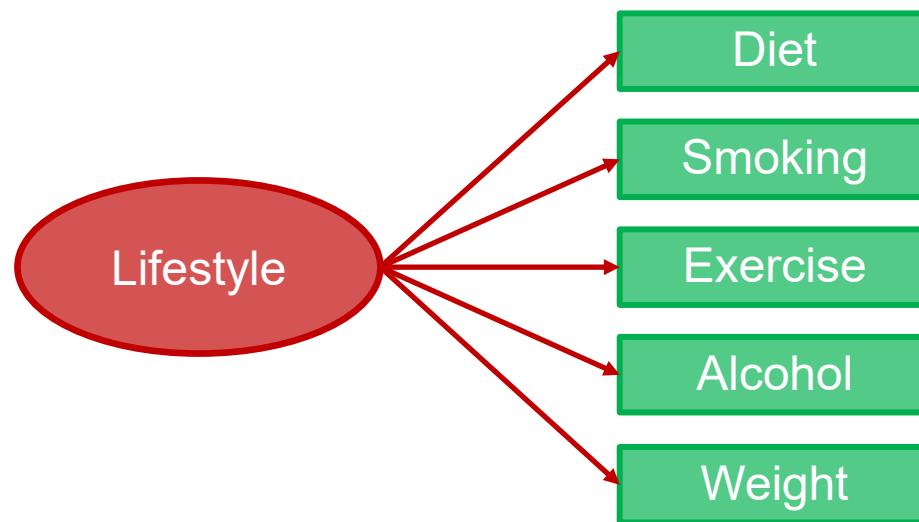


6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

C) What about interrelated variables that **crystallised** over a similar time?

When it is impossible to determine the **dominant direction of causality** we say the measurements '**crystallised together**'

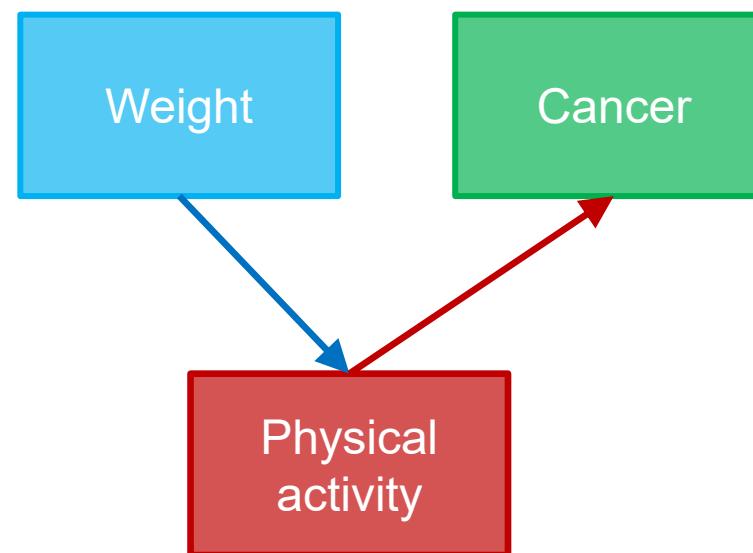
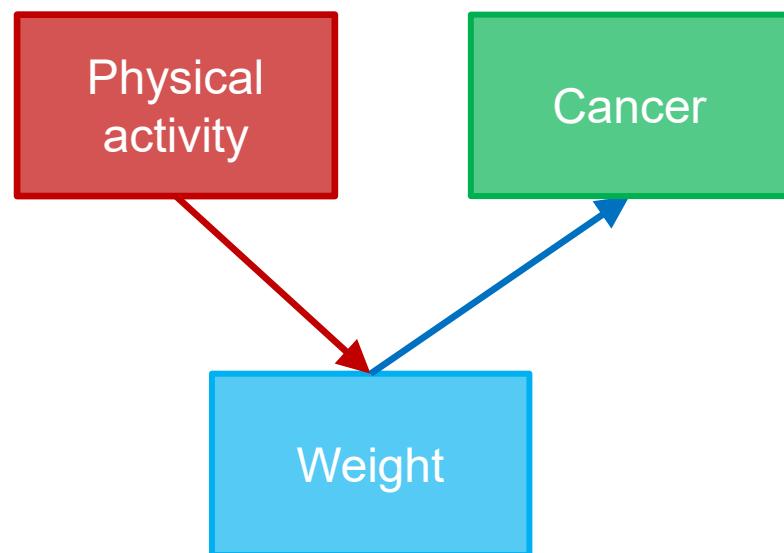
In many situations, this is because they have common origin(s), which we could indicate with a shared preceding **latent variable**



6) ARRANGE THE VARIABLES IN TEMPORAL ORDER

C) What about interrelated variables that **crystallised** over a similar time?

If you can't decide on the **dominant direction of causality**, and it matters to your focal relationship, consider drawing different DAGs to describe the different options, and analyse multiple ways ...



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

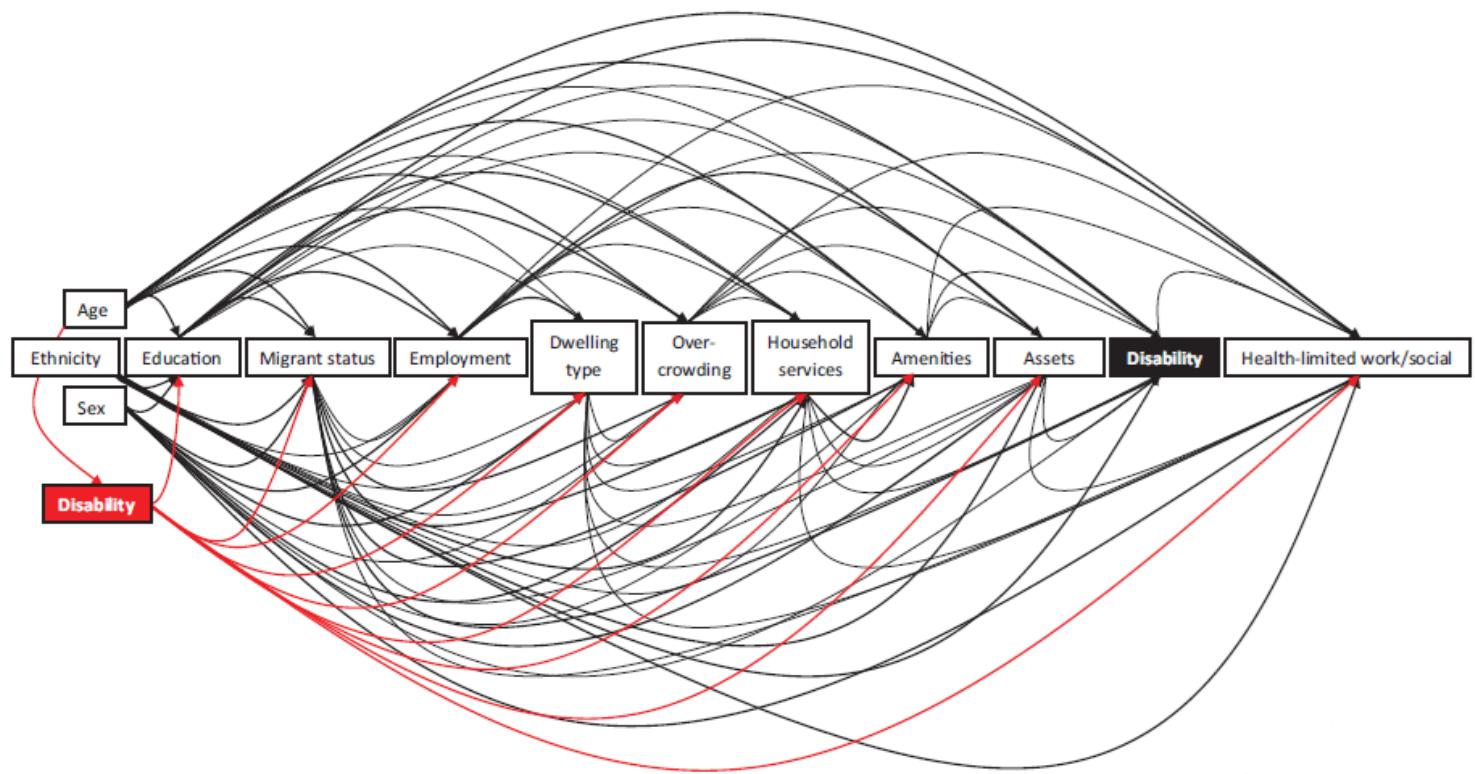
1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. **Draw forwards arcs, unless confident otherwise**
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



7) DRAW FORWARDS ARCS, UNLESS CONFIDENT OTHERWISE

Once you have arranged your variables in temporal order, add the arcs

It is useful to start by assuming that all variables are *potential* causes of all future variables



Saturated DAG

7) DRAW FORWARDS ARCS, UNLESS CONFIDENT OTHERWISE

Why?

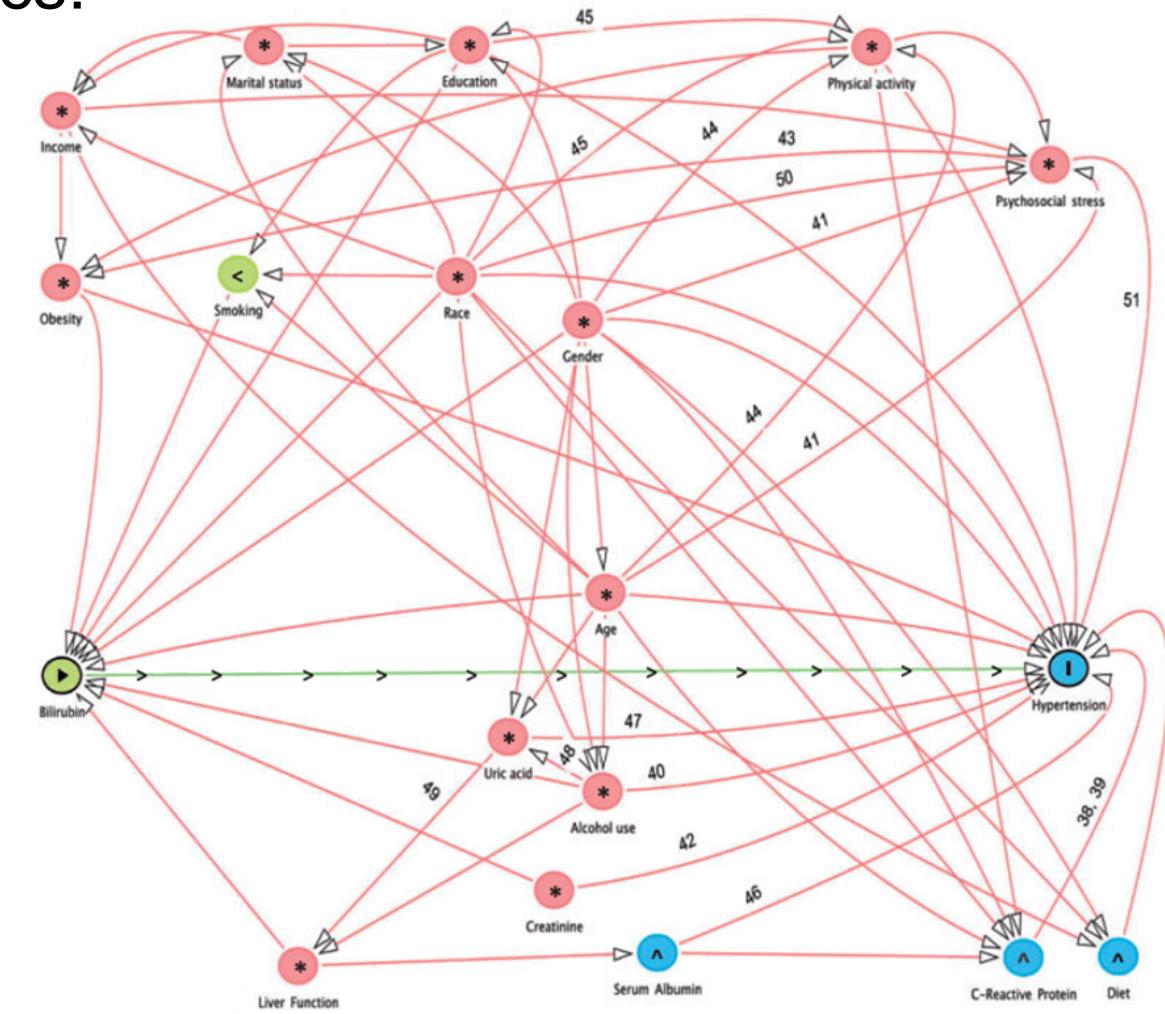
- Our reluctance to declare causality means analysts tend to leave out arcs unless they are pretty confident of causality
- But **removing an arc is a much stronger assumption** than including one
- Arcs allow for a causal relationship ***of any size***

No arc declares there must be NO causal relationship !

- If **X** occurs before **Y**, we say it has '**temporal precedence**' – by default is therefore a ***potential cause of Y***
- You should be confident that **X CANNOT** cause **Y** to remove its arc

6) DRAW ALL FORWARD ARCS

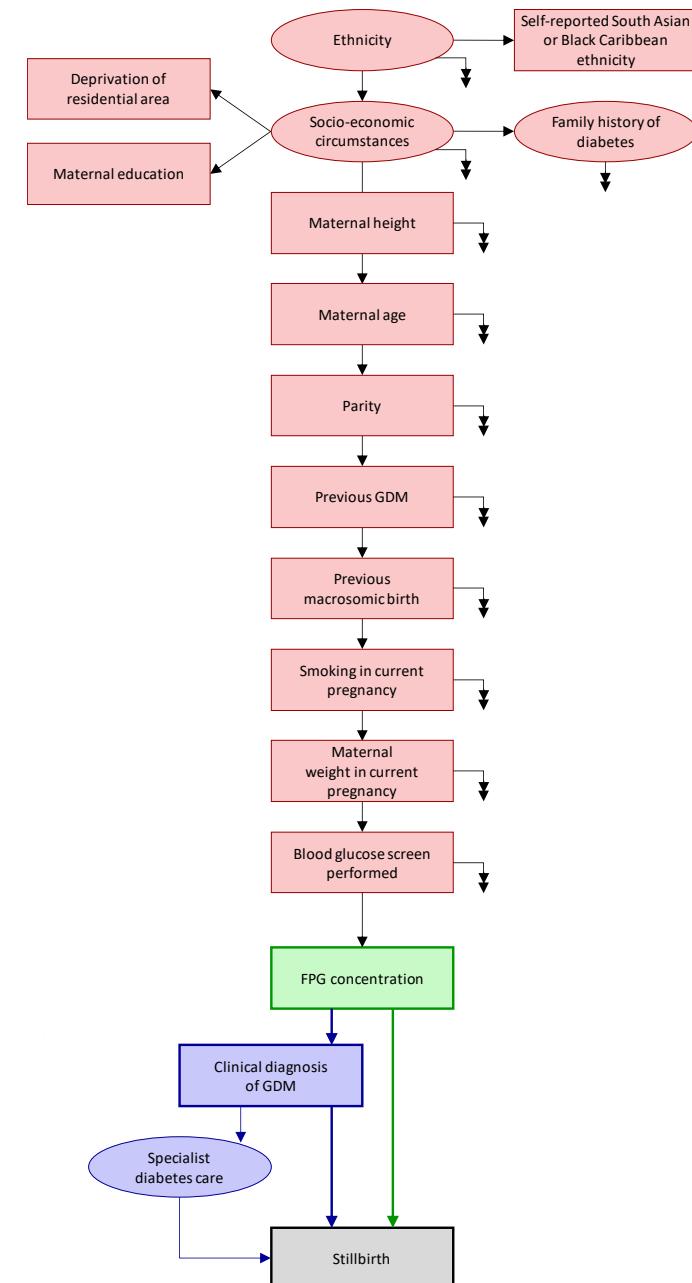
DON'T waste time finding evidence for arcs!



6) DRAW ALL FORWARD ARCS

Drawing all forward arcs can be messy!

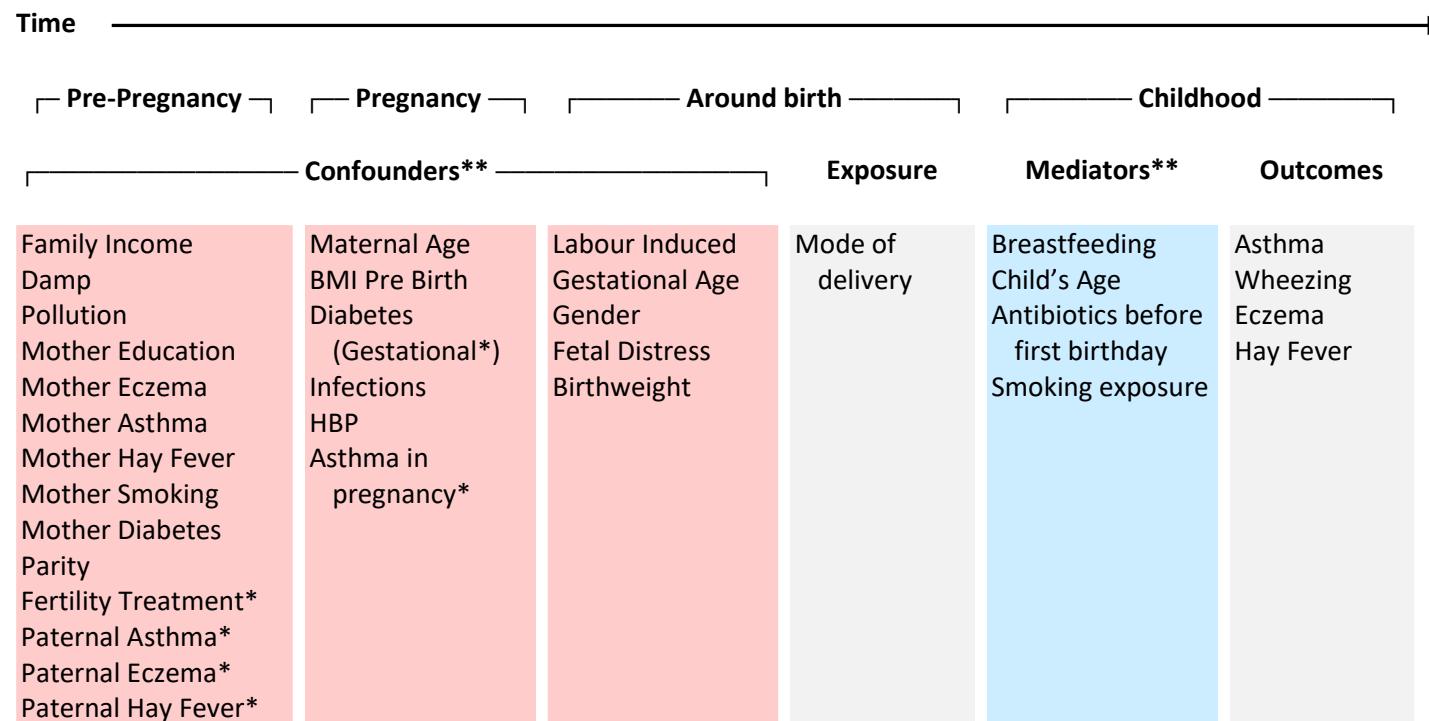
Alternative approaches include drawing
'double arrows' to denote e.g. 'and all future
variables'

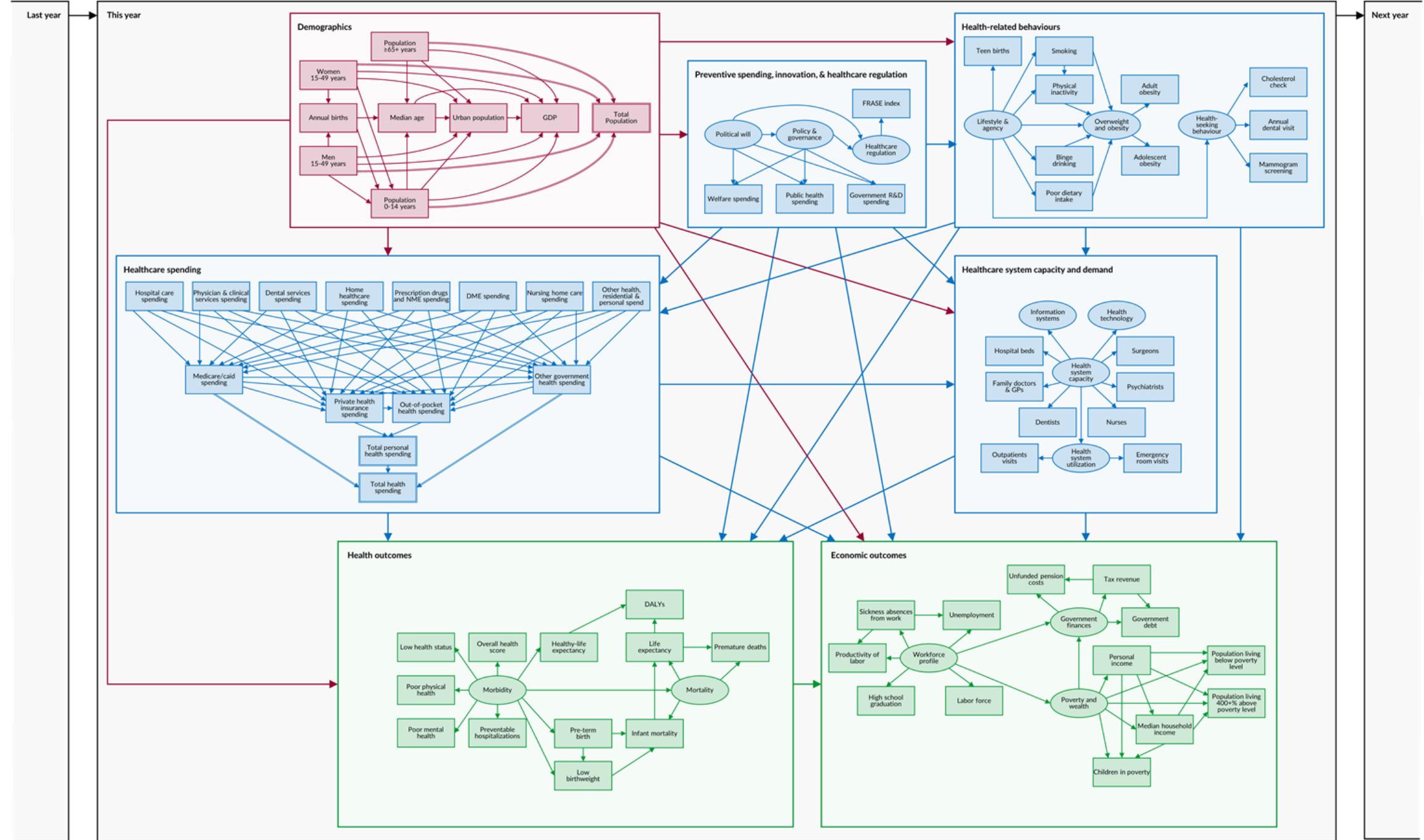


6) DRAW ALL FORWARD ARCS

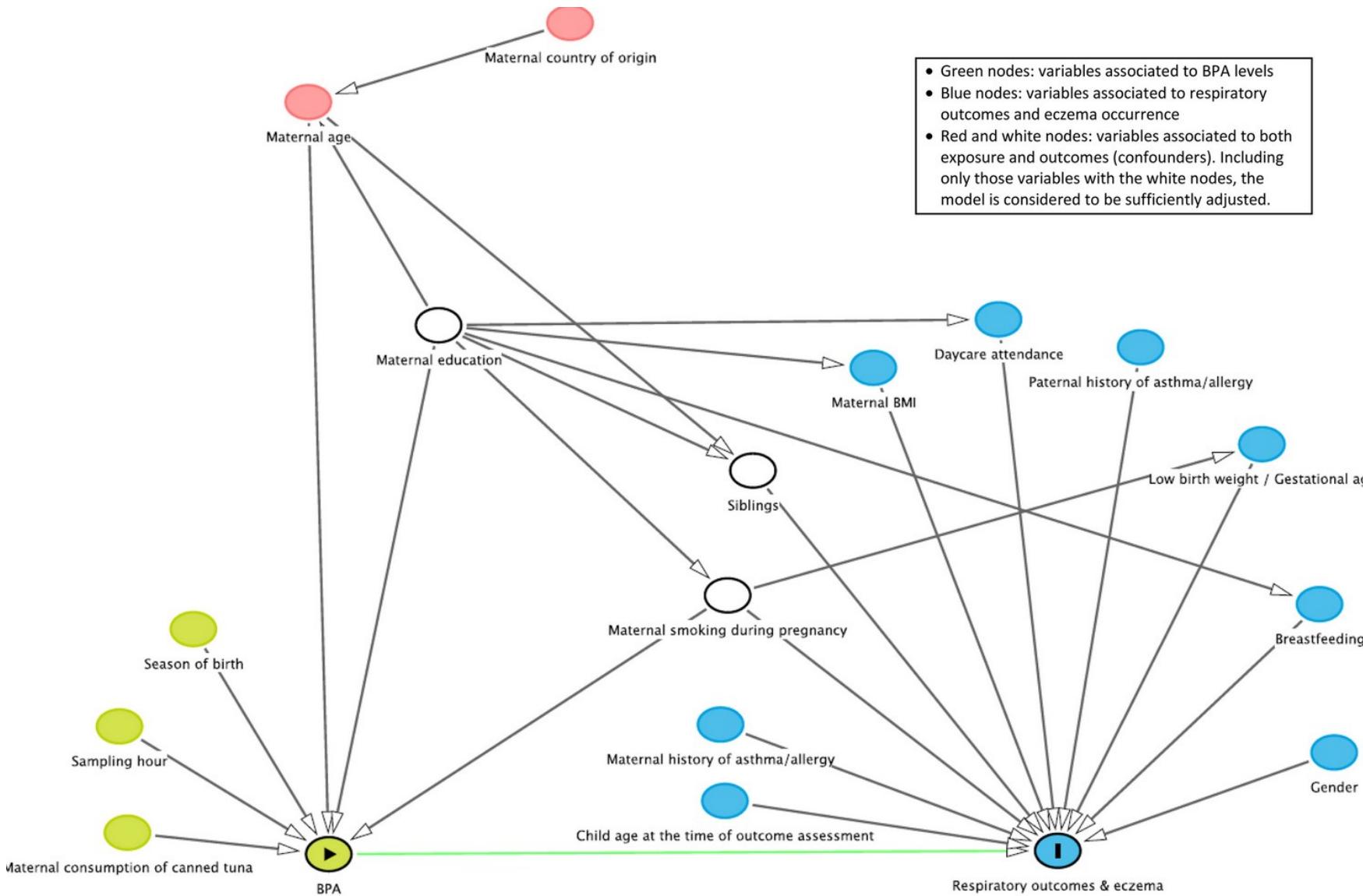
If you are only interested in one exposure-outcome relationship, then the temporal order of your confounders may not be so important

Here, a '**list-based DAG**' may be sufficient, but these are rare and cannot be used to identify other problems (described later)





DON'T BE LIKE THIS!



- Green nodes: variables associated to BPA levels
- Blue nodes: variables associated to respiratory outcomes and eczema occurrence
- Red and white nodes: variables associated to both exposure and outcomes (confounders). Including only those variables with the white nodes, the model is considered to be sufficiently adjusted.



THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forwards arcs, unless confident otherwise
8. **Check & update your DAG(s) against your data**
9. Use your DAG(s) to inform and interpret your model
10. Share & publish your DAG(s)



8) CHECK & UPDATE YOUR DAG(S) AGAINST YOUR DATA



Peter Tennant
@PWGTennant

We welcome comments & feedback on our new pre-print on "Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations"

[medrxiv.org/content/10.110...](https://www.medrxiv.org/content/10.1101/2020/01/07/2020.101101.full.pdf)

#DAGreview #EpiTwitter #StatsTwitter #CausalInference
#DataScience

cc @yudapearl @mendel_random

...



Johannes Textor @JohannesTextor · Jan 7, 2020

Replying to @PWGTennant and @yudapearl

Yes, I am still very surprised that Epidemiologists feel it's OK to use a model that has never been tested. Or it could be a kind of survivor bias since complex DAGs rarely survive testing in my experience.

1

1

2

↑



Johannes Textor @JohannesTextor · Jan 7, 2020

In other fields (physics, biology) models aren't generally trusted unless it's shown that they can at least explain some of the available data. Some of the best models have been tested, rejected, tweaked, tested, rejected, tweaked many times.

1

1

3

↑



Johannes Textor @JohannesTextor · Jan 7, 2020

Most (actually, I believe all) of the DAGs in the review are "one-off" models, built from scratch for one specific dataset. I hope to see more iterative and collaborative model building in the future that also incorporates rigorous model testing.

...

...

...

8) CHECK & UPDATE YOUR DAG(S) AGAINST YOUR DATA



International Journal of Epidemiology, 2020, 1–13
doi: 10.1093/ije/dyaa213
Original Article



Original Article

Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations

Peter WG Tennant ,^{1,2,3*} Eleanor J Murray,⁴ Kellyn F Arnold ,^{1,2} Laurie Berrie,^{1,5,6} Matthew P Fox,^{4,7} Sarah C Gadd,^{1,5} Wendy J Harrison ,^{1,2} Claire Keeble,¹ Lynsie R Ranker,⁴ Johannes Textor ,⁸ Georgia D Tomova,^{1,2,3} Mark S Gilthorpe^{1,2,3†} and George TH Ellison^{1,2,9†}

¹Leeds Institute for Data Analytics, University of Leeds, Leeds, UK, ²Faculty of Medicine and Health, University of Leeds, Leeds, UK, ³Alan Turing Institute, British Library, London, UK, ⁴Department of Epidemiology, School of Public Health, Boston University, Boston, MA, USA, ⁵School of Geography, University of Leeds, Leeds, UK, ⁶School of GeoSciences, University of Edinburgh, Edinburgh, UK, ⁷Department of Global Health, Boston University, Boston, MA, USA, ⁸Department of Tumour Immunology, Radboud University Medical Center, Nijmegen, The Netherlands and ⁹Centre for Data Innovation, Faculty of Science and Technology, University of Central Lancashire, Preston, UK

†Joint senior authors.

*Corresponding author. Peter WG Tennant, Leeds Institute for Data Analytics, University of Leeds, Level 11 Worsley Building, Clarendon Way, Leeds LS2 9NL, UK. E-mail:p.w.g.tennant@leeds.ac.uk

Editorial decision 14 September 2020; Accepted 12 October 2020

Section	Recommendation number	Recommendation details	Page
Introduction			
	1	The focal relationship(s) and estimand(s) of interest are stated in the study aims	
Methods			
	2	DAGs for all focal relationships and estimands of interest are provided	
	3	DAGs include all relevant variables, including those where direct measurements are unavailable	
	4	DAGs are arranged so that all constituent arcs flow in the same direction	
	5	Missing arcs have been carefully considered. Optionally, these are justified with theory and/or evidence.	
	6	DAG-implied adjustment set(s) for all estimand(s) of interest are clearly stated, including any unobserved confounders	
	8a	Alternative adjustment set(s) are clearly described and justified	
	8b	Optionally, the consistency of all DAGs with the observed data has been explored. Subsequently modified DAGs are reported separately	
Results			
	7a	Estimate(s) from unmodified DAG-implied adjustment set(s) - or the nearest approximation thereof - are provided	
	7b	Optionally, the impact of unobserved confounders has been estimated, and bias-adjusted estimates are reported	
	8c	Estimates from alternative adjustment set(s) are reported separately to those obtained from DAG-implied adjustment sets	

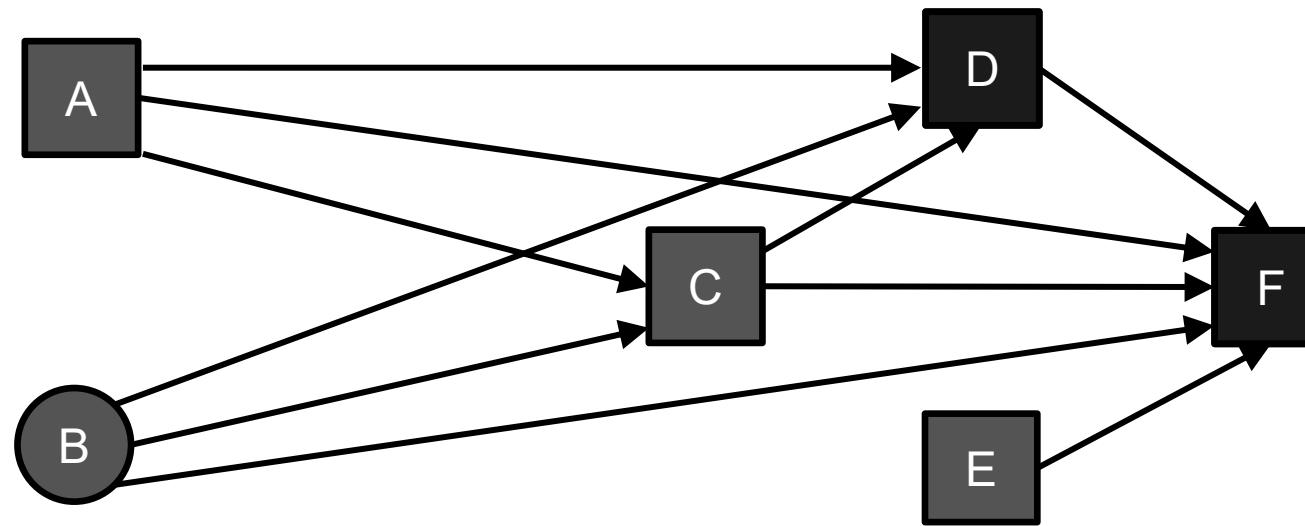
THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forwards arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. **Use your DAG(s) to inform and interpret your model**
10. Share & publish your DAG(s)



9) USE YOUR DAG(S) TO INFORM YOUR ANALYSES

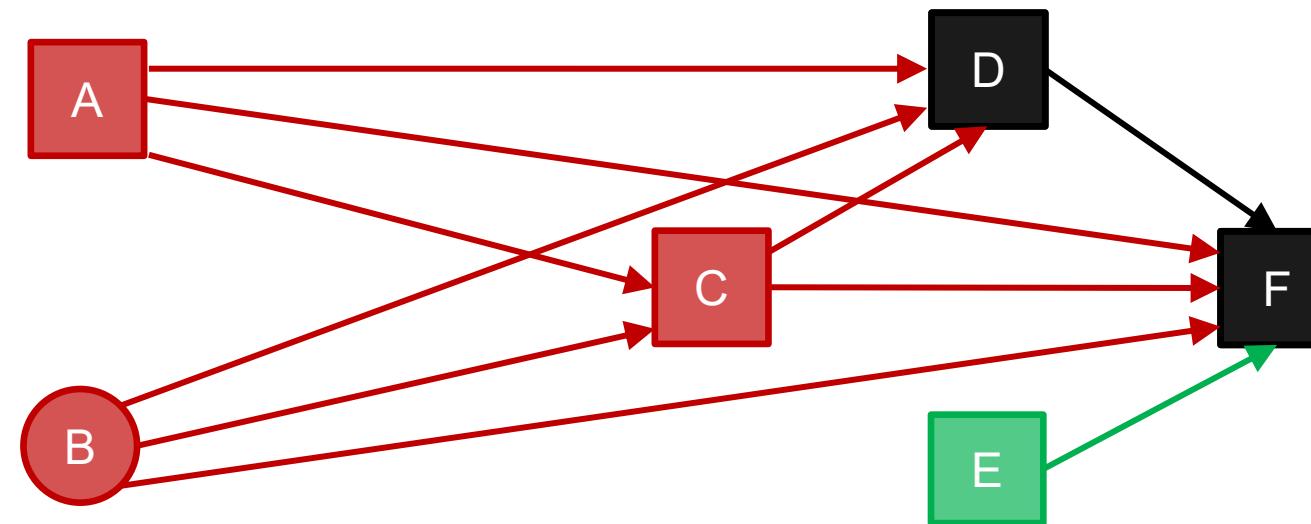
Focal relationship $D \rightarrow F$



9) USE YOUR DAG(S) TO INFORM YOUR ANALYSES

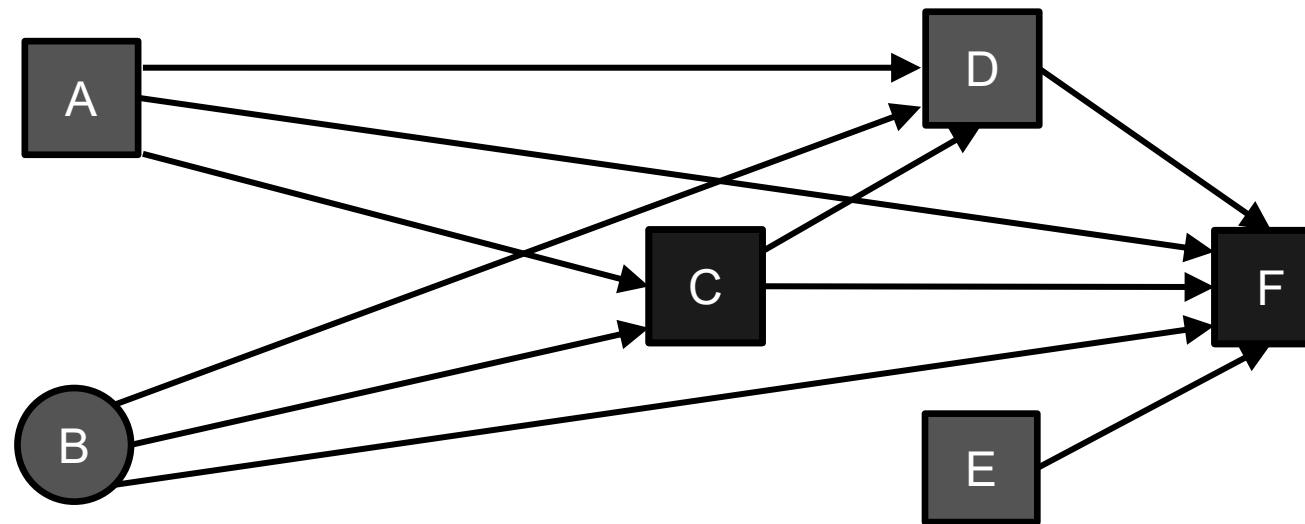
Focal relationship $D \rightarrow F$

- Condition on **A**, **B**, **C**, and maybe **E**



9) USE YOUR DAG(S) TO INFORM YOUR ANALYSES

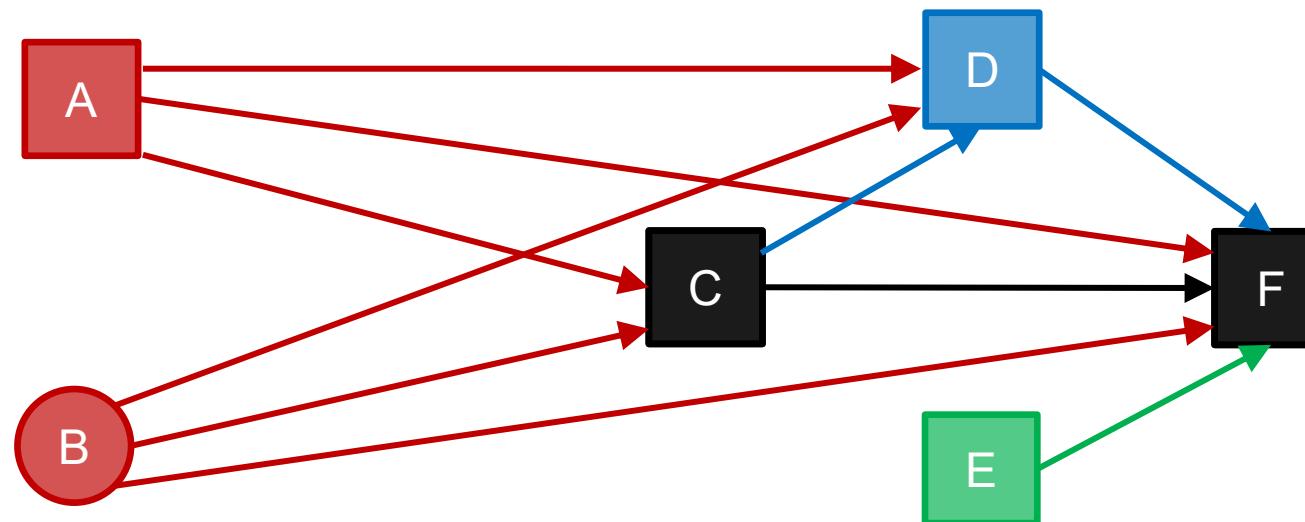
Focal relationship $C \rightarrow F$



9) USE YOUR DAG(S) TO INFORM YOUR ANALYSES

Focal relationship $C \rightarrow F$

- Condition on **A**, **B**, and maybe **E** but **NOT D**



9) USE YOUR DAG(S) TO INFORM YOUR ANALYSES

Clearly state the **adjustment set** for each exposure-outcome relationship:

DO SAY

“Our DAG implied that the following confounding variables required adjustment: maternal ethnicity, socio-economic circumstances, family history of GDM, height, weight, age, parity, previous histories of GDM and macrosomia, and smoking. Family history of GDM was however not known and is therefore a potential source of unobserved confounding.”

DON’T SAY

“Confounders for adjustment were identified by drawing a DAG”

9) USE YOUR DAG(S) TO INFORM YOUR ANALYSES

- If you have a large number of confounders, some variable reduction strategy may be necessary
- Various methods can be used, including:
 - **Regression-based** methods, such as **best-subsets regression**
 - **Other machine learning algorithms**, such as **gradient boosting**
 - **Ensemble methods**, such as **SuperLearner**
- We do not recommend the stepwise or ‘change-in-estimate’ procedures
- Variable reduction is best implemented using the two-step **propensity score approach** mentioned later
- Where possible, we still recommend reporting the results of the full adjustment set (that includes all confounders)

THE 10 STEPS TO DRAW YOUR DAG LIKE A PRO!

1. Develop and state a clear research question
2. Consider and state your context
3. Draw your DAG(s) as early as possible
4. Get help - don't draw it alone
5. Include all relevant variables
6. Draw your DAG(s) in temporal order
7. Draw forwards arcs, unless confident otherwise
8. Check & update your DAG(s) against your data
9. Use your DAG(s) to inform and interpret your model
- 10. Share & publish your DAG(s)**



8) SHARE AND PUBLISH YOUR DAG(S)

With pride!

International Journal of Epidemiology, 2015, Vol. 44, No. 1

221

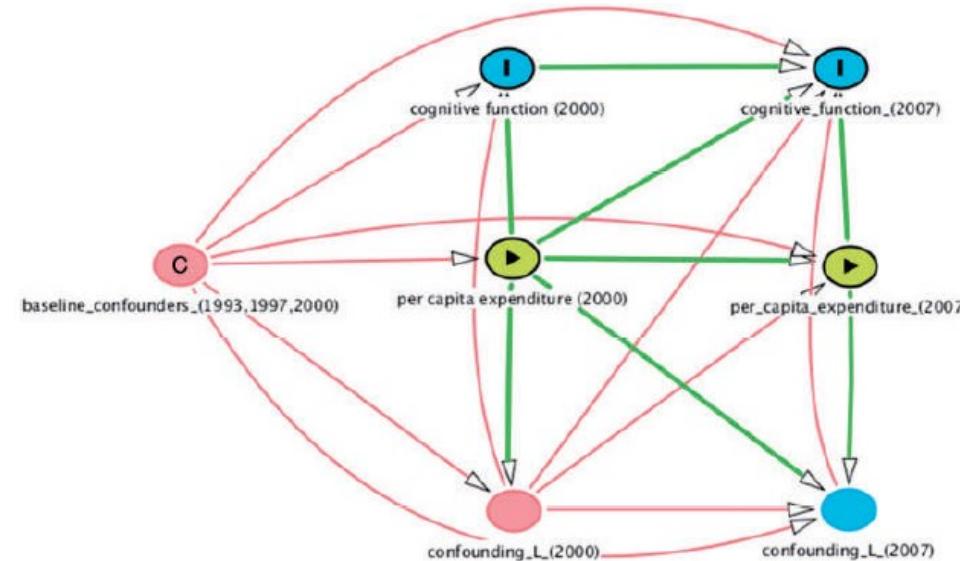


Figure 2. Direct acyclic graph (DAG) representing the relations between confounders, household per capita expenditure and child cognitive function. Exposure: household per capita expenditure. Outcome: cognitive function z-score. Ancestor of exposure and outcome (baseline confounders measured in 1993, 1997 and 2000): caregiver's age, education, employment status, household size, economic hardship, household had electricity, used piped or pumped well as the main drinking water source, owned toilet with septic tank, and residential area. Ancestor of exposure and outcome (confounding L measured in 2000): attending school and caregiver's mental health. Ancestor of outcome (confounding L measured in 2007): completed at least 8 years of education and caregiver's mental health. Causal path. Biasing path.

occasionally (3–4 days) and most of the time (5–7 days). For both measures, each item was scored ranging from 0 to 3 and summed as the total mental health score separately for 2000 (scores ranging 0–24) and 2007 (scores ranging 0–30). In the analysis, we used the total mental health score where a higher score indicated poorer mental health.

Missing data

Of the children in the IFLS who were administered the cognitive test, the response rate was 96% and 95% in 2000 and 2007, respectively. The proportion of children with missing information on the exposure was 0.7% in 2000 and 7% in 2007. Of the 6136 children, only 5305 (86%) were

Maika A, Mittinty MN, Brinkman S,
Lynch J. *International journal of
epidemiology*. 2015;44(1):218-28.

8) SHARE AND PUBLISH YOUR DAG(S)

With pride!

Open Access

6

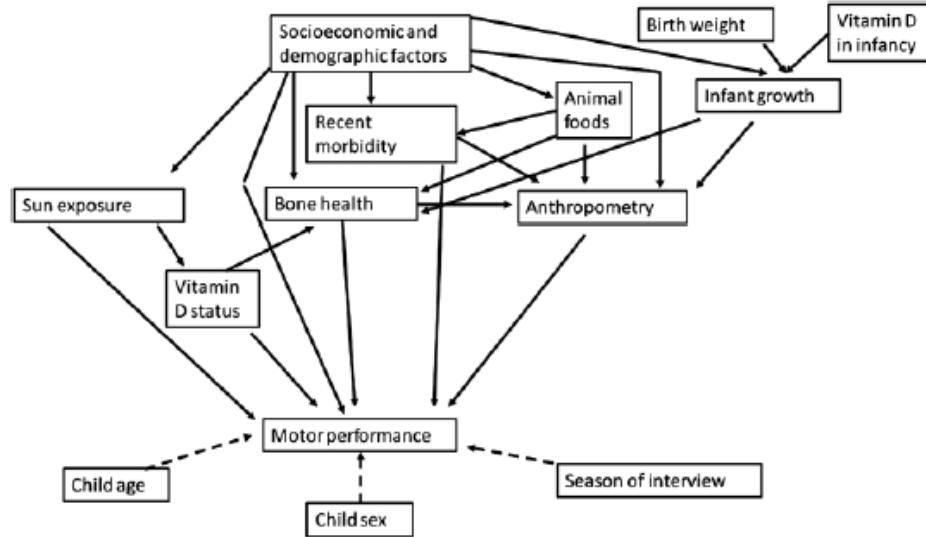


Figure 1 Conceptual framework for factors affecting motor development. Variables included under each of the group headings are: (1) motor performance: Ages and Stages Questionnaire pass/fail category, maximum grip strength, minimum run time, maximum number of squats in 15 s; (2) anthropometry: height and body mass index Z scores, arm muscle area; (3) vitamin D status: serum 25-hydroxyvitamin D; (4) bone health: radius and tibia quantitative ultrasound Z scores; (5) sun exposure: h/day; (6) diet: animal food groups; (7) infant growth: birth weight tertile, length-for-age Z score at 6 months, change in length-for-age Z score from birth to 6 months; (8) recent morbidity: reported symptoms in the past 3 days; (9) season of interview: 4-month divisions; (10) sociodemographic factors: quintiles from principle components analysis and (11) vitamin D in infancy: treatment group allocation in Delhi Infant Vitamin D Supplementation (DIVIDS)-1 trial.

other outcomes were: maximum grip strength, 2.49 kg (n=830, SD 0.93), minimum time to run 20 m, 7.1 (n=861, SD 1.8) s and maximum number of squats within 15 s, 12 (n=840, SD 3).

Table 2 shows crude associations of key potential mediators—25OHD, anthropometry and bone health—

In view of these results, and because tibia and radius QUS Z scores were correlated ($r=0.46$, $p<0.001$), further analyses used HAZ to represent current anthropometry, AMA to represent lean body mass, length-for-age tertile at 6 months to represent early growth and tibia Z score to represent bone health.

Filteau S et al. *BMJ open*.
2016;6(1):e009268.

8) SHARE AND PUBLISH YOUR DAG(S)

With pride!

204

N. Röhrig et al. / Journal of Clinical Epidemiology 67 (2014) 199–206

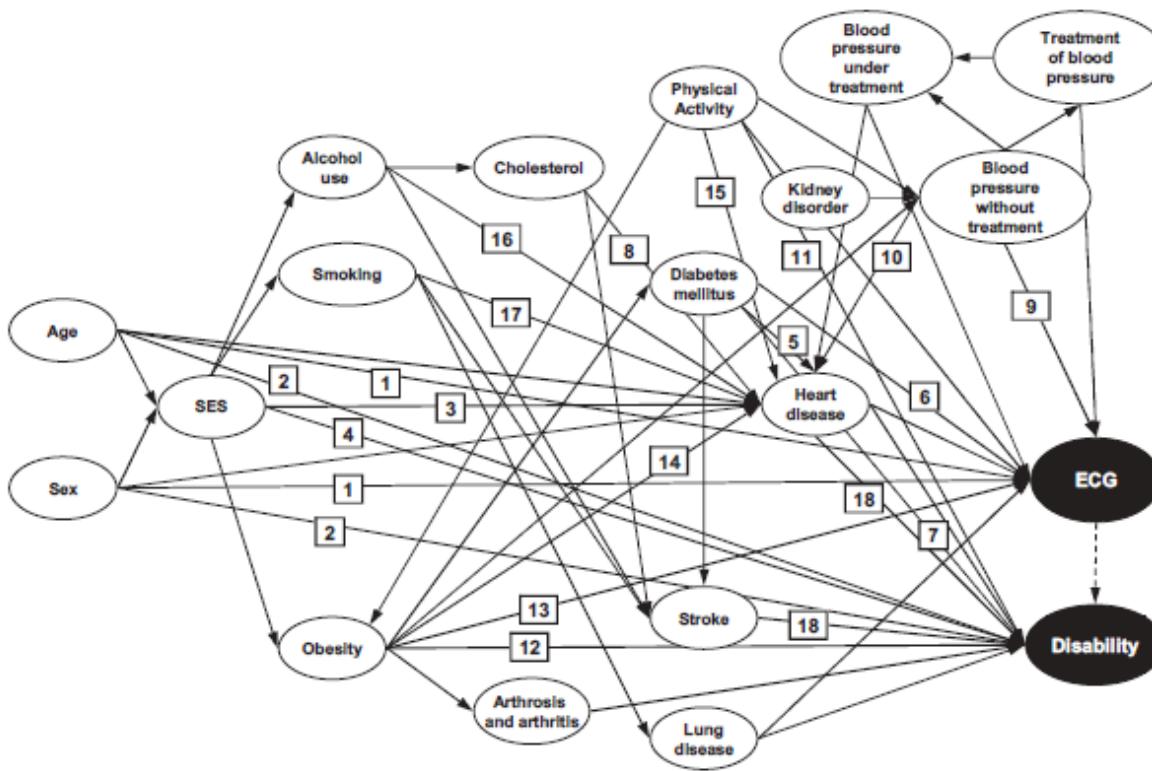


Fig. 1. DAG derived from literature and expert knowledge. Nodes represent variables and arrows represent causal associations. Dark-colored nodes label ECG findings and disability, representing exposure and outcome, respectively. Pale-colored nodes represent possible confounding factors. Numbers represent available information from the literature (see Table 1 at www.jclinepi.com for full references). SES, socioeconomic status; ECG, electrocardiography.

Röhrig N, et al. *Journal of clinical epidemiology*. 2014 Feb 1;67(2):199-206.

selection methods increase the risk of selecting the wrong covariates [35]. In our study, stepwise selection resulted in a data set reduced by 174 participants without

information on alcohol consumption. Using stepwise regression yielded a significant association of ECG with disability. Arguably, this result is biased because it cannot be

**IN CONCLUSION...
DAGS ROCK!**



RECOMMENDED READING

- Tennant, P.W.G, Murray, E.J., Arnold, K.F., Berrie, L., Fox, M.P., Gadd, S.C., Harrison, W.J., Keeble, C., Ranker, L.R., Textor, J. and Tomova, G.D., 2021. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International journal of epidemiology*, 50(2), pp.620-632.
- Digitale, J.C., Martin, J.N. and Glymour, M.M., 2022. Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology*, 142, pp.264-267.
- Laubach, Z.M., Murray, E.J., Hoke, K.L., Safran, R.J. and Perng, W., 2021. A biologist's guide to model selection and causal inference. *Proceedings of the Royal Society B*, 288(1943), p.20202815.
- Rohrer, J.M., 2018. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in methods and practices in psychological science*, 1(1), pp.27-42.

SUMMARY

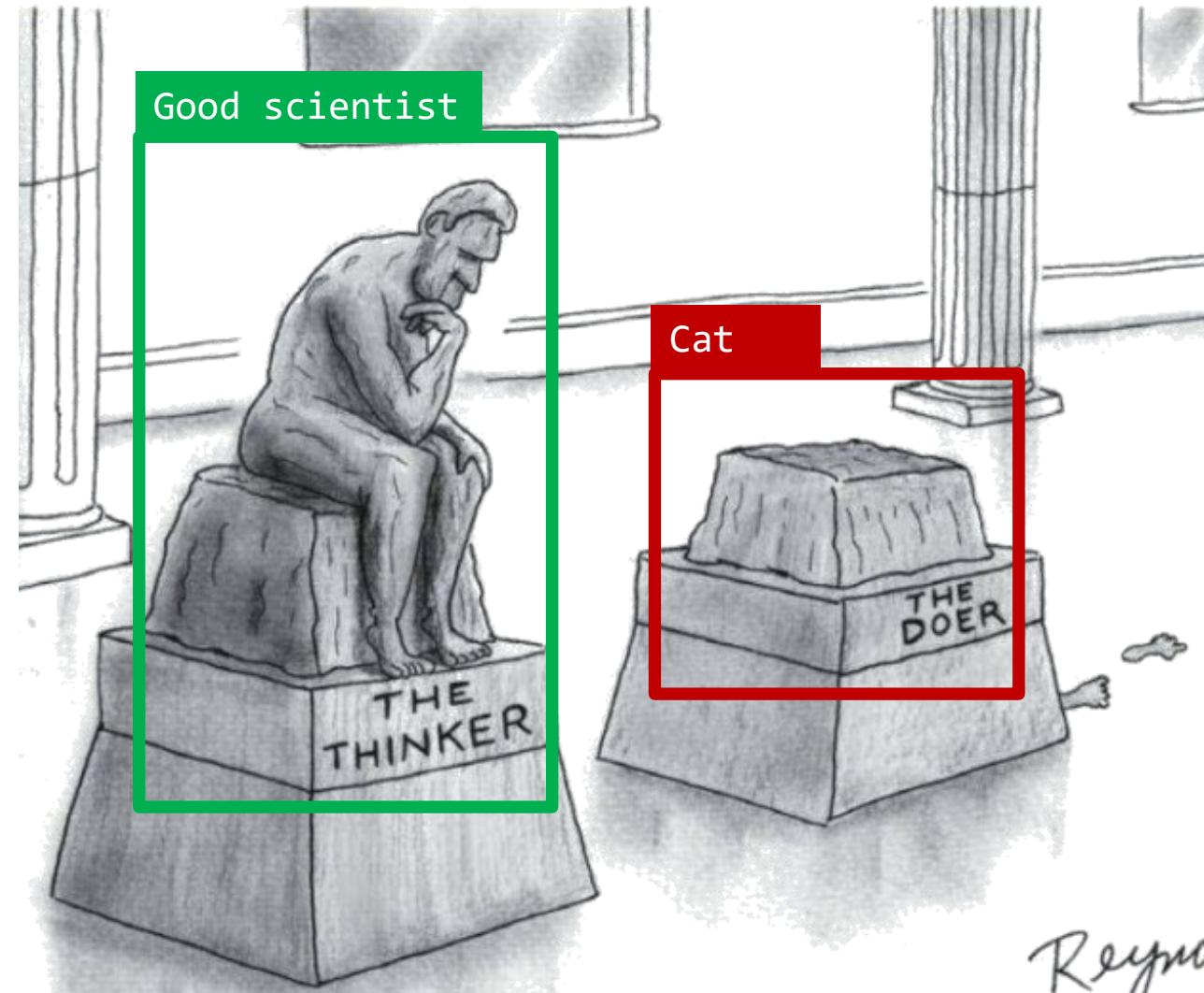
DAGs are easy to learn but difficult to master

The theoretical rules can seem particularly difficult to apply to complex 'real world' scenarios

Our 10-step approach should help you to draw the best DAG for almost any scenario

BUT: DAGs are only an aid !

They cannot replace the need for thinking and scholarship !



2.3 - WRIGHT'S PATH RULES AND PARAMETRIC CONSIDERATIONS

GEORGIA



@GEORGIATOMOVA

MARK



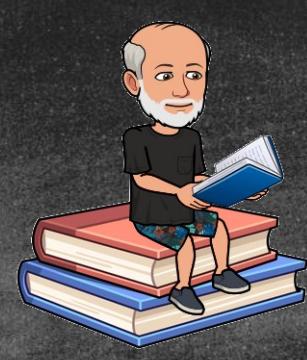
@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 2

09:30-10:15 LECTURE 2.1

10:15-11:00 ACTIVITY 2-A

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 2.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:00 ACTIVITY 2-B

15:00-15:30 LECTURE 2.3

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 2.4

17:00-17:45 ACTIVITY 2-C

17:45-18:00 Q&A

LEARNING OBJECTIVES

Learn about **Wright's path rules**

- ✓ understand how these underpin **Structural Equation Models** (SEMs), **Structural Causal Models** (SCMs), & parametrised **Directed Acyclic Graphs** (DAGs)

Appreciate how to evaluate **DAG-data consistency** while avoiding data-driven model building

Understand how simulate data that respects the underlying **data generating mechanism** (DMG) using DAGs

Understand key differences between **DAG-informed simulation** of the DGM and **covariance-matrix simulation** of the consequences of the DMG

INFORMING SIMULATIONS FROM A CAUSAL PERSPECTIVE

Simulation studies are useful to:

- evaluate performance, adequacy, & properties of **statistical methods**
- explore & explain potential **biases** and **analytical artifacts**

Validity of a simulation depends on **data generation mechanism** (DGM)

- NOT just their consequences \Rightarrow overlooks **temporal precedence**

Simulation informed by a DAG

- fundamentally different from simulation using a covariance/correlation
- explicitly uses temporal (i.e. longitudinal) processes

Distinction: '**solution space**' of all possible simulated datasets may differ
 \Leftrightarrow there is no 1-to-1 map of DAG to covariance/correlation & vice versa

WRIGHT'S 'PATH TRACING' RULES - 1921

Pioneering precursors to SEMs, SCMs, & DAGs

Note: deriving causal structure from data is (generally) not feasible

↔ involves *a priori* knowledge or external theory

- Wright assumed *a priori* **causal structure** plus **multivariate normality** & **no interactions** (extension exist for SEMs)

Simplifying complex real-world scenarios can provide powerful insights

- 'toy' simulations extremely valuable & **aids understanding** of context

Can also improve how we approach drawing DAGs

↔ **for your DAG to be meaningful, you must be able to simulate data from it**

WRIGHT'S 'PATH TRACING' RULES - 1921

Wright proposed a set of rules that integrate

- a **linear, multivariate normal** path diagram
- standardised **path coefficients** (standardised partial regression coefficients in linear model)
- the **Pearson correlation** between any two variables
- numerical contribution of an indirect path is the product of path coefficients for each arc along the path (this does not hold with interactions or nonlinearity)

Path tracing rules:

- loops not allowed (cannot pass through the same variable twice when following a particular route)
- no going forward then backward (once backward then forward is okay)
- only one **bidirectional** arrow is allowed (allows for correlation)

BIDIRECTIONAL ARCS IN PATH DIAGRAMS

Path diagrams may sometimes have **bidirectional** arcs

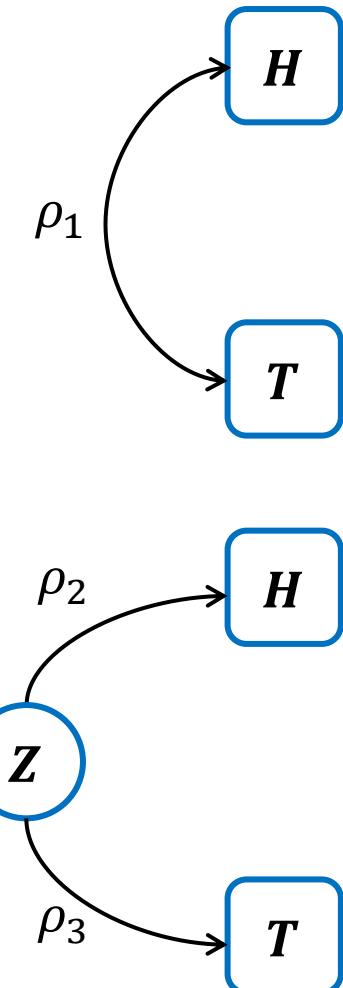
- ↔ correlations between two variables
- standardised path coefficient is the correlation $\rho_1 \leq 1$

DAGs do not permit bidirectional arcs, but easy to fix

- simply introduce a **latent variable** as a common ancestor
- two new path coefficients: ρ_2 & ρ_3 where $\rho_1 = \rho_2 \times \rho_3$
- reasonable to set: $\rho_2 = \rho_3 = \sqrt{\rho_1}$

Thus, all path diagrams can be made into a DAG

- **with very simplified parametric assumptions**



WRIGHT'S EXAMPLE

Wright investigated causal factors that determine **wet bulb depression B**

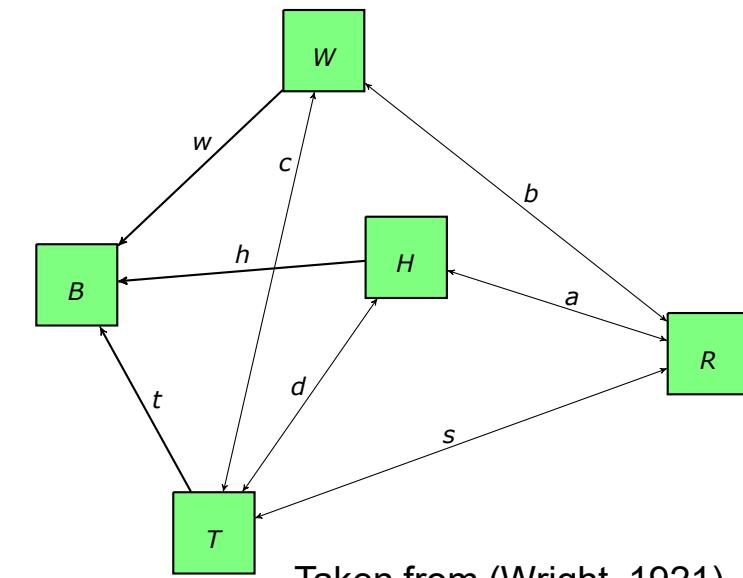
- difference between dry-bulb temperature & wet-bulb temperature of a thermometer
- factors: temperature T , absolute humidity H , wind velocity W , and radiation R

He combined his path tracing rules and the available **bivariate correlations**:

- ρ_{BW} , ρ_{BR} , ρ_{BT} , ρ_{WR} , ρ_{WT} , and ρ_{RT}

From these he inferred **direct**, **indirect** & **total** causal effects

But the key revelation is the relationship between causal **path coefficients** & **bivariate correlations**



Taken from (Wright, 1921)

WRIGHT'S EXAMPLE

correlation ≠ causation

- ✓ well-known but now quantified explicitly

causation ≠ correlation

- ✓ less well known & widely overlooked
- ⇒ non-zero causal link between two variables can have zero (or near zero) correlation due to causal structure

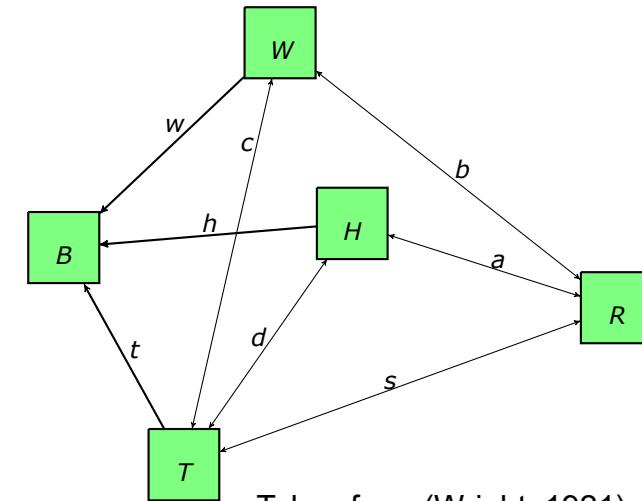
Wright's assumptions (external theory) clear:

- ⇒ W & H have no causal link (strong assumption)
- ⇒ W & H have no common (unobserved) ancestors
since W is wind velocity, H is humidity ⇒ is this tenable?

True underlying DGM not known but ...

... we can now evaluate DAG-data consistency

Causal path coefficient	Bivariate (Pearson) correlation
$W \rightarrow B = w$	$w + tc$
$H \rightarrow B = h$	$h + dt$
$T \rightarrow B = t$	$t + dc + wc$
$R \rightarrow B = ah$	$ts + bw + ah$
$H \rightarrow W = 0$	0
$H \rightarrow T = \text{unspecified}$	d
$W \rightarrow T = \text{unspecified}$	c
$R \rightarrow T = \text{unspecified}$	s
$R \rightarrow W = \text{unspecified}$	b
$R \rightarrow H = \text{unspecified}$	a



Taken from (Wright, 1921)

EVALUATING DAG-DATA CONSISTENCY

The **dagitty** the *R* package does many things

- simulates data from prespecified DAG
- evaluates DAG-data consistency ... & more

Assesses every **implied constraint** in DAG

- absence of arcs (strong assumption)
- ‘local’ tests – each need not involve all variables in the DAG
- **underutilised** ⇒ much needed uptake of this

But ... if dataset not consistent with DAG

- something is clearly wrong ⇒ first, take a step back & reflect

DAG may be incomplete \Leftrightarrow possibly missing key variable(s)

- if unclear what might be missing, probably related to **data provenance**
- selection bias (aka **collider bias**)



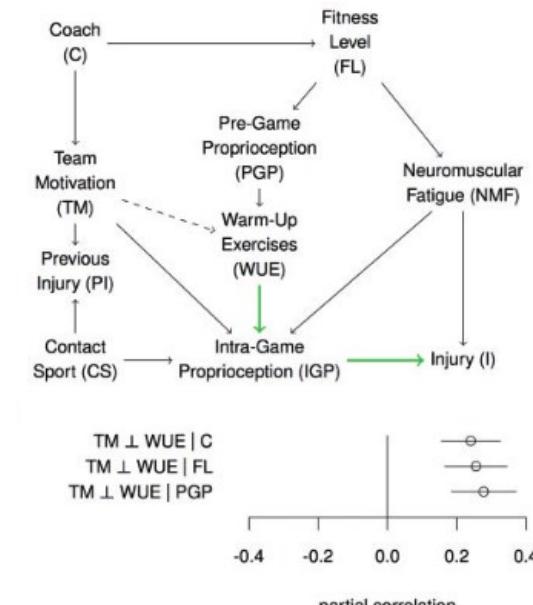
International Journal of Epidemiology, 2016, 1887–1894
doi: 10.1093/ije/dyw341
Advance Access Publication Date: 14 January 2017
Original article

Software Application Profile

Robust causal inference using directed acyclic graphs: the R package ‘dagitty’

Johannes Textor,^{1*} Benito van der Zander,² Mark S Gilthorpe,^{3,4}
Maciej Liśkiewicz² and George TH Ellison^{3,4}

¹Department of Tumour Immunology, Radboud University Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands, ²Institute for Theoretical Computer Science, University of Luebeck, Luebeck, Germany, ³Leeds Institute of Cardiovascular and Metabolic Medicine and ⁴Leeds Institute for Data Analytics, University of Leeds, Leeds, UK



DATA SIMULATION USING A DAG

Simulation studies are best kept relatively simple (while also realistic)

When you have a DAG consistent with your data \Rightarrow **simplify**

- one variable each for:

- ✓ outcome
- ✓ exposure
- ✓ confounder
- ✓ mediator
- ✓ competing exposure
- ✓ selection variable

Okay to omit some complexities (context-specific)

- e.g., ‘no confounding’ or ‘no competing exposure’, or ‘no selection variable’, etc.
- DAG must capture (simplified) underlying **DGM**

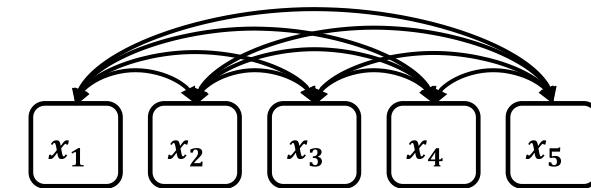
DATA SIMULATION USING A DAG

Stages of simulating a ‘toy’ dataset:

1. **Theorise**: draw your DAG – simplify as much as is reasonable
2. Adopt naïve assumptions of **linearity** & **multivariate normality** (others are possible with specific *R* routine) using variable transformations if necessary
3. Select initial (plausible) path coefficients to achieve a **positive definite covariance matrix** (*dagitty R* is excellent for multivariate normal data)
4. Simulate **standardised data** & transform to appropriate scales
5. **Evaluate statistics**: prevalences, correlations vs. observed data, regression coefficients vs. results in the literature
6. For discrepancies from real data, **tweak path coefficients** & repeat 3 to 5

IMPLICATIONS OF DAG-INFORMED SIMULATIONS

You might ‘think’ DAG, but statistical software does not!



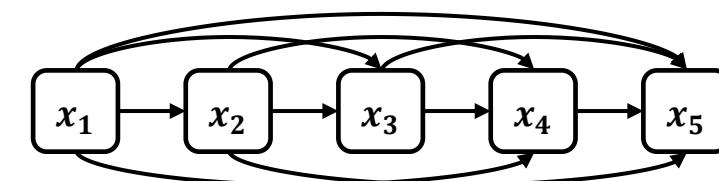
This has implications for **bivariate correlations** in the data

Simulated data from covariance/correlation matrix have few constraints
whereas introducing temporality generates many constraints

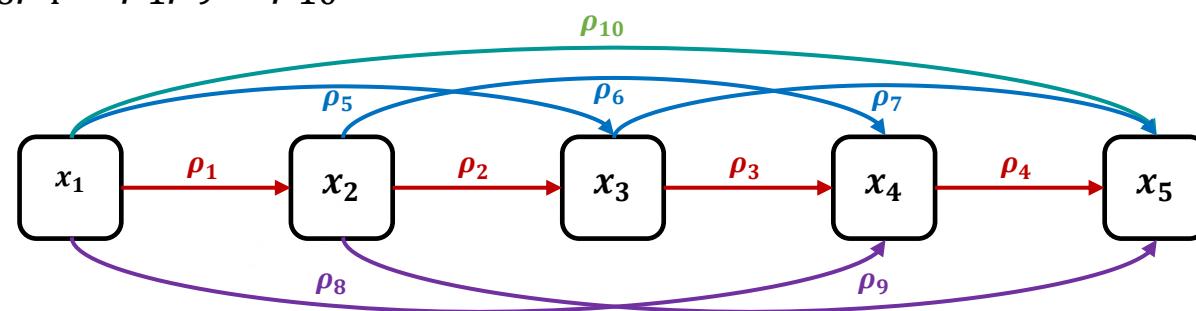
A parameterised DAG reveals the extent of this:

the bivariate correlation between x_1 and x_5 is

$$\rho_{x_1 x_5} = \rho_1 \rho_2 \rho_3 \rho_4 + \rho_5 \rho_3 \rho_4 + \rho_5 \rho_7 + \rho_1 \rho_2 \rho_7 + \rho_1 \rho_6 \rho_4 + \rho_8 \rho_4 + \rho_1 \rho_9 + \rho_{10} \leq \pm 1$$



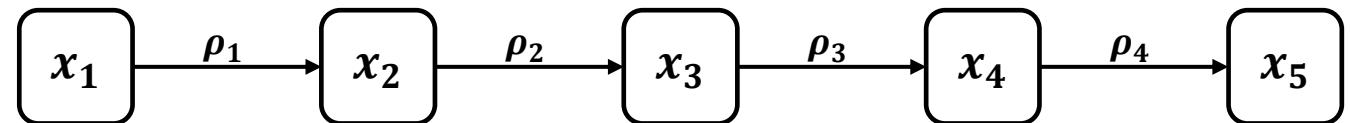
This is incredibly convoluted & constraining!



IMPLICATIONS OF DAG-INFORMED SIMULATIONS: EXAMPLE

Let x be a **longitudinal measure** with modest within-subject variation over time

- as with many growth measures & homeostatic biological variables
 - consider path coefficients between successive measures where $\rho_1 = \rho_2 = \rho_3 = \rho_4 \approx 0.8$
- $\Rightarrow \rho_{x_1 x_5} = 0.8 \times 0.8 \times 0.8 \times 0.8 \approx 0.4$



With no other causal influences on x_1 & x_5 & no other mediating variables

$$\Rightarrow \rho_{x_1 x_5} \text{ is sizable at } 0.4$$

With other variables in the DAG (each requiring their own path coefficients)

- most path coefficients must be small to avoid unrealistic bivariate correlations
- the DGM can impose considerable constraint on possible datasets

SUMMARY

Simulating data is not straightforward

Capturing the underlying **DGM** is essential for sound simulation studies

Possible datasets to be considered (e.g. parameter ranges) are more restrictive when informed by a DAG \Rightarrow the evaluation '**solution space**' is constrained

Failure to realise this may lead to **misleading simulation studies**

Even if simulation evaluation not needed, it is invaluable to still simulate data:

- **improved DAG** – meaningless DAG if unable to inform a simulation of similar data
- provides confirmation that methods do as suspected
- useful for **open access** where (sensitive) study data are not in public domain
- many methodological insights can emerge

2.4 - PROPENSITY SCORE APPROACHES

GEORGIA



@GEORGIATOMOVA

MARK



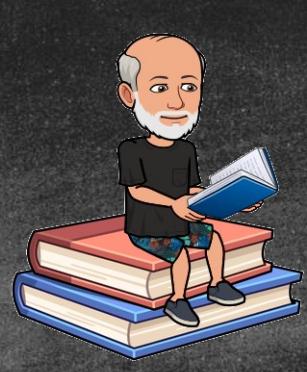
@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 2

09:30-10:15 LECTURE 2.1

10:15-11:00 ACTIVITY 2-A

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 2.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-15:00 ACTIVITY 2-B

15:00-15:30 LECTURE 2.3

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 2.4

17:00-17:45 ACTIVITY 2-C

17:45-18:00 Q&A

LEARNING OBJECTIVES

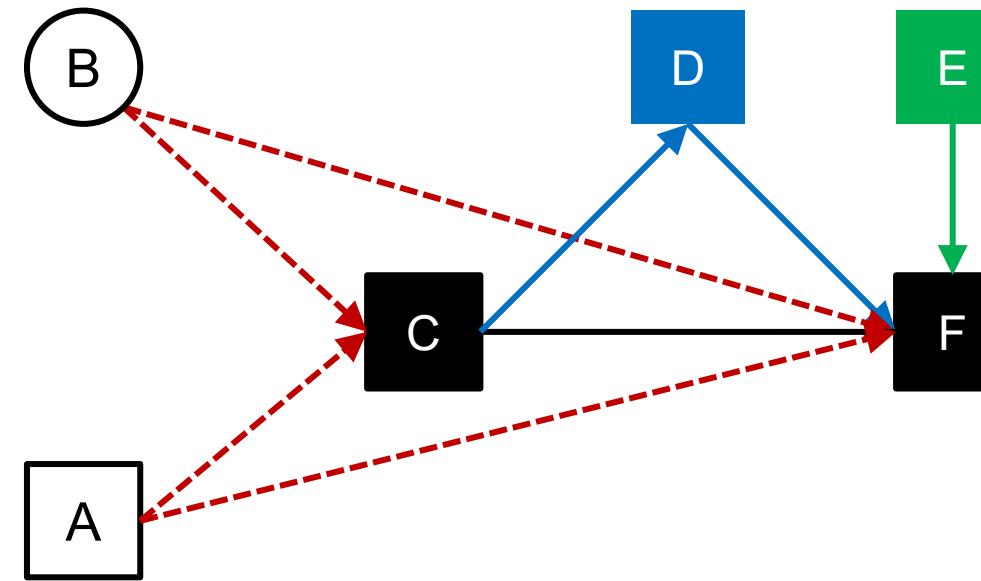
By the end of this lecture, you will be able to:

- Define a **propensity score**
- Discuss the similarities and differences between a propensity score approach and a traditional approach of '**directly adjusting for confounding**'
- Identify which covariates should and should not be included in a propensity score
- Explain 4 different ways in which propensity scores can be used
- Discuss the benefits and drawbacks of a propensity score approach
- Explain why a propensity score approach offers an easy framework for integrating machine learning methods into causal inference

RECAP: ESTIMATING THE TOTAL CAUSAL EFFECT

To estimate the total causal effect of **C** on **F**:

- Aim for all **causal paths** to be **open**
- Aim for all **confounding paths** to be **closed**

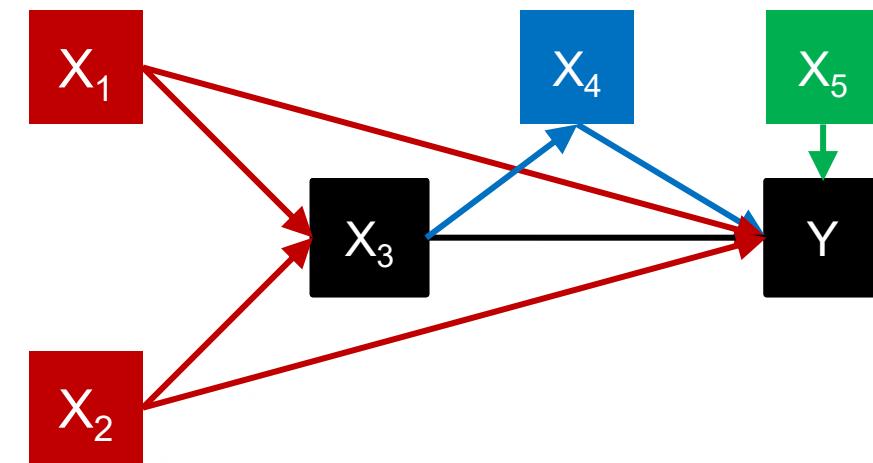
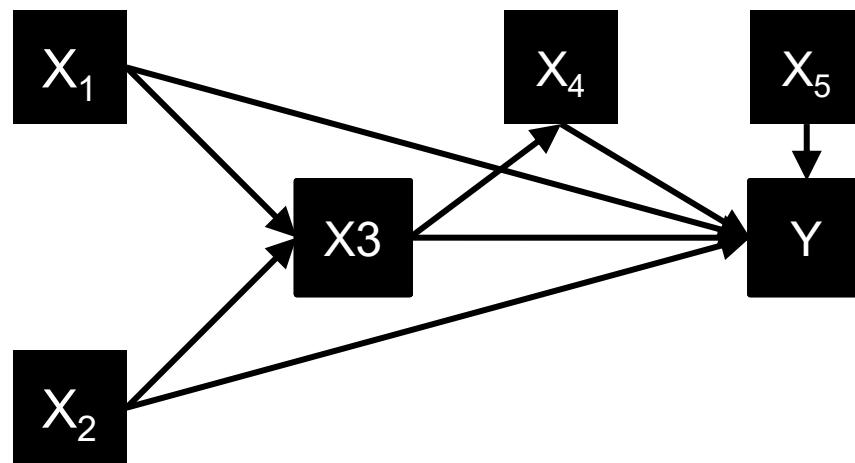


RECAP: CONDITIONAL EXCHANGEABILITY

Example: Total causal effect of X_3 on Y :

So far, have talked about building a single model that includes **confounders** (X_1, X_2) and **competing exposures** (X_5)

- $Y \sim f(X_3, X_1, X_2, X_5)$

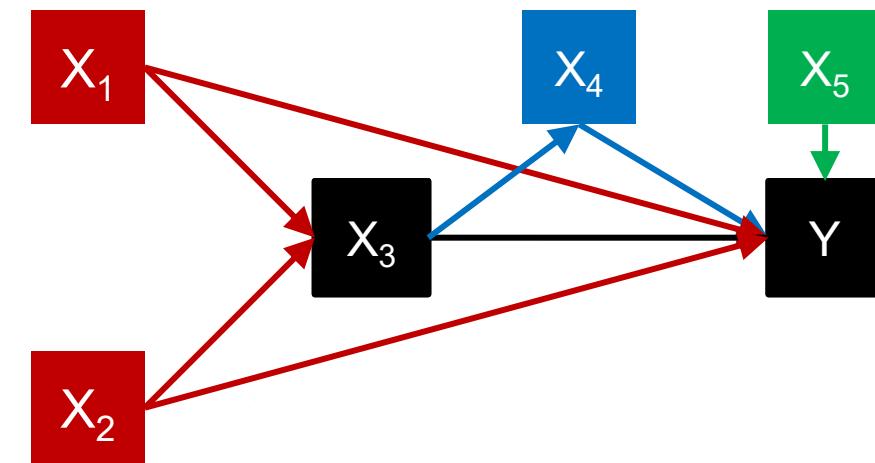
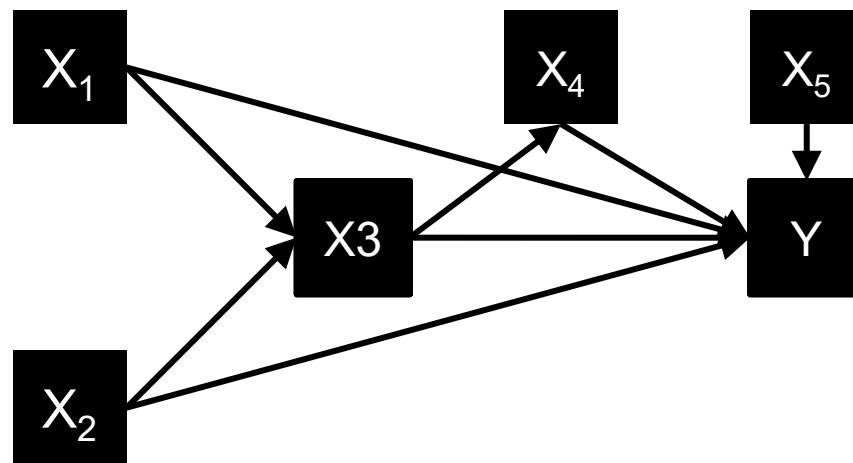


DIRECTLY ADJUSTING FOR CONFOUNDERS

E.g. covariate adjustment within a generalised linear model (GLM)

- $Y = \beta_0 + \beta_1 X_3 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_5 + \varepsilon$
- $Y = \beta_0 + \beta_1 X_3 + \beta_2 X_3^2 + \beta_3 X_1 + \beta_4 X_2 + \beta_5 X_1 * X_2 + \beta_6 X_5 + \varepsilon$

You might call this the ‘directly adjusting for confounders’ approach



PROPENSITY SCORE METHODS

Propensity score methods provide a (subtle) alternative to including the confounders (and competing exposures) in your **outcome model** directly

Two stage approach:

- predict the **propensity of exposure** (known as the **propensity score**)
- estimate the **exposure-outcome relationship**, conditional on **propensity score**

Instead of:

- $Y \sim f(X_3, X_1, X_2, X_5, \varepsilon)$

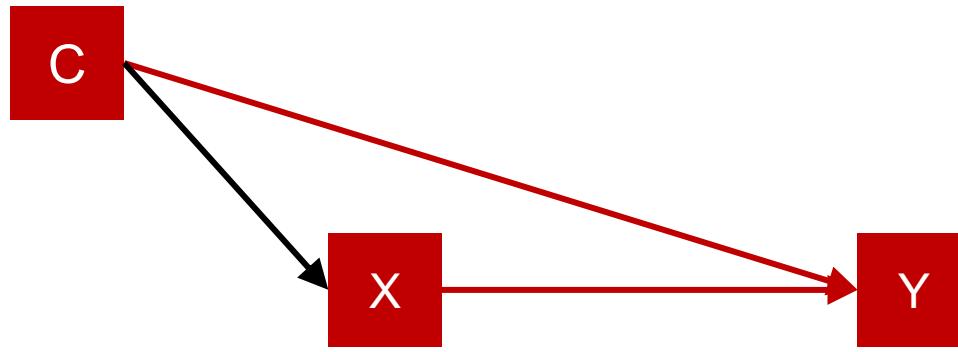
Do:

- $X_3 = g(X_1, X_2, X_5, u) = PS + u$
- $Y \sim h(X_3, PS, \varepsilon)$

PROPENSITY SCORE METHODS

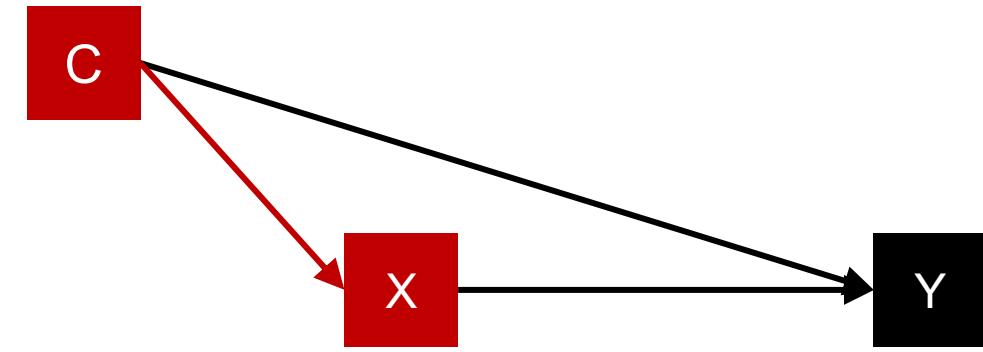
Propensity score methods provide a (subtle) alternative to including the confounders (and competing exposures) in your **outcome model** directly

Instead of modelling this:



$$E(Y|X, C)$$

We model this:



$$E(X|C)$$

DEFINITION OF THE PROPENSITY SCORE

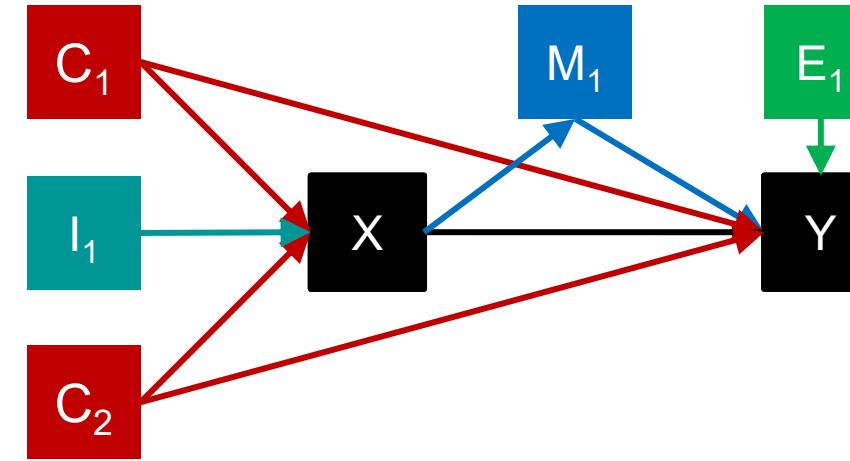
- The conditional probability that an individual was exposed, given his/her covariate history:

$$\textcolor{violet}{PS} = P(X = 1|C)$$

- Unique properties of the propensity score:
 - It is a single number, regardless of the number of confounders in C
 - If an exposed and unexposed person have the same value of the propensity score, they were equally likely to have received the exposure (i.e. they are exchangeable)
 - If conditional exchangeability holds given C, then conditional exchangeability also holds given PS

ESTIMATING A PROPENSITY SCORE

Example: Total causal effect of **X** on **Y**:



Which covariates should we include when estimating the propensity score?

SAME VARIABLE SELECTION RULES APPLY!

DO

- Include **confounders** to block confounding paths

DO NOT

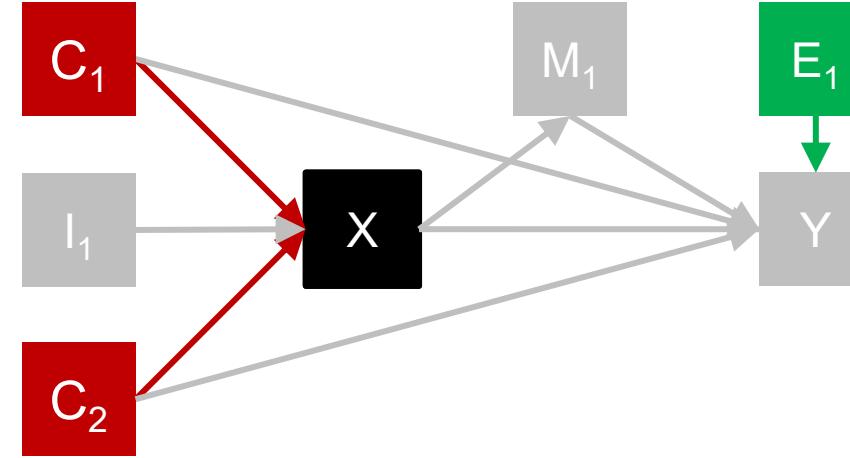
- Include **mediators**, as this would block true causal paths (and risk **collider bias**)
- Include **instrumental variables**, as this *reduces precision and may amplify biases from unmeasured confounders (i.e. **bias amplification**)*

DO (OPTIONALLY)

- Include **competing exposures**, as this *improves precision*
 - ✓ **NOTE:** competing exposures that occur *before* the exposure may be **confounders** (include)
 - ✓ **WARNING:** competing exposures that occur after the exposure may be **mediators** (exclude)

SAME VARIABLE SELECTION RULES APPLY!

Example: Total causal effect of **X** on **Y**:



$$PS = P(X|C_1, C_2, E_1)$$

$$\text{logit}(PS) = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 E_1$$

WAYS OF USING PROPENSITY SCORES

Two stage approach:

- predict the **propensity of exposure** (i.e. the **propensity score**)
- estimate the **exposure-outcome relationship**, conditional on **propensity score**

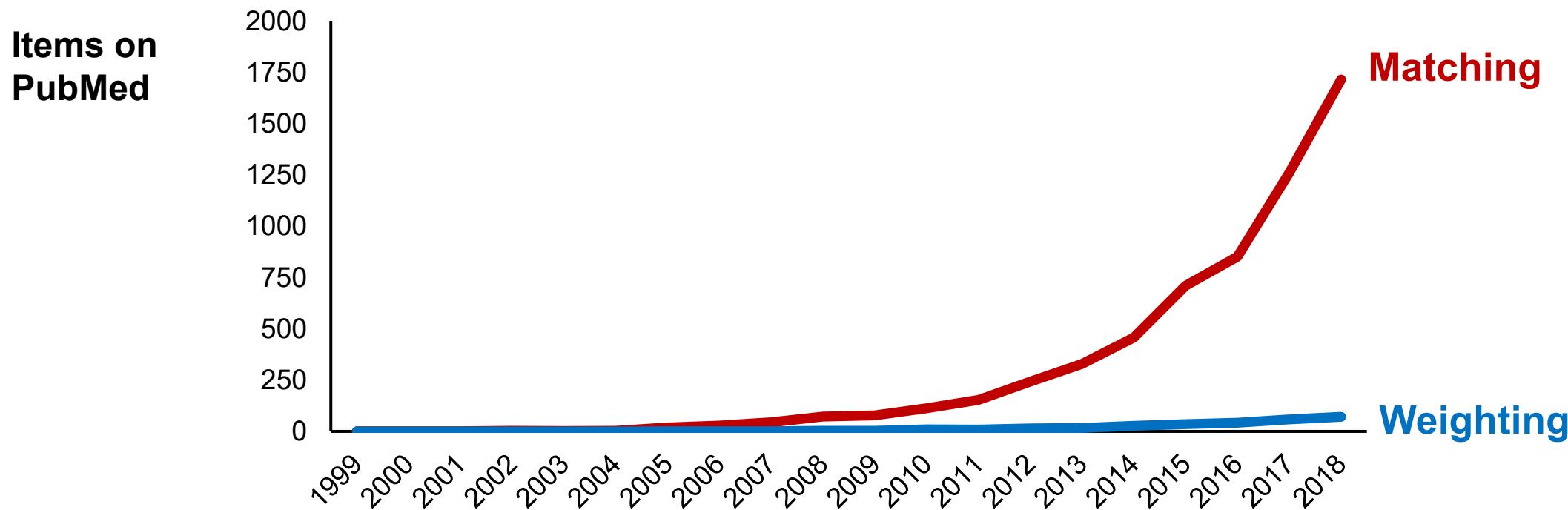


1. **Stratification**: basic comparisons within subgroups of patients with similar propensity scores
2. **Matching**: pairing exposed and unexposed patients with similar propensity scores
3. **Weighting**: reweighting the data using weights defined by the propensity score (i.e. inverse probability weighting)
4. **Adjustment**: adjusting for propensity scores within a regression model

MATCHING

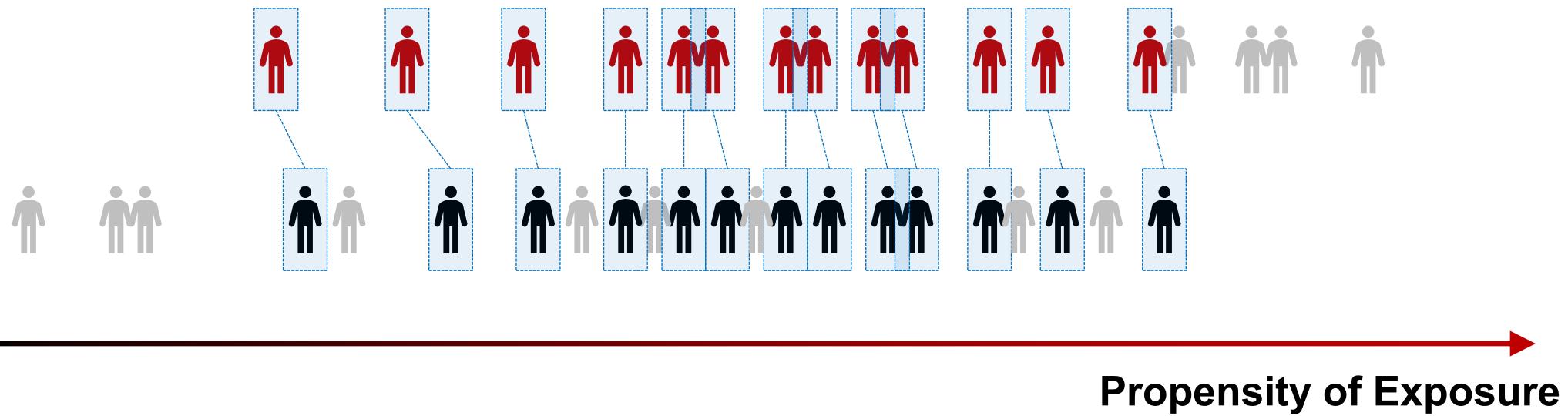
Propensity score matching was traditionally the most popular approach to conditioning on the propensity score in the exposure-outcome model

Popular in clinically-led research teams due to conceptual similarities with randomised experimental studies



MATCHING

A '**balanced**' analytical sample is created by selecting pairs (or groups) of exposed and unexposed participants with similar propensity scores, and unmatched participants are discarded



Have to decide on an appropriate 'matching' algorithm (e.g. nearest neighbour) and how close the match must be

CAVEATS TO MATCHING

- Not conceptually suitable for continuous exposures
- Imperfect matching leads to residual confounding
- Discards unmatched participants...
- Doesn't generally provide estimate of total causal effect
- Standard errors are tricky to calculate
- Can increase imbalance, inefficiency, and bias

The screenshot shows a journal article page from the *Political Analysis* journal. The title of the article is "Why Propensity Scores Should Not Be Used for Matching". It was published online by Cambridge University Press on 07 May 2019. The authors are Gary King and Richard Nielsen. The article has sections for Article, Figures, Supplementary materials, and Metrics. Below the article title, there are buttons for Save PDF, Share, Cite, and Rights & Permissions. The abstract section discusses the limitations of propensity score matching, stating that it often increases imbalance, inefficiency, model dependence, and bias. The journal logo "PA POLITICAL ANALYSIS" is visible on the left.

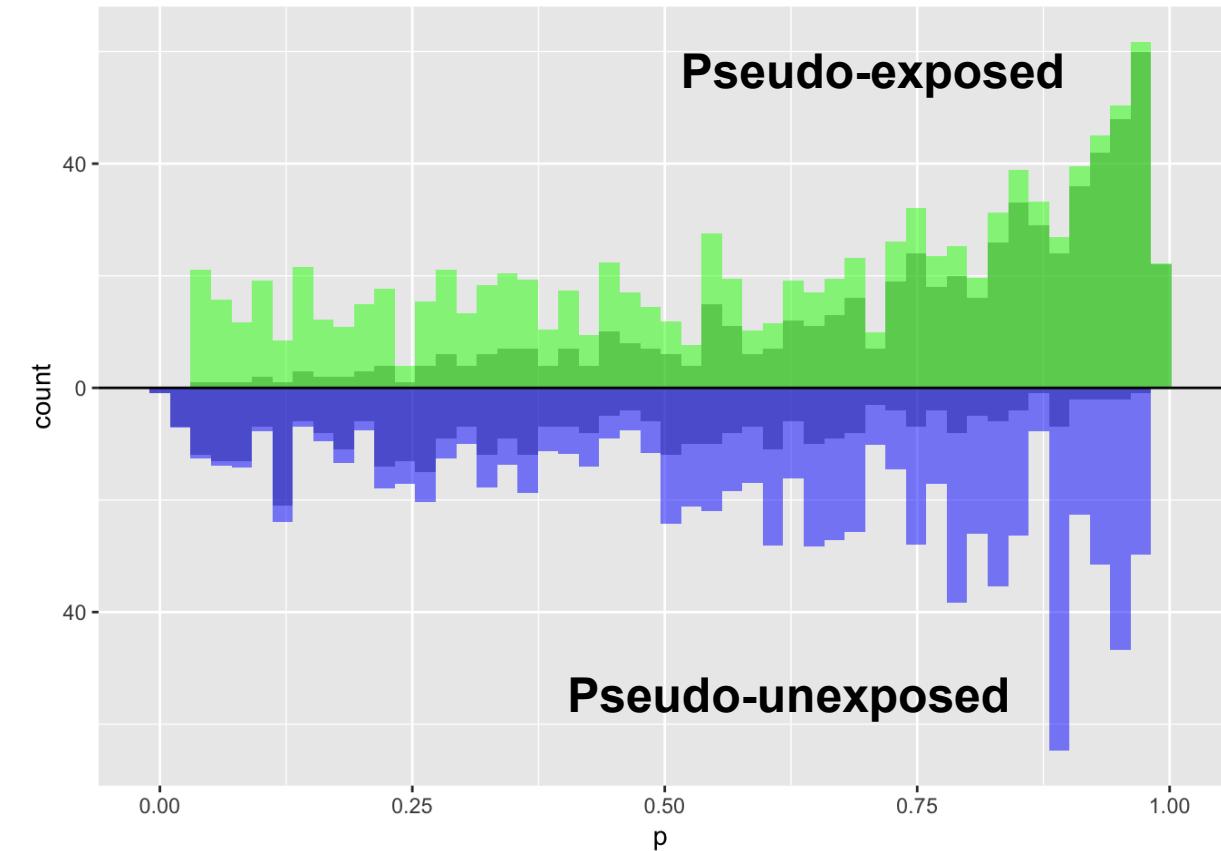
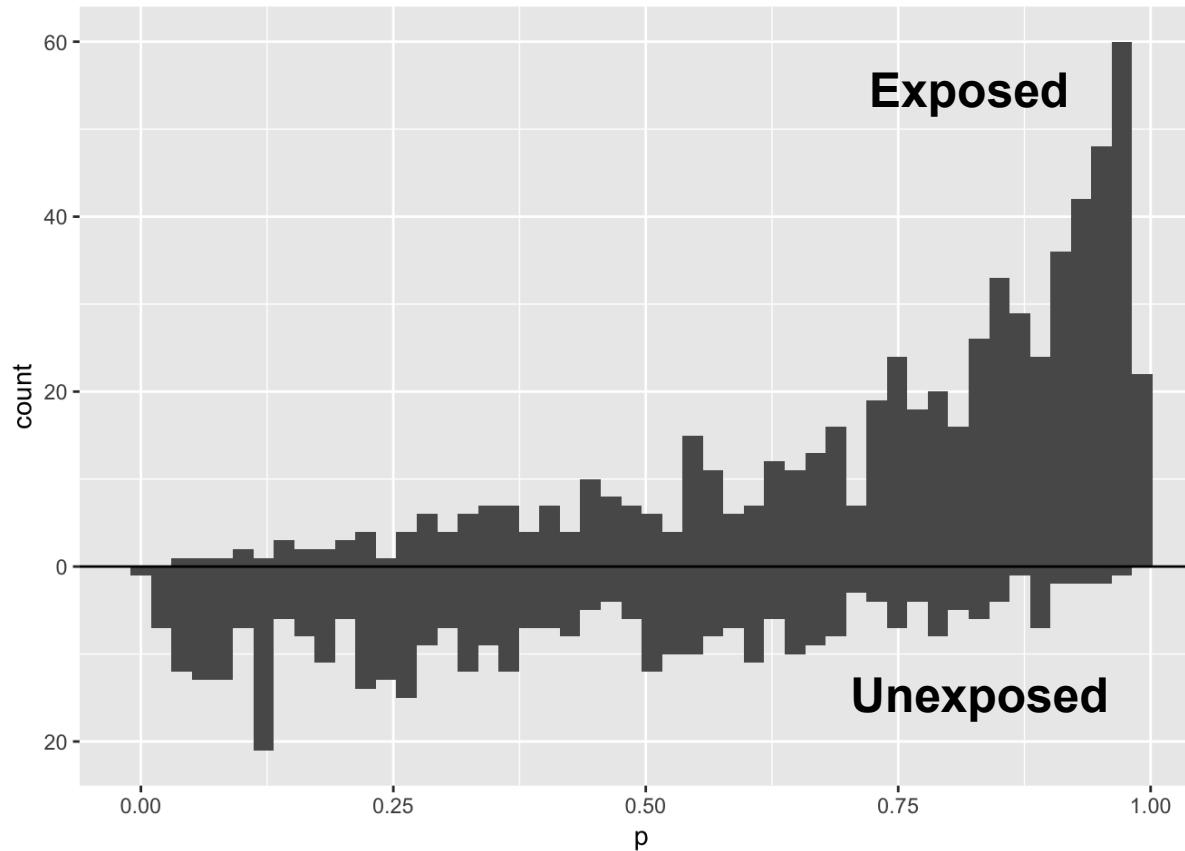
King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 27(4), 435-454. doi:10.1017/pan.2019.11

WEIGHTING

- Weight each individual by the inverse of the conditional probability of receiving the treatment that they actually received (this is referred to as **inverse probability weighting**)
 - for exposed individuals: $\frac{1}{PS}$
 - for unexposed individuals: $\frac{1}{1-PS}$
- Combine all individuals into a new '**pseudo-population**', in which confounding doesn't exist
- Association between exposure and outcome in the '**pseudo-population**' can be interpreted as an average total causal effect
- Retains information from all participants, but can become unstable with extreme weights

WEIGHTING

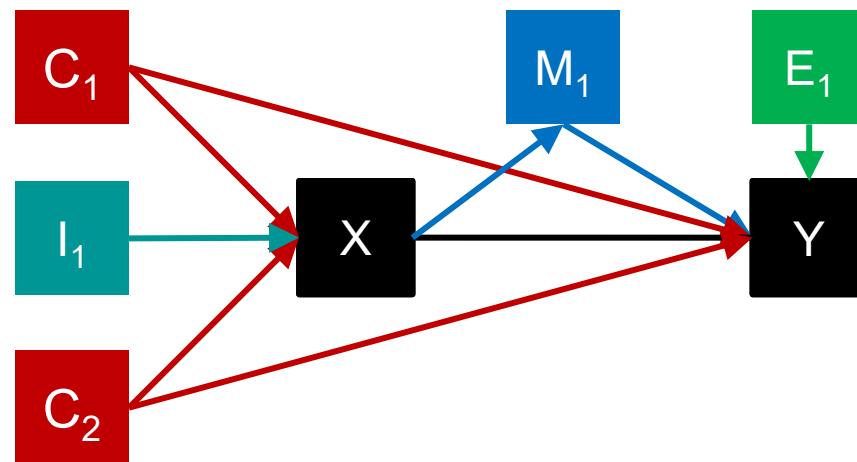
Evaluate-outcome model in a balanced **pseudo-population** generated by **inverse probability weighting**



Figures from very useful blog by Lucy D'Agostino McGowan : <https://livefreeordichotomize.com/2019/01/17/understanding-propensity-score-weighting/>

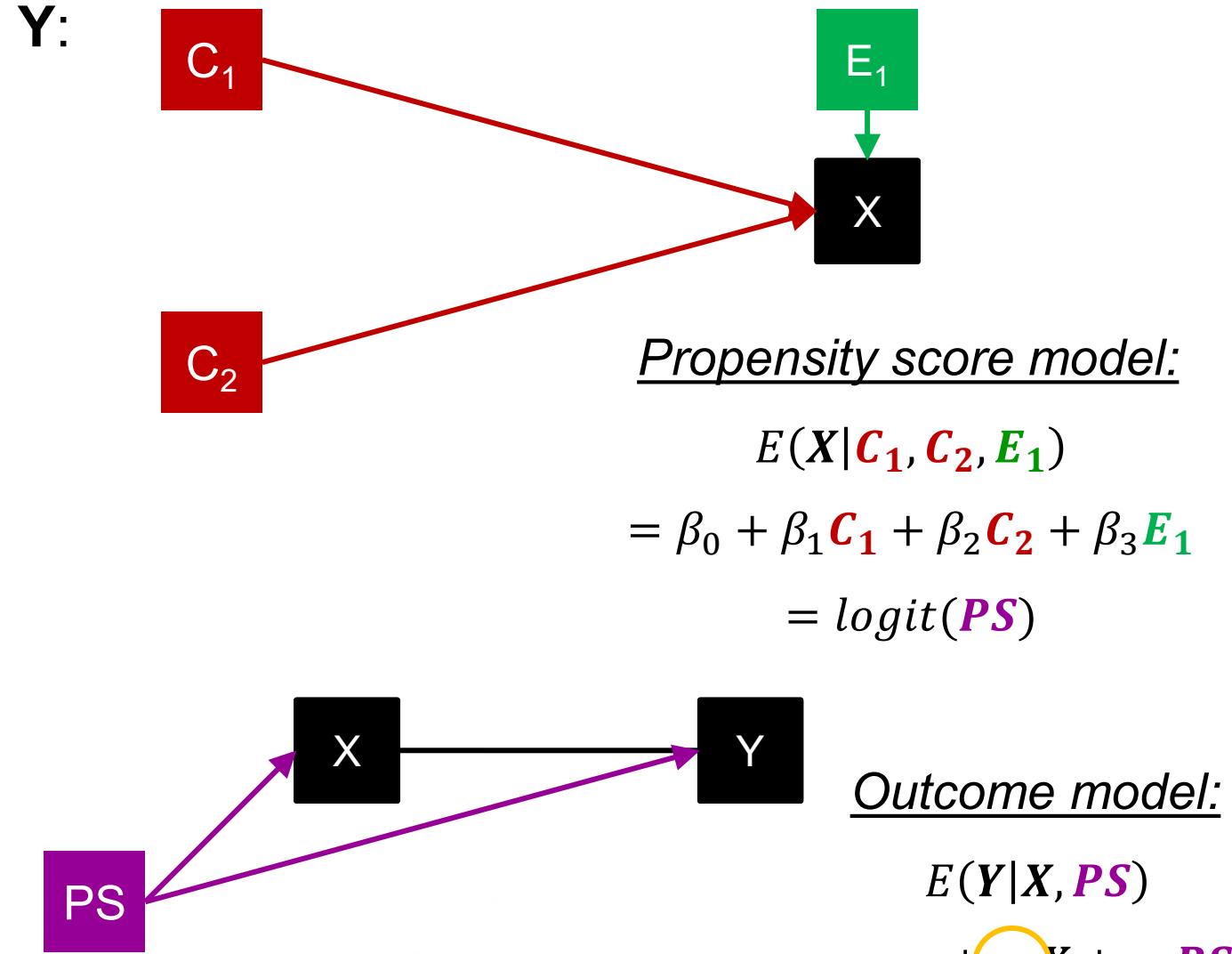
ADJUSTMENT

Example: Total causal effect of X on Y :



Exposure-outcome model:

$$E(Y|X, \mathbf{C}_1, \mathbf{C}_2, E_1) \\ = \alpha_0 + \alpha_1 X + \alpha_2 C_1 + \alpha_3 C_2 + \alpha_4 E_1$$



WHY BOTHER?!

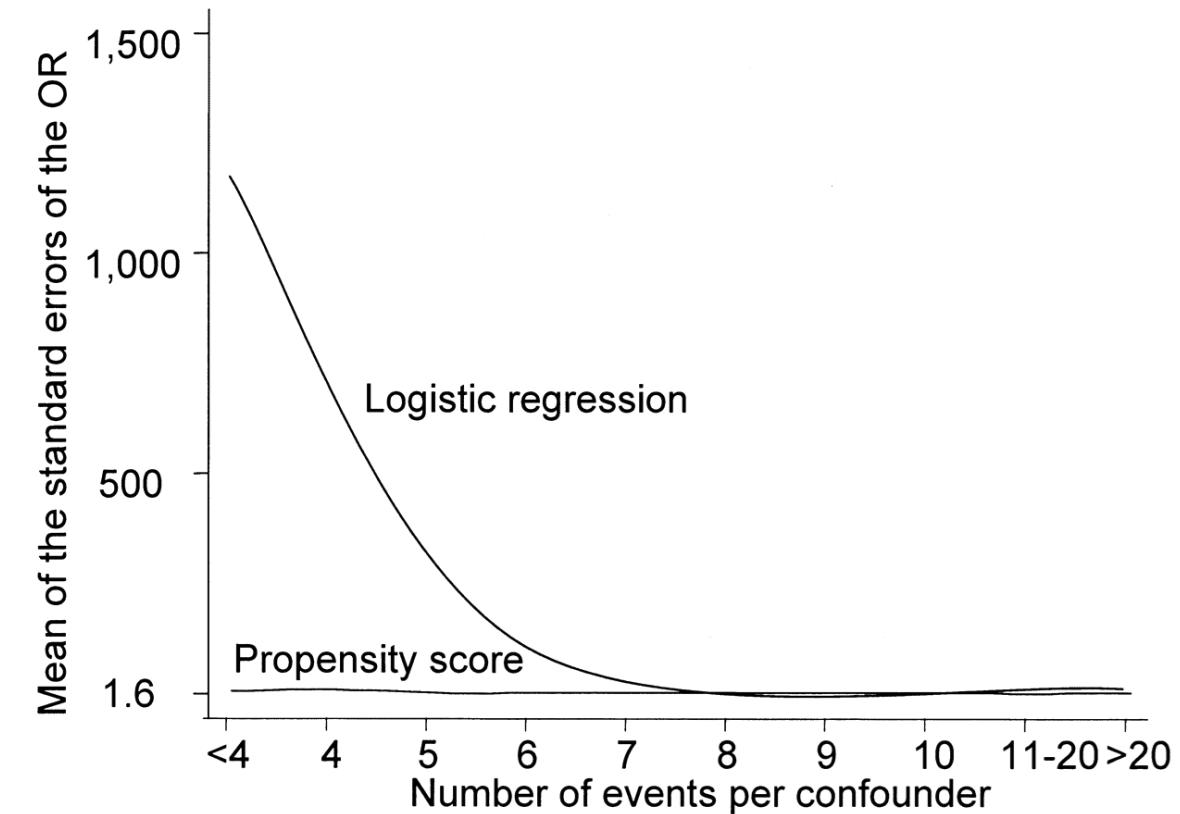
Benefits of a PS approach:

- discourages **Table 2 Fallacy**
- more efficient when **outcome is rare**
- encourages checking for **positivity violations** among confounders
- more suitable for combination with machine learning



BENEFITS FOR RARE OUTCOMES

- With rare outcomes it can be difficult to accurately model confounder-outcome relationships due to **data sparsity**
- Propensity score approach reduces this problem, since PS score is modelling in relation to exposure
- The PS approach is less biased, more robust, and more precise than outcome regression when there are few events per confounder.



Cepeda, M. S., Boston, R., Farrar, J. T., & Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American journal of epidemiology*, 158(3), 280–287. <https://doi.org/10.1093/aje/kwg115>

TABLE 2 FALLACY

Standard regression model:

$$Y = \alpha_0 + \alpha_1 X + \alpha_2 \textcolor{red}{C}_1 + \alpha_3 \textcolor{red}{C}_2 + \alpha_4 \textcolor{green}{E}_1$$



Propensity score model:

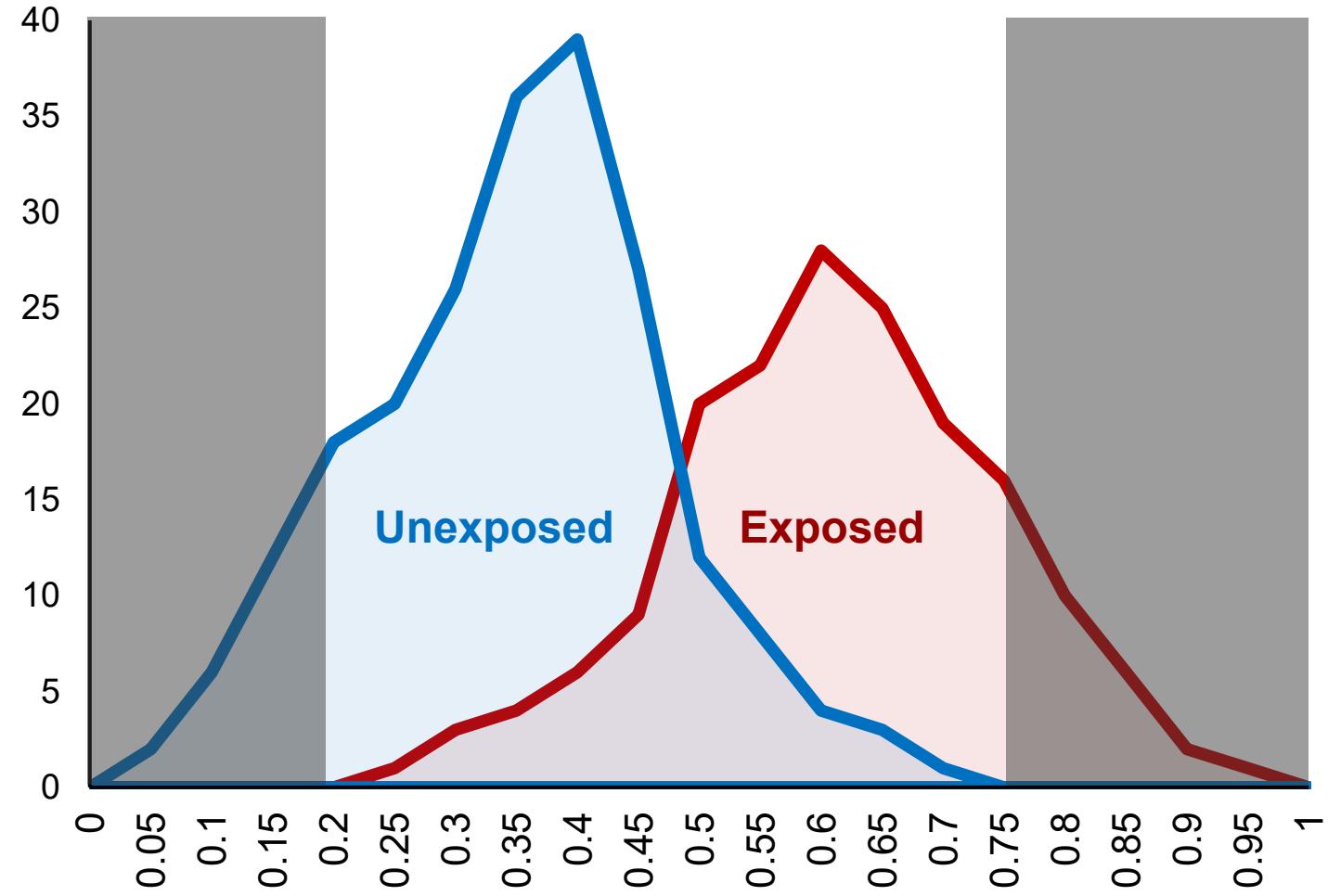
$$Y = \gamma_0 + \gamma_1 X + \gamma_2 \textcolor{purple}{PS}$$



POSITIVITY VIOLATIONS

A propensity score approach allows you to (explicitly) check that all levels of exposure are possible at all propensities

For those propensities where *everyone* is exposed or *everyone* is unexposed, you technically have a positivity violation



DEALING WITH POSITIVITY VIOLATIONS

You could **trim** (i.e. restrict) your analytical sample to focus only on areas of **overlap**

- necessary when non-overlap is due to **structural positivity violation**
- but less clear whether it is necessary for **random positivity violations**

Furthermore, **trimming** will result in a non-representative sample

- perhaps best resolved using sensitivity analysis?

ESTIMATORS: MACHINE LEARNING

Machine learning methods are predominantly used for prediction and may not be as suitable for causal inference, but generally allow for more **complex parameterisations**

Methods for incorporating the theory into machine learning algorithms have been developed but their use in practice is extremely limited

However, ML can be helpful since the first stage of PS can be reframed as a **predictive activity**

- **CAVEAT**: *The goal of a propensity score isn't to predict the exposure, but to adjust for confounding*

PROPENSITY SCORE METHODS

In summary, the effectiveness of PS methods rests critically on the choice of covariates, X , and that choice cannot be left to guesswork; it requires that we understand, at least figuratively, what relationships may exist between observed and unobserved covariates and how the choice of the former can bring about strong ignorability or a reasonable approximation thereof [6].

Pearl, J. Remarks on the method of propensity score. (2009). *Statistics in Medicine* 2009, 28, 1415–1424.

COMPARING DIFFERENT APPROACHES

Method	Pros	Cons
Exposure-outcome model (‘directly adjusting for confounding’)	<ul style="list-style-type: none">• Can explore confounder-effect interactions• Can estimate population effect estimate and in any strata	<ul style="list-style-type: none">• Less precise when outcome rare• Risk of Table 2 Fallacy
PS covariate adjustment	<ul style="list-style-type: none">• More precise when outcome rare• Reduced risk of Table 2 Fallacy• Can estimate true population effect estimate and in strata of PS	<ul style="list-style-type: none">• Typically similar results to exposure-outcome model• More difficult to explore confounder-effect interactions
PS matching	<ul style="list-style-type: none">• Minimises positivity violations• Easy to interpret (particularly for those familiar with RCTs)	<ul style="list-style-type: none">• Sensitive to matching approach• Losses information from unmatched• Less precise• Effect not estimated in full population
PS stratification	<ul style="list-style-type: none">• Retains more participants• Explicitly models PS-outcome interactions	<ul style="list-style-type: none">• Less effective when outcomes rare• Residual confounding possible
PS inverse probability weighting	<ul style="list-style-type: none">• Retains all participants• Can estimate true population effect and in strata of PS	<ul style="list-style-type: none">• Unstable for with extreme weights

RECOMMENDED READING

- D'Agostino, R.B., Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17, 2265-2281. [https://doi.org/10.1002/\(SICI\)1097-0258\(19981015\)17:19<2265::AID-SIM918>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1097-0258(19981015)17:19<2265::AID-SIM918>3.0.CO;2-B)
- Pearl, J. (2012). On a class of bias-amplifying variables that endanger effect estimates. *arXiv*: 1203.3503.
- Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American journal of epidemiology*, 163(12), 1149–1156.
<https://doi.org/10.1093/aje/kwj149>
- Cepeda, M. S., Boston, R., Farrar, J. T., & Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American journal of epidemiology*, 158(3), 280–287. <https://doi.org/10.1093/aje/kwg115>

SUMMARY

- Propensity score methods provide a (subtle) alternative to conditioning on confounders
- The primary methods for using propensity scores are: **stratification, matching, weighting, and adjusting**
- Propensity score adjustment offers improvements over traditional regression adjustment in settings with rare outcomes and high dimensional data
- Machine learning methods can be used to model propensity

09:30-10:15 ACTIVITY 3-A

10:15-11:00 LECTURE 3.1

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 3.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 3.3

14:45-15:30 ACTIVITY 3-B

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 3.4

17:00-17:45 ACTIVITY 3-C

17:45-18:00 Q&A

3.1 - INTRODUCTION TO COLLIDER BIAS

GEORGIA



MARK



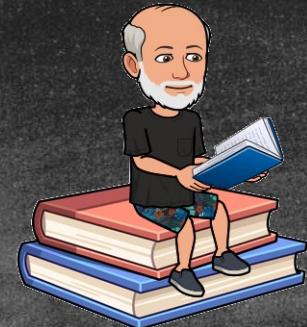
@GEORGIATOMOVA

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

LEARNING OBJECTIVES

By the end of this session, you will be able to:

- Define a **collider**, a **collider path**, and **collider bias**
- Explain how **collider bias** and **conditional dependencies** occur
- Use DAGs to detect (and avoid) potential collider biases
- Recognise **mediator-outcome confounding**
- Explain the **Birthweight Paradox** and how conditioning on mediators can lead to **collider bias**

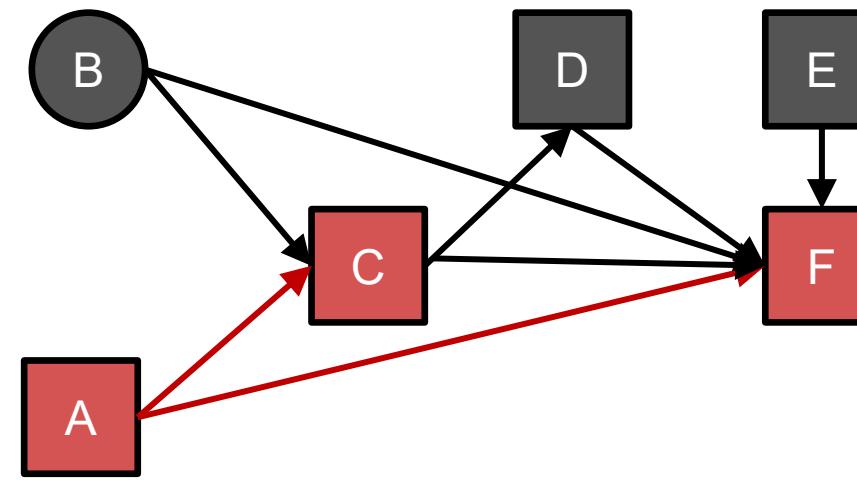
You may also need:

- ice-cream, alcohol, a hot towel to wrap your head in !

RECAP: CONFOUNDING

A **confounding path** (AKA **backdoor path**) is one where the arrows do not flow in the same direction – initially backwards, then forwards

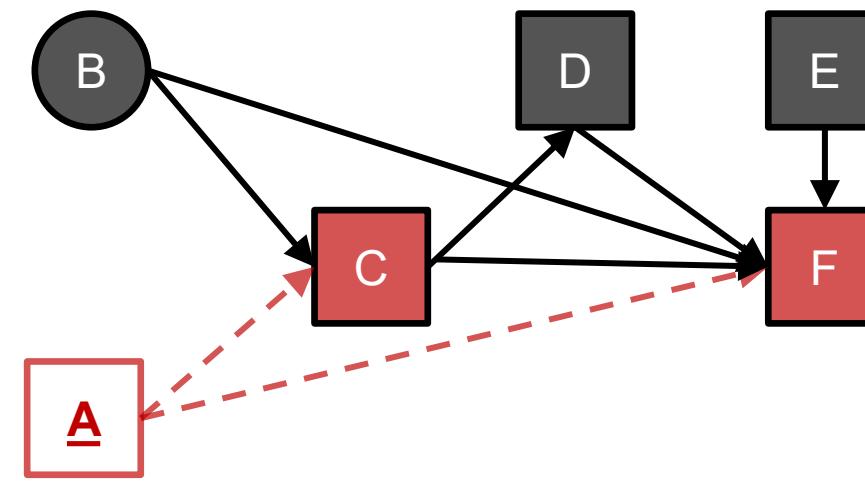
- e.g. $C \leftarrow A \rightarrow F$



Without conditioning, **confounding paths** are **open** and will transmit dependencies between the variables caused by the confounder (e.g. A)

RECAP: CONFOUNDING

A **confounding path** can be **closed** by conditioning on the confounding node(s) – or a proxy thereof



Conditioning on **A** closes **$C \leftarrow A \rightarrow F$**

RECAP: ADJUSTMENT SETS

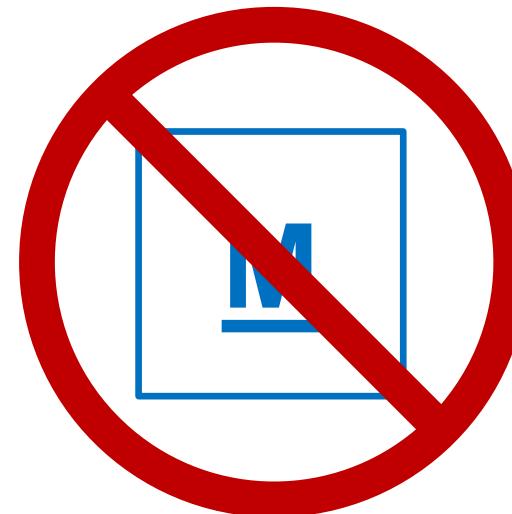
To estimate causal effects:

DO

- Condition on **confounders** to block confounding paths

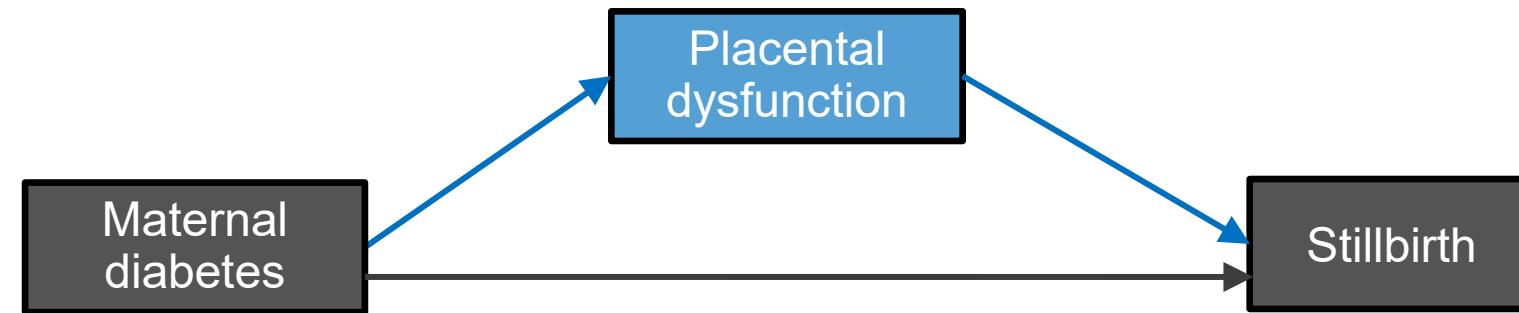
DO NOT

- Condition on **mediators** as *this would block true causal paths*



RECAP: ADJUSTMENT SETS

e.g. Maternal diabetes and stillbirth

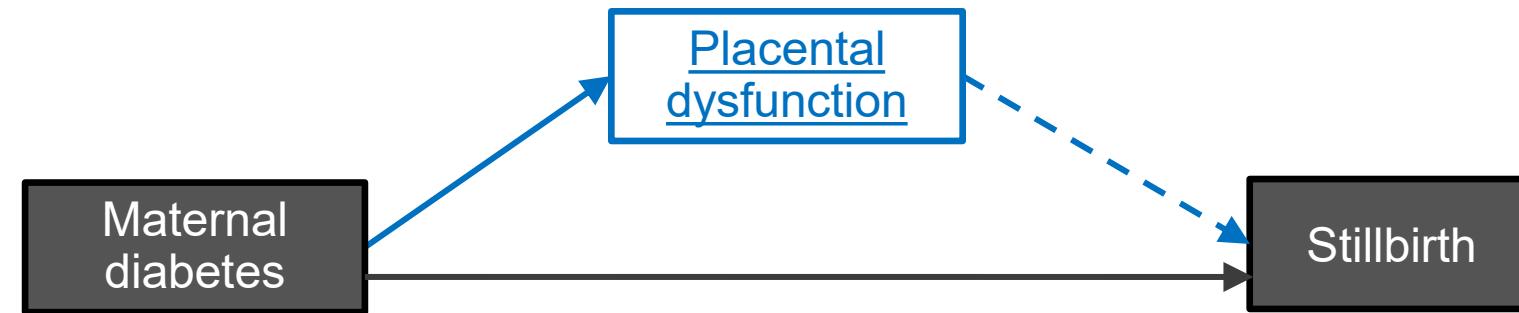


Maternal diabetes causes vascular damage to the placenta, which greatly increases the chance of stillbirth

Maternal diabetes → Placental dysfunction → Stillbirth

RECAP: ADJUSTMENT SETS

e.g. Maternal diabetes and stillbirth



If we study effect of maternal diabetes on stillbirth ...

... in mothers with similar clinical-rated placental pathology ...

... we would block causal path – and (large) part of total effect would be obscured

COLLIDER BIAS

But:

- Conditioning on mediators also risks summoning '**Collider bias**'



Collider bias (artist's impression)

COLLIDER BIAS

WARNING!

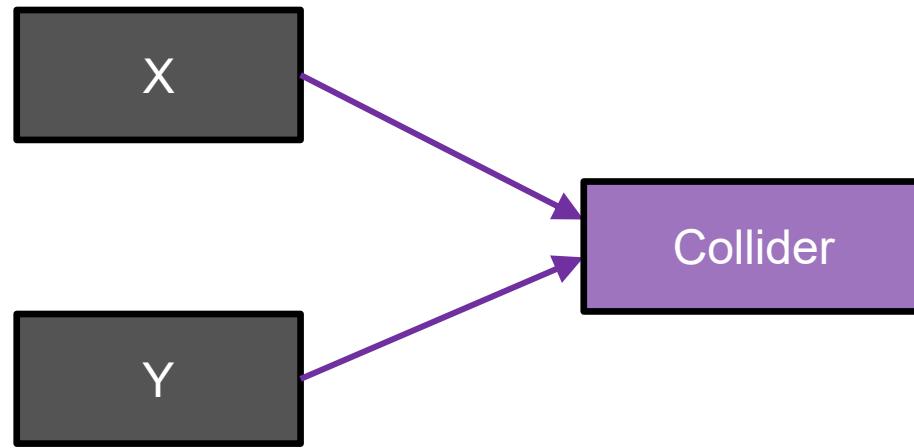
- Collider bias is **NOT** intuitive!
- It takes a long time to grasp intuitively (if ever...)
- Causal diagrams are extremely useful!



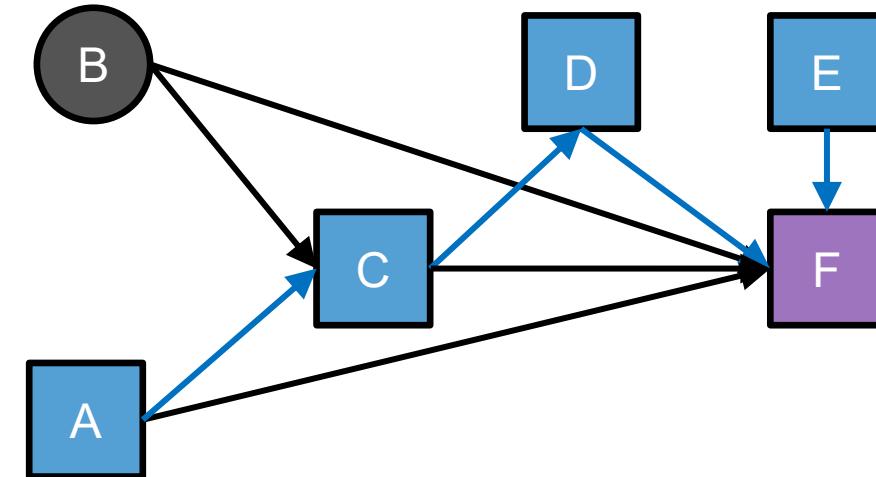
COLLIDERS

A **collider** is a variable that is caused by ≥ 2 independent causal paths

A **collider path** is a path that includes a **collider**



e.g. **X** → **Collider** ← **Y**

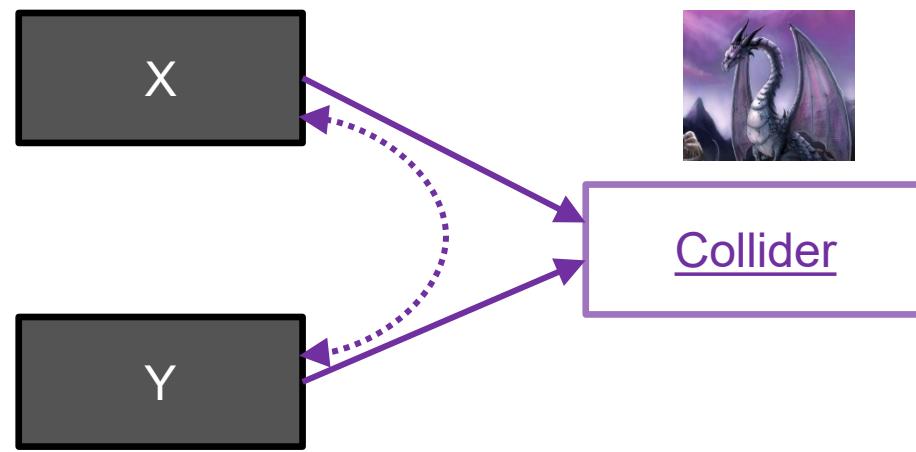


e.g. **A** → **C** → **D** → **F** ← **E**

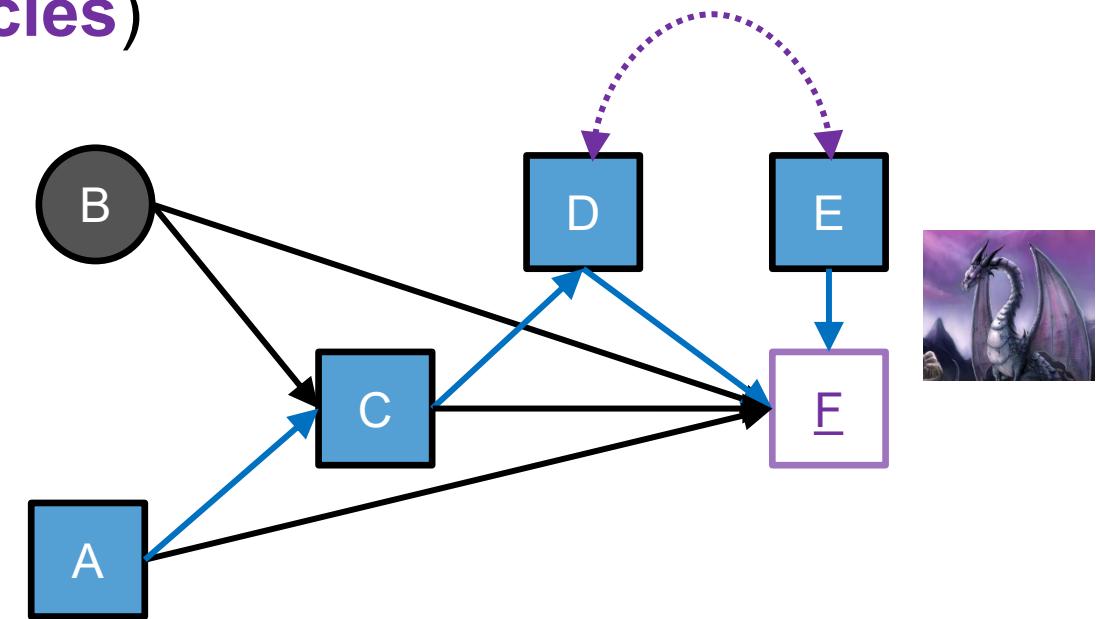
Without conditioning, **collider paths** are **closed** and will **NOT** transmit **dependencies** between the variables either side of the collider (e.g. **F**)

COLLIDERS

Collider bias is a form of **non-causal association** created by conditioning on a collider (AKA **conditional dependencies**)



e.g. $X \rightarrow \text{Collider} \leftarrow Y$
 $X <...> Y$ is opened by conditioning on C



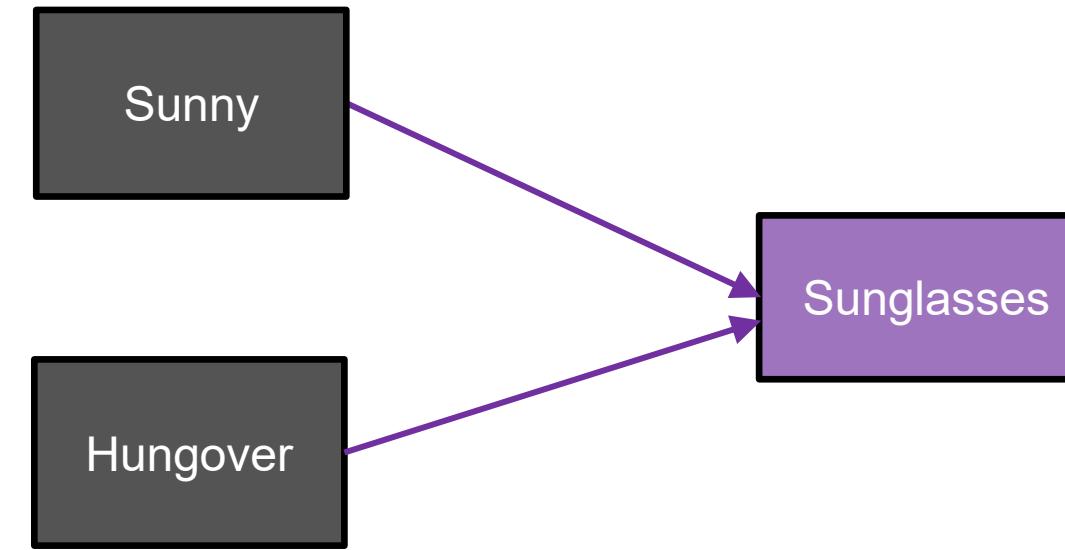
e.g. $A \rightarrow C \rightarrow D \rightarrow E \leftarrow E$
 $D <...> E$ is opened by conditioning on F

Conditional dependencies can be **mind boggling**...

EXAMPLE: COLLIDER BIAS

Example: The Sunglasses Revelation

- British people typically wear sunglasses when the **sun is shining**
- They sometimes also wear them when they are **very hungover**

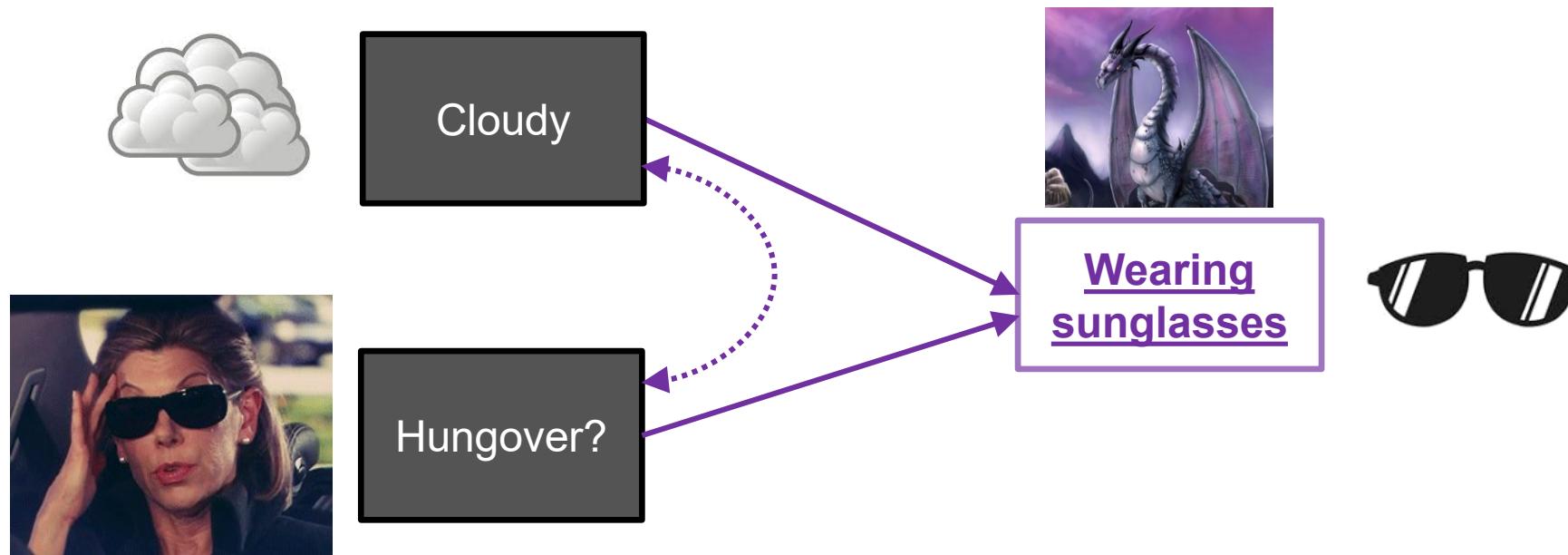


- Suppose hangovers occur equally all year round, regardless of weather

EXAMPLE: COLLIDER BIAS

Example: The Sunglasses Revelation

- You see someone wearing sunglasses on a **cloudy day**
- What does this tell us about their likelihood of being **hungover**?

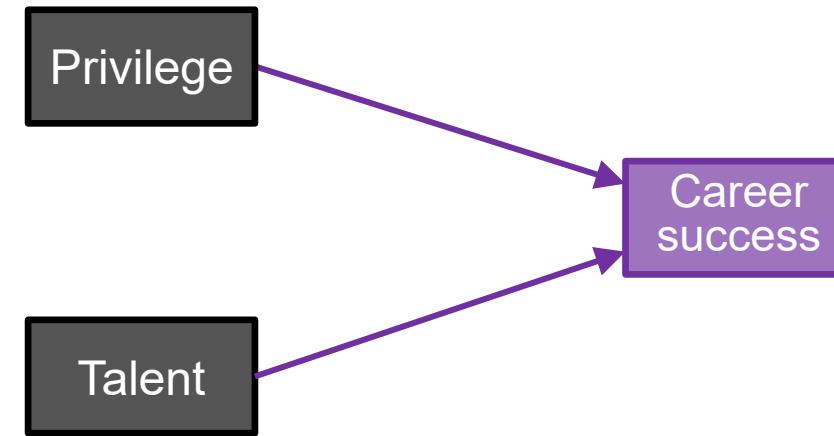


- Conditioning on sunglasses opens $\text{Cloudy} \longleftrightarrow \text{Hungover}$

EXAMPLE 2: COLLIDER BIAS

Example: The ‘Mediocre White Male Problem’

- **Privilege** and **talent** are two competing reasons why someone might achieve **career success** (e.g. becoming **Tenured Faculty**)



- In the general population, it is unlikely that **privilege** and **talent** are related

EXAMPLE: COLLIDER BIAS

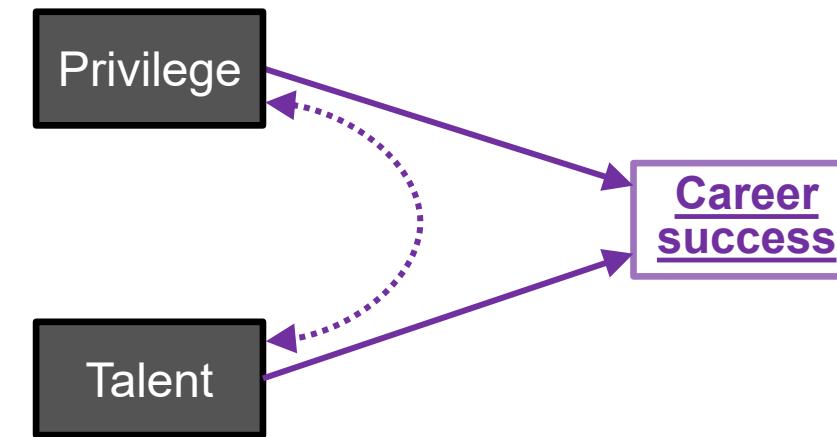
Example: The ‘Mediocre White Male Problem’

- If you look only at people who have achieved career success you can expect a **conditional dependency** – so **privilege** will be *inversely* associated with **talent**

Piers Morgan



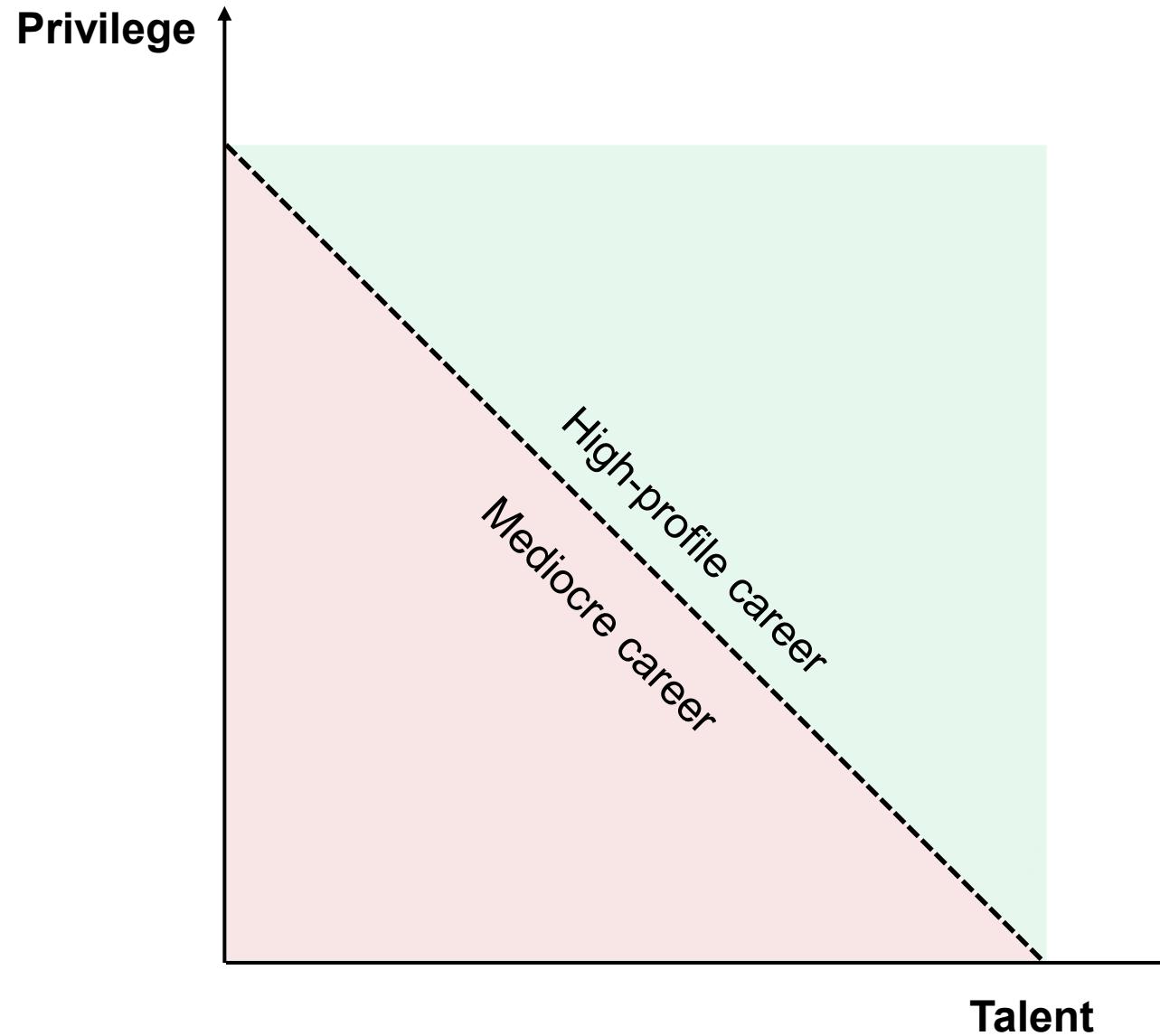
Oprah Winfrey



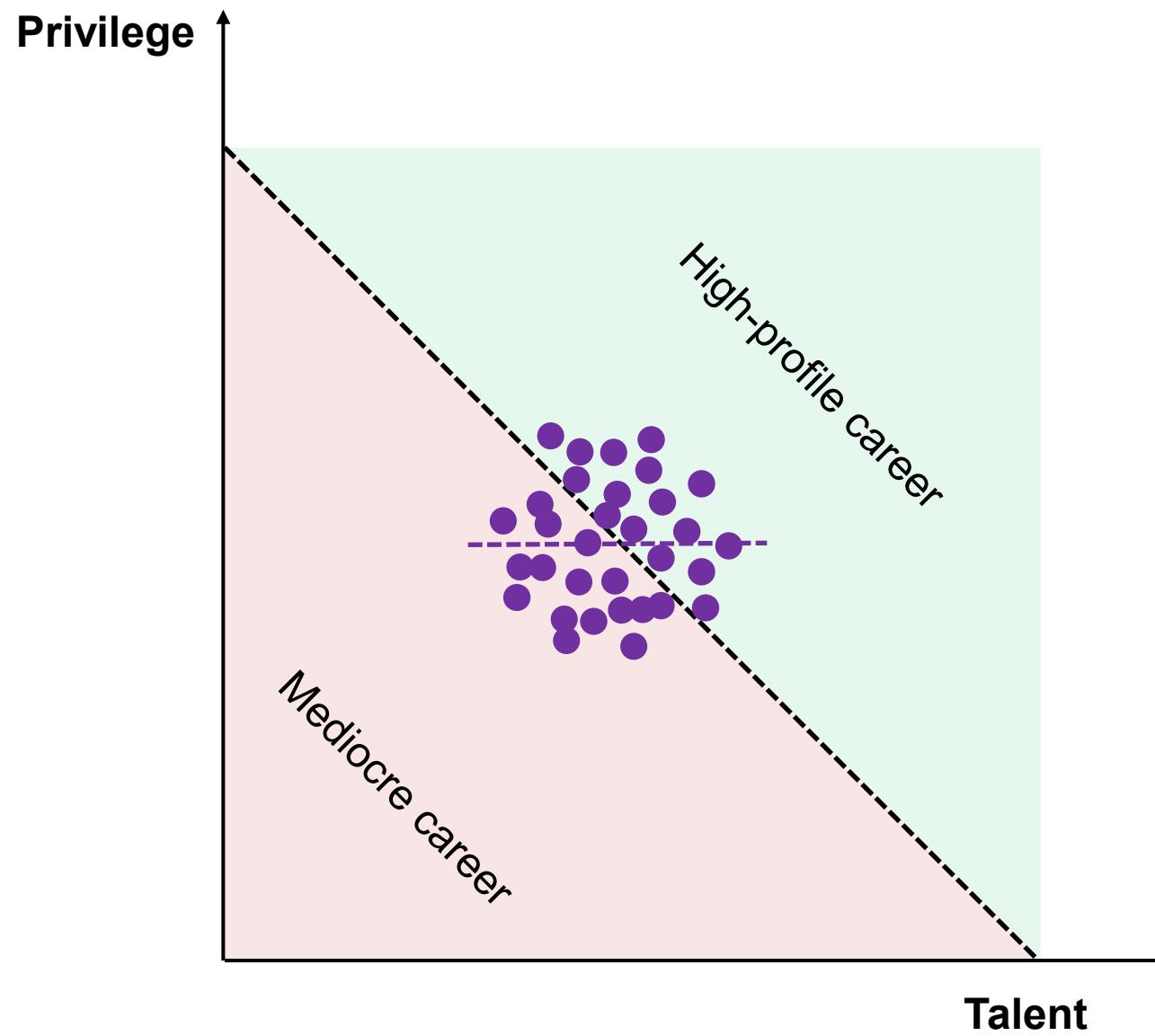
- Conditioning on career success opens **Privilege <...> Talent**

PARADOXICAL

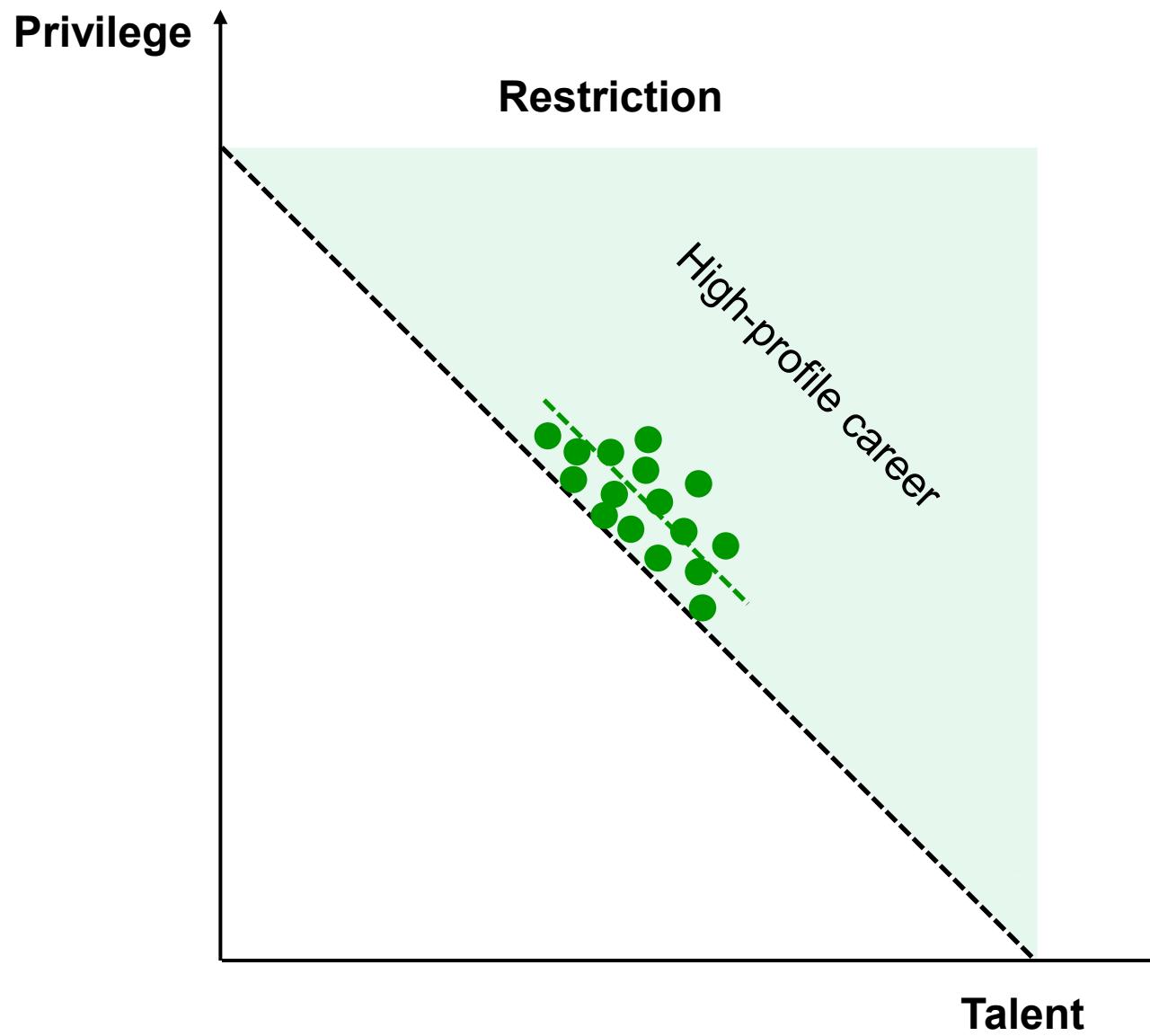
The nature of this association can appear quite **paradoxical**...



PARADOXICAL



PARADOXICAL



PARADOXICAL



EXAMPLE: COLLIDER BIAS

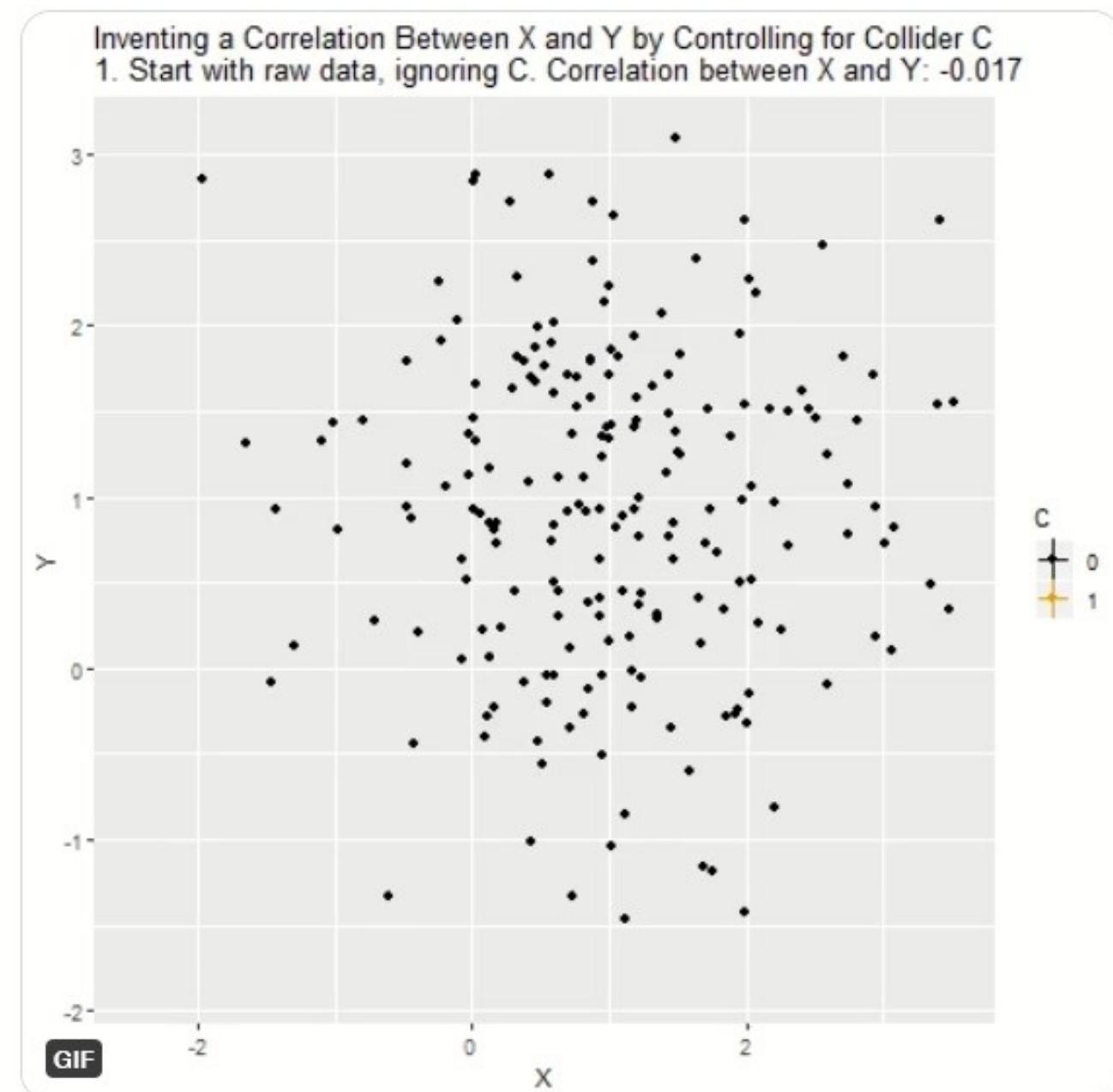


Nick HK
@nickchk

As requested, slower graphs! Also added a graph on collider bias, the webpage explanation helps there.

These graphs are intended to show what standard causal inference methods actually **do** to data, and how they work.

<https://twitter.com/nickchk/status/1068215492458905600>



CONDITIONING APPROACH

Collider bias can / will occur regardless of the conditioning approach

- **Restriction**

- ✓ Restricting the sample to exclude certain values of the collider
 - ✓ This most commonly occurs during study selection – see next lecture !

- **Stratification**

- ✓ Estimating effects for different categories of a collider

- **Covariate adjustment**

- ✓ Adjusting for the collider in multivariable regression

Collider bias will occur even when the conditioning is incomplete !

ALTERNATIVE NAMES

The capacity to cause such bizarre associations means collider bias has many names and has caused many so-called ‘paradoxes’

Simpson's paradox

- Collider bias in **categorical** data

Reversal paradox

- General term for extreme collider biases that result in an effect reversal

Berkson's paradox

- Special case of reversal paradox due to studying hospital-based sample

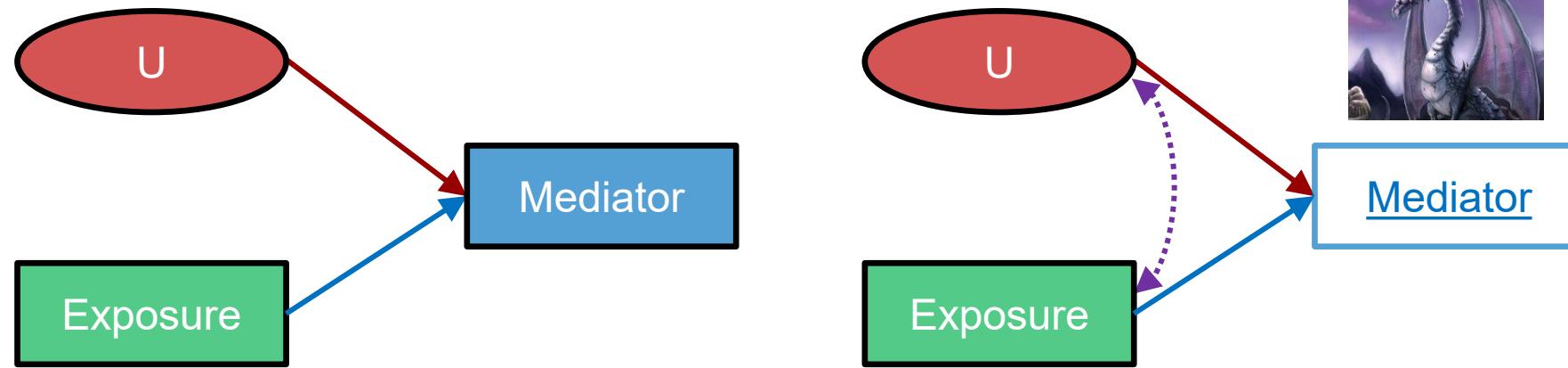
Obesity paradox

- Apparent paradox in effect of obesity due to studying older-aged sample

CONDITIONING ON MEDIATORS

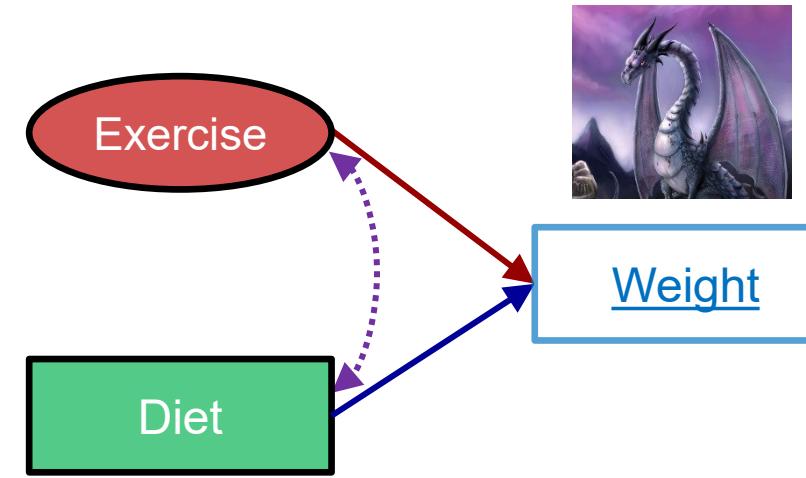
Conditioning on **mediators** can also introduce **collider bias**

This is because **mediators** are **colliders** for the effect of the **exposure** and any other **(unobserved) causes of the mediator**



CONDITIONING ON MEDIATORS

e.g. Association between **exercise** and **diet** and conditional on weight

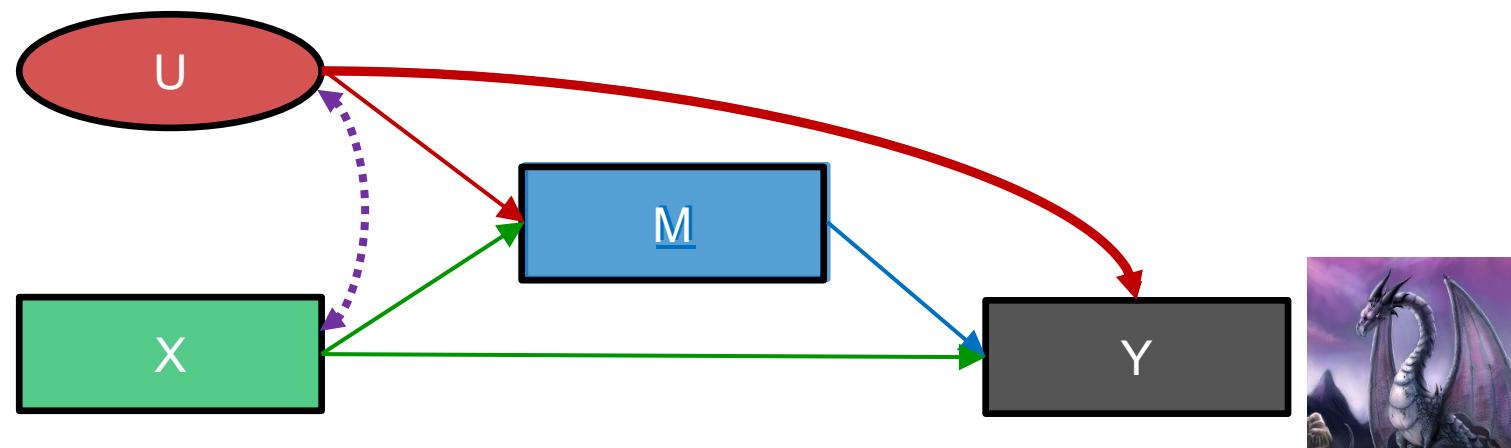


Conditioning on weight opens **Diet <…> Exercise**

CONDITIONING ON MEDIATORS

Mediator-outcome confounders are variables that cause both mediator and outcome – they likely include most unobserved causes of the mediator

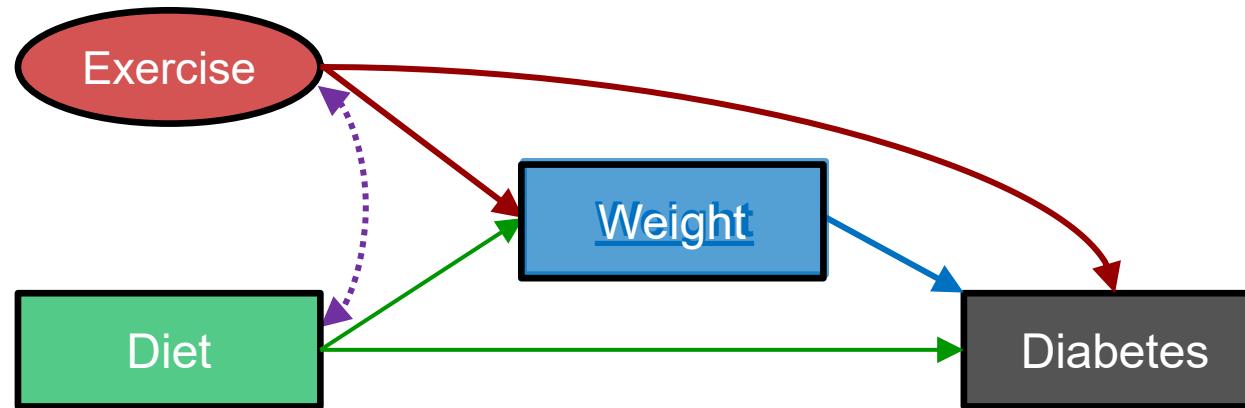
If you create a **conditional dependency** with a **mediator-outcome confounder** it will pass on the **non-causal association** to the outcome



Conditioning on **M** has opened **X** \longleftrightarrow **U** \rightarrow **Y**

CONDITIONING ON MEDIATORS

e.g. Association between **diet** and **diabetes** conditional on weight



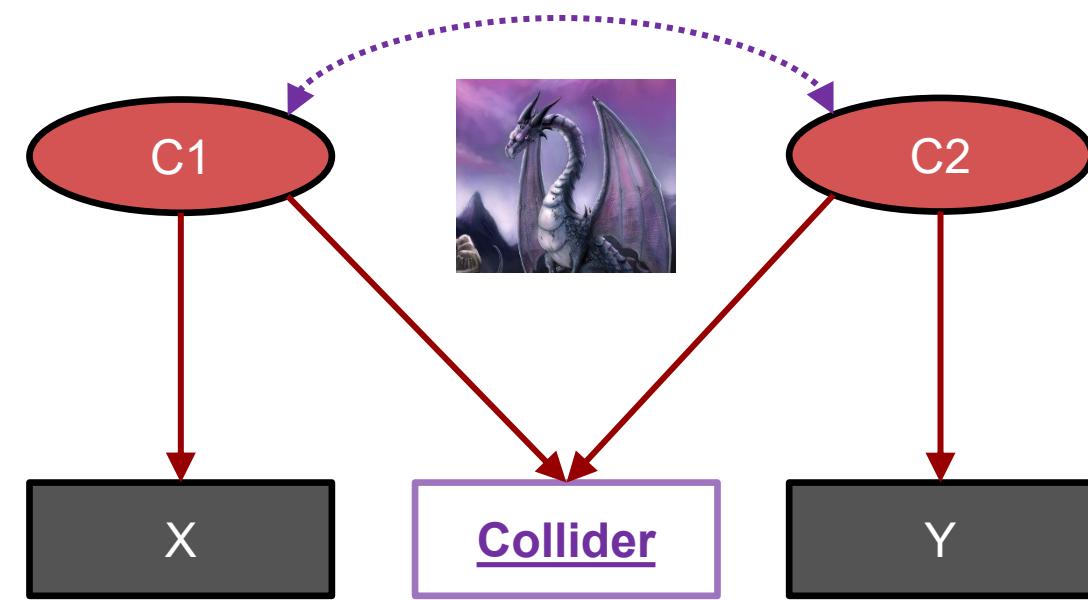
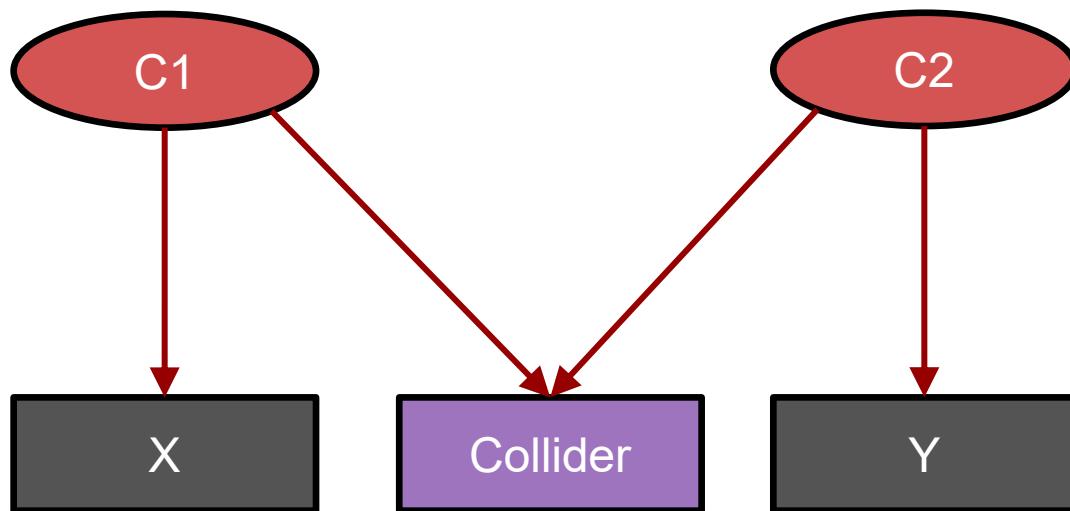
Conditioning on weight opens **Diet** <...> **Exercise** → **Diabetes**

		Traditional diet		Exotic diet		RR
		No diabetes	Diabetes	No diabetes	Diabetes	
<40kg/m ²	19	1	37	3	1.50	
≥40kg/m ²	28	12	12	8	1.33	
Total	47	13	49	11	0.85	

TRANSMITTING COLLIDER BIAS

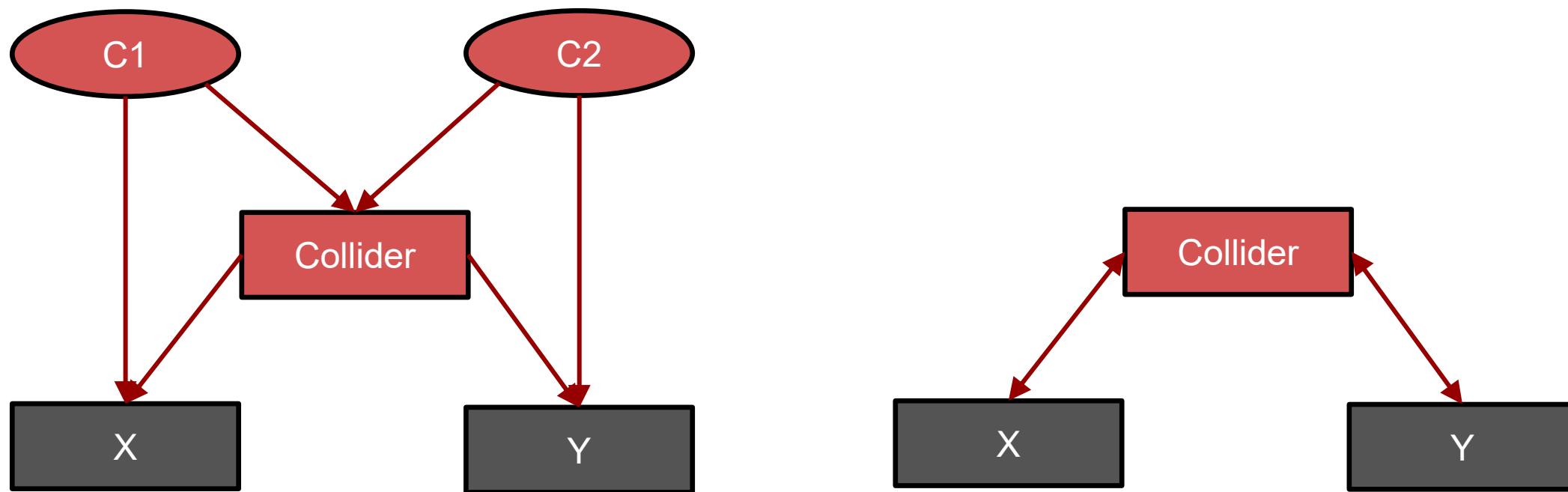
Conditional dependencies can be **transmitted** in this way between **all variables on all open paths** either side of the collider

In this way, a non-causal association can be transmitted between two variables that may themselves have no causal relationship with the collider!



This scenario is known as **M-bias**

Without knowledge of the unobserved confounders (**C1,C2**), this collider might appear as a harmless confounder...



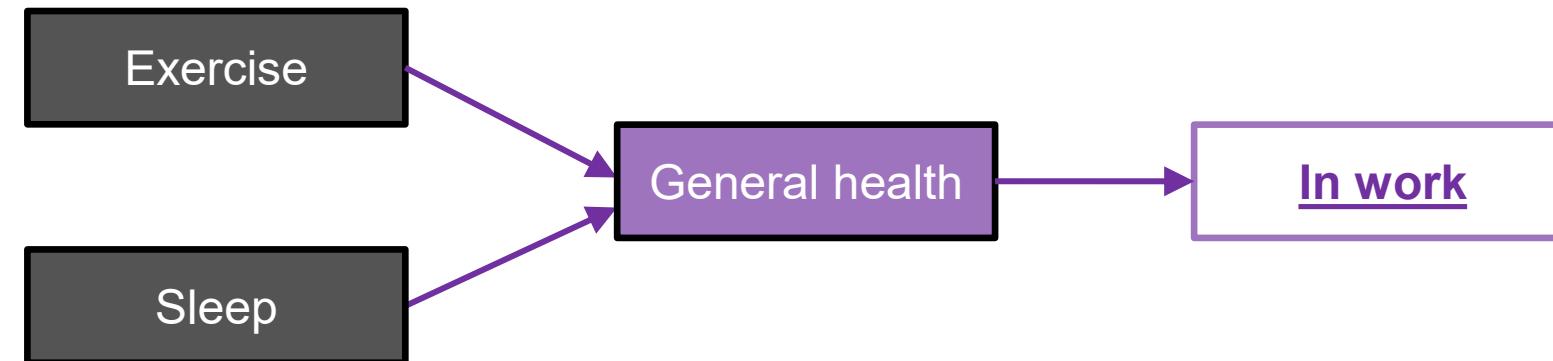
Hence it's vital to consider **all relevant variables** in your DAG

TRANSMITTING COLLIDER BIAS

The collider also needn't be directly conditioned on

- Conditioning on a **descendent** is equivalent to partial conditioning

e.g.



is equivalent to:



TRANSMITTING COLLIDER BIAS

Why?



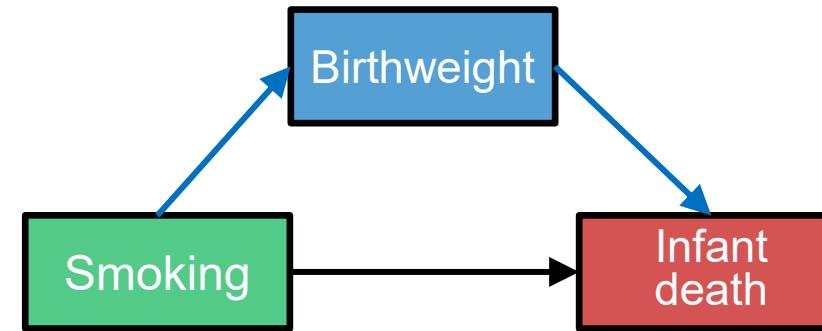
If participants are all in work, they are likely to have reasonable health, as maintaining employment is increasingly difficult with poor health

EXAMPLE: THE BIRTHWEIGHT PARADOX

One of the most well-known examples of **collider bias / reversal paradox** arising from **mediator conditioning** is the **Birthweight Paradox**

Context:

- **smoking** causes **infant death**
- **smoking** causes **lower birthweight**
- **lower birthweight** causes **infant death**



THE BIRTHWEIGHT PARADOX

The ‘paradox’ was first observed in 1971 by Yerushalmy

Yerushalmy was interested in the effect of **smoking** on **infant death**, stratified by ethnicity and **birthweight**



J Yerushalmy 1971. *Am J Epidemiol*; 93(6): 443-56.

EXAMPLE: THE BIRTHWEIGHT PARADOX

Yerushalmy analysed the rate of **neonatal death** in **low birth weight infants** (restricted), by maternal ethnicity and **smoking status**

Yerushalmy reported a ‘puzzling’ finding

The rate of **neonatal death** in **low birthweight infants** was *lower* if the mother had *smoked* than not

Smoking Status	Single liveborn infants weighing ≤ 2500 gm at birth					
	No		Neonatal Deaths			
	White	Black	White	Black	White	Rate per 1000
All mothers	434	261	70	41	161.3	157.1
Nonsmoker	197	129	43	26	218.3*	201.6
Smoker	237	132	27	15	113.9*	113.6
Nonsmoker:						
Never smoked	134	106	28	22	209.0	207.5
Past smoker	63	23	15	4	238.1	173.9

EXAMPLE: THE BIRTHWEIGHT PARADOX

Could this be a survival artefact?

- If stillbirths were more common in smokers, maybe LBW babies that survived to birth were healthier?
- Yerushalmy analysed the rate of neonatal death – by maternal smoking status – in all categories of birthweight (stratified) ...

Birth weight (grams)	Neonatal mortality rates per 1000							
	White		Black		Rate ratio			
	Nonsmoker	Smoker	Nonsmoker	Smoker	White	Black		
≤1500	791.7	565.2	740.7	666.7	0.71	0.90		
1501–2000	406.3	346.2	166.7	45.5	0.85	0.27		
2001–2500	78.0	26.6	25.6	21.7	0.34	0.85		
2501–3000	11.6	6.1	4.2	12.6	0.53	3.00		
3001–3500	2.2	4.5	4.3	4.8	2.05	1.12		
3501+	3.8	2.6	8.7	9.8	0.68	1.13		
Total	11.0	11.3	17.1	21.5	1.03	1.26		

EXAMPLE: THE BIRTHWEIGHT PARADOX

Yulshamy concludes ...

the phenomenon of smoking and its relation to health is complex and calls for continued exploration and vigorous and energetic investigations. It calls primarily for the development of study designs and methods which would overcome some of the inherent difficulties in investigations on humans.

EXAMPLE: THE BIRTHWEIGHT PARADOX

40 years of head scratching

On the importance—and the unimportance—of birthweight

Allen J Wilcox

Birthweight is one of the most accessible and most misunderstood variables in epidemiology. A baby's weight at birth is strongly associated with mortality risk during the first year and, to a lesser degree, with developmental problems in childhood and the risk of various diseases in adulthood. Epidemiological analyses often regard birthweight as on the causal pathway to these health outcomes. Under this assumption of causality, birthweight is used to explain variations in infant mortality and later morbidity, and is also used as an intermediate health endpoint in itself. Evidence presented here suggests the link between birthweight and health outcomes may not be causal. Methods of analysis that assume causality are unreliable at best, and biased at worst. The category of 'low birthweight' in particular is uninformative and seldom justified. The main utility of the birthweight distribution is to provide an estimate of the proportion of small preterm births in a population (although even this requires special analytical methods). While the ordinary approaches to birthweight are not well grounded, the links between birthweight and a range of health outcomes may nonetheless reflect the workings of biological mechanisms with implications for human health.

Birthweight, fetal growth, gestational age, low birthweight, intrauterine growth retardation, small for gestational age, infant mortality, analytical methods

30 July 2001

Is it time to abandon adjustment for birth weight in studies of infant mortality?

Irva Hertz-Pannier

University of California, Davis, CA, USA

Birthweight and its relationship to adverse perinatal outcomes has been the subject of a great many public health investigations. Sometimes birthweight is considered an outcome in its own right, and other times it is used as an indicator of 'risk', e.g. for mortality. This latter use of birthweight generated a critique,^{1–3} methodological research^{2–5} and continued controversy.^{6–10} Whether or not low birthweight merits being seen as an adverse outcome worthy of study, its use as an indicator of increased risk of mortality is of high practical concern, particularly in the field of quality of care.

In this context, Joyce and Peacock¹¹ attempt to eval-

health services or about potential environmental exposures, these comparisons across ethnic groups or geographical areas need to be unconfounded with respect to background risk or pre-existing health status.

Those of us who might endeavour to make these comparisons are confronted by two difficulties: one is a conceptual/philosophical question, and the other is a matter of technique. First, why adjust for birthweight? Secondly, if birthweight adjustment is the correct way to proceed, then how should one do it?

The philosophical issue is more fundamental. In an effort to achieve a valid comparison of perinatal mor-

EXAMPLE: THE BIRTHWEIGHT PARADOX

2006: Hernan & colleagues explain the ‘paradox’

- Birthweight is a **collider** for **other (serious) unobserved causes of infant death**

The Birth Weight “Paradox” Uncovered?

Sonia Hernández-Díaz^{1,2}, Enrique F. Schisterman³, and Miguel A. Hernán¹

¹ Department of Epidemiology, Harvard School of Public Health, Boston, MA.

² Slone Epidemiology Center, Boston University, Boston, MA.

³ Epidemiology Branch, National Institute of Child Health and Human Development, Bethesda, MD.

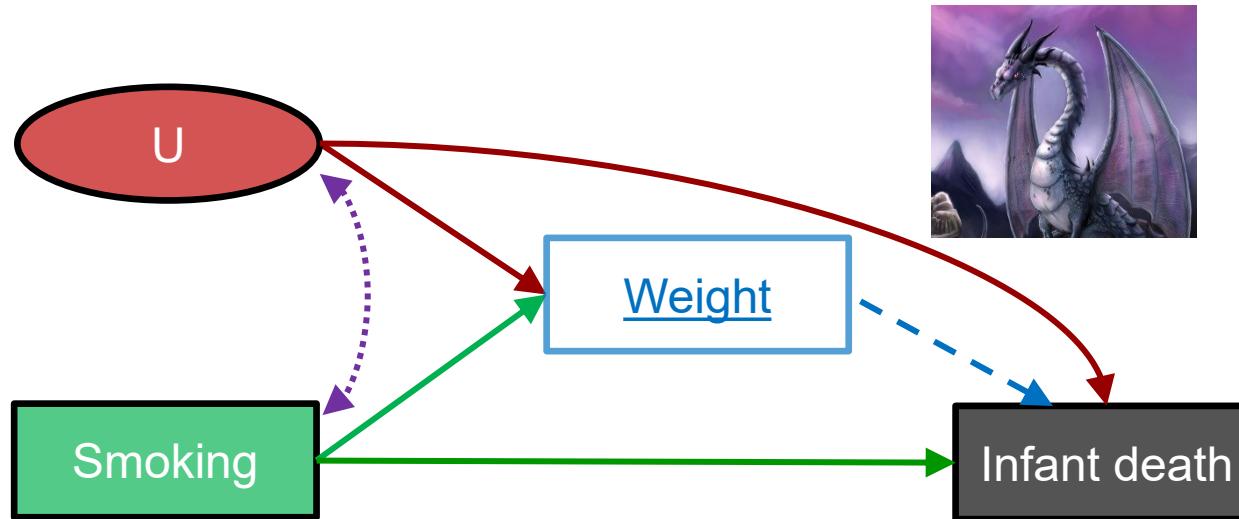
Received for publication February 7, 2005; accepted for publication January 23, 2006.

Low birth weight (LBW) infants have lower infant mortality in groups in which LBW is most frequent. For example, in 1991, US infants born to smokers had higher risks of both LBW and infant mortality than infants born to non-smokers. However, among LBW infants, infant mortality was lower for infants born to smokers (relative rate = 0.79). There are competing theories regarding this so-called “paradox.” One is that maternal smoking is beneficial for LBW infants. The authors use causal diagrams to show that, even in the absence of any beneficial effect of smoking, an inverse association due to stratification on birth weight can be found. This variable is affected by the exposure of interest and shares common causes with the outcome. That is, LBW infants born to smokers may have a lower risk of mortality than other LBW infants whose LBW is due to causes associated with high mortality (e.g., birth defects). Under realistic causal diagrams, adjustment for birth weight is unwarranted when the analytical goal is to estimate overall effects of prenatal variables on infant mortality. Even for estimating direct effects of prenatal variables, adjustment for birth weight may be invalid when there is an unmeasured common cause of LBW and mortality. An appropriate justification for conditioning on birth weight requires specifying 1) the causal question motivating this analytical approach and 2) the assumptions regarding the proposed underlying biologic mechanisms.

EXAMPLE: THE BIRTHWEIGHT PARADOX

Smoking is a *weak cause* of infant death

The **unobserved factors** are *strong causes* of infant death



Low birth weight in the absence of **smoking** implies more serious reasons

Conditioning on LBW opens **Smoking** \longleftrightarrow U \rightarrow **Infant death**

EXAMPLE: THE BIRTHWEIGHT PARADOX

Yerushalmy's explanation ...

Among the interpretations which imply a *causative* hypothesis, perhaps the most attractive is that mother's smoking causes a displacement from higher to lower birth weights, and this displacement, or shift, is responsible for the observed excess of infants of "low-birth-weight" and for their favorable early mortality. It may, therefore, be of interest to review in greater detail the consequences of this hypothesis to determine whether they are reasonable.

REVERSAL NOT NECESSARY

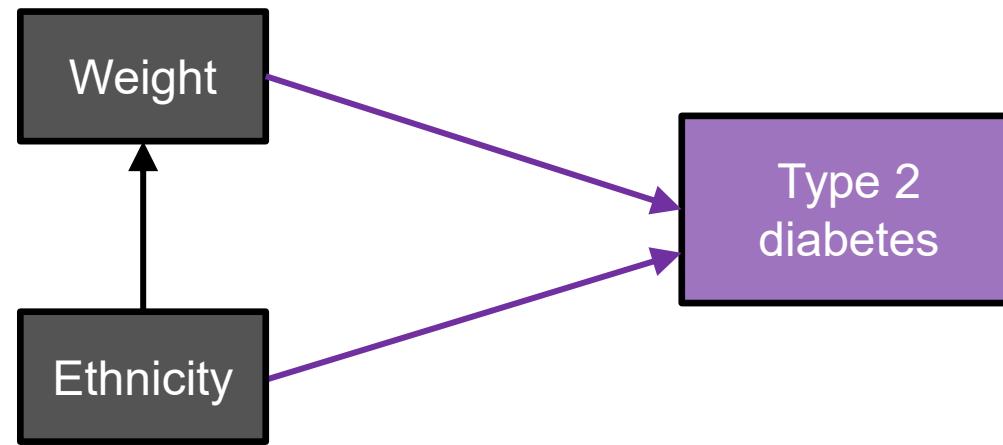
Note: Collider bias needn't cause the association to reverse

- Reversal paradox examples are some of the best known because they are the most visible
- People have been trying to understand the **Birthweight paradox** for 40-years
- Modest collider biases are likely unnoticed
- The nature and extent of bias will depend on the underlying relationships between the variables involved
- Don't assume it's OK to condition

REVERSAL NOT NECESSARY

Example: weight, ethnicity, and type 2 diabetes

- Obesity and South Asian ethnicity are both associated with higher risks of diabetes

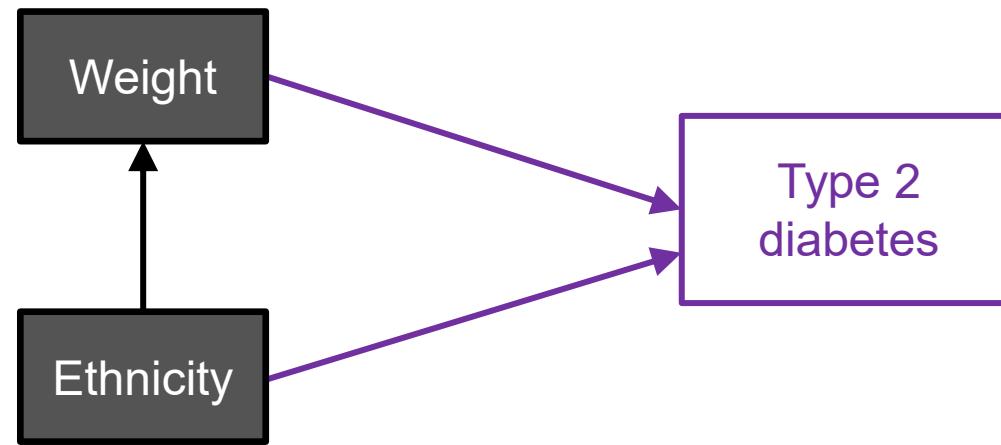


- In general UK population, **average weight is lower in South Asians** than White

REVERSAL NOT NECESSARY

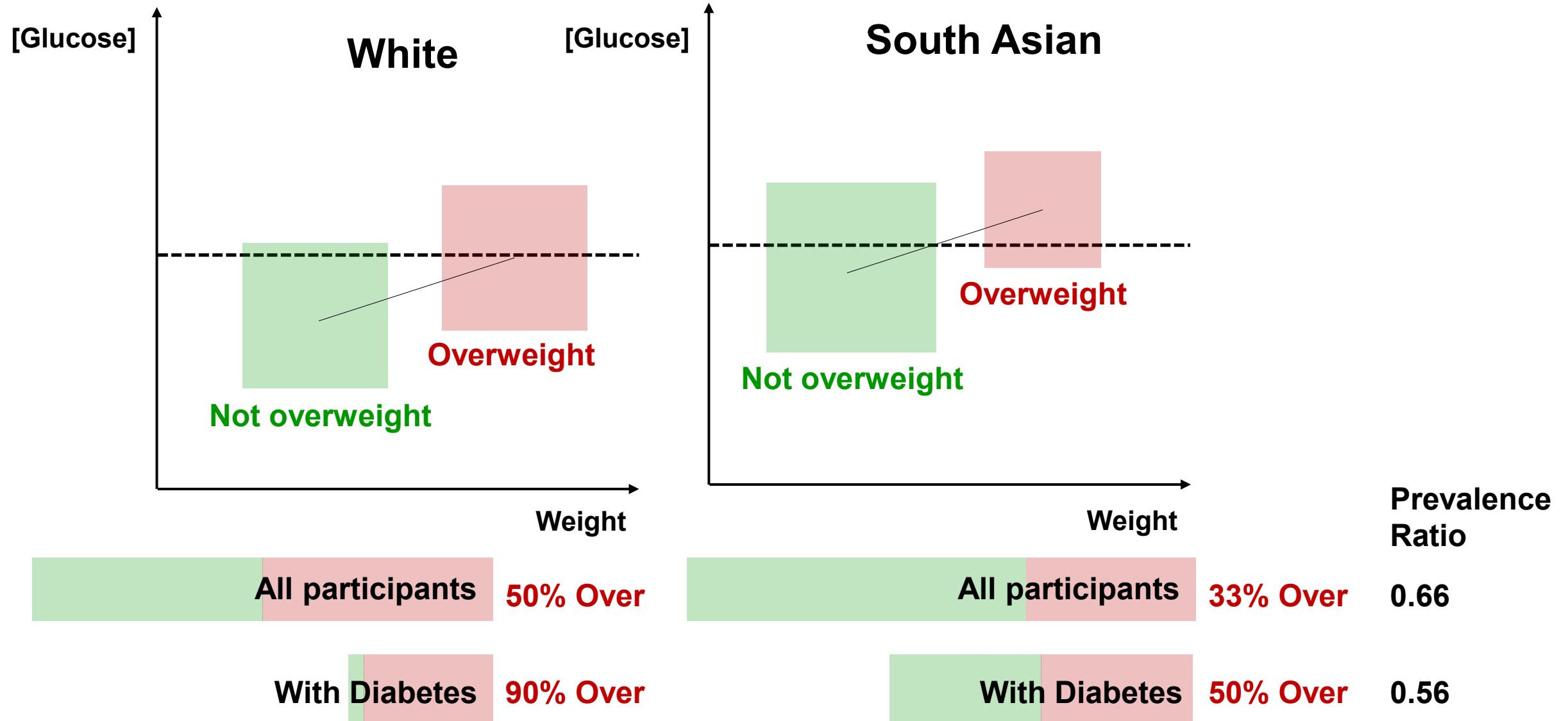
Example: weight, ethnicity, and type 2 diabetes

- But ... if the relationship was studied in people with diabetes



- **Ethnicity <...> Weight** is opened, and difference in weight between South Asian and White appears *larger*

REVERSAL NOT NECESSARY



RECOMMENDED READING

- Elwert, F. and Winship, C., 2014. Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology*, 40, pp.31-53.
- Cole, S.R., Platt, R.W., Schisterman, E.F., Chu, H., Westreich, D., Richardson, D. and Poole, C., 2010. Illustrating bias due to conditioning on a collider. *International journal of epidemiology*, 39(2), pp.417-420.
- Holmberg, M.J. and Andersen, L.W., 2022. Collider Bias. *JAMA*, 327(13), pp.1282-1283.

SUMMARY

- **Collider bias** is a common form of bias that occurs due to opening non-causal paths by **inappropriate** and/or **inadvertent conditioning**
- It is challenging to understand and almost impossible to grasp intuitively
- Impacts can be dramatic and/or confusing, giving rise to many '*paradoxes*'
- Impacts needn't however be dramatic; every observational study is likely complicated by some level of collider bias
- Thoughtful and conscientious use of DAGs are the best way to detect, avoid, and mitigate against examples of collider bias
- With familiarity, you may start seeing examples all over your discipline !

09:30-10:15 ACTIVITY 3-A

10:15-11:00 LECTURE 3.1

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 3.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 3.3

14:45-15:30 ACTIVITY 3-B

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 3.4

17:00-17:45 ACTIVITY 3-C

17:45-18:00 Q&A

3.2 - SELECTION BIAS

GEORGIA



MARK



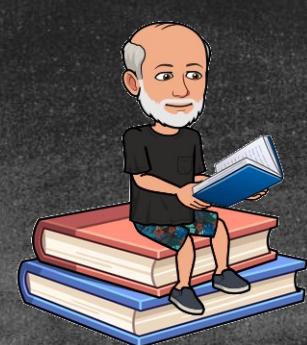
@GEORGIATOMOVA

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

LEARNING OBJECTIVES

By the end of this session, you will be able to:

- Define (**differential selection bias**)
- Define **informed presence bias**
- Explain how **sampling** and **participation** lead to **collider bias**
- Explain why such biases are especially common in retrospective data
- Explain how **M-bias** can mean that even **prospective studies** are not immune
- Suggest solutions to differential selection bias

You may also need to:

- Rest your sore head

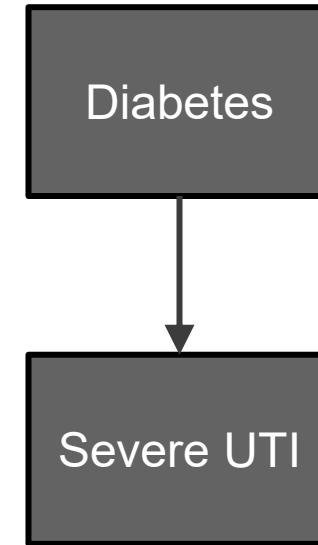
RECAP: COLLIDER BIAS

- **Collider bias** is a form of **non-causal association** created by conditioning on a **collider** that can cause **mind-boggling paradoxes**
- It commonly affects adjusted multivariable regression models, when several variables are included and their coefficients interpreted (**Table 2 Fallacy**)
- This can be avoided by appropriately selecting the adjustment set for each exposure-outcome relationship
- Unfortunately, there is another circumstances where collider bias commonly occurs, and the solutions are rather more difficult ...
- There are nevertheless still big benefits to identifying collider bias, for improved transparency, more accurate interpretation, and potentially seeking solutions or conducting sensitivity analyses

EXAMPLE: BERKSON'S BIAS

Example: Berkson's paradox

- Diabetes causes frequent and **severe urinary tract infections**
- We want to understand how much more common severe UTIs are in people with **diabetes** than those without

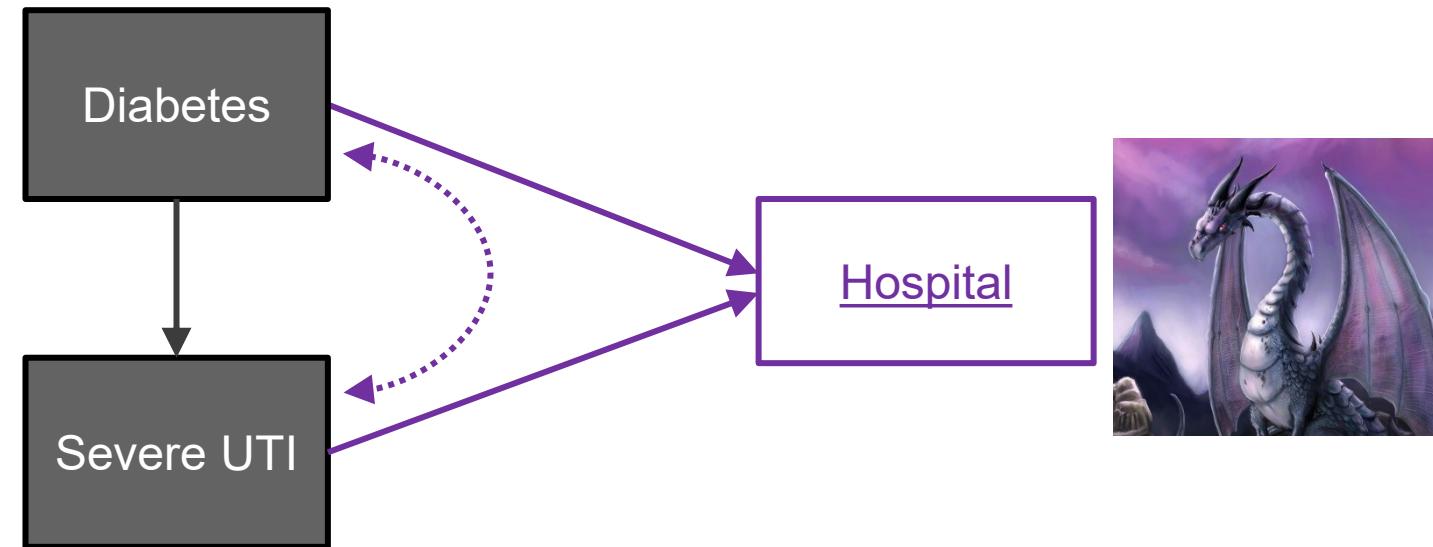


- Embracing '**big data**' we decide to estimate this in a **hospital admission data**

EXAMPLE: BERKSON'S BIAS

Example: Berkson's paradox

- Diabetes complications and severe UTIs are two competing reasons for attending hospital
- By looking in hospital records, we open Diabetes <…> UTI



- In this scenario, diabetes appears **protective** of **severe UTI** – hence the **paradox**

SELECTION BIAS

Berkson's paradox is a classic example of **collider bias** where the collider is selection into the study

Because a sample is restricted to include only those who participate then selection is implicitly conditioned creating **(differential) selection bias**

Selection bias can arise from:

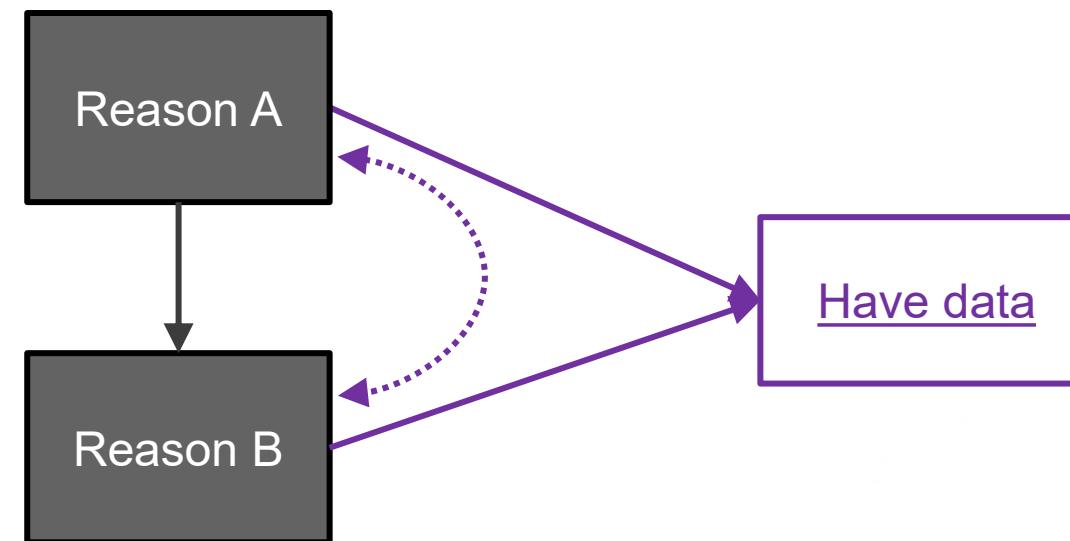
- the **sampling strategy** - i.e. how have the participants been identified?
- **differential participation** - i.e. different sorts of people agree to participate
- **differential attrition** – i.e. different sorts of people drop out of your study
- **(missing data)** – i.e. different sorts of people provide complete data creating bias if you only analyse those with complete data

INFORMED PRESENCE BIAS

In '**big data**' contexts, we may not think of people being 'selected' into a dataset, as we just collect everything available

But there will always be *reasons* why we have (more) data for some people than we do others, these reasons will collide at data collection !

This is sometimes known as **informed presence bias**

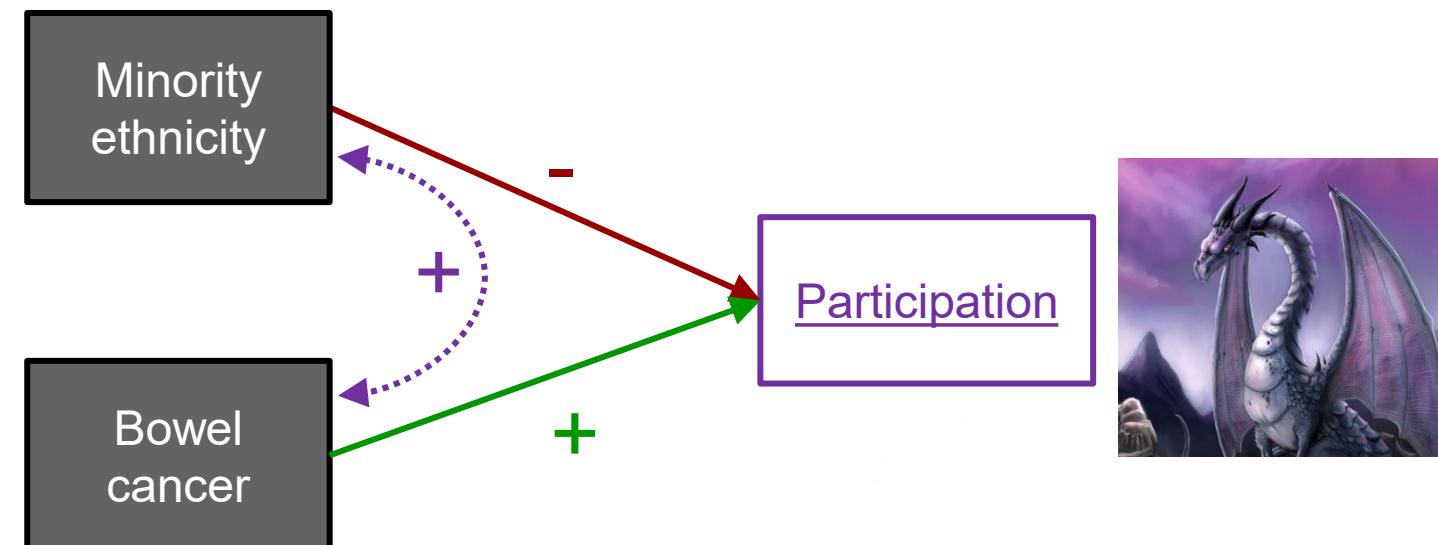


COLLIDER BIAS IN RETROSPECTIVE STUDIES

Selection bias is a particular threat in **case-control studies**, because cases are often (always!) more interested in participating

e.g. Ethnicity and bowel cancer

- Minority ethnic groups are **less likely** to participate
- Bowel cancer cases are **more likely** to participate
- Minority ethnic groups are **spuriously associated** with higher risk of bowel cancer !



SOURCES OF PARTICIPATION BIAS

Willingness and ability to participate varies by:

- **Health**
 - ✓ Physical or psychological
- **Education**
 - ✓ Interest, curiosity etc.
- **Beliefs**
 - ✓ Religious, spiritual, political, etc.
- **Psychology & personality**
 - ✓ Self-efficacy, openness, scepticism, etc
- **Economics**
 - ✓ Time, cost



SOURCES OF PARTICIPATION BIAS

Willingness to volunteer (**self-selection bias**), respond to invitations (**non-response bias**), provide consent (**consent bias**), remain involved in a study (**attrition bias**), and participate completely (**missing data bias**) are generally lower in those who are:

- male
- older
- poorer
- less educated
- in poorer health
- from an ethnic minority
- marginalised (e.g. migrants, asylum seekers, travellers)

Except: In **retrospective studies, cases** (i.e. with outcome) are more likely to participate than **non-cases** (i.e. without outcome)

COLLIDER BIAS IN PROSPECTIVE STUDIES

What about **prospective studies**?

Example: Biobank UK



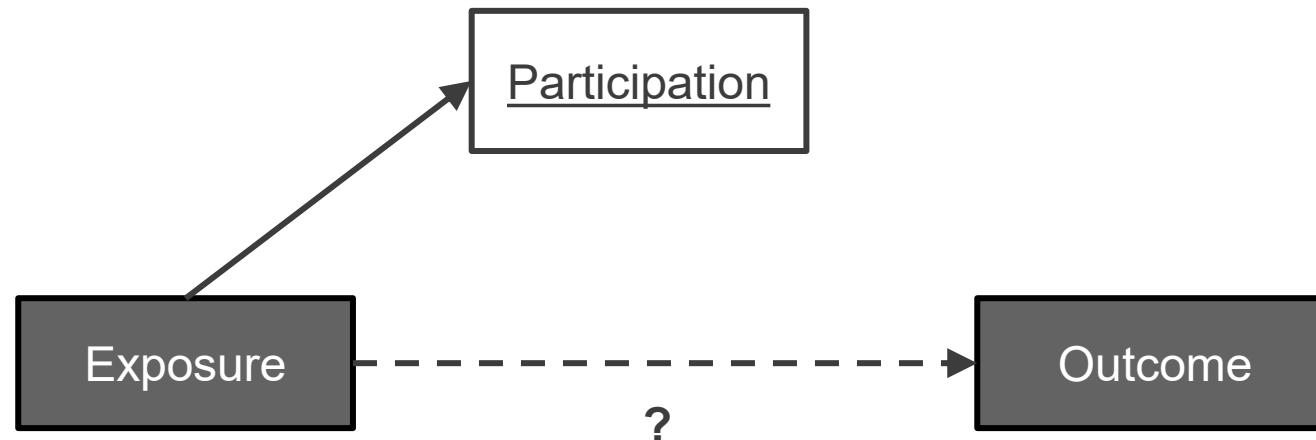
UK Biobank is not representative of the general population on a variety of sociodemographic, physical, lifestyle and health-related characteristics, with evidence of a ‘healthy volunteer’ selection bias. As a result, UK Biobank is not a suitable resource for deriving generalizable disease prevalence and incidence rates. However, the large sample size and heterogeneity of exposure measures allow for valid scientific inferences of associations between exposures and health outcomes that are generalizable to the wider population.

We advise that, where appropriate, publications that use UK Biobank data include a statement clarifying that “while UK Biobank participants are not representative of the general population (and hence cannot be used to provide representative disease prevalence and incidence rates), valid assessment of exposure-disease relationships are nonetheless widely generalizable and do not require participants to be representative of the population at large.”

COLLIDER BIAS IN PROSPECTIVE STUDIES

In **prospective observational studies**, participation precedes the outcome

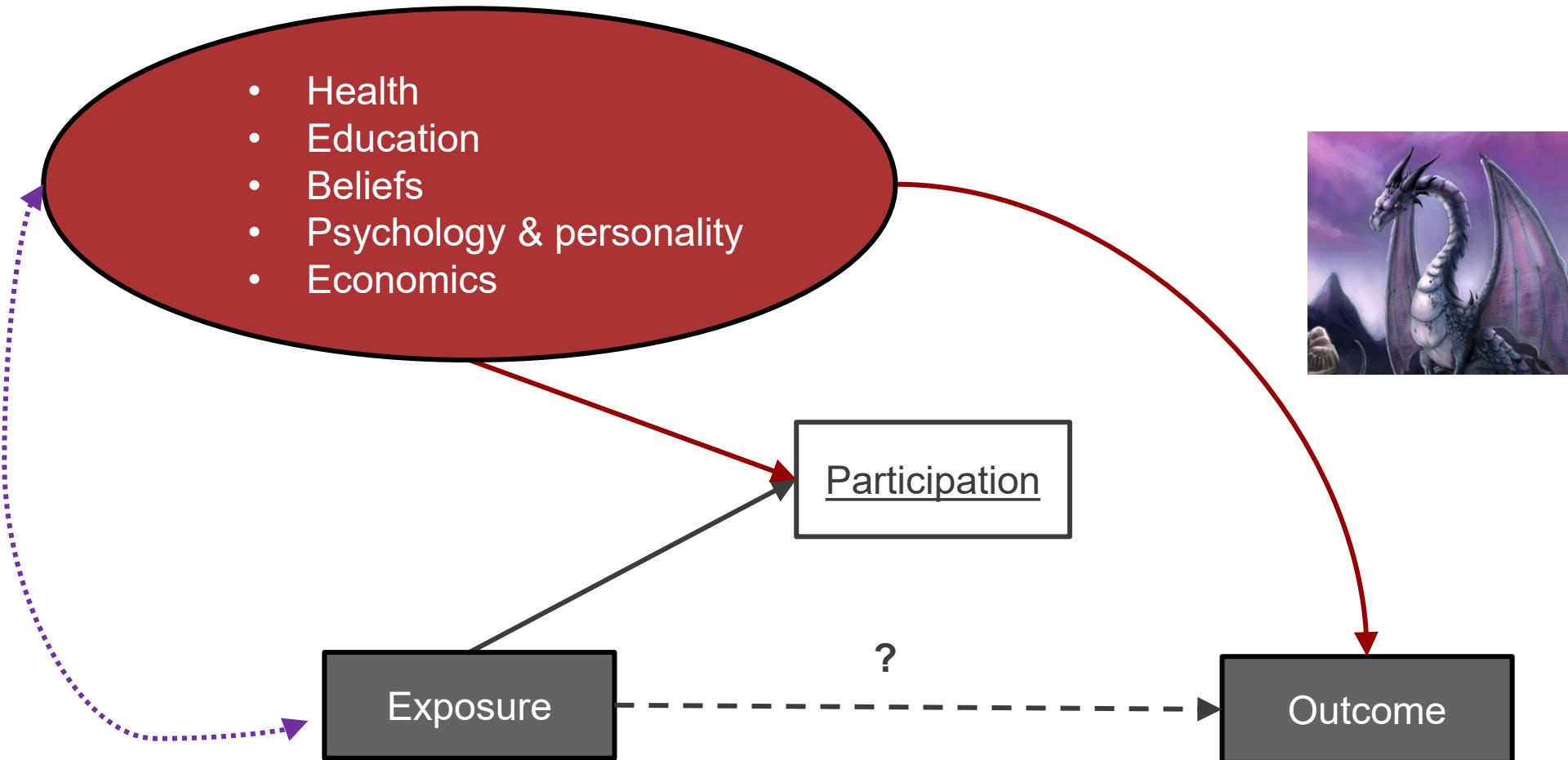
If the exposure precedes participation, we should assume it is a potential cause



COLLIDER BIAS IN PROSPECTIVE STUDIES

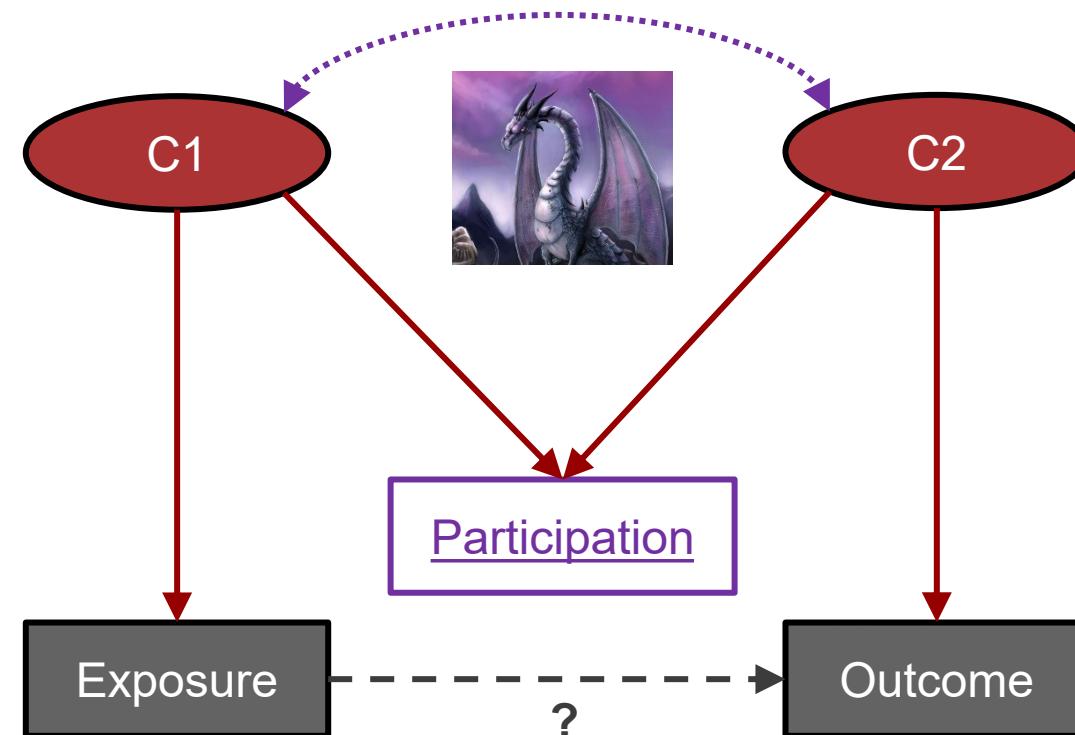
We've just said there are **many determinants of participation**

If any of these also cause the outcome...



In fact, the exposure needn't cause participation...

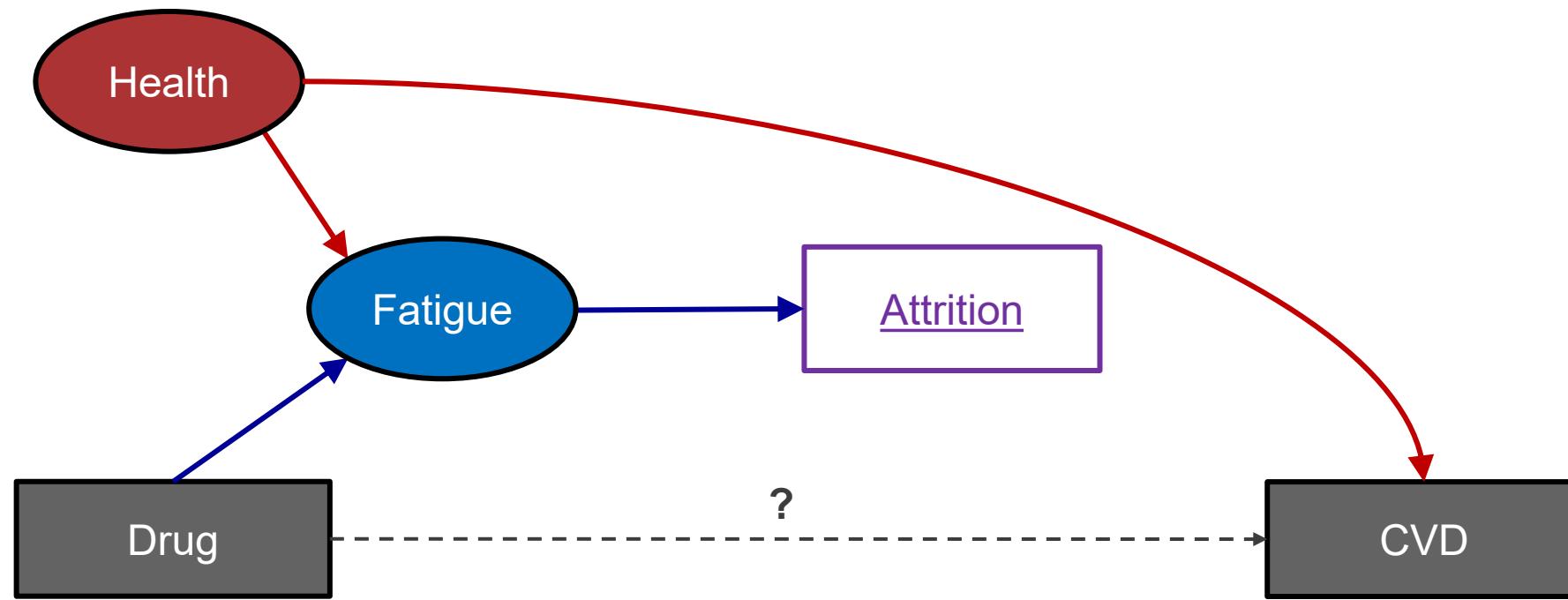
If **C1** or **C2** include any of **health, education, beliefs, psychology & personality**, or **economics** then there's the potential for collider bias



COLLIDER BIAS IN RANDOMISED EXPERIMENTS

Even randomised experiments are not immune if assignment causes attrition

- e.g. drug for cardiovascular disease

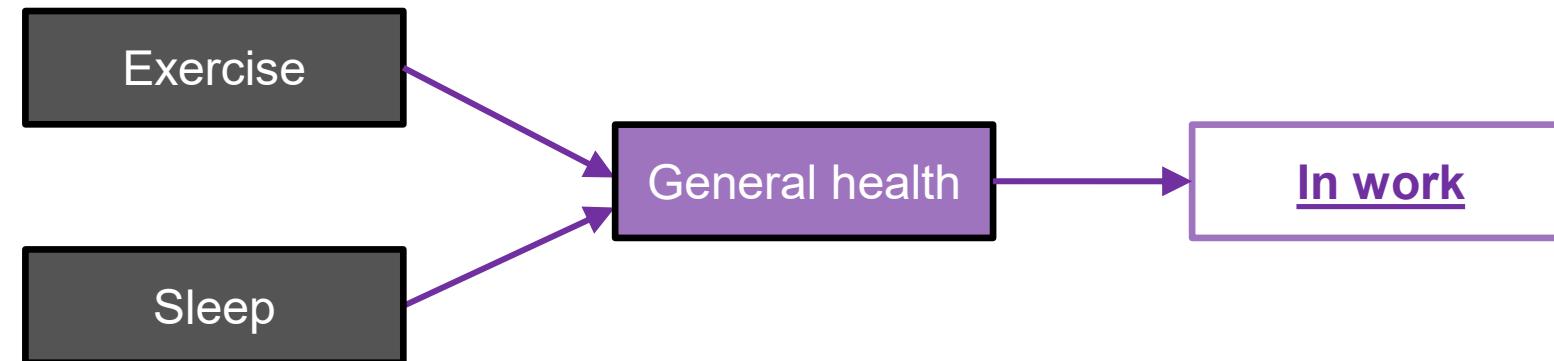


COLLIDER BIAS IN RANDOMISED EXPERIMENTS

Recall: A collider needn't be directly conditioned on

- **Conditioning on a descendant** is equivalent to conditioning directly

e.g.



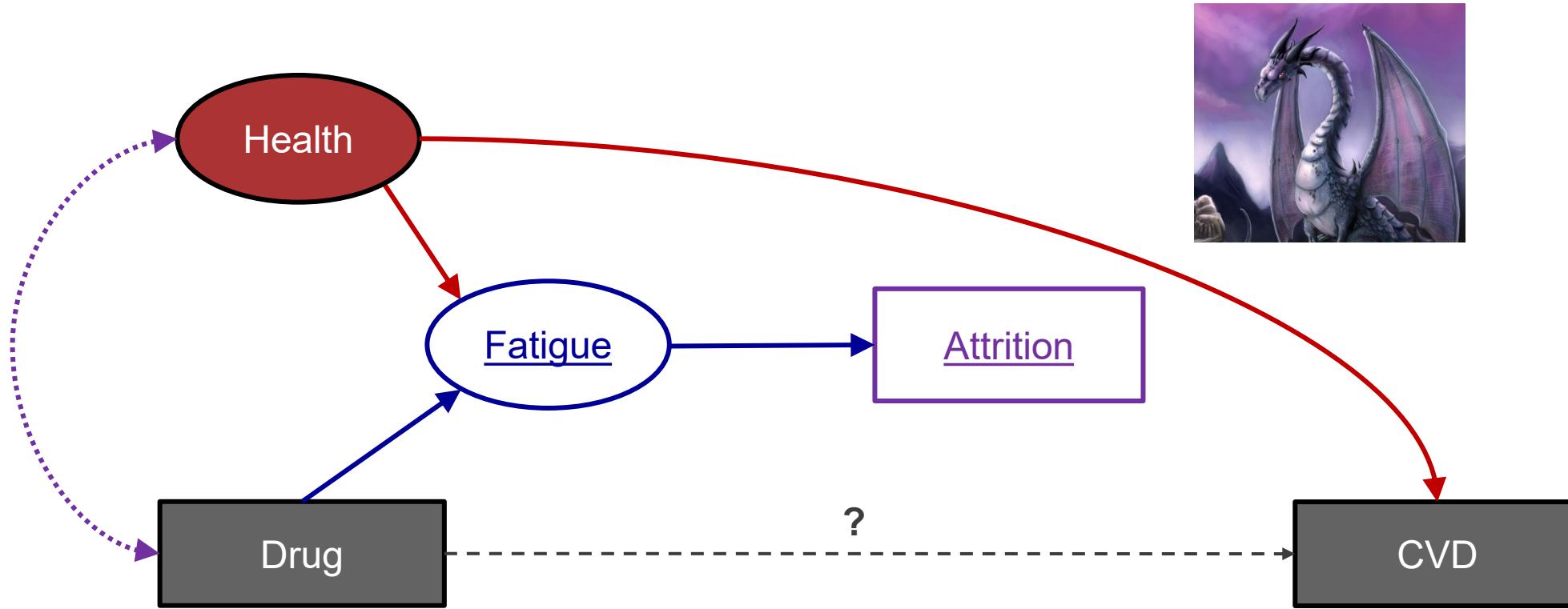
is equivalent to:



COLLIDER BIAS IN RANDOMISED EXPERIMENTS

Even randomised experiments are not immune if **assignment** causes **attrition**

- e.g. drug for cardiovascular disease



COLLIDER BIAS FROM DIFFERENTIAL SELECTION

If:

- **Exposure & participation** are correlated in population

And:

- **Outcome & participation** are correlated in population

Then:

- Your estimate of the **exposure-outcome relationship** will likely be biased

Thus:

- **Collider bias from differential selection** is probably ubiquitous
- Very unrepresentative samples are likely the worst affected

SELECTION BIAS IN PROSPECTIVE STUDIES

Given this...

Do you agree with this statement?



UK Biobank is not representative of the general population on a variety of sociodemographic, physical, lifestyle and health-related characteristics, with evidence of a 'healthy volunteer' selection bias. As a result, UK Biobank is not a suitable resource for deriving generalizable disease prevalence and incidence rates. However, the large sample size and heterogeneity of exposure measures allow for valid scientific inferences of associations between exposures and health outcomes that are generalizable to the wider population.

We advise that, where appropriate, publications that use UK Biobank data include a statement clarifying that "while UK Biobank participants are not representative of the general population (and hence cannot be used to provide representative disease prevalence and incidence rates), valid assessment of exposure-disease relationships are nonetheless widely generalizable and do not require participants to be representative of the population at large."

1000+ studies indexed on Pubmed

£100m+ has been spent on the cohort

The screenshot shows the UK Biobank homepage. At the top, there is a navigation bar with links for "Researcher log in" (purple), "Participant log in" (orange), and "Contact us" (grey). Below the navigation bar, there are three buttons: "Enable your research" (purple), "Explore your participation" (orange), and "Learn more about UK Biobank" (grey) with a magnifying glass icon. The main content area features a large image of a robotic arm moving between rows of sample tubes in a storage facility. To the left of this image, a callout box contains the text "Enabling your vision to improve public health" and a paragraph about the database's purpose. It also includes two buttons: "Data Showcase" and "Future data releases". Below this section, a larger text block provides a detailed overview of what UK Biobank is and its impact.

Enabling your vision to improve public health

Data drives discovery. We have curated a uniquely powerful biomedical database that can be accessed globally for public health research. Explore data from half a million UK Biobank participants to enable new discoveries to improve public health.

Data Showcase **Future data releases**

UK Biobank is a large-scale biomedical database and research resource, containing in-depth genetic and health information from half a million UK participants. The database is regularly augmented with additional data and is globally accessible to approved researchers undertaking vital research into the most common and life-threatening diseases. It is a major contributor to the advancement of modern medicine and treatment and has enabled several scientific discoveries that improve human health.

<http://www.ukbiobank.ac.uk>

Example: Kyle et al 2017 *Sleep Medicine* 38:85-91

Examines sleep and cognitive performance



Contents lists available at [ScienceDirect](#)

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep



Original Article

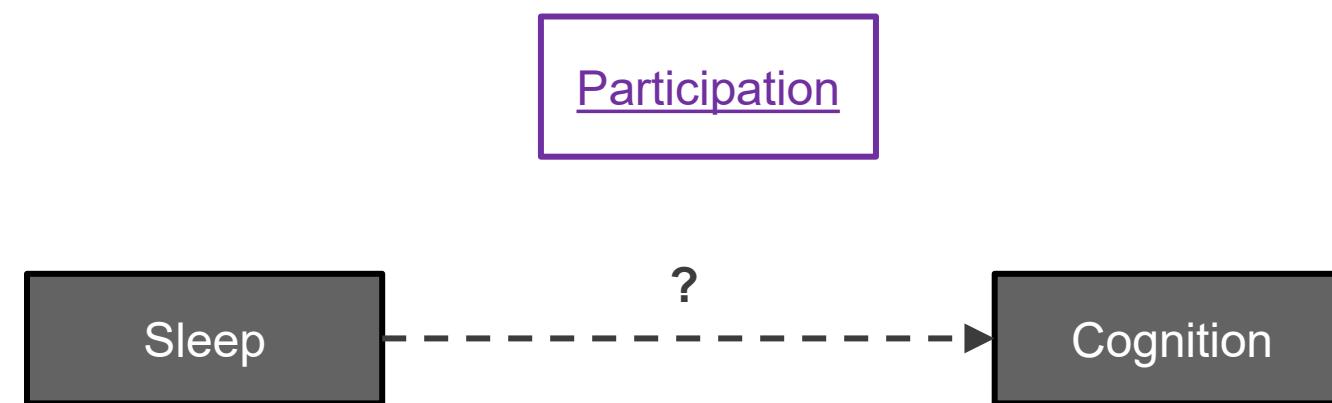
Sleep and cognitive performance: cross-sectional associations in the UK Biobank



Simon D. Kyle ^{a,*}, Claire E. Sexton ^b, Bernd Feige ^c, Annemarie I. Luik ^a, Jacqueline Lane ^{d,e,f}, Richa Saxena ^{d,e,f,g}, Simon G. Anderson ^h, David A. Bechtold ⁱ, William Dixon ^j, Max A. Little ^{k,l}, David Ray ^m, Dieter Riemann ^c, Colin A. Espie ^a, Martin K. Rutter ^{m,n}, Kai Spiegelhalder ^c

Example: Kyle et al 2017 *Sleep Medicine* 38:85-91

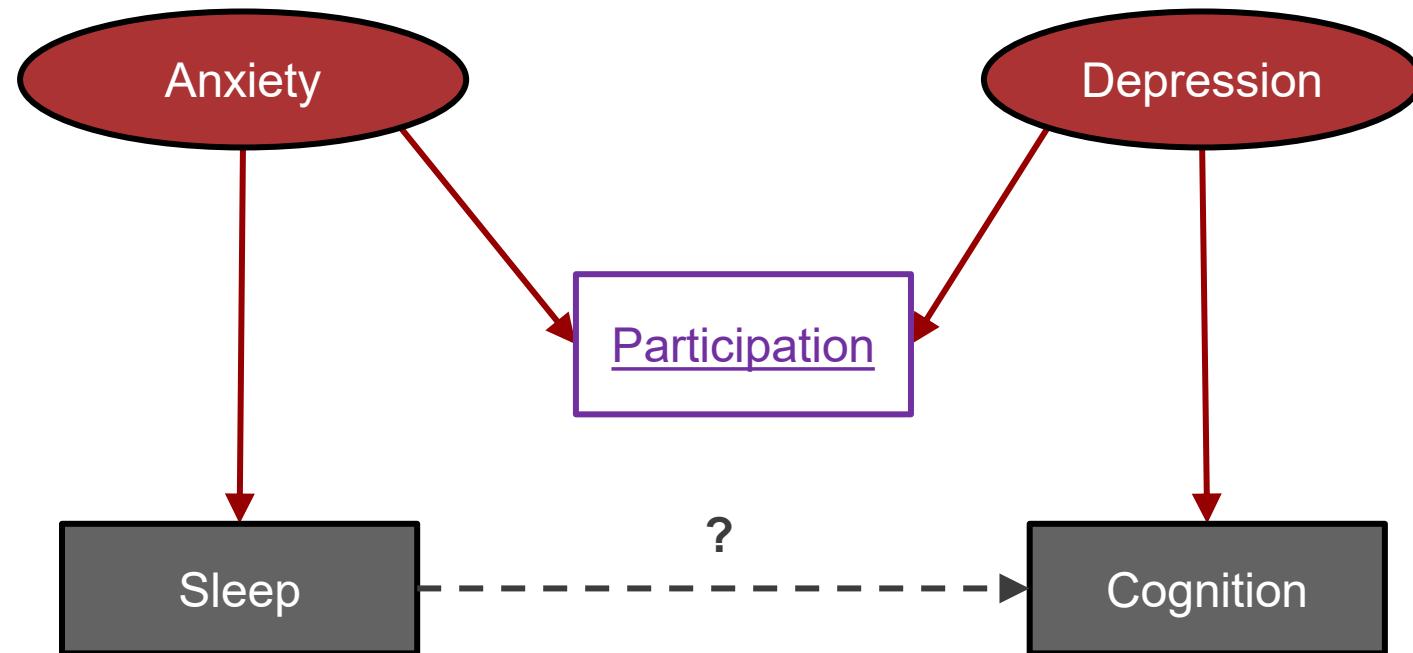
Examines **sleep** and **cognitive performance**



Example: Kyle et al 2017 *Sleep Medicine* 38:85-91

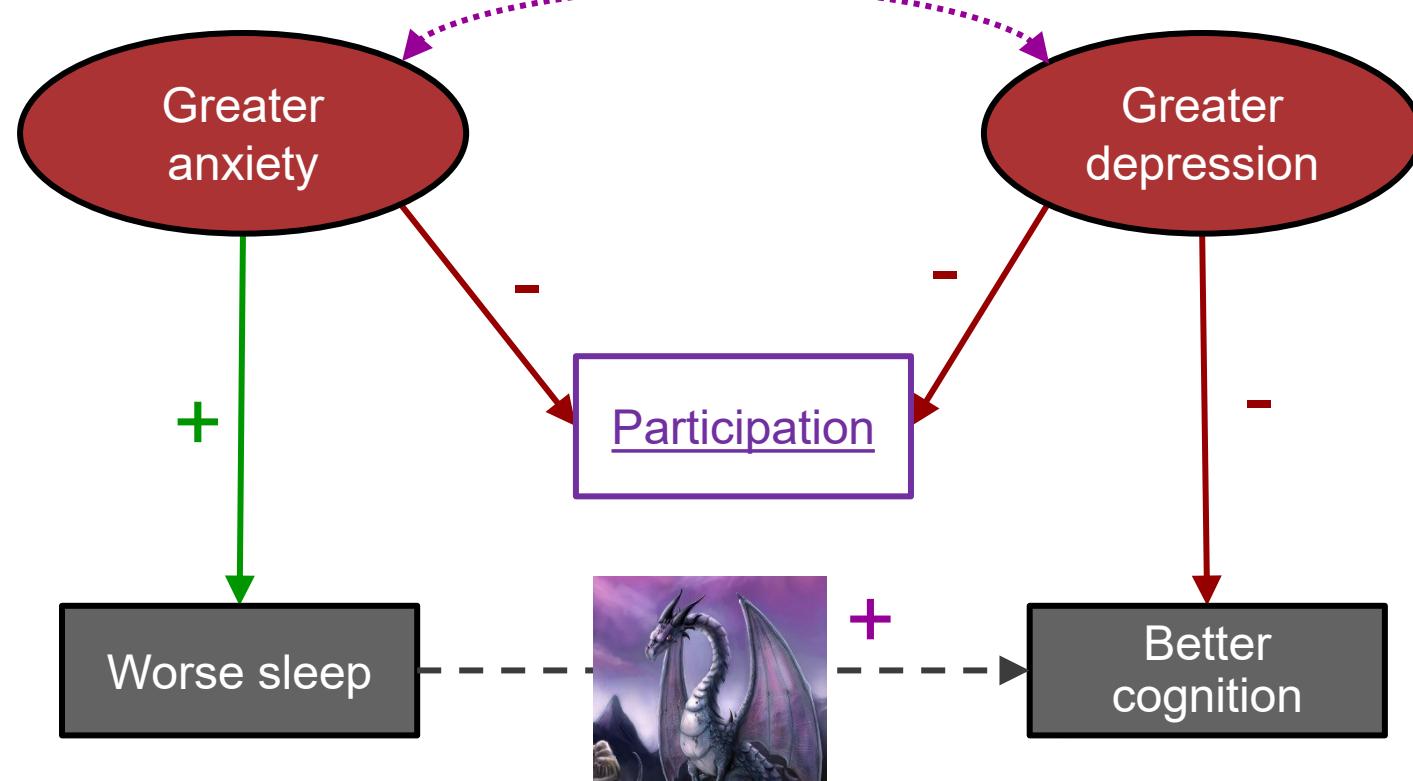
Examines **sleep** and **cognitive performance**

Seems plausible these would be affected by **anxiety** and **depression**



How might **anxiety** and **depression** effect **sleep** and **cognition**?

But if this was true, it would create a **bizarre association!**



Conclusions: Our results suggest that after adjustment for potential confounding variables, frequent insomnia symptoms may be associated with a small statistical advantage, which is unlikely to be clinically meaningful, on simple neurocognitive tasks. Further work is required to examine the mechanistic underpinnings of an apparent evening chronotype advantage in cognitive performance and the impairment associated with morning chronotype, sleep medication use, and sleep duration extremes.

Cognitive Domain	Model 1 [#]	Model 2 ^{##}	Model 3 ^{###}	Model 4 ^{####}
Reasoning	$\beta = -1.2 \times 10^{-1} ***$	$\beta = 1.9 \times 10^{-2}$	$\beta = 5.9 \times 10^{-2} ***$	$\beta = 9.5 \times 10^{-2} ***$
	(n = 158,180)	(n = 134,295)	(n = 115,935)	(n = 115,668)
Reaction Time	$\beta = 1.4 \times 10^{-2} ***$	$\beta = -2.2 \times 10^{-3} **$	$\beta = -5.9 \times 10^{-3} ***$	$\beta = -6.7 \times 10^{-3} ***$
	(n = 473,144)	(n = 386,588)	(n = 333,095)	(n = 332,303)
Numeric memory	$\beta = -1.0 \times 10^{-1} ***$	$\beta = -1.8 \times 10^{-2}$	$\beta = 4.9 \times 10^{-3}$	$\beta = 1.6 \times 10^{-2}$
	(n = 48,091)	(n = 40,151)	(n = 34,844)	(n = 34,753)
Visual Memory	$\beta = 1.4 \times 10^{-2} ***$	$\beta = -9.0 \times 10^{-3} **$	$\beta = -1.3 \times 10^{-2} ***$	$\beta = -1.4 \times 10^{-2} ***$
	(n = 473,955)	(n = 372,318)	(n = 334,779)	(n = 333,966)
Prospective Memory	$\beta = -2.7 \times 10^{-2}$	$\beta = 6.1 \times 10^{-2} ***$	$\beta = 1.1 \times 10^{-1} ***$	$\beta = 1.5 \times 10^{-1} ***$
	(n = 163,077)	(n = 136,770)	(n = 117,721)	(n = 117,438)

[#]unadjusted analyses;

^{##}adjusting for age, sex, socioeconomic status and education;

^{###}adjusting for age, sex, socioeconomic status, education, BMI, hypertension, cardiovascular disease, antihypertensive medication, depressive symptoms, psychotropic medication, sleep medication, chronotype;
^{####}adjusting for Model 3 variables plus sleep duration.



International Journal of Epidemiology, 2018, 226–235

doi: 10.1093/ije/dyx206

Advance Access Publication Date: 27 September 2017

Original article



Methods

Collider scope: when selection bias can substantially influence observed associations

Marcus R Munafò,^{1,2*} Kate Tilling,^{1,3} Amy E Taylor,^{1,2} David M Evans,^{1,4} and George Davey Smith^{1,3}

¹MRC Integrative Epidemiology Unit, ²UK Centre for Tobacco and Alcohol Studies, ³School of Social and Community Medicine, University of Bristol, Bristol, UK and ⁴University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD, Australia

*Corresponding author. School of Experimental Psychology, University of Bristol, Bristol BS8 1TU, UK. E-mail: marcus.munaf@bristol.ac.uk

Editorial decision 27 August 2017; Accepted 1 September 2017

“The magnitude of effects we observed in our simulations...are comparable to many reported associations derived from large but selected samples, such as between personality and cognitive function... a range of physical and mental health outcomes...and between chronotype (i.e. ‘morningness’) and years of education.. Such associations could therefore plausibly be generated by selection bias...”

“Studies in samples with unknown selection/attrition mechanisms run the risk of providing biased and misleading results. In our opinion these important caveats should be borne in mind when interpreting the results of such studies.”



Peter Tennant
@PWGTennant

Hi Maria,

Could you help with a confusion?

From @_MiguelHernan et al, I thought an X-Y relationship could be (collider) biased if X & Y are correlated with selection

But @uk_biobank say this on their website (ukbiobank.ac.uk/scientists-3/)

What've I missed?

#GlymourTwitterTakeover



SER
@societyforepi

Replies to @PWGTennant @_MiguelHernan and @uk_biobank

Suspect there's no justification for the claim, just hope. Maybe it's true or bias is small, but it's an empirical question. Collider bias not induced if X & Y influence selection multiplicatively (Hernan 2004 SA.3). Got evidence @uk_biobank? See @epi_kerrykeyes work for examples



Peter Tennant
@PWGTennant

Replies to @societyforepi @_MiguelHernan and 2 others

Thanks Maria, hopefully @uk_biobank will be able to provide evidence for this hopeful claim that collider bias is not generally a problem for studies in their sample.



MarkSG @statsmethods · Nov 2, 2018

Replies to @PWGTennant @societyforepi and 3 others

I have it on good authority that they know this is problematic. They just need to take this off their website because it's seriously misleading.

UK Biobank, big data, and the consequences of non-representativeness

UK Biobank is an unparalleled resource of extensive health information from 500 000 individuals and with more than 400 peer-reviewed publications to date. The sampling population is volunteer-based and is not representative of the UK population.¹ Investigators state that although the estimates of prevalence and incidence should be interpreted with caution, valid measures of association and estimates of causal effect can be more readily interpreted as they do “not require participants to be representative of the population at large”.²

This statement is a puzzling claim: sample selection can indeed influence

3 times the risk compared with those who are unexposed; among those without A, there is no association between X and Y. We hypothetically recruit 500 000 participants (approximately the size of the UK Biobank sample). But, during recruitment, those unexposed to A (healthier volunteers) are more likely to join,¹ than those exposed to A are (by about 2:1, appendix). In the study sample, the risk ratio is 1·67 (appendix). The association between those exposed to X and the disease differs between the study sample and the target population because the prevalence of A differs; indeed, the association would differ in any population with a different prevalence of A. The magnitude of an exposure’s association with an outcome depends on the prevalence of other factors that interact with the exposure; in our case,

*Katherine M Keyes, Daniel Westreich
kmk2104@columbia.edu

Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA (KMK); Center for Research on Society and Health, Universidad Mayor, Santiago, Chile (KMK); and Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, NC, USA (DW)

- 1 Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186**: 1026–34.
- 2 UK Biobank. Researchers. <http://www.ukbiobank.ac.uk/scientists-3/> (accessed Dec 6, 2018).
- 3 Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. *Epidemiology* 2017; **28**: 553–61.
- 4 Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. *Am J Epidemiol* 2019; **188**(2): 438–43.
- 5 Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol* 2017; **186**: 1010–14.

Keyes KM, Westreich D. *The Lancet*. 2019;393(10178):1297.



UNIVERSITY OF
OXFORD

Causal Inference Seminar

Acknowledging the third pillar of contemporary
data science, Dr Peter Tennant

21 March 2019

*“Please remove this statement off your website, because
it’s not scientifically true ... and I don’t think that’s good”*

Please note: UK Biobank is not representative of the general population on a variety of sociodemographic, physical, lifestyle and health-related characteristics, with evidence of a 'healthy volunteer' selection bias, details of which are published elsewhere (Fry et al, Am J Epidemiol 2017;186:1026-34. PMID 28641372). As a result, UK Biobank is not a suitable resource for deriving generalizable disease prevalence and incidence rates.

Is Cohort Representativeness Passé? Poststratified Associations of Lifestyle Risk Factors with Mortality in the UK Biobank

Emmanuel Stamatakis,^a Katherine B. Owen,^b Leah Shepherd,^b Bradley Drayton,^b Mark Hamer,^c and Adrian E. Bauman^b

ARTICLE

<https://doi.org/10.1038/s41467-020-19478-2>

OPEN



Collider bias undermines our understanding of COVID-19 disease risk and severity

Gareth J. Griffith^{1,2,4}, Tim T. Morris^{1,2,4}, Matthew J. Tudball^{1,2,4}, Annie Herbert^{1,2,4}, Giulia Mancano^{1,2,4}, Lindsey Pike^{1,2}, Gemma C. Sharp^{1,2}, Jonathan Sterne², Tom M. Palmer^{1,2}, George Davey Smith^{1,2}, Kate Tilling^{1,2}, Luisa Zuccolo^{1,2}, Neil M. Davies^{1,2,3} & Gibran Hemani^{1,2,4✉}

COMMENTARY

Representativeness Is Not Representative Addressing Major Inferential Threats in the UK Biobank and Other Big Data Repositories

Jonathan Yinhao Huang

“Our results suggest that future UKB (and analogous cohort) users should exercise caution when examining associations between established risk factors and mortality outcomes as poor cohort sample representativeness might influence materially some estimates”

“Results from samples that are likely not representative of the target population should be treated with caution by scientists and policy makers.”

“There are no shortcuts: identifying the causes of selection biases are mandatory.”

Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis

G David Batty,^{1,2} Catharine R Gale,^{3,4} Mika Kivimäki,¹ Ian J Deary,⁴ Steven Bell^{5,6,7}

ABSTRACT

OBJECTIVE

To compare established associations between risk factors and mortality in UK Biobank, a study with an exceptionally low rate of response to its baseline survey, against those from representative studies that have conventional response rates.

DESIGN

Prospective cohort study alongside individual participant meta-analysis of other cohort studies.

SETTING

United Kingdom.

PARTICIPANTS

Analytical sample of 499 701 people (response rate 5.5%) in analyses in UK Biobank; pooled data from the Health Surveys for England (HSE) and the Scottish Health Surveys (SHS), including 18 studies and 89 895 people (mean response rate 68%). Both

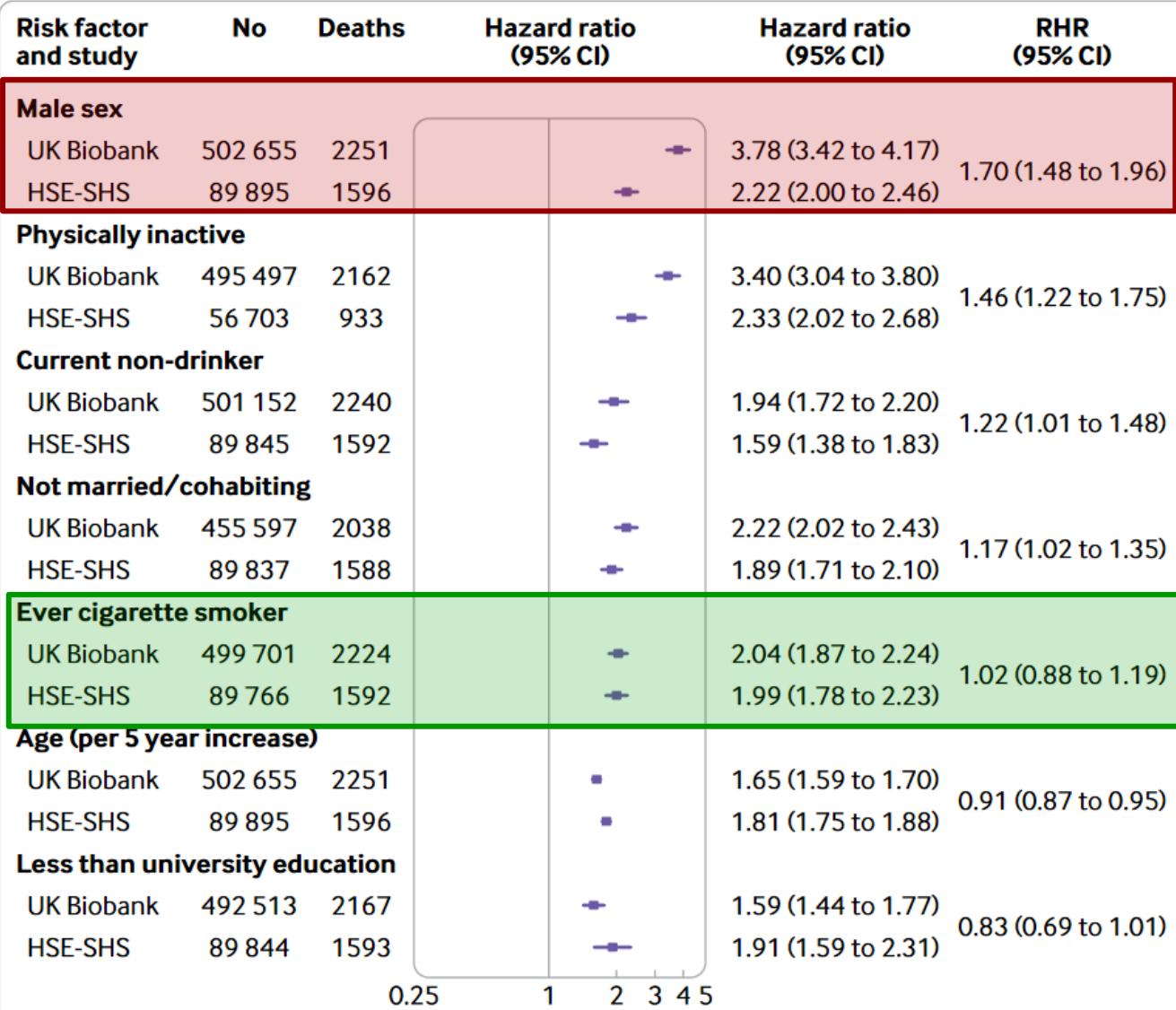
Based on 14 288 deaths during an average of 7.0 years of follow-up in UK Biobank and 7861 deaths over 10 years of mortality surveillance in HSE-SHS, for cardiovascular disease mortality, for instance, the age and sex adjusted hazard ratio for smoking cigarettes (versus never) was 1.87 to 2.24 in UK Biobank and 2.23 in HSE-SHS, yielding close to unity (1.02, 0.88 to 1.12). The level of agreement between studies was unchanged when results were stratified by sex and when baseline year was aligned.

CONCLUSION

Despite a very low response rate, risk factor associations in the UK Biobank seem to be generalisable.

CONCLUSION

Despite a very low response rate, risk factor associations in the UK Biobank seem to be generalisable.



Maria Glymour
@MariaGlymour

I hope everyone reads the results in this paper but ignores the conclusions, since conclusions do not seem to reflect the results Batty et al: [bmj.com /content/368/bm...](https://bmj.com/content/368/bm...) @PWGTenant @epi_kerrykeyes @EpiEllie @EpidByDesign @MarcusMunafo @MikaKivimaki 1/n



Comparison of risk factor associations in UK Biobank again...



Maria Glymour
@MariaGlymour

Replies to @MariaGlymour

The good news from Batty: for every comparison they present the effect estimates were the same sign. The bad news: if you used the UKB estimates to make public health decision or prioritization, you'd often make the wrong decision. 8/n

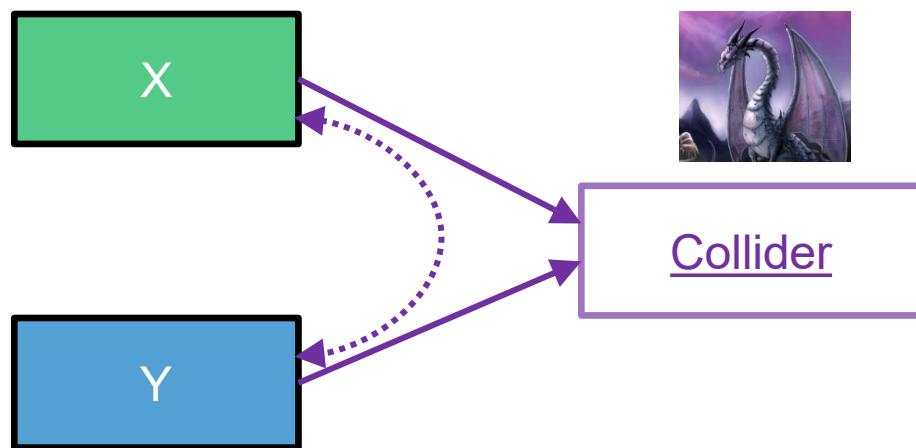
10:41 PM · Feb 17, 2020 · Twitter Web App



MAN, I'M GLAD THAT'S OVER.

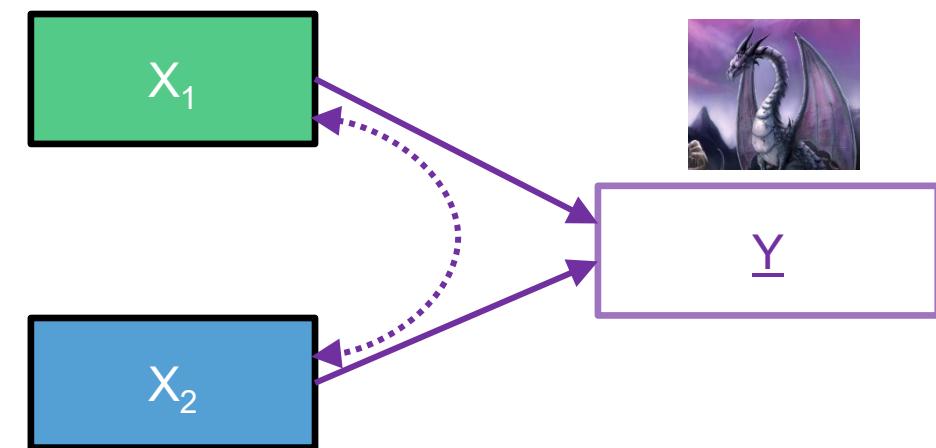
OUTCOME SELECTION BIAS

We have considered the effect of **collider bias** on the relationship between mutual causes of a collider



e.g. between **X** and **Y**

Where the collider is your outcome, **collider bias** will also distort the apparent effect of all causes of that outcome



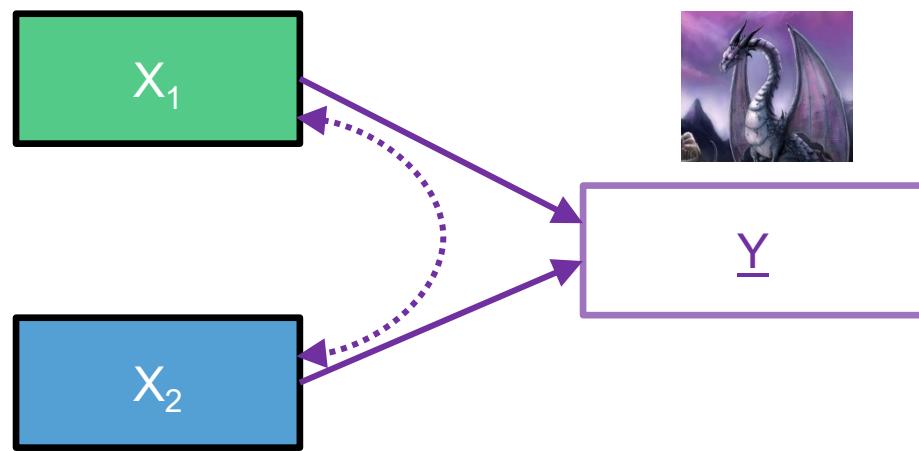
e.g. between **X_1** and **Y**
e.g. between **X_2** and **Y**

OUTCOME SELECTION BIAS

You can condition on your outcome when you analyze an extreme (non-random) sample of values; this is known as **conditioning-on-the-outcome**

- e.g. your outcome is **weight** – and you only study **obese adults**

The apparent effect of all causes of the outcome will now be biased by the dependencies between them – this is known as **outcome selection bias**



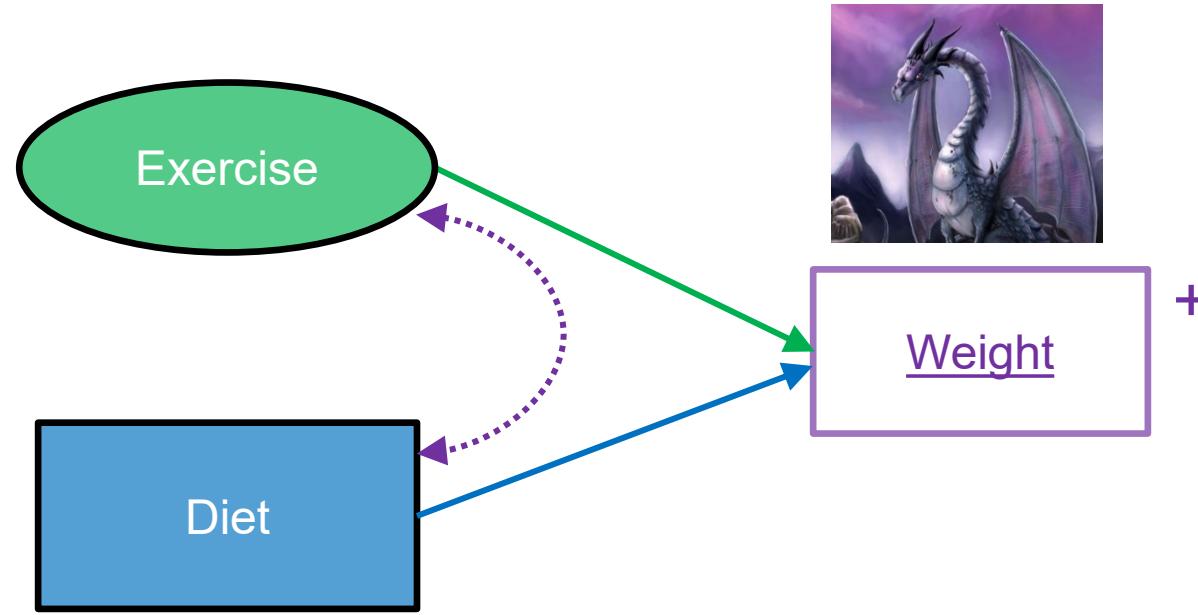
The effect of X_1 on Y is biased by
 $X_1 < \dots > X_2 > Y$

The effect of X_2 on Y is biased by
 $X_2 < \dots > X_1 > Y$

EXAMPLE: OUTCOME SELECTION BIAS

The effect of **diet** on weight in obese adults (denoted by +)

Exercise also causes **weight**, but suppose there was no overall effect of **exercise** on **diet**



Conditioning on weight opens **Diet <...> exercise**, meaning the effect of diet on weight is now biased by **diet <...> exercise > weight**

EXAMPLE: OUTCOME SELECTION BIAS

METHODOLOGY

Open Access



The implications of outcome truncation in reproductive medicine RCTs: a simulation platform for trialists and simulation study

Jack Wilkinson^{1*} , Jonathan Y. Huang², Antonia Marsden¹, Michael O. Harhay³, Andy Vail¹ and Stephen A. Roberts¹

Abstract

Background: Randomised controlled trials in reproductive medicine are often subject to outcome truncation, where the study outcomes are only defined in a subset of the randomised cohort. Examples include birthweight (measurable only in the subgroup of participants who give birth) and miscarriage (which can only occur in participants who become pregnant). These outcomes are typically analysed by making a comparison between treatment arms within the subgroup (for example, comparing birthweights in the subgroup who gave birth or miscarriages in the subgroup who became pregnant). However, this approach does not represent a randomised comparison when treatment influences the probability of being observed (i.e. survival). The practical implications of this for the design and interpretation of reproductive trials are unclear however.

Methods: We developed a simulation platform to investigate the implications of outcome truncation for reproductive medicine trials. We used this to perform a simulation study, in which we considered the bias, type 1 error, coverage, and precision of standard statistical analyses for truncated continuous and binary outcomes. Simulation settings were informed by published assisted reproduction trials.

Results: Increasing treatment effect on the intermediate variable, strength of confounding between the intermediate and outcome variables, and the presence of an interaction between treatment and confounder were found to adversely affect performance. However, within parameter ranges we would consider to be more realistic, the adverse effects were generally not drastic. For binary outcomes, the study highlighted that outcome truncation could cause separation in smaller studies, where none or all of the participants in a study arm experience the outcome event. This was found to have severe consequences for inferences.

Conclusion: We have provided a simulation platform that can be used by researchers in the design and interpretation of reproductive medicine trials subject to outcome truncation and have used this to conduct a simulation study. The study highlights several key factors which trialists in the field should consider carefully to protect against erroneous inferences. Standard analyses of truncated binary outcomes in small studies may be highly biased, and it remains to identify suitable approaches for analysing data in this context.

Keywords: Outcome truncation, Truncation by death, Competing risks, Censoring, Hypothetical estimand, Infertility, Assisted reproduction, In vitro fertilisation, Randomised controlled trials, Reproductive medicine

Reproductive studies suffer outcome selection because only healthy offspring can be observed

- e.g. effect of pollution on birthweight... in live born babies
- LBW offspring are less likely to be liveborn, so only looks at larger babies
- Williamson et al 2021 showed that major biases could occur!

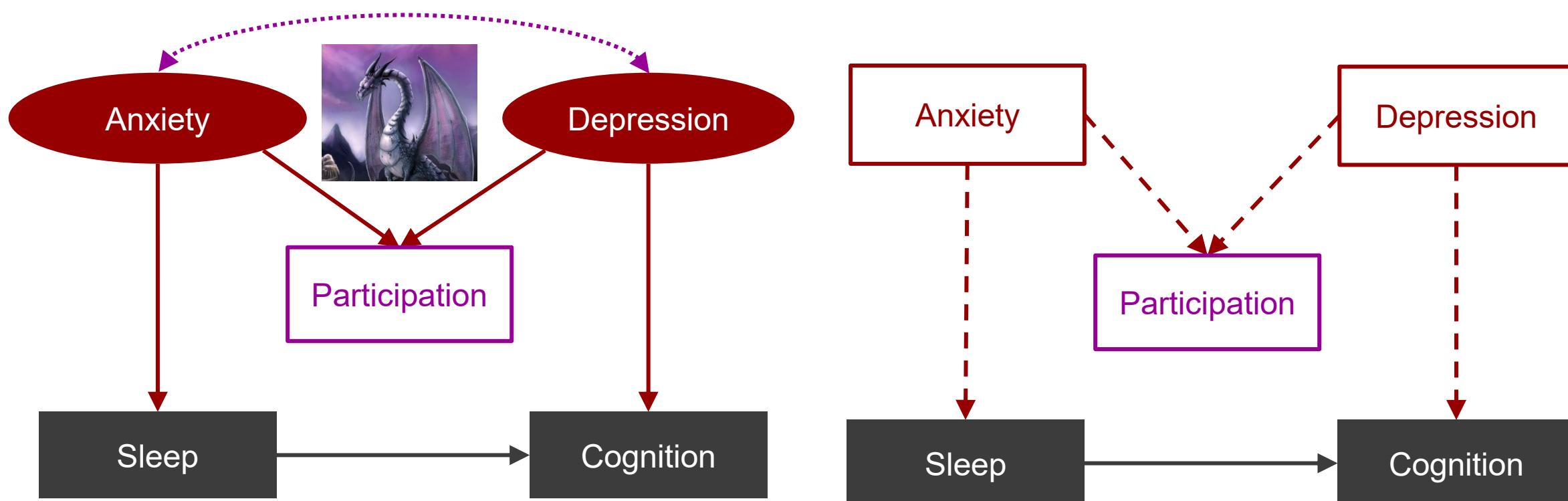
We return to the problems with outcome truncation when we consider RTM

AVOIDING SELECTION BIAS

How can you avoid **differential selection bias**?

1. Block the spurious path(s) with **conditioning**

Limitation: need to have measured the confounders



AVOIDING SELECTION BIAS

In electronic health records, conditioning for number of healthcare encounters reduces selection bias, although it also *introduces* some residual (albeit smaller) collider bias



American Journal of Epidemiology
© The Author 2016. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Vol. 184, No. 11
DOI: 10.1093/aje/kww112
Advance Access publication:
November 16, 2016

Practice of Epidemiology

Controlling for Informed Presence Bias Due to the Number of Health Encounters in an Electronic Health Record

Benjamin A. Goldstein*, Nrupen A. Bhavsar, Matthew Phelan, and Michael J. Pencina

* Correspondence to Dr. Benjamin A. Goldstein, Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, 2424 Erwin Road, Suite 1105, Room 11041, Durham, NC 27705 (e-mail: ben.goldstein@duke.edu).

Initially submitted September 29, 2015; accepted for publication September 15, 2016.

Electronic health records (EHRs) are an increasingly utilized resource for clinical research. While their size allows for many analytical opportunities, as with most observational data there is also the potential for bias. One of the key sources of bias in EHRs is what we term *informed presence*—the notion that inclusion in an EHR is not random but rather indicates that the subject is ill, making people in EHRs systematically different from those not in EHRs. In this article, we use simulated and empirical data to illustrate the conditions under which such bias can arise and how conditioning on the number of health-care encounters can be one way to remove this bias. In doing so, we also show when such an approach can impart M bias, or bias from conditioning on a collider. Finally, we explore the conditions under which number of medical encounters can serve as a proxy for general health. We apply these methods to an EHR data set from a university medical center covering the years 2007–2013.

Goldstein 2016 Am J Epidemiol;
184(11): 847–855.

AVOIDING SELECTION BIAS

2. Reweight the sample using **inverse probability weighting**

Selection should be ‘as random’ \Rightarrow all units of analysis should have an equal probability of selection

When selection not random, need to equalise probabilities by re-weighting units

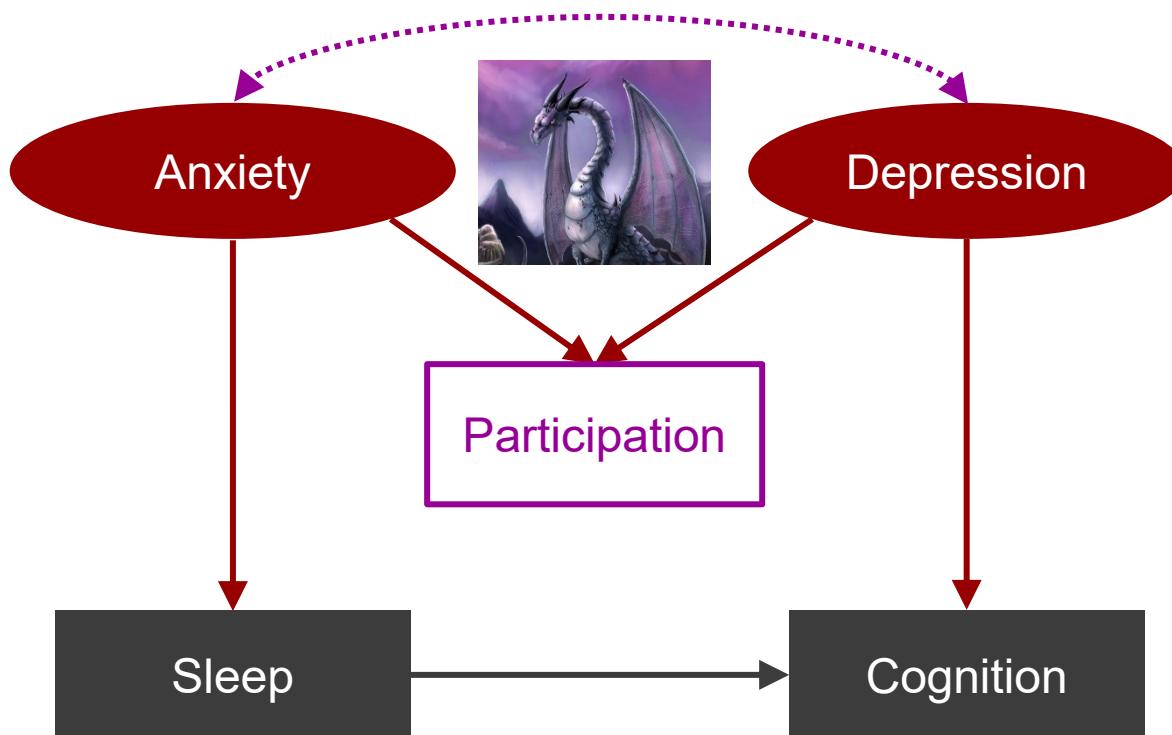
Adopt ‘two stage’ approach (similar to ‘**propensity score**’ analyses)

- predict **participation propensity** in the total population
- calculate **participation propensity** for units in the sample
- weight analysis by inverse of **participation propensity**

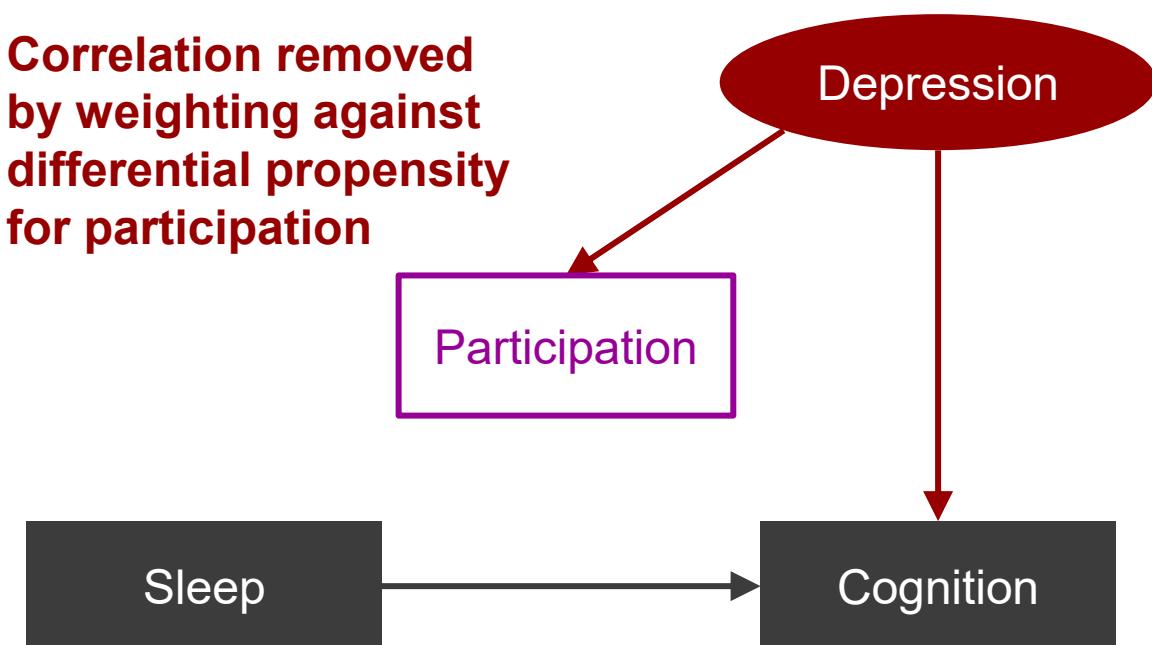
AVOIDING SELECTION BIAS

2. Reweight the sample using **inverse probability weighting**

- e.g. remove **association** between **sleep** and **participation**
- e.g. if people with insomnia were half as likely to take part – participants with insomnia are reweighted to count for double!



**Correlation removed
by weighting against
differential propensity
for participation**



AVOIDING SELECTION BIAS

2. Reweight the sample using **inverse probability weighting**

Limitation: you need to have some information on **participation propensity**

- Some studies (particularly those using '**double consent**') collect this – you should try to do this wherever possible
- For existing datasets (or routinely collected data) this may not be possible
- You could **estimate** the weights by comparing the demographics of your participants with the general population

Note: IPW is particularly good for dealing with bias from **differential attrition**, provided you have good information at baseline

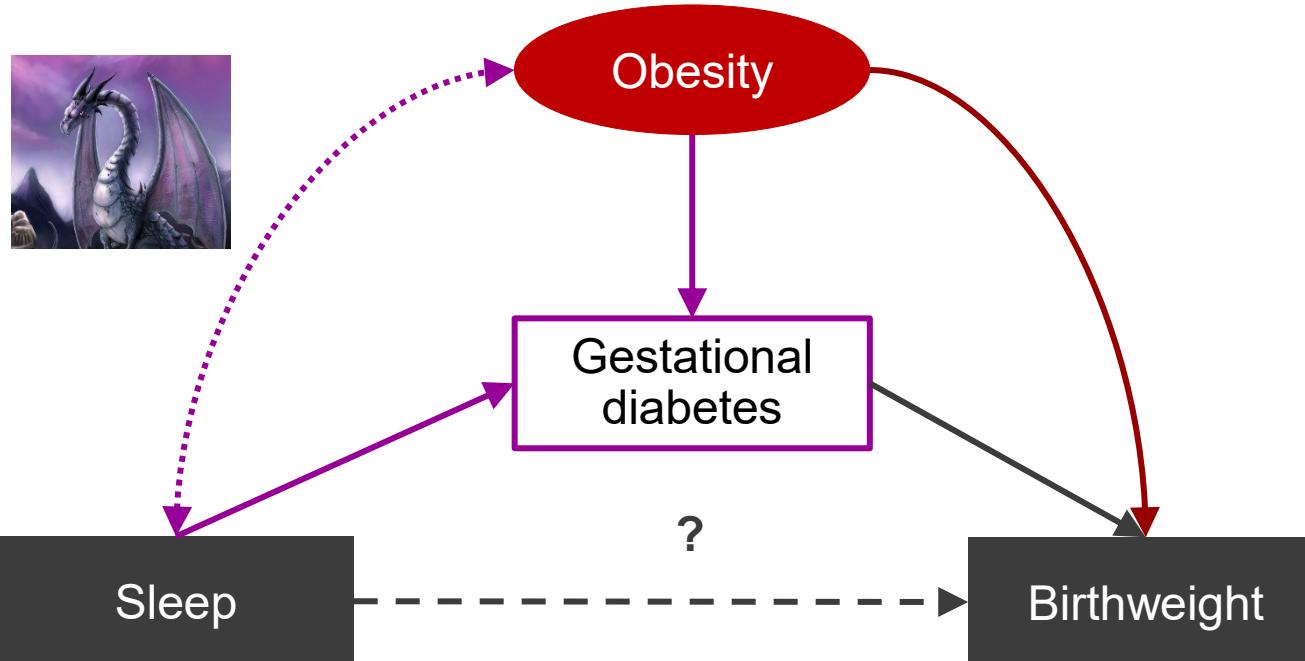
We will see later where **inverse probability weighting** can be used to solve problems that 'standard methods' fail

AVOIDING SELECTION BIAS

Sometimes the bias is unsolvable, therefore ...

3. Be transparent and recognise where your effect estimates may be biased

e.g. **Sleep** and **birthweight** in **women with gestational diabetes**



Best to restrict inferences
to similar populations, and
recognise they may have
no causal interpretation

SUMMARY

Collider bias is **evil**



← **Evil!**

Collider bias (evil)

RECOMMENDED READING

- Elwert, F. and Winship, C., 2014. Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology*, 40, pp.31-53.
- Cole, S.R., Platt, R.W., Schisterman, E.F., Chu, H., Westreich, D., Richardson, D. and Poole, C., 2010. Illustrating bias due to conditioning on a collider. *International journal of epidemiology*, 39(2), pp.417-420.
- Munafò, M.R., Tilling, K., Taylor, A.E., Evans, D.M. and Davey Smith, G., 2018. Collider scope: when selection bias can substantially influence observed associations. *International journal of epidemiology*, 47(1), pp.226-235.
- Griffith, G.J., Morris, T.T., Tudball, M.J., Herbert, A., Mancano, G., Pike, L., Sharp, G.C., Sterne, J., Palmer, T.M., Davey Smith, G. and Tilling, K., 2020. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature communications*, 11(1), pp.1-12.

SUMMARY

- Differential selection bias is a form of **common** (universal) collider bias that occurs due to non-random sampling or participation into a study sample
- You should consider potential sources of differential selection bias when planning your study and/or analyses; DAGs are essential for this !
- It is a huge scientific advance to recognise and highlight when your data may be affected by non-trivial selection bias
- If you have the information, you may be able to reduce differential selection biases by clever conditioning or using inverse probability weights

DON'T WORRY IF YOU DON'T UNDERSTAND!

- Collider bias is **NOT** intuitive!
- It takes a long time to grasp intuitively (if ever...)
- This is true for **everyone** who learns about it!



3.3 - RTMA AND CONDITIONING- ON-THE-OUTCOME

GEORGIA



MARK



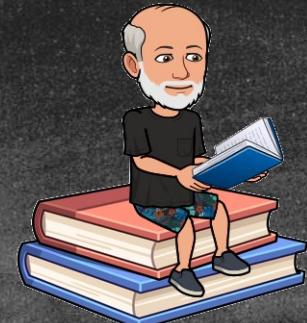
@GEORGIATOMOVA

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 3

09:30-10:15 ACTIVITY 3-A

10:15-11:00 LECTURE 3.1

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 3.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 3.3

14:45-15:30 ACTIVITY 3-B

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 3.4

17:00-17:45 ACTIVITY 3-C

17:45-18:00 Q&A

LEARNING OBJECTIVES

By the end of this session, you will be able to:

- Describe **regression-to-the-mean** (RTM) as a universal consequence of variation and randomness
- Draw **DAGs** to depict variation and randomness, and explain how these help us understand regression-to-the-mean
- Explain how conditioning – such as **conditioning-on-the-outcome** - can create distorting and misleading **collider error**
- Explain how the **placebo effect** may be understood as a consequence of regression-to-the-mean and collider error

CENTRIPETAL DRIFT

CENTRIPETAL DRIFT: A FALLACY IN THE EVALUATION OF THERAPEUTIC RESULTS

THE fallacy to be described here has been observed twice in recent numbers of carefully edited medical journals,¹ and for that reason alone deserves the attention of investigators. It is likely that a search of the literature of therapeutics would yield numerous instances.

Using an instrument whose readings are affected by large chance errors, an investigator examines, say, 100 subjects. He selects the 10 people whose performance on the test happens to be the poorest, and gives them some kind of treatment. Next day he reexamines the 100 subjects. The average performance of the 100 is exactly what it was the day before, but the 10 who did most poorly then are now found to have improved strikingly.

Jung FT 1938

REGRESSION-TO-THE-MEAN

Regression-to-the-mean is a universal phenomenon that arises whenever two variables are imperfectly correlated

Commonly discussed with respect to repeated measures of the same variable, but need not involve the same variable or even repeated measurements !

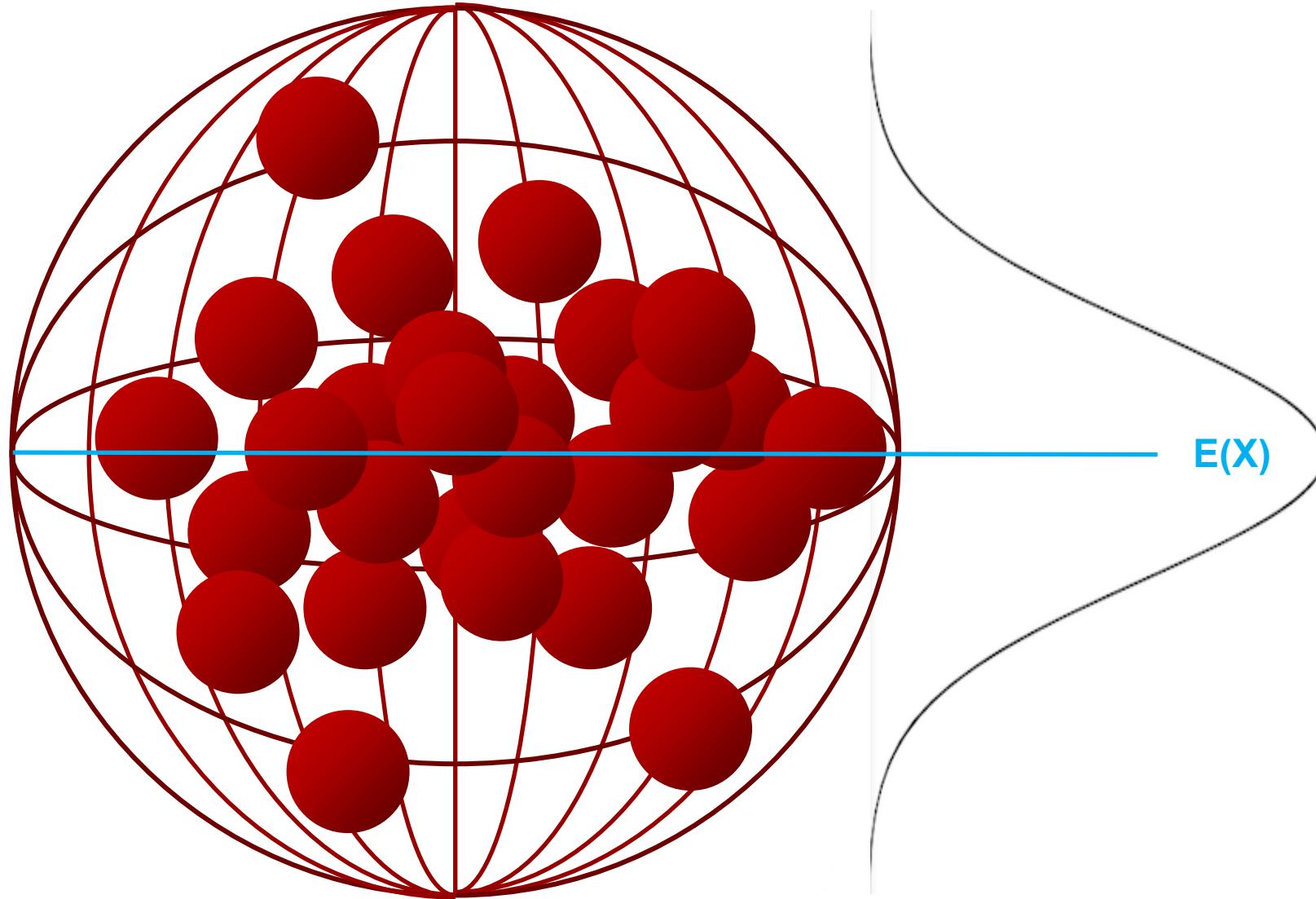
“Given an extreme random event, the next random event is likely less extreme”

Consider two independent (normally distributed) variables X and Y

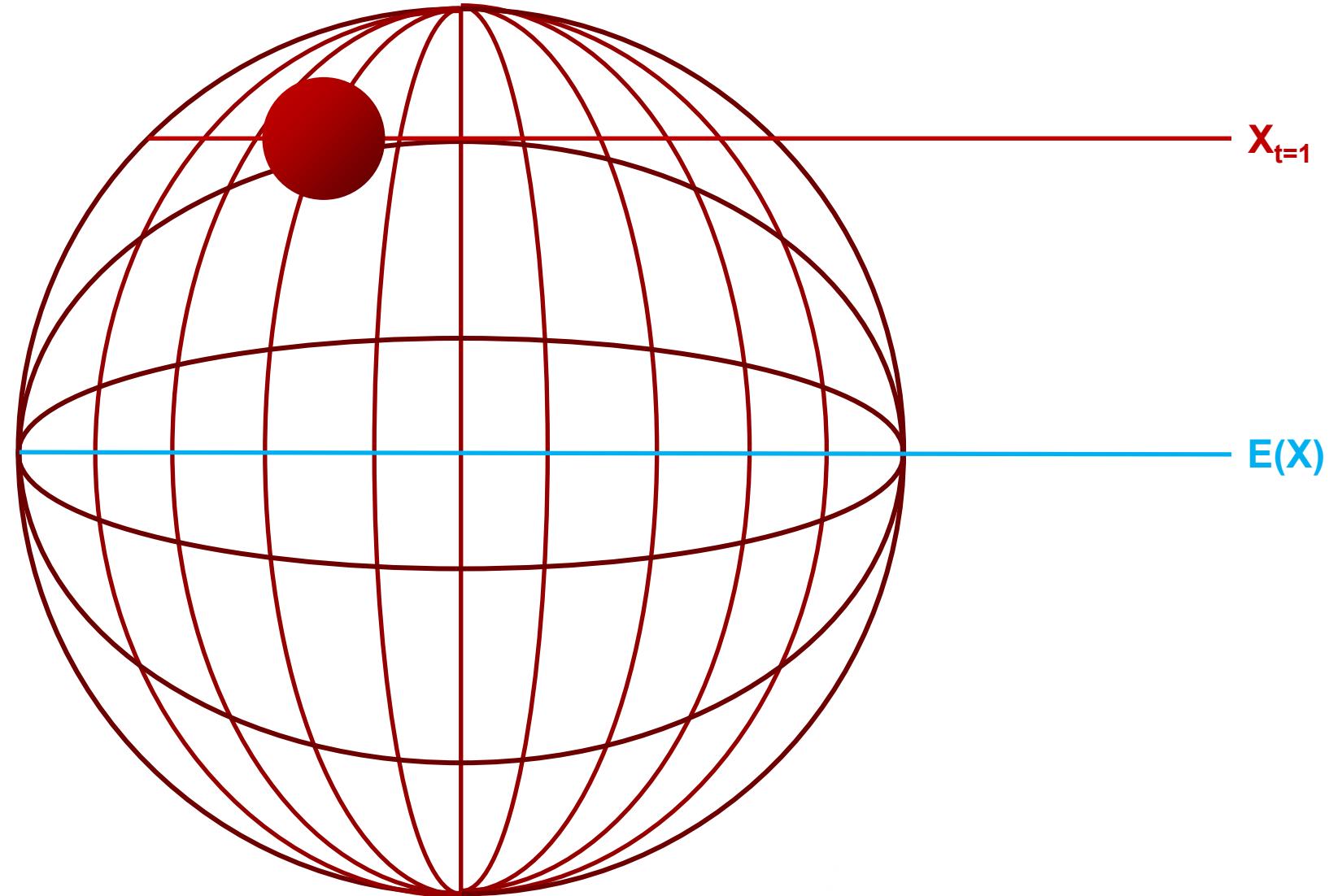
- select a value (or group) for X that is away from the mean – this is ‘extreme’
- now (independently) select Y: the likely value for Y is $E(Y)$ – this is not extreme

RTM occurs due to all possible reasons why two variables may be different, including **natural variation**, **enigmatic variation**, **error** and **randomness**

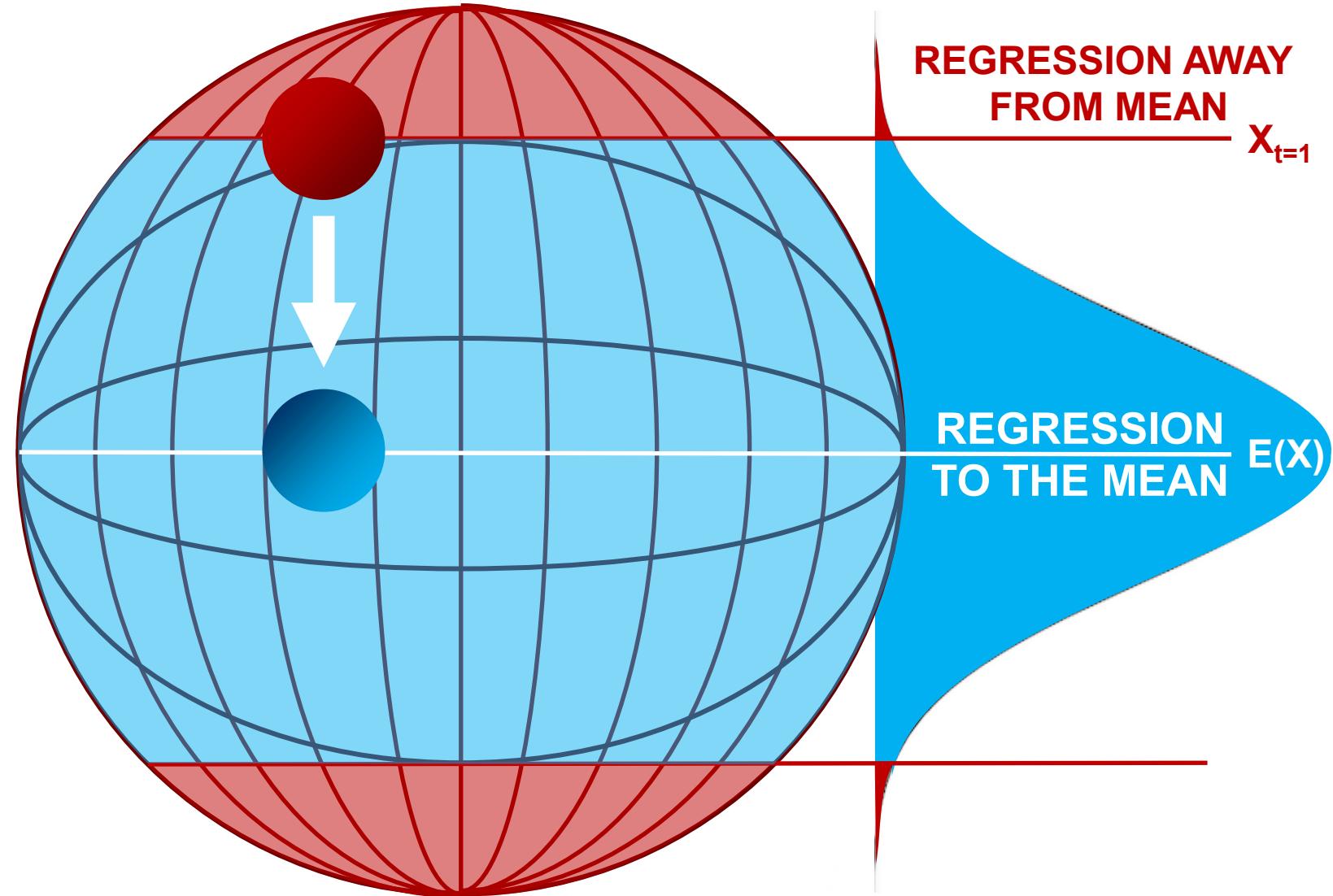
REGRESSION-TO-THE-MEAN



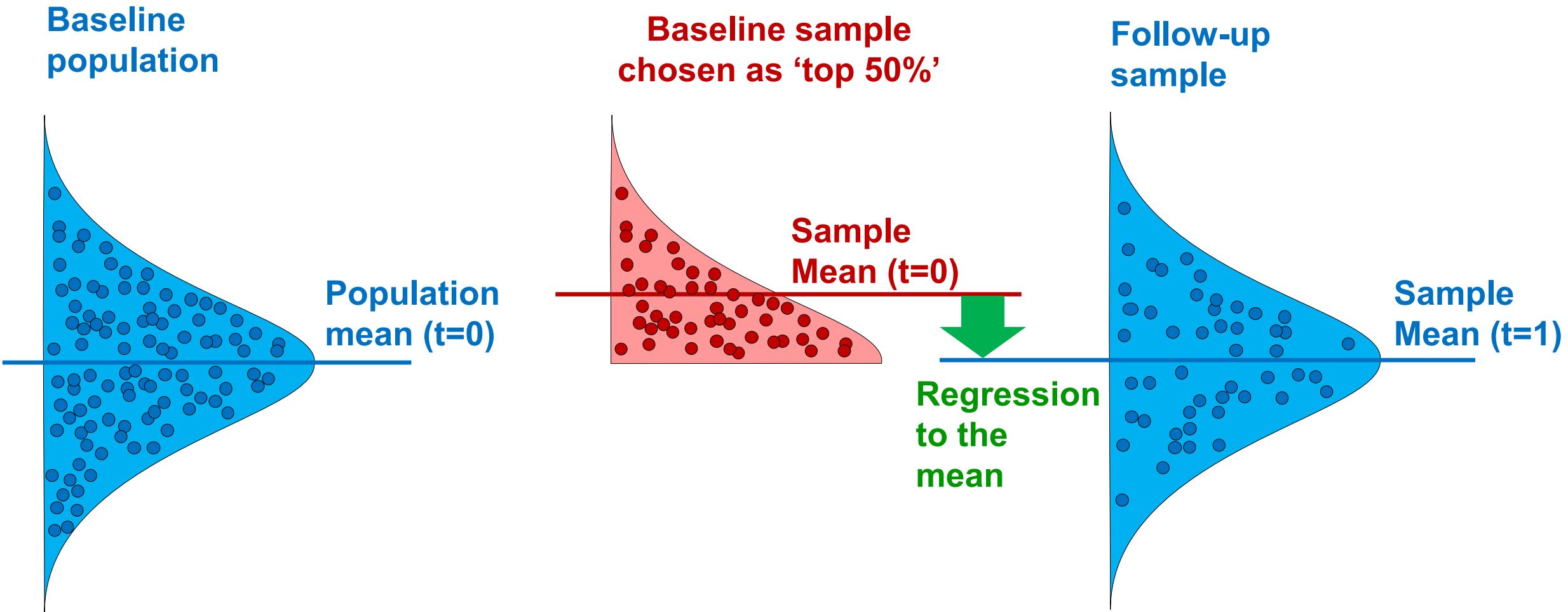
REGRESSION-TO-THE-MEAN



REGRESSION-TO-THE-MEAN



REGRESSION-TO-THE-MEAN



REGRESSION-TO-THE-MEAN

Baseline

Follow-up

Population
mean ($t=0$)


Population
change

Population
mean ($t=1$)

Baseline sample
chosen as 'top 50%'

Sample
Mean ($t=0$)


Population
change

Regression
to the
mean


Sample
Mean ($t=1$)

Follow-up
sample

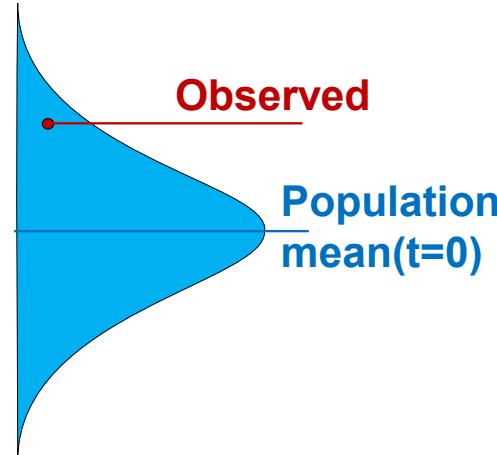
REGRESSION-TO-THE-MEAN

The weaker the correlation between two variables, the greater the RTM

For repeated measures: the greater the variability → the greater the RTM

May be understood as an ‘information compromise’ between stable part (that predicts similarly extreme value) and random part (that predicts the mean)

Population distribution (prior) and first observation

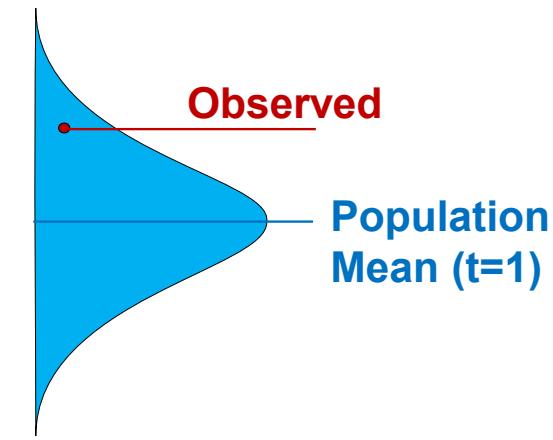


Low variability for observed variable (e.g. height)

Posterior probability distribution mostly determined by observation



Population distribution and second observation



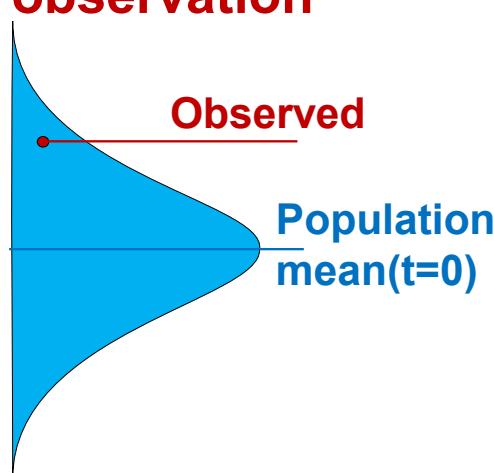
REGRESSION-TO-THE-MEAN

The weaker the correlation between two variables, the greater the RTM

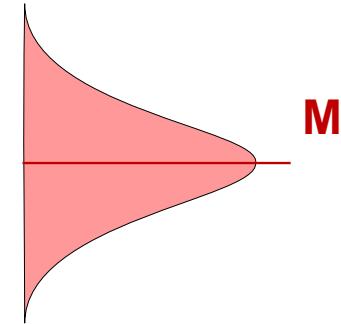
For repeated measures: the greater the variability → the greater the RTM

May be understood as an ‘information compromise’ between stable part (that predicts similarly extreme value) and random part (that predicts the mean)

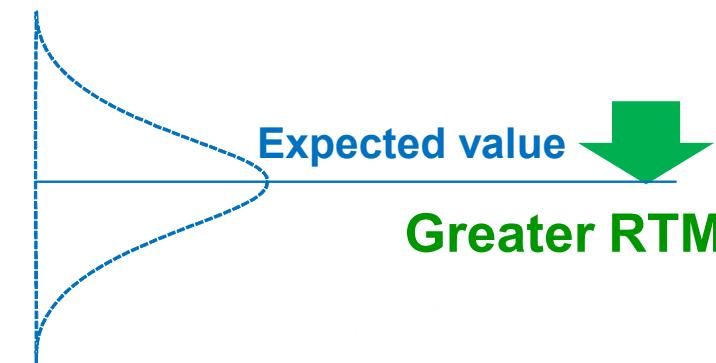
Population distribution (prior) and first observation



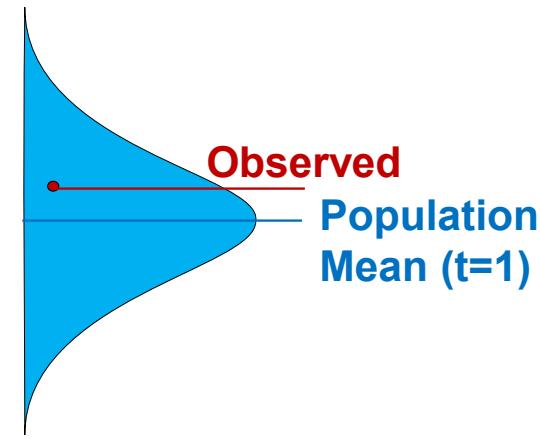
High variability for observed variable (e.g. blood pressure)



Posterior probability distribution determined by balance of observation & prior distribution



Population distribution and second observation



REGRESSION-TO-THE-MEAN

SPORT

[Home](#) [Football](#) [Cricket](#) | [Formula 1](#) | [Rugby U](#) | [Rugby L](#) | [Tennis](#) | [Golf](#) | [Boxing](#) | [Athletics](#) |

[Leicester City](#) > [Scores & Fixtures](#) | [Table](#) | [Top Scorers](#)

Leicester City win Premier League title after Tottenham draw at Chelsea

SPORT

[Home](#) [Football](#) [Cricket](#) | [Formula 1](#) | [Rugby U](#) | [Rugby L](#) | [Tennis](#) | [Golf](#) | [Boxing](#) | [Athletics](#) | [Paralympics](#)

[Leicester City](#) > [Scores & Fixtures](#) | [Table](#) | [Top Scorers](#)

Claudio Ranieri: Leicester City sack Premier League-winning manager

⌚ 23 February 2017 | [Leicester](#) | 📺 1403

ET THE ECONOMIC TIMES

Opinion → Times View Readers' Blog Times Evoke City India World Entertainment Sports Spirituality Business •••



NEWS / BLOGS / SPORTS / 'Regression to the Mean': The reason why Leicester FC are underperforming

SPORTS

'Regression to the Mean': The reason why Leicester FC are underperforming

September 28, 2016, 8:37 AM IST / Economic Times in ET Commentary, Sports, ET



FACEBOOK



TWITTER



LINKEDIN



EMAIL



Economic Times
Quick takes, analyses
and macro-level views
on all contemporary
economic, financial and
political events.

By Joy Bhattacharjya

Just a few days ago, I was watching a Manchester United team still clearly not at their best demolish Leicester FC in the first half of a match. And as the players walked off at half time, the commentator remarked that Manchester United were leading the Champions of England 4-0.

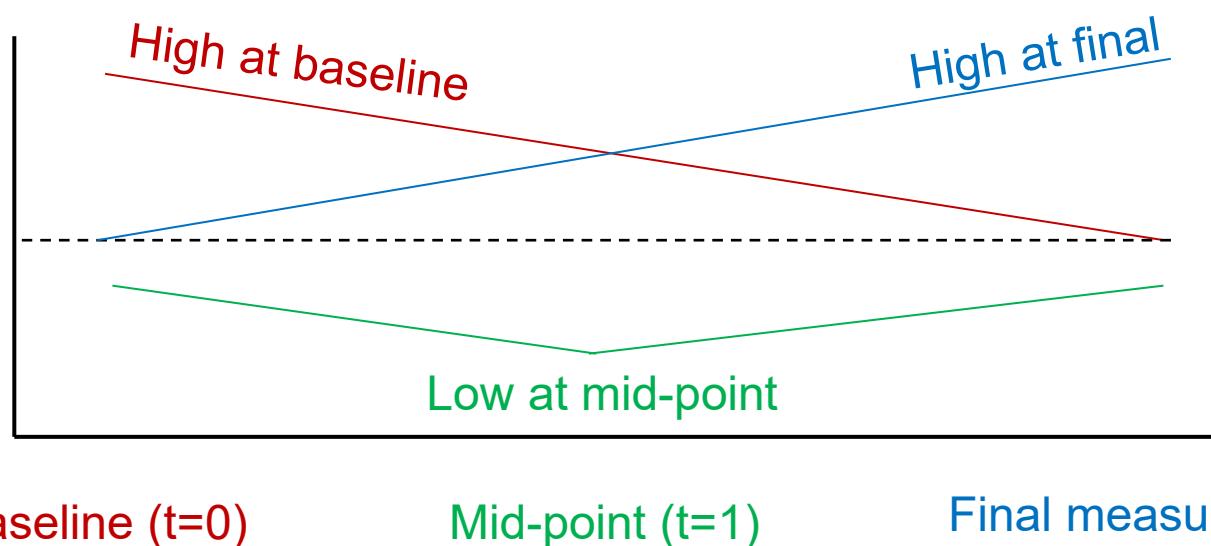
Champions of England, that's what this Leicester FC team was last season, and within a month of this season commencing, they have lost as many Premier League games as they did all season last year. Poster boys for what is often called the 'Regression to The Mean' theory.

WHY DOES THIS MATTER

Regression-to-the-mean can be seriously misleading if you don't learn to recognise and understand it

Naïve analyses can suggest change is taking place where none is involved

In longitudinal data: if you examine any extreme group at one point in time, they will appear less extreme in any direction across time

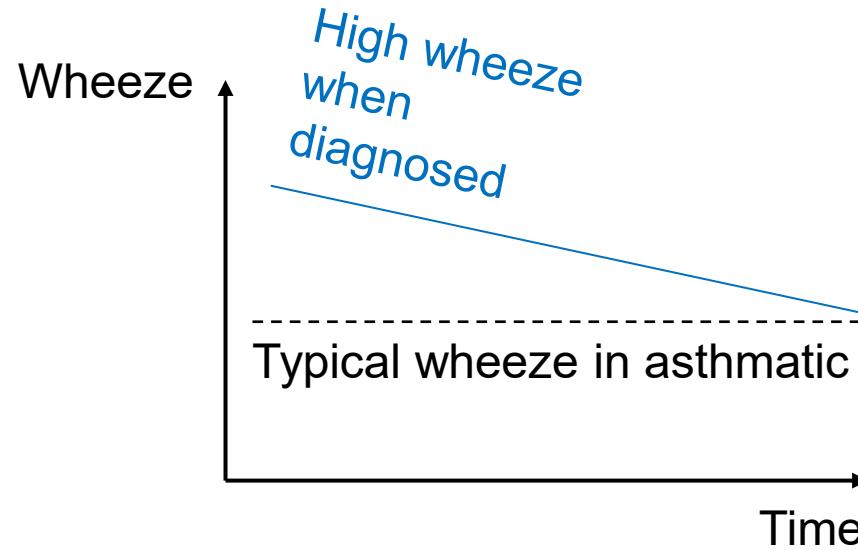


This commonly happens when we **condition-on-the-outcome**

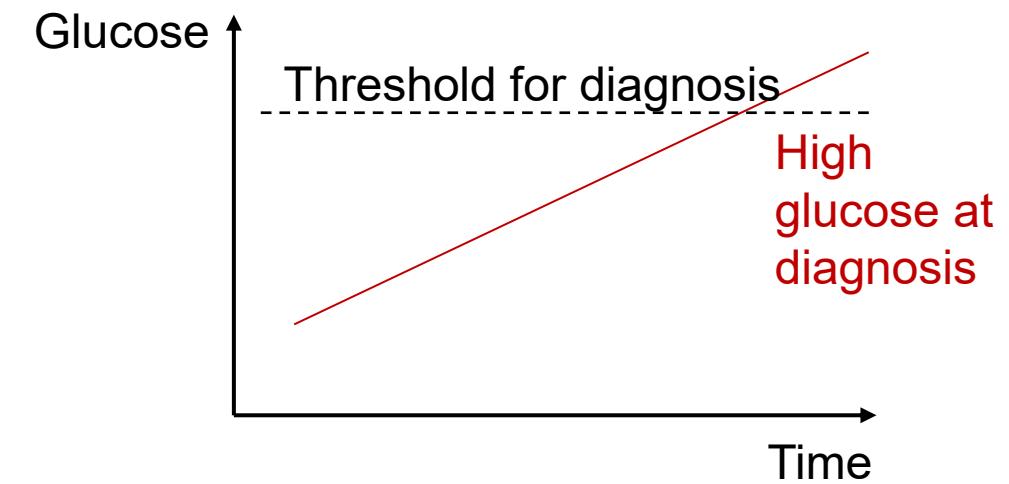
The ‘lines’ we see here represent RTM, but are sometimes misinterpreted as ‘*trajectories*’

MISLEADING ‘TRAJECTORIES’

What is the trajectory of wheezing symptoms for people who are newly diagnosed with **asthma**?



What is the trajectory of blood glucose for people before they get diagnosed with **diabetes**?



MISLEADING ‘TRAJECTORIES’

Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study

Adam G Tabák, Markus Jokela, Tasnime N Akbaraly, Eric J Brunner, Mika Kivimäki, * Daniel R Witte*

Summary

Background Little is known about the timing of changes in glucose metabolism before occurrence of type 2 diabetes. We aimed to characterise trajectories of fasting and postload glucose, insulin sensitivity, and insulin secretion in individuals who develop type 2 diabetes.

Methods We analysed data from our prospective occupational cohort study (Whitehall II study) of 6538 (71% male and 91% white) British civil servants without diabetes mellitus at baseline. During a median follow-up period of 9.7 years, 505 diabetes cases were diagnosed (49.1% on the basis of oral glucose tolerance test). We assessed retrospective trajectories of fasting and 2-h postload glucose, homeostasis model assessment (HOMA) insulin sensitivity, and HOMA β-cell function from up to 13 years before diabetes diagnosis (diabetic group) or at the end of follow-up (non-diabetics).

Findings Multilevel models adjusted for age, sex, and ethnic origin confirmed that all metabolic measures followed linear trends in the group of non-diabetics (10 989 measurements), except for insulin secretion that did not change during follow-up. In the diabetic group (801 measurements), a linear increase in fasting glucose was followed by a steep quadratic increase (from 5.79 mmol/L to 7.40 mmol/L) starting 3 years before diagnosis of diabetes. 2-h postload glucose showed a rapid increase starting 3 years before diagnosis (from 7.60 mmol/L to 11.90 mmol/L), and HOMA insulin sensitivity decreased steeply during the 5 years before diagnosis (to 86.7%). HOMA β-cell function increased between years 4 and 3 before diagnosis (from 85.0% to 92.6%) and then decreased until diagnosis (to 62.4%).

Interpretation In this study, we show changes in glucose concentrations, insulin sensitivity, and insulin secretion as much as 3–6 years before diagnosis of diabetes. The description of biomarker trajectories leading to diabetes diagnosis could contribute to more-accurate risk prediction models that use repeated measures available for patients through regular check-ups.

Funding Medical Research Council (UK); Economic and Social Research Council (UK); British Heart Foundation (UK); Health and Safety Executive (UK); Department of Health (UK); National Institute of Health (USA); Agency for Health Care Policy Research (USA); the John D and Catherine T MacArthur Foundation (USA); and Academy of Finland (Finland).



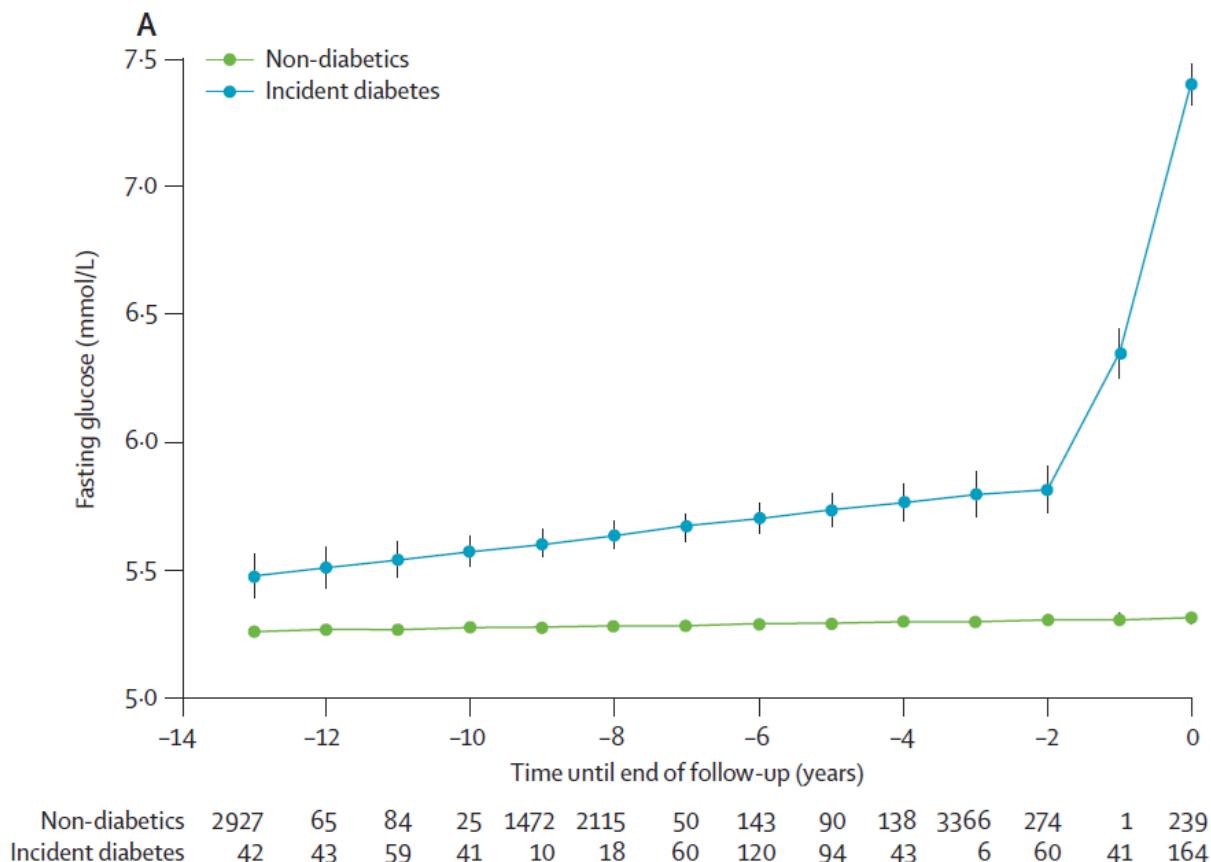
Lancet 2009; 373: 2215-21

Published Online
June 8, 2009
DOI:10.1016/S0140-
6736(09)60619-X
See Comment page 2178

*These authors contributed
equally

Department of Epidemiology
and Public Health, University
College London, London, UK
(A G Tabák MD, M Jokela PhD,
T N Akbaraly PhD,
E J Brunner PhD,
Prof M Kivimäki PhD,
D R Witte MD); Semmelweis
University Faculty of Medicine,
1st Department of Medicine,
Budapest, Hungary (A G Tabák);
Department of Psychology,
University of Helsinki, Helsinki,
Finland (M Jokela);
INSERM U888 and University
Montpellier 1, Montpellier,
France (T N Akbaraly); Finnish
Institute of Occupational
Health, Helsinki, Finland
(Prof M Kivimäki); and Steno
Diabetes Center, Gentofte,
Denmark (D R Witte)

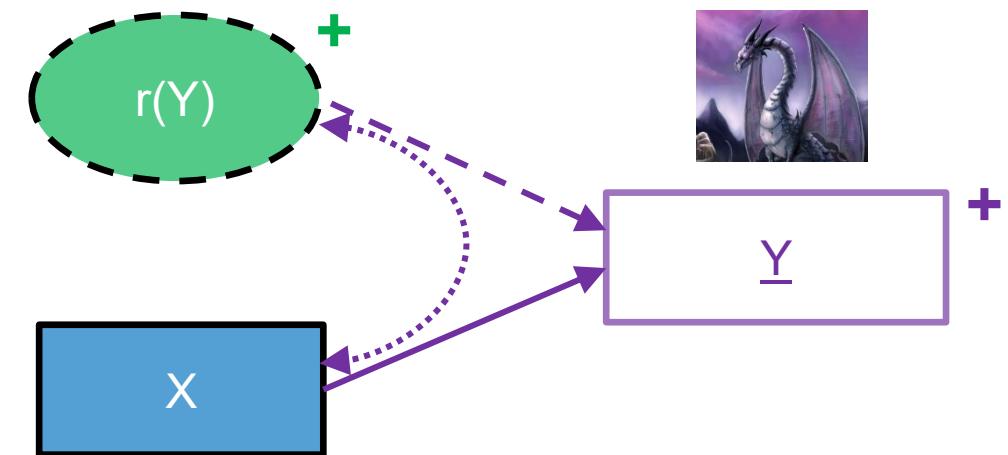
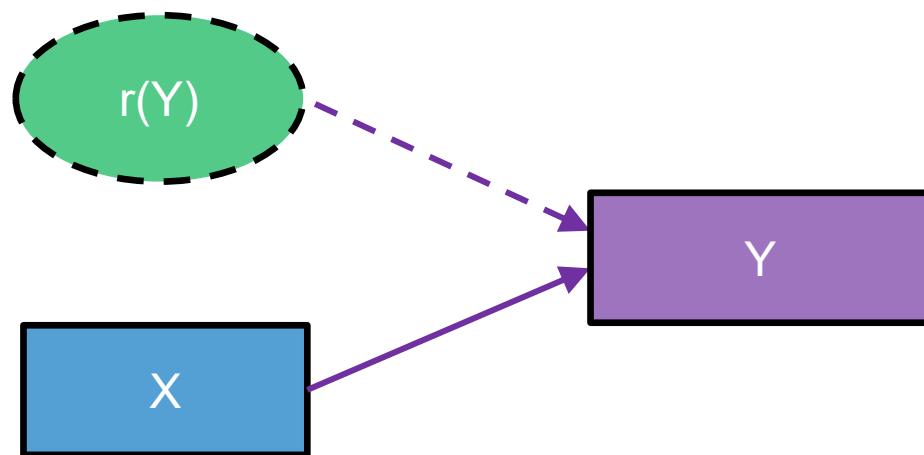
Correspondence to:
Dr Adam G Tabák, Department of
Epidemiology and Public Health,
University College London.



RTM IN DAGS

Regression-to-the-mean – and **conditioning-on-the-outcome** – can be understood as ‘**collider error**’ by including **random variation** terms

We call it ‘collider error’ not **collider bias** because the distortion is caused by random rather than structural causes

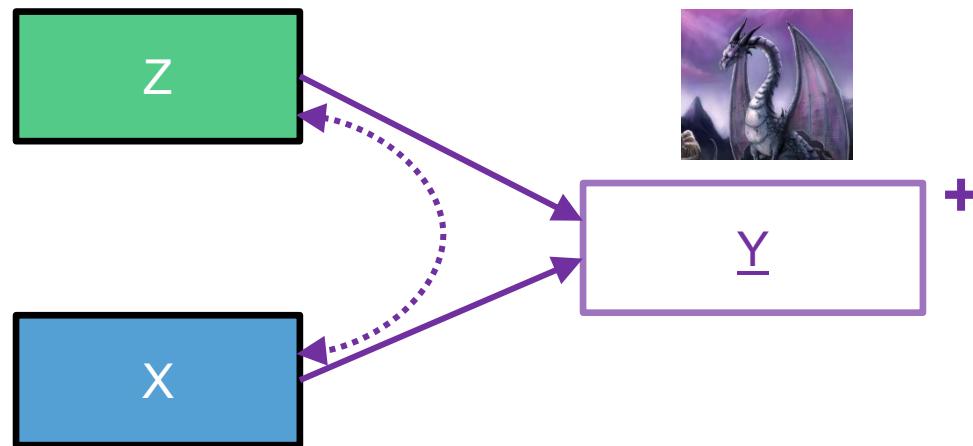


Problems like RTM – which stem from *error* not bias – are often overlooked by statisticians because they would not be observed with repeated sampling

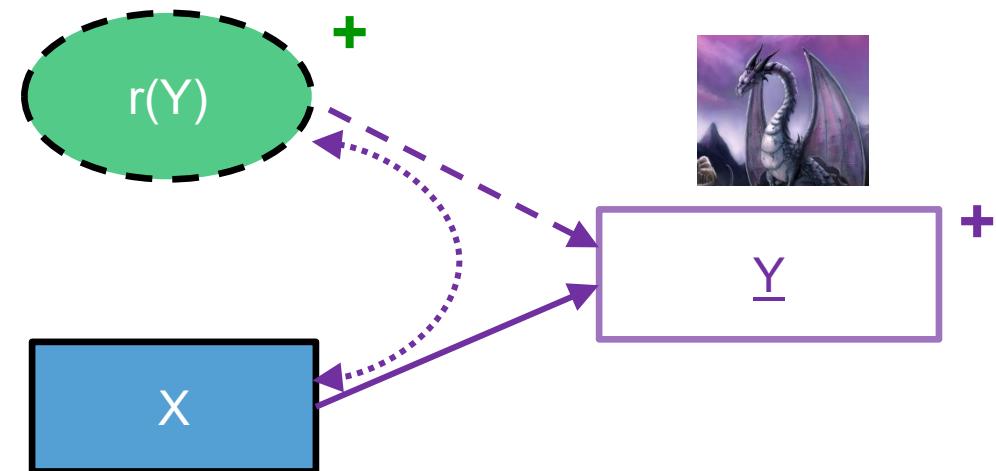
CONDITIONING-ON-THE-OUTCOME

Recall how **outcome selection bias** occurred when **conditioning-on-the-outcome**

The effect of our exposure was biased by other mutual causes of the outcome



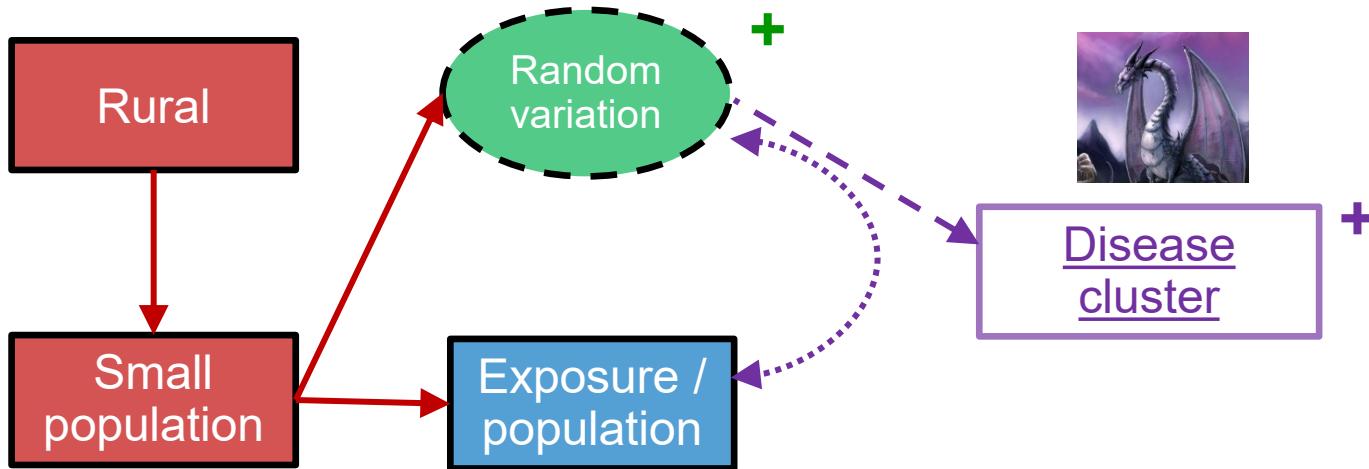
Because outcome selection involves studying *extreme* values of the outcome, we also create a dependency with all **random determinants** of the outcome!



SELECTION-ON-THE-OUTCOME

Studies of extreme values of Y, will therefore be biased by ‘all other reasons’ for extreme values of Y, including causes of random variation

- e.g. in geographical studies, areas with smaller populations have greater variation and are more likely to have random (high) spikes in the outcome
- if only areas with high disease occurrence are examined, anything related to small population size will appear as a cause ...



The Association Between Childhood Leukemia
and Population Mixing
An Artifact of Focusing on Clusters?

Laurie Berrie,^{a,b} George T.H. Ellison,^{a,b} Paul D. Norman,^{b,c} Paul D. Baxter,^{a,b}
Richard G. Feltbower,^{a,b} Peter W.G. Tennant,^{a,b} Mark S. Gilthorpe^{a,b}

EXAMPLE: REGRESSION-TO-THE-MEAN

Selecting a sample based on extreme values of a variable creates a conditional dependency with the random variation for that variable

- This cannot be expected when you look at other variables in the same sample



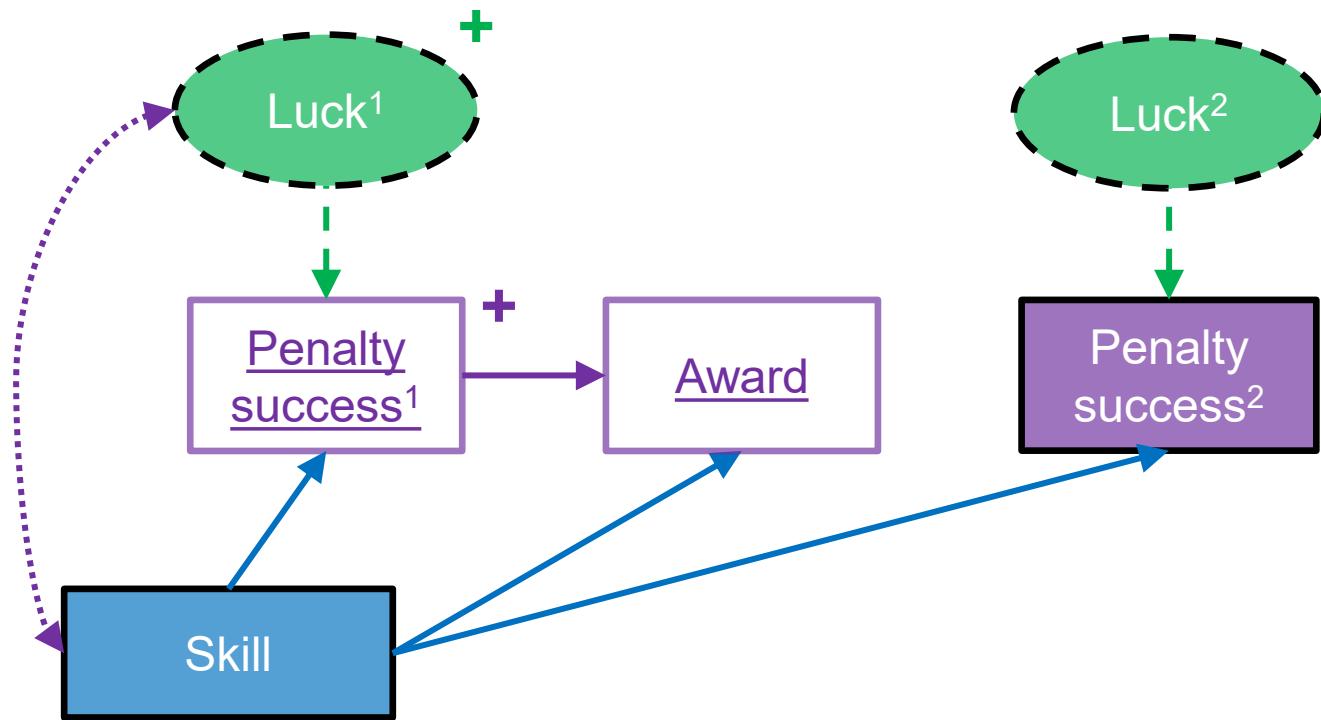
Geir Jordet
@GeirJordet

Replies to @GeirJordet

"Superstars" can be a liability in a penalty shootout.
AFTER receiving a prestigious individual award, players score 65% of their shots. PRIOR to receiving an award, they score 89%. Status adds pressure to an already high-pressure event! (4/13) tandfonline.com/doi/abs/10.108...



Jordet, G. (2009). When superstars flop: Public status and "choking under pressure" in international soccer penalty shootouts. *Journal of Applied Sport Psychology*, 21, 125-130.

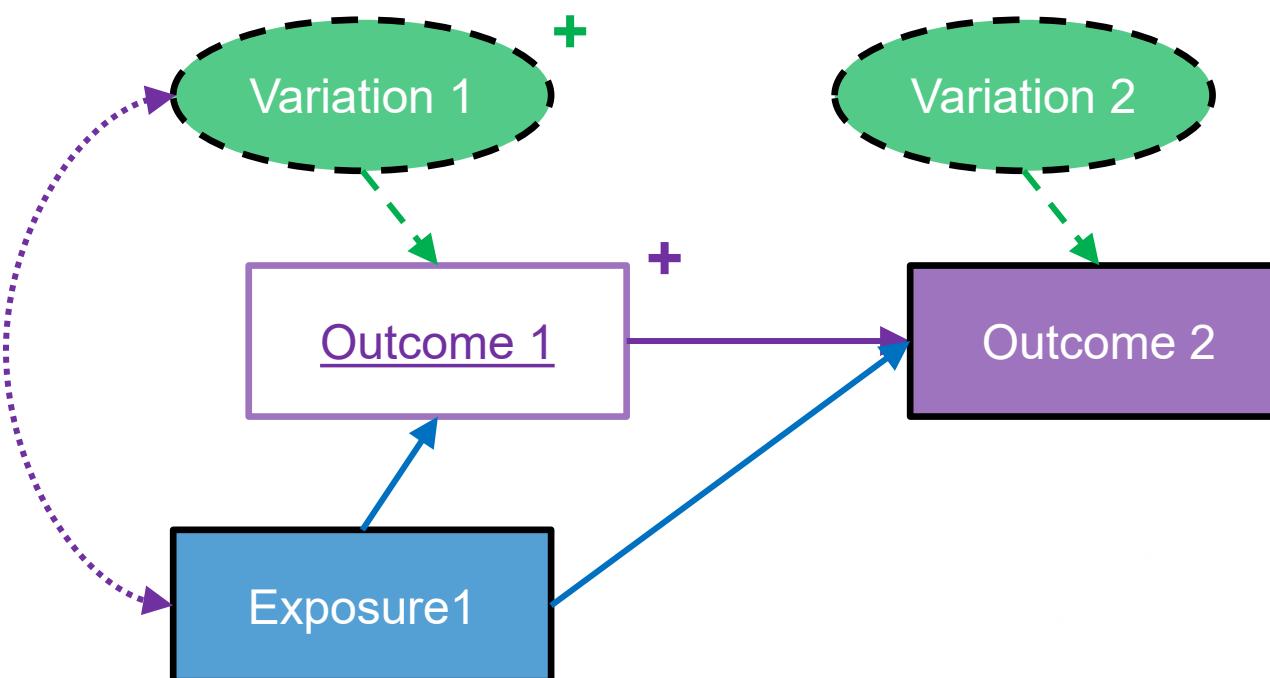


Skill → Penalty¹ is errored by **Skill** <...> **Luck¹** > Penalty¹
Skill → Penalty² has no such error bonus !

DAG EXPLANATION OF REGRESSION-TO-THE-MEAN

For any two outcome variables:

- when we select extreme values of **Outcome 1**, we select participants with high random variation for **Outcome 1**
- there is no conditioning on **Outcome 2**, so the mean of expected variation for **Outcome 2** is zero – i.e. less than for **Outcome 1**

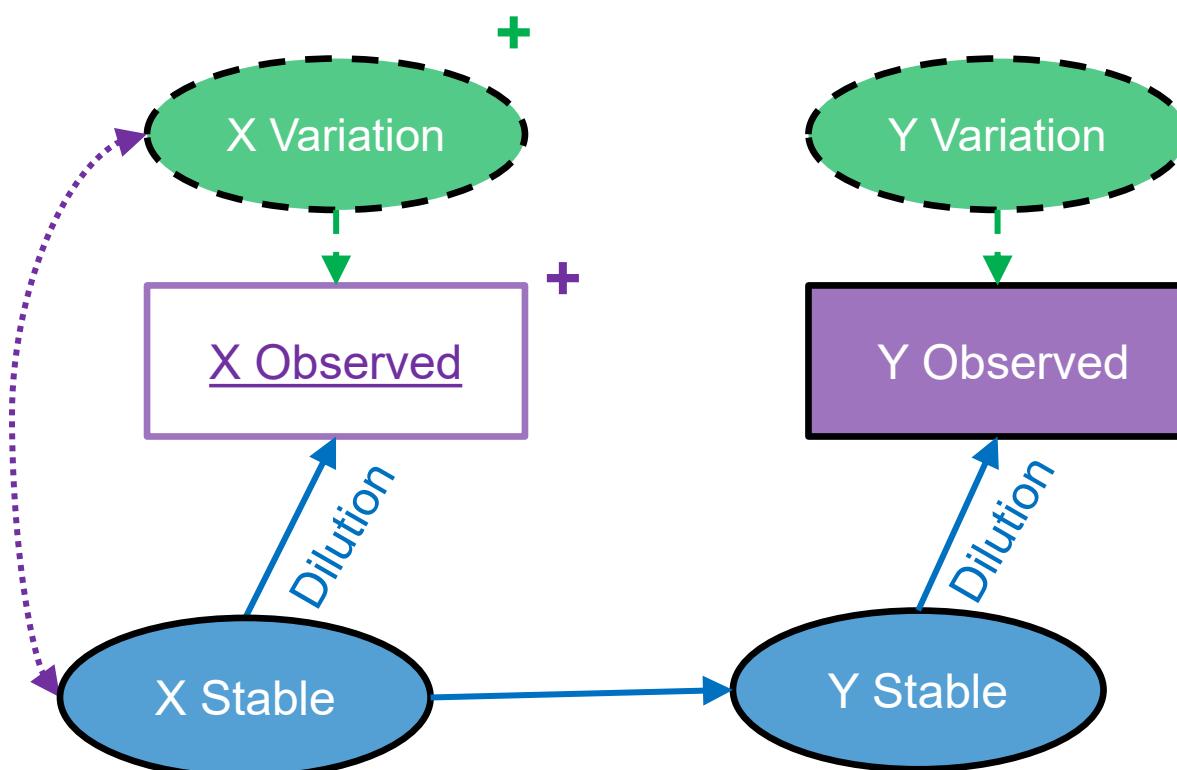


LATENT AND OBSERVED VALUES

'Measurement error DAGs' distinguish between the latent 'true' or 'stable' value of a variable and the 'errored' observed value:

- when we study extreme values of a variable, we invoke RTM
- but regardless, our estimate of $X \rightarrow Y$ is diluted because we must observe

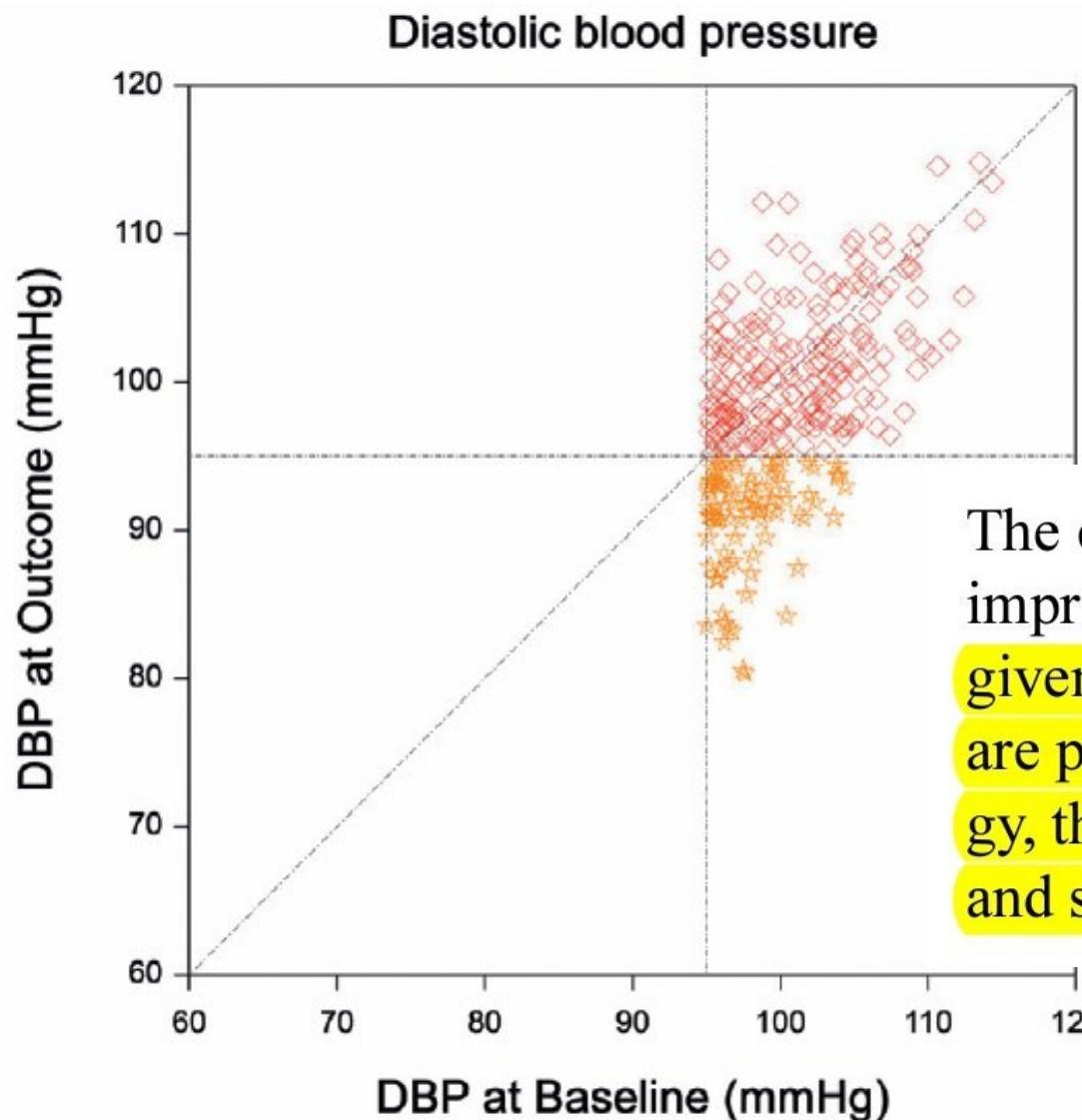
$X_O \leftarrow X_S \rightarrow Y_S \rightarrow Y_O$ – this is known as **regression dilution**



This diluting effect can be reduced by taking additional measurements of X and Y

More intensive data collection is the only really 'solution' to RTM

RTM AND THE PLACEBO EFFECT



The consequence is that on average patients will appear to improve even if the treatment is ineffective. In fact, patients given placebo can be expected to improve for reasons that are purely statistical. There is no need to invoke psychology, the healing hands of the physician, the white coat effect and so forth. The way that the data are collected suffices.

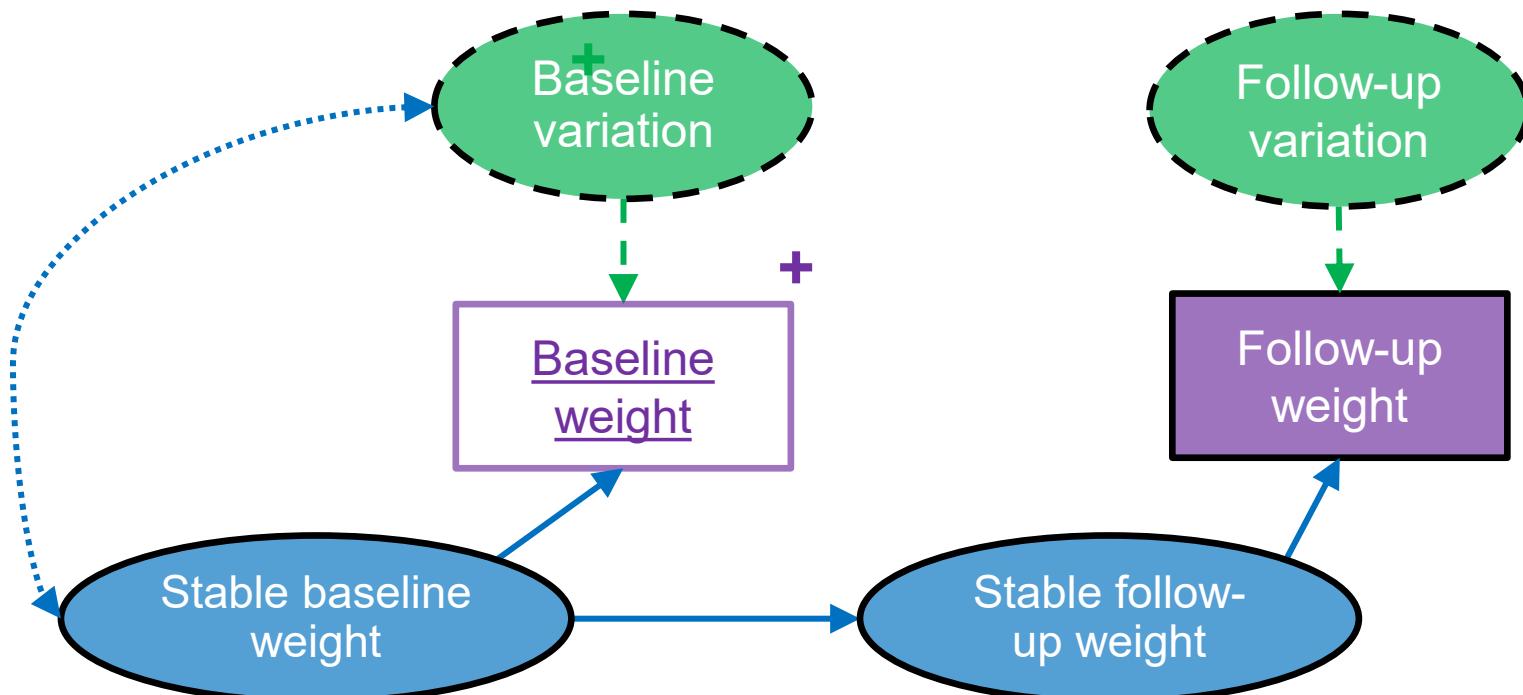
RTM AND THE PLACEBO EFFECT

A study recruits individuals with obese **baseline weight (+)**

Some will be higher than usual by chance (i.e. **baseline variation = +**)

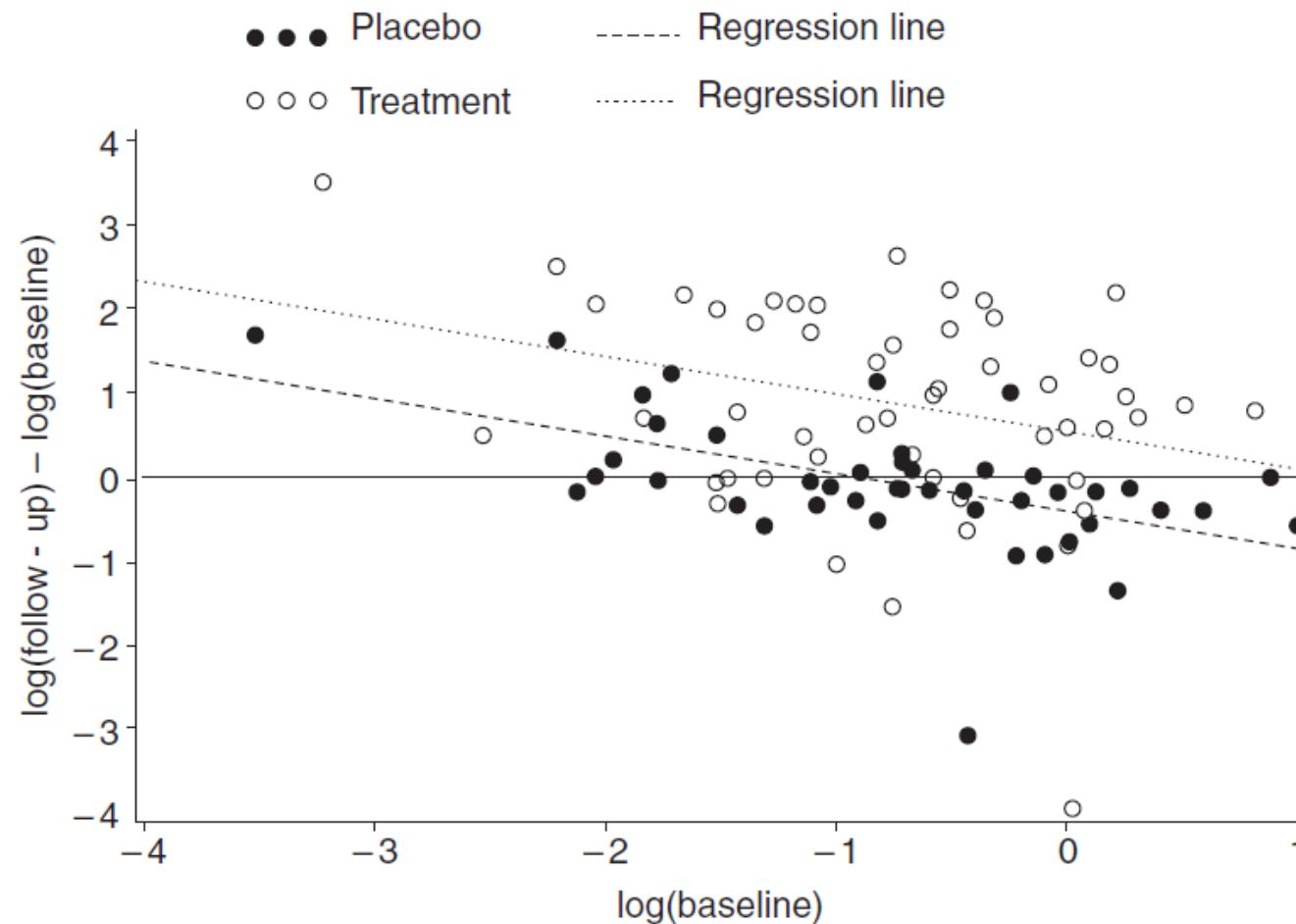
At follow-up, random variation will return to zero (i.e. **follow-up variation = 0**)

We attribute this miraculous weight loss to the **PLACEBO EFFECT**



WHY ARE RCTS IMMUNE?

Because RTM occurs in *both* the treatment and placebo group



RECOMMENDED READING

- Barnett, A.G., Van Der Pols, J.C. and Dobson, A.J., 2005. Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, 34(1), pp.215-220.
- Morton, V. and Torgerson, D.J., 2005. Regression to the mean: treatment effect without the intervention. *Journal of Evaluation in Clinical Practice*, 11(1), pp.59-65.
- Linden, A., 2013. Assessing regression to the mean effects in health care initiatives. *BMC medical research methodology*, 13(1), pp.1-7.
- Gadd, S.C., Tennant, P.W., Heppenstall, A.J., Boehnke, J.R. and Gilthorpe, M.S., 2019. Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in lifecourse research. *PloS one*, 14(12), p.e0225217.

SUMMARY

- **Regression-to-the-mean** is a universal phenomenon that arises whenever two variables are imperfectly correlated
- Regression-to-the-mean can be misleading, suggesting change where only random variation is taking place
- Examining extreme samples – by **conditioning-on-the-outcome** – can cause extremely misleading **collider-error**
- We can understand RTM, conditioning-on-the-outcome, and collider error using DAGs that include **random variation terms**
- Hopefully, this will help people to understand RTM, which remains extremely poorly understood

3.4 - NATURAL EXPERIMENT APPROACHES

GEORGIA



@GEORGIATOMOVA

MARK



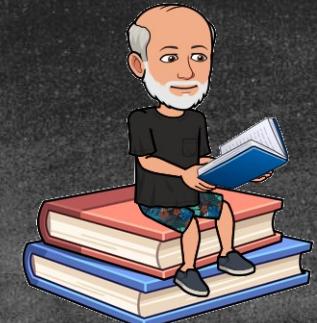
@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 3

09:30-10:15 ACTIVITY 3-A

10:15-11:00 LECTURE 3.1

11:00-11:30 TEA & COFFEE

11:30-12:45 LECTURE 3.2

12:45-13:00 Q&A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 3.3

14:45-15:30 ACTIVITY 3-B

15:30-16:00 TEA & COFFEE

16:00-17:00 LECTURE 3.4

17:00-17:45 ACTIVITY 3-C

17:45-18:00 Q&A

LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- Discuss the potential of natural experiments for estimating causal effects
- Explain the core principles and caveats behind a range of natural experiment approaches including:
 - ✓ **Instrumental variable** approaches (including **Mendelian randomisation**)
 - ✓ **Difference-in-differences**
 - ✓ **Regression discontinuity designs**
 - ✓ **Interrupted time series**

NATURAL EXPERIMENTS

Natural experiments are observational studies where ‘nature’ (i.e. **exogenous forces**) approximates the conditions of an experiment or **quasi-experiment**

(True) natural experiment: exposure assigned ‘as random’ (i.e. independent of the propensity of the outcome) creating **unconditional exchangeability**

Quasi natural experiment: exposure assignment not strictly random, hence units are not unconditionally exchangeable

Natural experiments may occur due to:

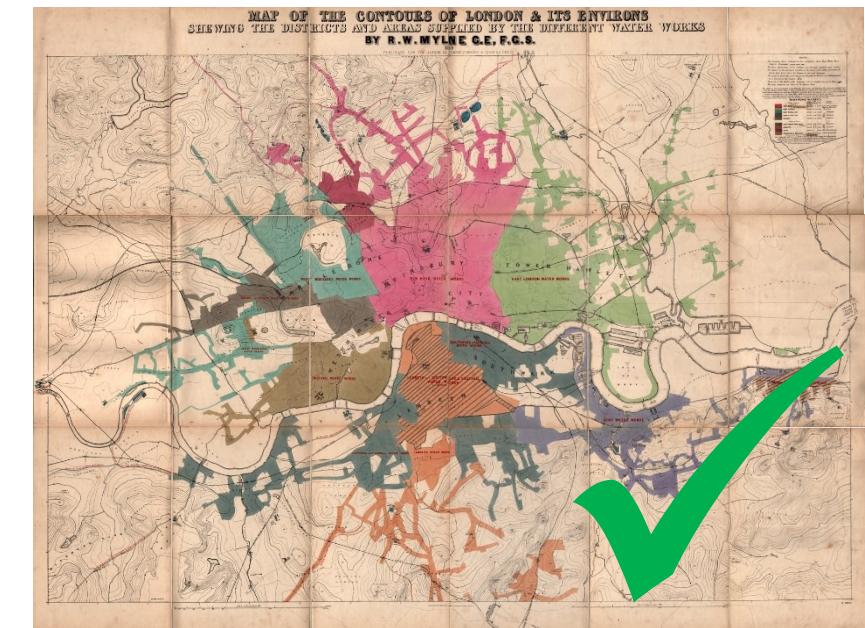
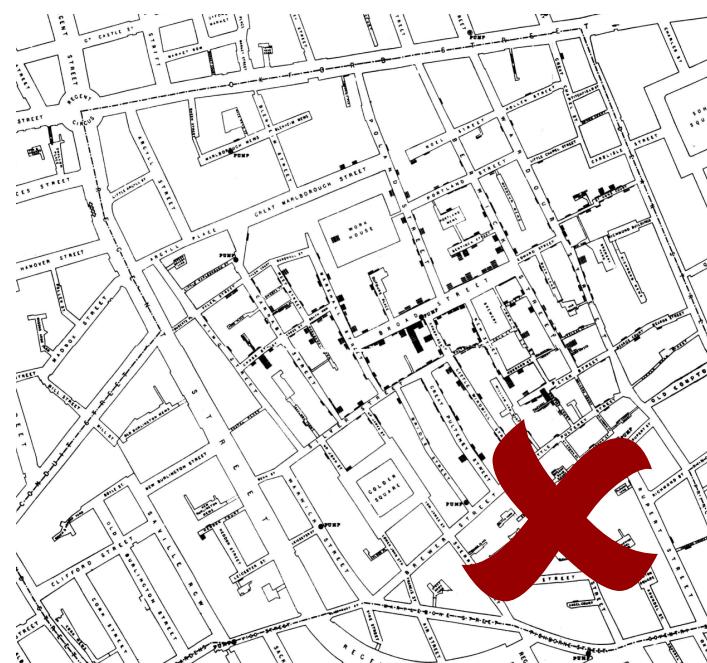
- ‘acts of God’ (e.g. weather, climate, disasters); which involve ‘no human agency’
- geo-political events (e.g. war, famine, recession, Brexit)
- government or policy changes (e.g. smoking ban, sugar tax, austerity)
- other ‘exogenous’ changes (e.g. staff moved to new open-plan office)

THE GRAND EXPERIMENT



John Snow, often described
as the 'first Epidemiologist'

Snow's **Grand Experiment** (August 1854) remains the finest example of a true natural experiment
It was this – not his subsequent study of the 1854 Soho cholera outbreak – that was his real masterpiece!



THE GRAND EXPERIMENT



John Snow, often described
as the 'first Epidemiologist'

Snow believed cholera was contracted by ingesting 'Cholera poison' from water contaminated with human sewerage

But there's no way he could have performed an experiment !

- impractical
- unethical

THE GRAND EXPERIMENT



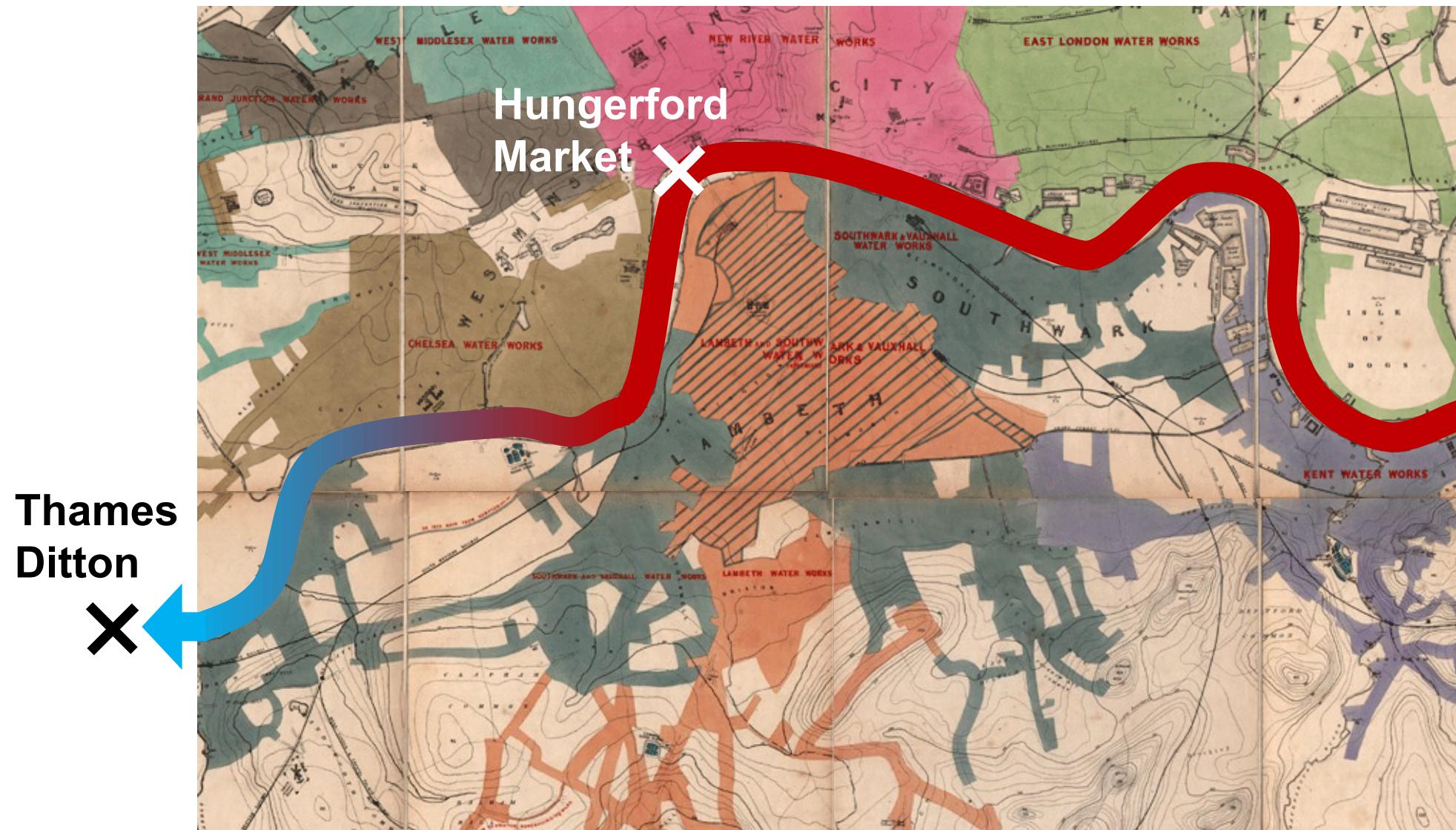
John Snow, often described
as the 'first Epidemiologist'

Snow noticed an opportunity for a natural experiment

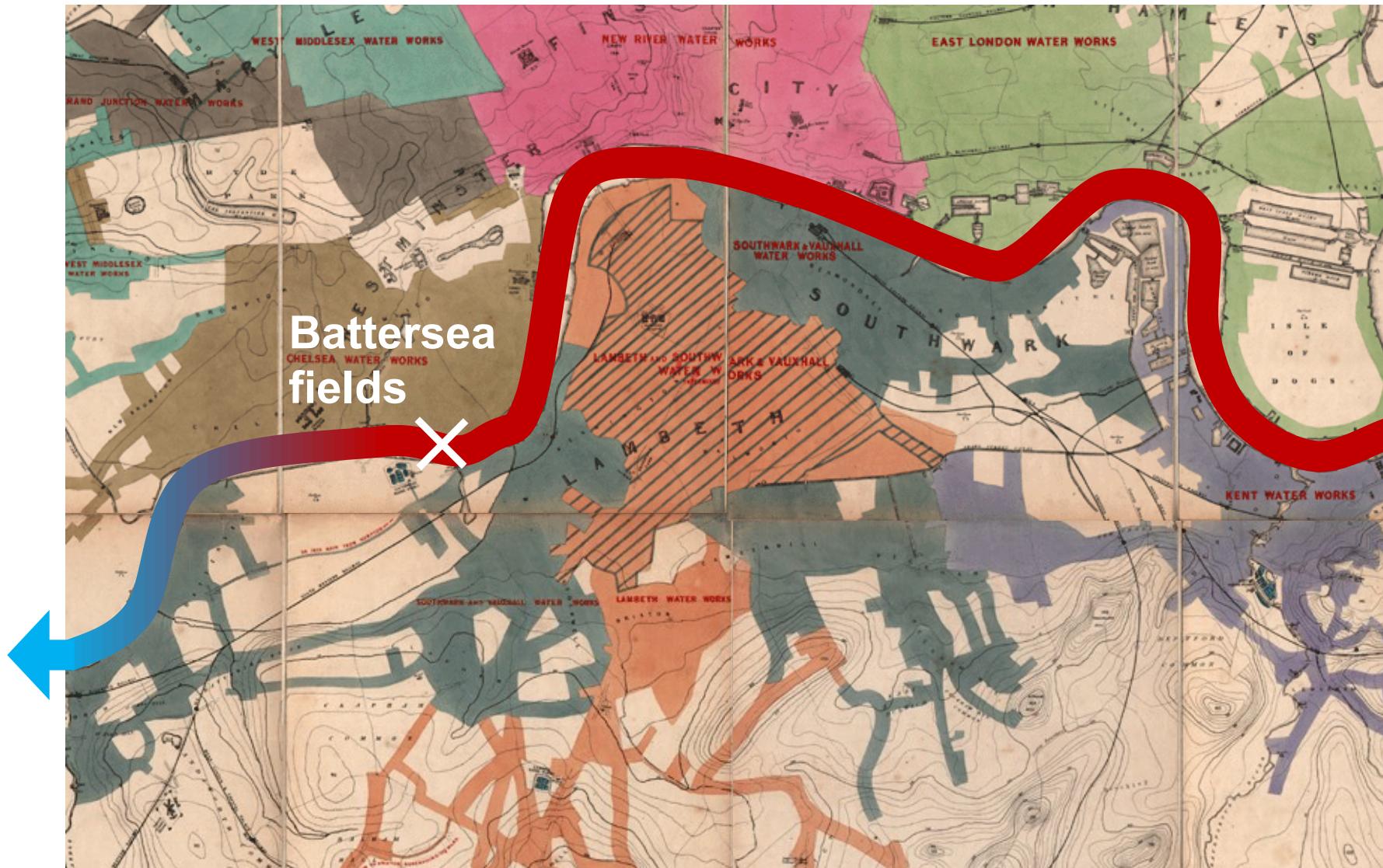
"London was without cholera from the latter part of 1849 to August 1853..."

"During this interval an important change had taken place in the water supply of several of the south districts of London"

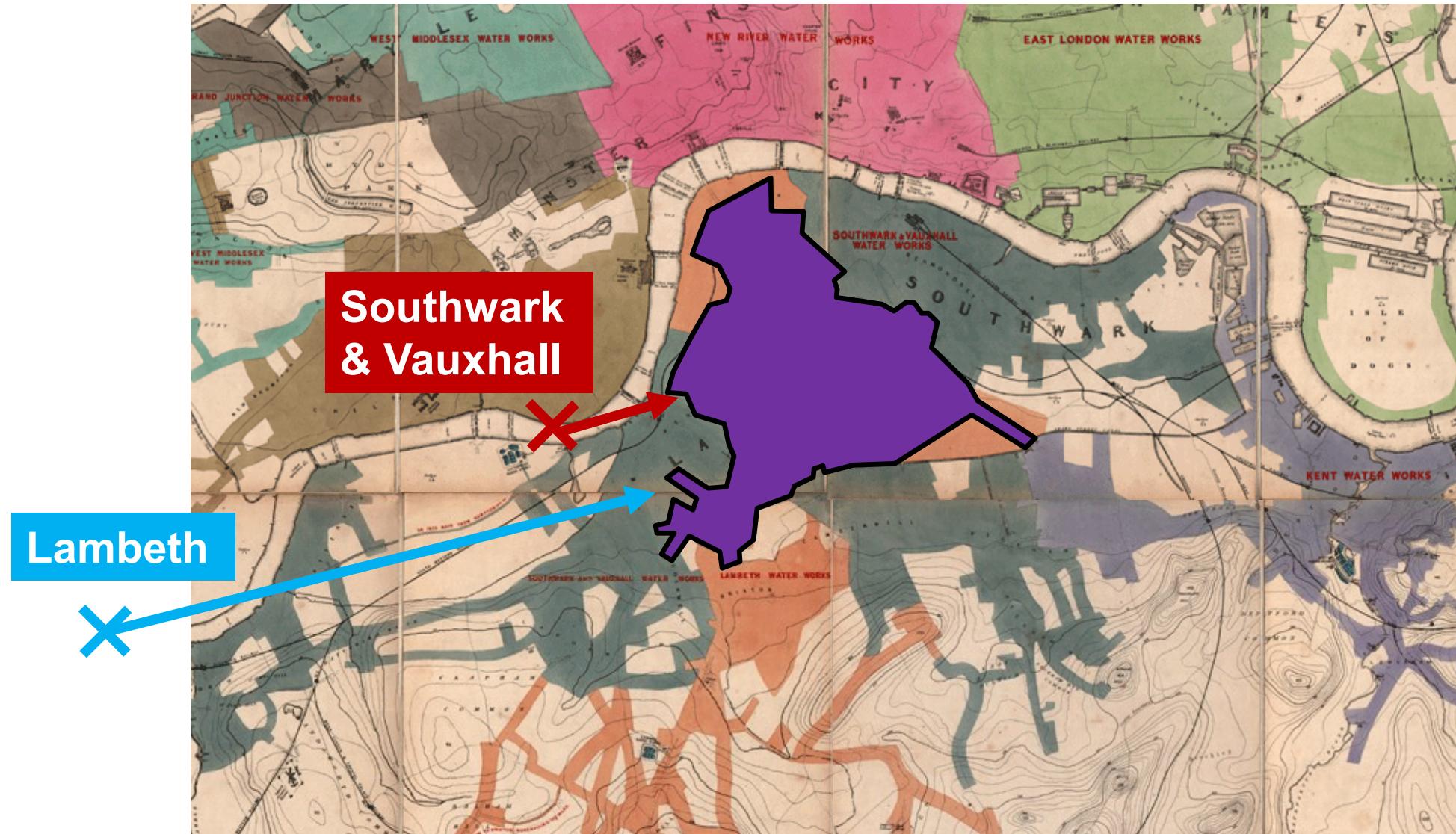
"The Lambeth Company removed their water works from opposite Hungerford Market to Thames Ditton; thus obtaining a supply of water quite free from the sewage of London"



"The Southwark and Vauxhall Company ... derived their supply from the Thames at Battersea Fields"



"The two Companies were in active competition ... the pipes of the Lambeth Water Company and those of the Southwark and Vauxhall Company pass together down all the streets of several of the south districts."



THE GRAND EXPERIMENT



John Snow, often described
as the 'first Epidemiologist'

Units were **unconditionally exchangeable**:

- “*The pipes of each Company go down all the streets, and into ... all the courts and alleys ... Each company supplies both rich and poor ... large houses and small ... there is no difference ... in the condition or occupation of the persons receiving the water of ... (two) Companies*”
- “*Three hundred thousand people of both sexes ... every age and occupation ... every rank and station... were divided into two groups without their choice and... (or) their knowledge (into those) supplied with water containing ... sewage ... (or) water ... free from such impurity*”

THE GRAND EXPERIMENT



John Snow, often described
as the 'first Epidemiologist'

Able to directly compare mortality ratios between
'exposed' (S&V) and 'unexposed' (Lambeth) to
estimate **causal risk ratio**:

	Number of houses.	Deaths from Cholera.	Deaths in each 10,000 houses.
Southwark and Vauxhall Company	40,046	1,263	315
Lambeth Company	96,107	98	37
Rest of London	956,423	1,422	59

$$315 / 37 = 8.5$$

DUTCH HUNGER WINTER FAMILIES STUDY



Dutch ‘Hunger Winter’ (1944-45)

During October 1944-1945, people living in the Western Netherlands suffered extreme hardship under Nazi occupation

Extreme famine: calorie intake averaged 500-600 per person per day

~20,000 people died from malnutrition and starvation

Babies were born ~300g smaller

DUTCH HUNGER WINTER FAMILIES STUDY

- The '**Dutch hunger winter**' was recognised as an (unfortunate) '**natural experiment**' of undernutrition in utero
- 1987: **Dutch Hunter Winter Families Study** established from >1000 women born during August 1944 - April 1946
- The cohort includes a mix of people with different exposures to the famine, including some born before and some born after

TRUE NATURAL EXPERIMENTS ARE RARE

The wrong exposure

- We're often not really interested in the 'natural exposure':
 - ✓ out of our control
 - ✓ unsuitable for intervention
- Dutch Hunter Winter Families study was not interested in 'famine', but the effect of nutrient and calorie intake during pregnancy

Positivity violations

- Some exposures are only given to certain groups or subgroups

Imperfect counterfactuals

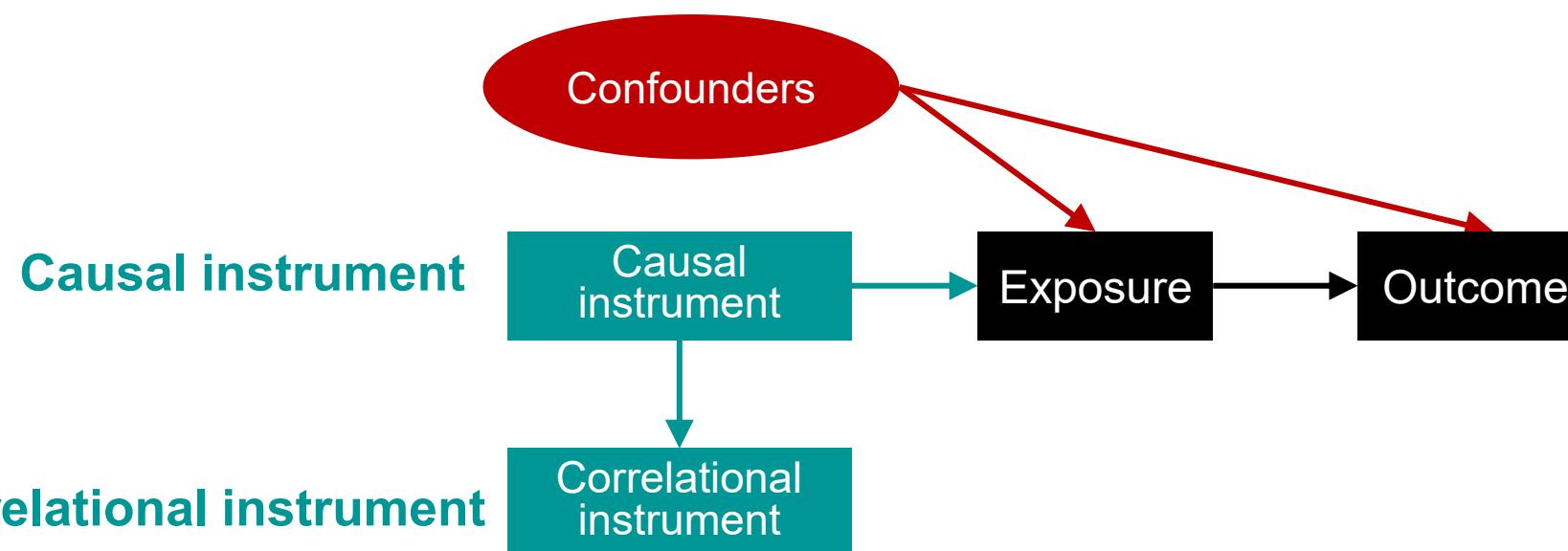
- Few exposures are sufficiently random or short / reversable to provide unconditional exchangeability; many 'natural experiments' are therefore simply **pre-post studies**

INSTRUMENTAL VARIABLES

'Natural exposures' can still (potentially) provide useful estimates of causal effects of an exposure of interest if they behave as an **instrumental variable**

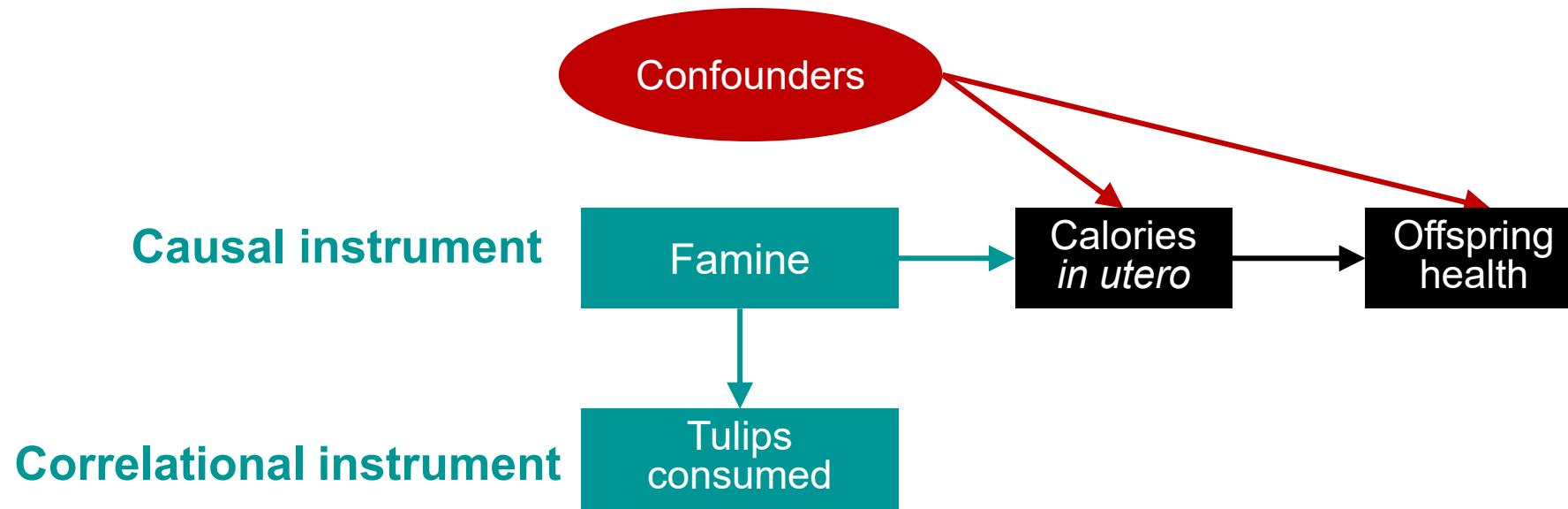
Instrumental variable:

- associated with **exposure**
- **NO** (residual) association with **outcome**, other than through exposure
- **NOT** associated with **ANY** confounders



INSTRUMENTAL VARIABLES

By contrasting levels of the instrumental variable, we can examine how (corresponding) changes in the exposure effect the outcome ...

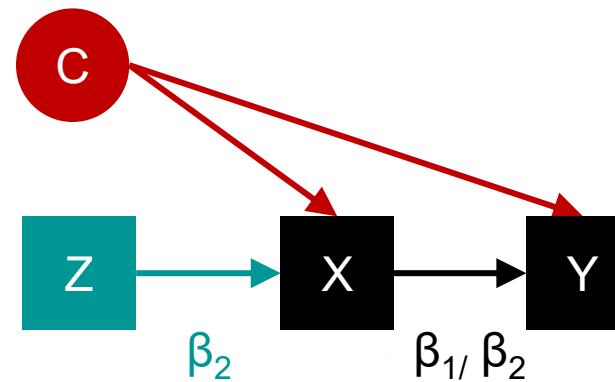
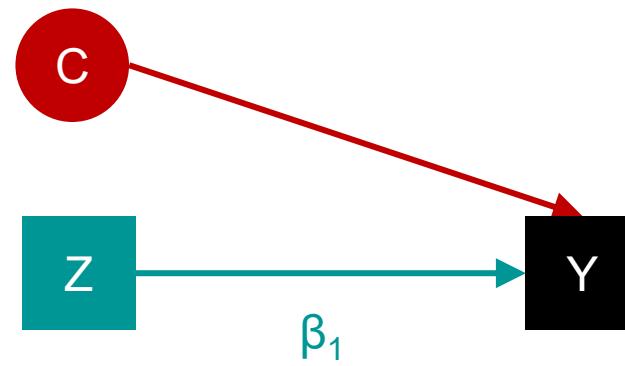


INSTRUMENTAL VARIABLES

The **total causal effect** of an exposure on an outcome can be estimated from the ratio of the (apparent) effect of the instrument on the outcome divided by (apparent) effect of the instrument on the exposure

If: $Y \sim \beta_1 Z$ and $X \sim \beta_2 Z$

Then: $Y \sim (\beta_1 / \beta_2)X$

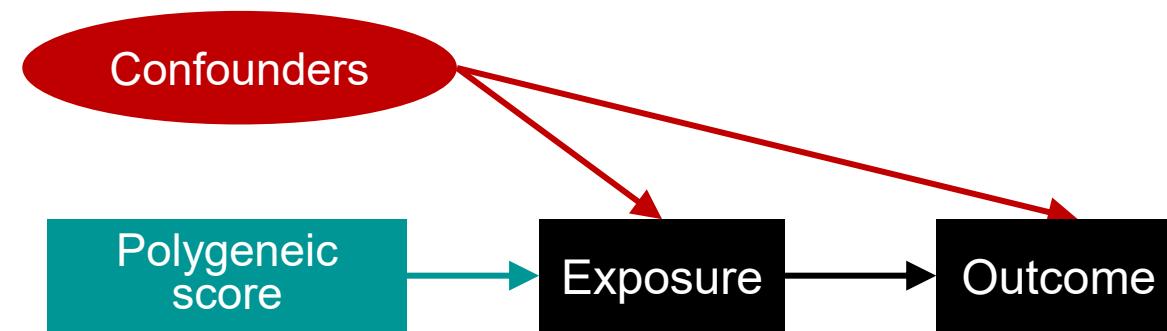


MENDELIAN RANDOMISATION

Genetic variants are also increasingly used as instrumental variables in an approach termed '**Mendelian randomisation**'

Individual genes are identified from **genome-wide association studies (GWAS)** that are associated with the exposure but have no residual association with the outcome (or marker thereof)

Instrumental variables (**polygenic scores**) are then constructed from multiple genes that are *associated* with the **propensity of exposure**



IV ASSUMPTIONS AND CAVEATS

Instrumental variable estimates are very sensitive to their assumptions !

Rare to find variables associated with exposure *and* no residual association with outcome / confounders; impossible to know these are true causal effects

- e.g. genes are pleiotropic, i.e. have multiple effects
- bias can be reduced by testing multiple instruments made from different variables

If instrument is ‘weak’ (i.e. weakly associated with exposure), then signal can be overwhelmed by modest assumption violations

- genetic associations are particularly weak

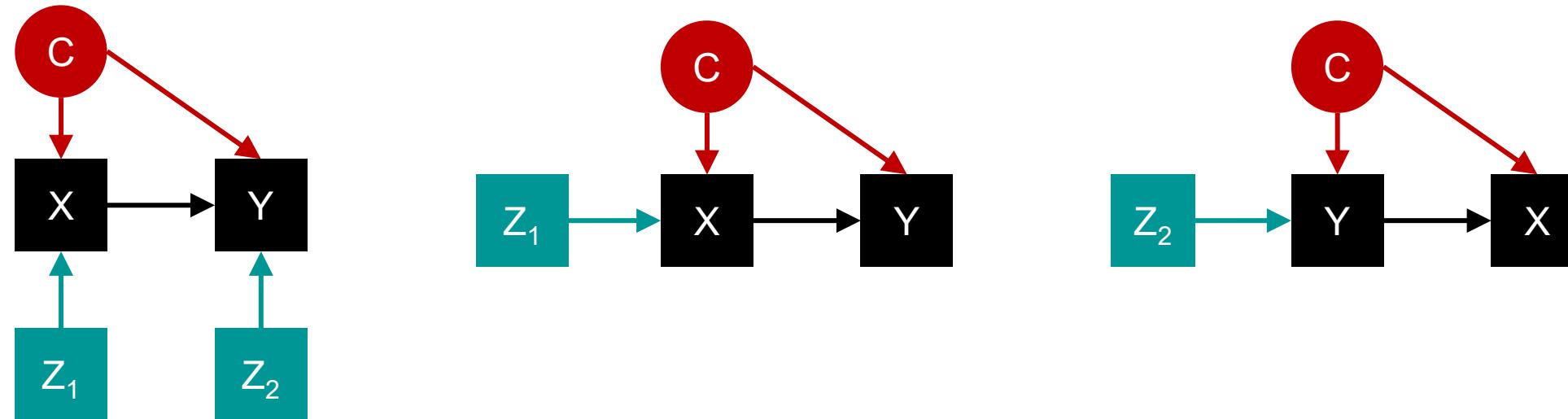
Selection bias can distort true association of instrument with exposure/outcome

- should test assumptions in different populations and/or correct for selection bias

IV ASSUMPTIONS AND CAVEATS

Doesn't guarantee the direction of causality (i.e. that X precedes Y)

- This can be examined by identifying an instrumental variable for the outcome and testing the association in reverse



Not suitable for **time-varying relationships**

Vanderweele: IV studies may be stronger for *refuting* suspected associations, because it is unlikely that various biases will balance to zero

MENDELIAN RANDOMISATION EXAMPLE

Example: Cannabis use and schizophrenia risk

SNP = Single nucleotide polymorphism



Gage et al 2017 *Psychol Med.* 2017;47(5):971-980

Cannabis → Schizophrenia: OR=1.04 (1.01-1.07)

Schizophrenia → Cannabis: OR=1.10 (1.05-1.14)

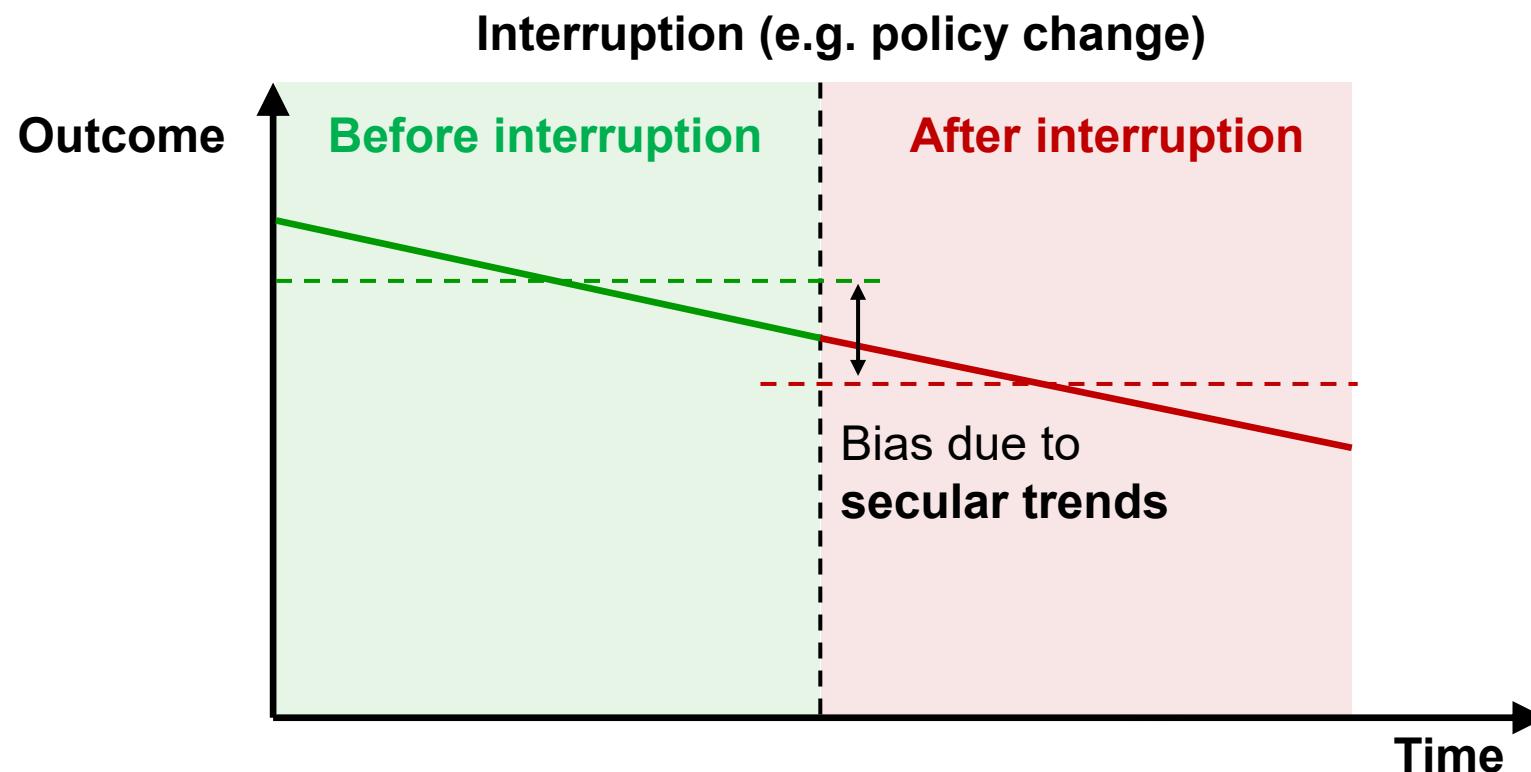
Interpreted as evidence of small causal effect of Cannabis on Schizophrenia

But: larger effect of Schizophrenia on Cannabis attributed to '*strong instruments for schizophrenia*'

IMPERFECT COUNTERFACTUALS

Many 'natural experiments' are **pre-post studies**

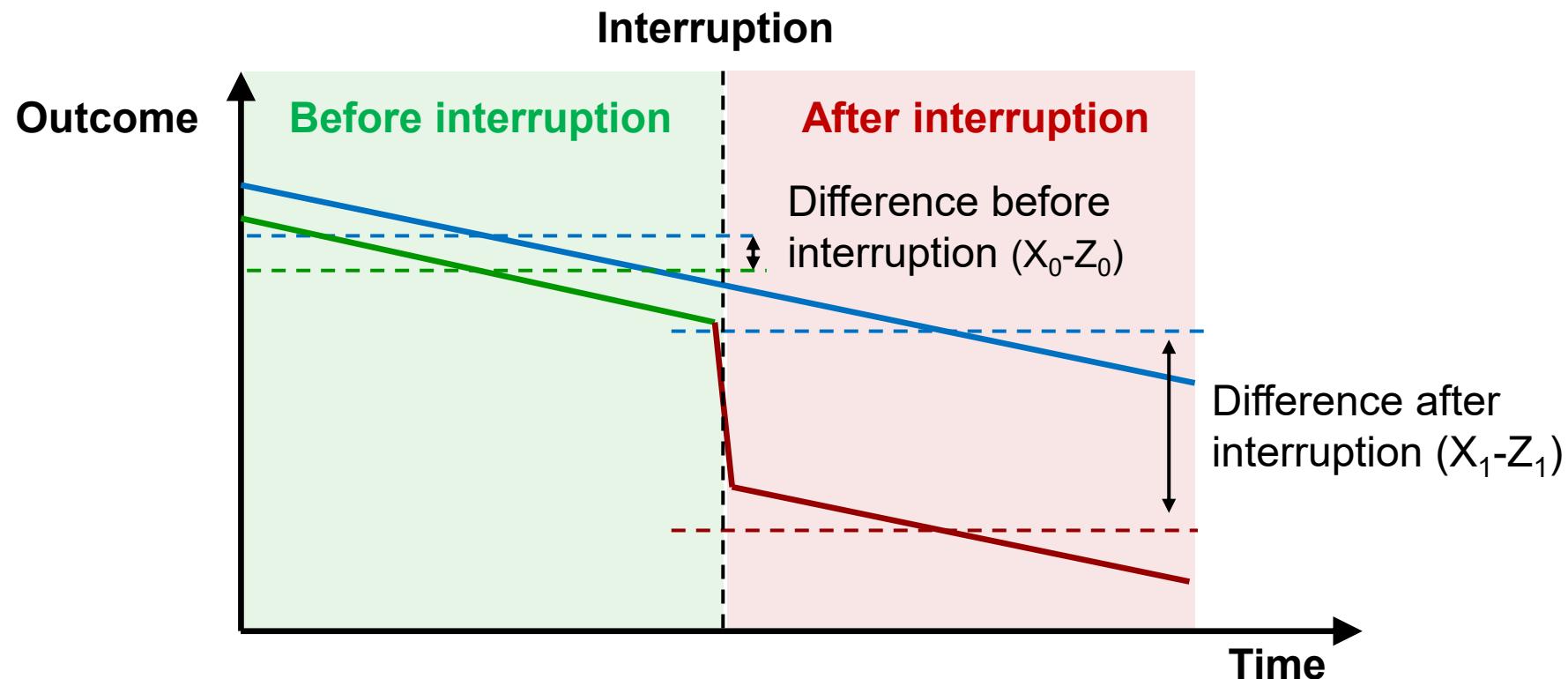
Problem: pre-post averages are rarely exchangeable, due to secular trends, seasonal trends, and other changes



DIFFERENCE-IN-DIFFERENCES

Better if you have a comparison (e.g. Z) to serve as your counterfactual

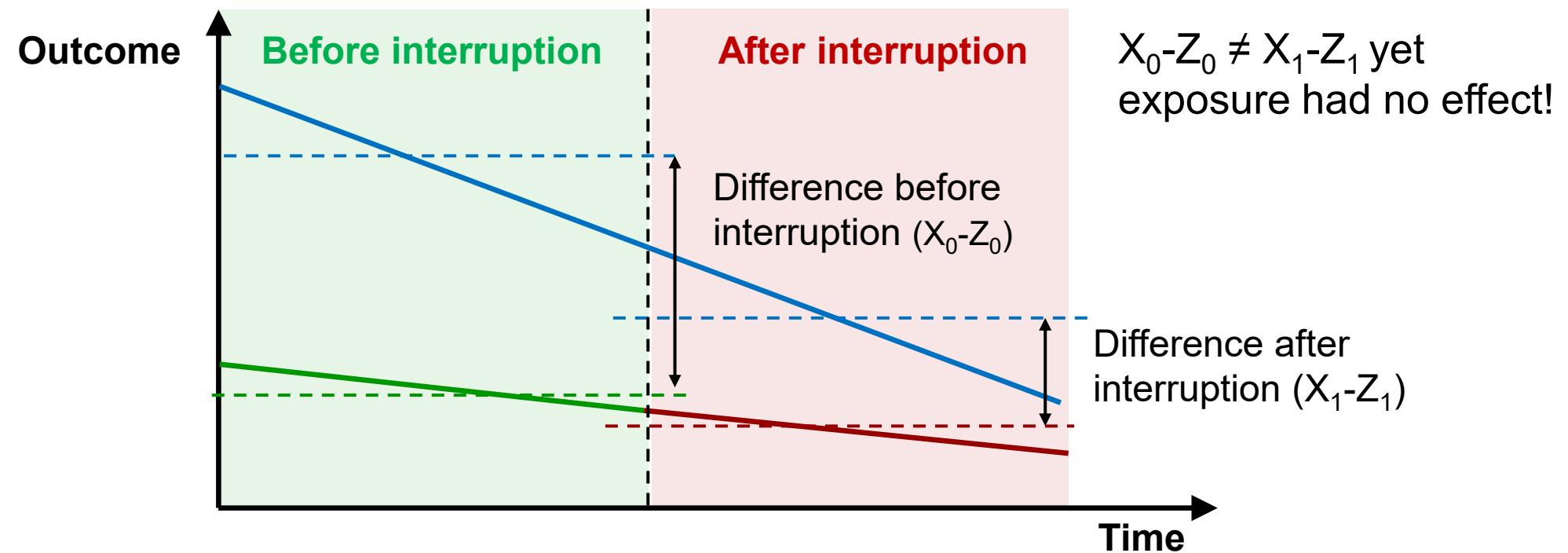
- then compare the ‘difference-in-differences’ between ‘**exposed**’ & ‘**unexposed**’
- i.e. $[X_1 - Z_1] - [X_0 - Z_0]$



DID ASSUMPTIONS & CAVEATS

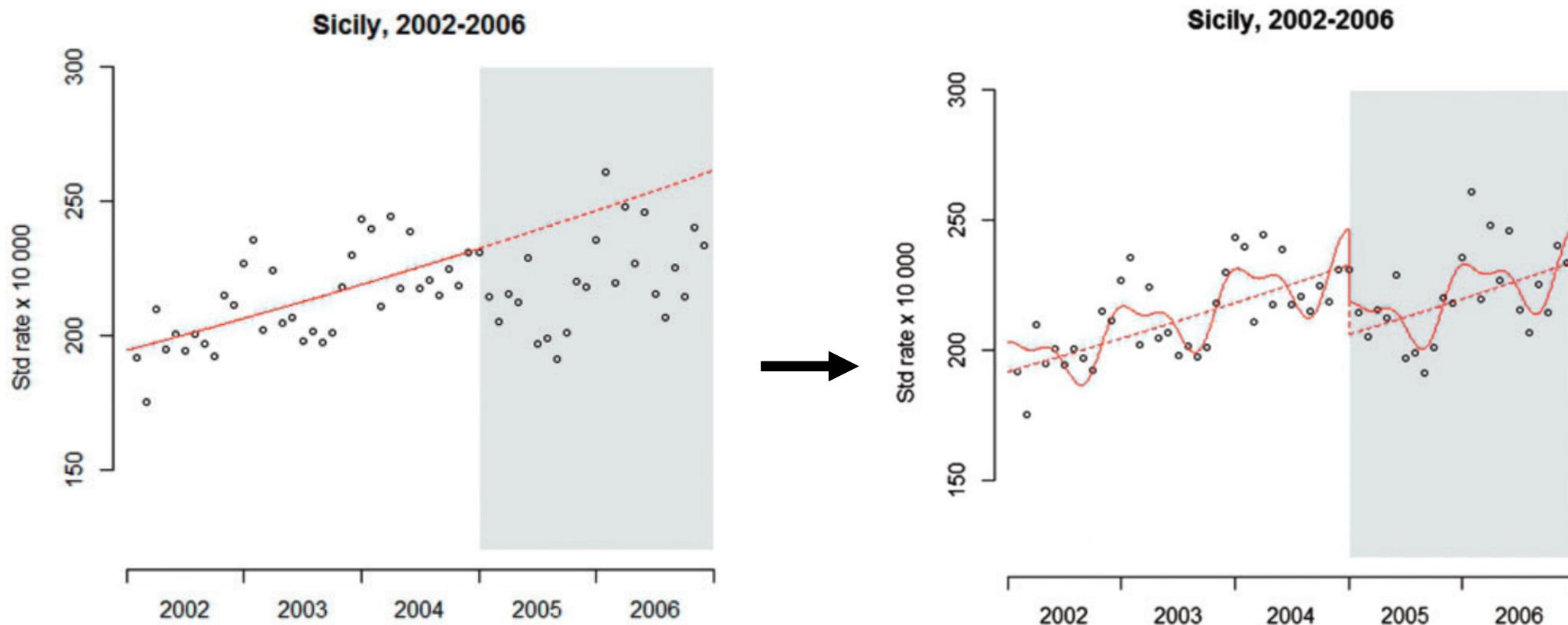
DiD assumes **common trends** in exposed & unexposed

This is often a naive assumption



INTERRUPTED TIME-SERIES

In **time-series data** (i.e. intensive repeated-measure data), counterfactual estimates are obtained by modelling **secular trends** and **seasonal trends** and examining if the outcome changes in line with exposure ‘interruptions’

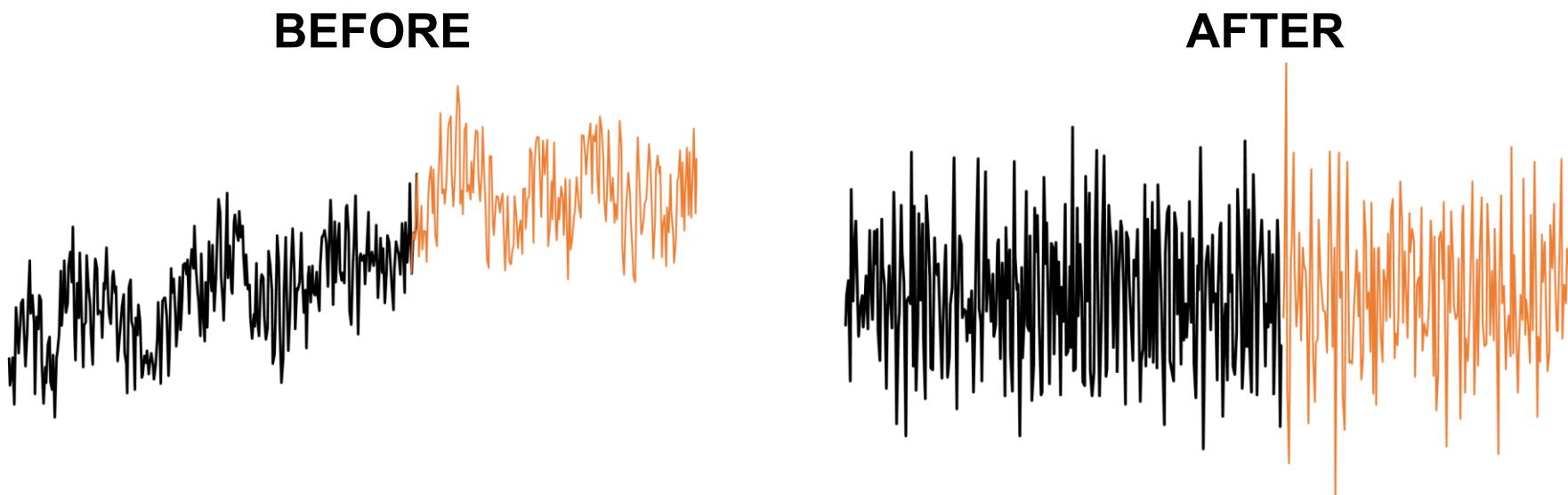


INTERRUPTED TIME-SERIES

Interrupted time-series are classically analysed using autoregressive-moving average (ARMA) regression models

STAGE 1: Model secular trends and seasonal trends by adding autoregressive (AR) and moving average (MA) terms

- as well as providing an improved estimate of the counterfactual, this ensures successive observations are independent



INTERRUPTED TIME-SERIES

STAGE 2: The effect of the exposure ‘interruption’ is often examined with a simple exposure-time interaction; which describes an instantaneous ‘step change’ and subsequent change in trend

- e.g. assuming linearity, X = exposure / intervention, and T = time

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X + \beta_3 TX + [\text{ARMA terms}] + \varepsilon_t$$



Trend in outcome over
time (before exposure
interruption)



Instant effect
(step change)
of exposure



Lagged
effect (trend
change) of
exposure



Autoregressive terms;
interpretation not important

ITT ASSUMPTIONS AND CAVEATS

- Needs lots of data points on both sides of the exposure interruption
- Autoregressive structure must be accurately modelled
- Trends in pre- and post-time periods must be similar and possible to model
- Assumes ‘stable’ confounding in pre- and post-time periods (i.e. no major changes other than primary exposure)

Controlled time-series:

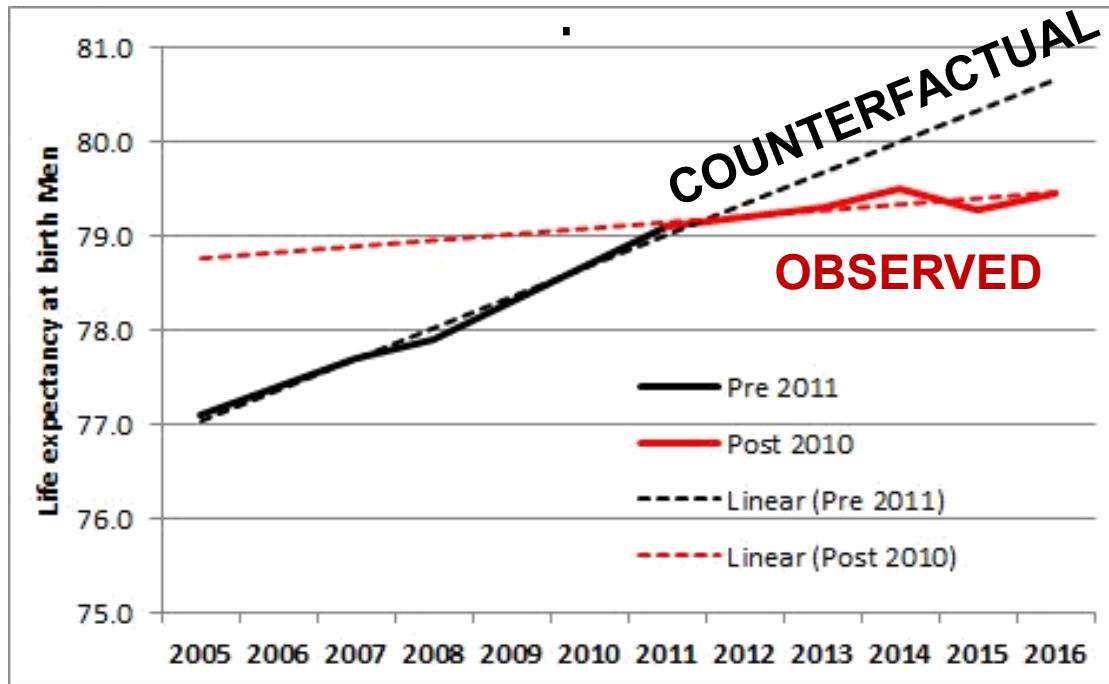
- greater confidence can be achieved by finding comparison time-series data
- a ‘synthetic control’ could be constructed from an external population (e.g. different country) that did not experience exposure interruption using a representative selection of units

INTERRUPTED TIME-SERIES

Example: Austerity politics and life expectancy



INSTITUTE *of*
HEALTH EQUITY



<http://www.instituteofhealthequity.org/>

the guardian

Rise in life expectancy has stalled since 2010, research shows

Sir Michael Marmot, a former government adviser, highlights 'miserly' levels of spending on health and social care



Sir Michael Marmot says the needs of the ageing population will place the NHS and social care under strain in the near future. Photograph: Alamy Stock Photo

POSITIVITY VIOLATIONS

Many social exposures and interventions are deliberately assigned according to the probability of the outcome, introducing a **positivity violation**

Example: selective schools

- difficult to compare performance compared with non-selective schools, because children must ‘pass’ a certain performance threshold for admittance

Example: statins and cardiovascular disease

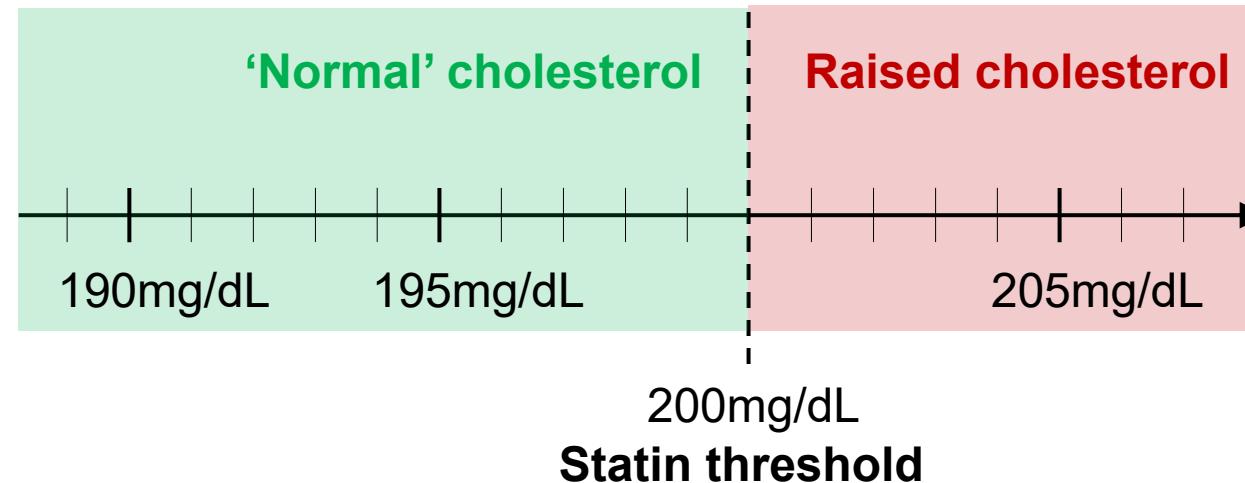
- difficult to examine the effect of statins on cardiovascular disease risk, because people must have ‘raised risk’ (e.g. high cholesterol) for prescription

Example: ‘fat letters’ and adult weight

- difficult to examine the effect of ‘fat letters’ on adult weight, because children must be ‘obese’ to receive them

WINDOW OF OPPORTUNITY

If we believe individuals either side of threshold are exchangeable, we could compare outcomes between those just below and just above



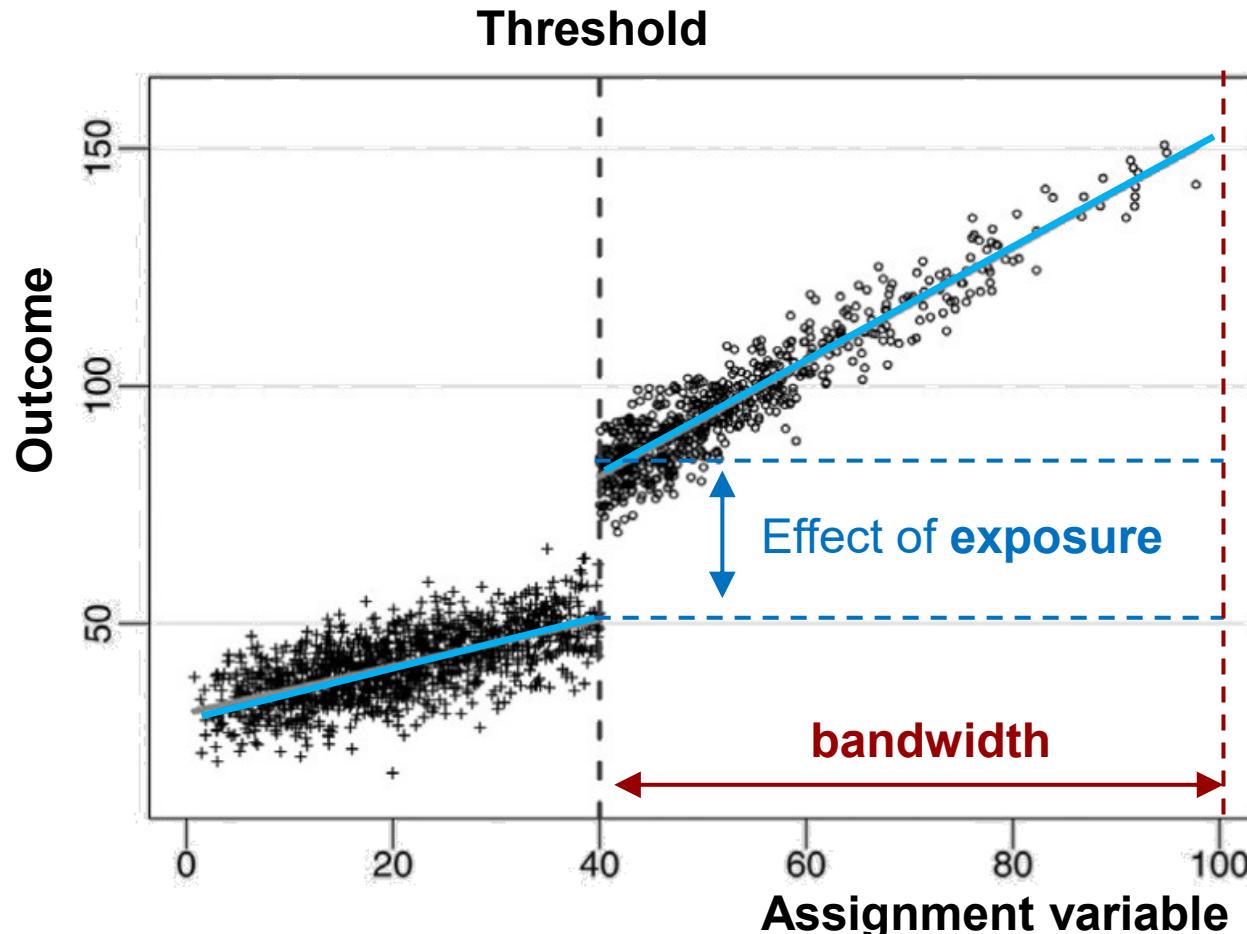
Suppose cholesterol blood test had $\pm 5\text{mg/dL}$ random error:

- we could compare outcomes between **195-199.9 mg/dL** with **200-205 mg/dL**
- we would need a **VERY** large sample size!
- but are those with **195 mg/dL** and **205 mg/dL** entirely 'exchangeable'? Often implausible!

REGRESSION DISCONTINUITY DESIGN

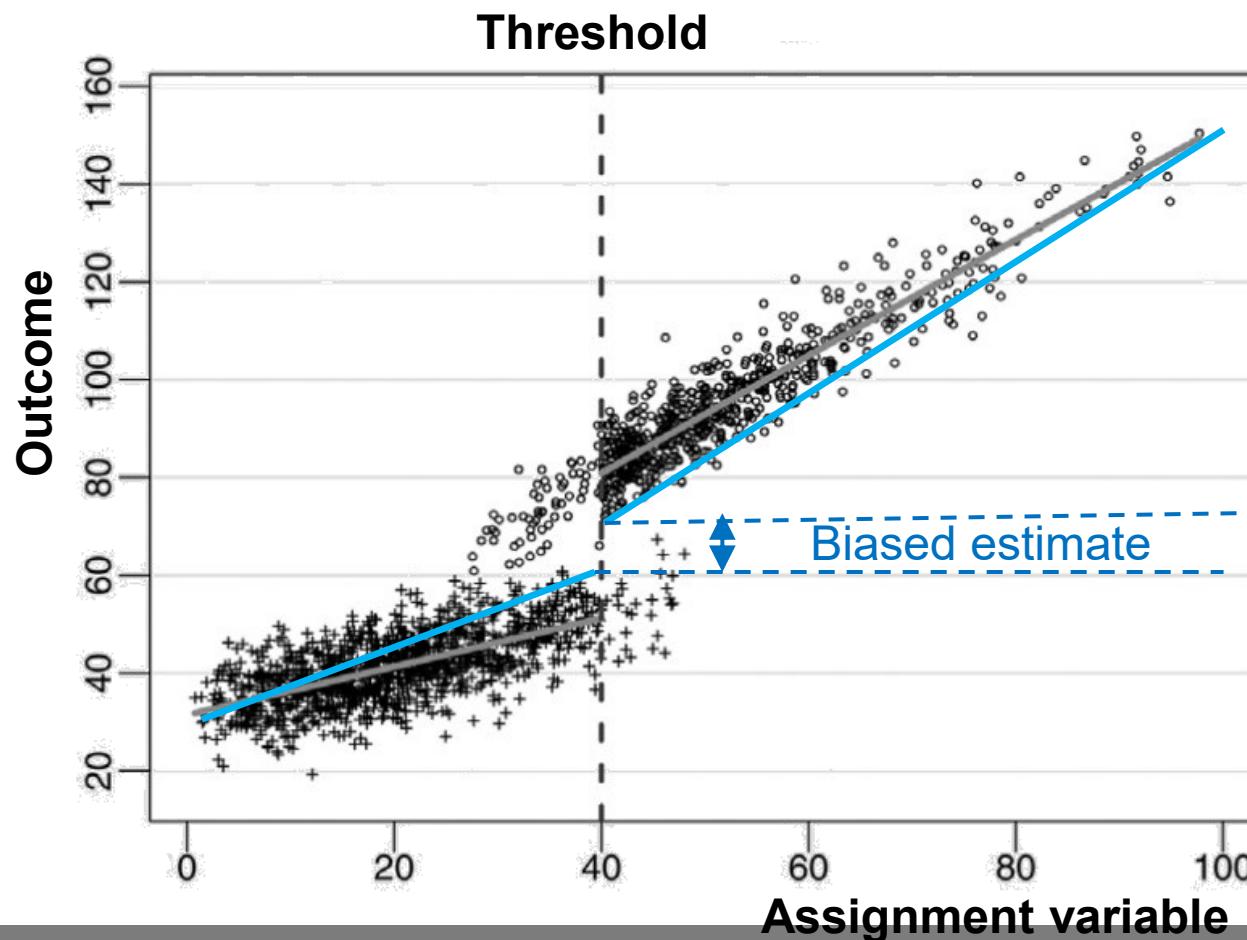
Better to model any small differences in outcome probability using regression

This approach is known as a **regression discontinuity design**



FUZZY ASSIGNMENT

Problem: The assignment of the exposure is not always ‘sharp’. Often, there is variation above and below the threshold creating ‘fuzzy’ assignment

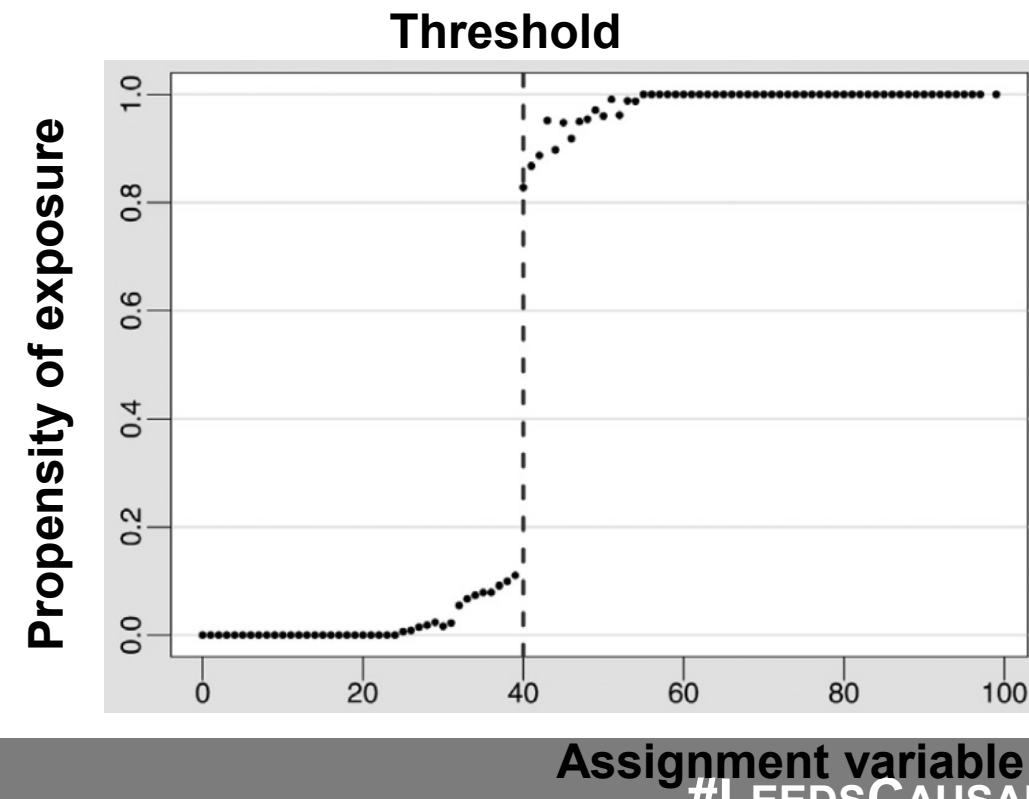
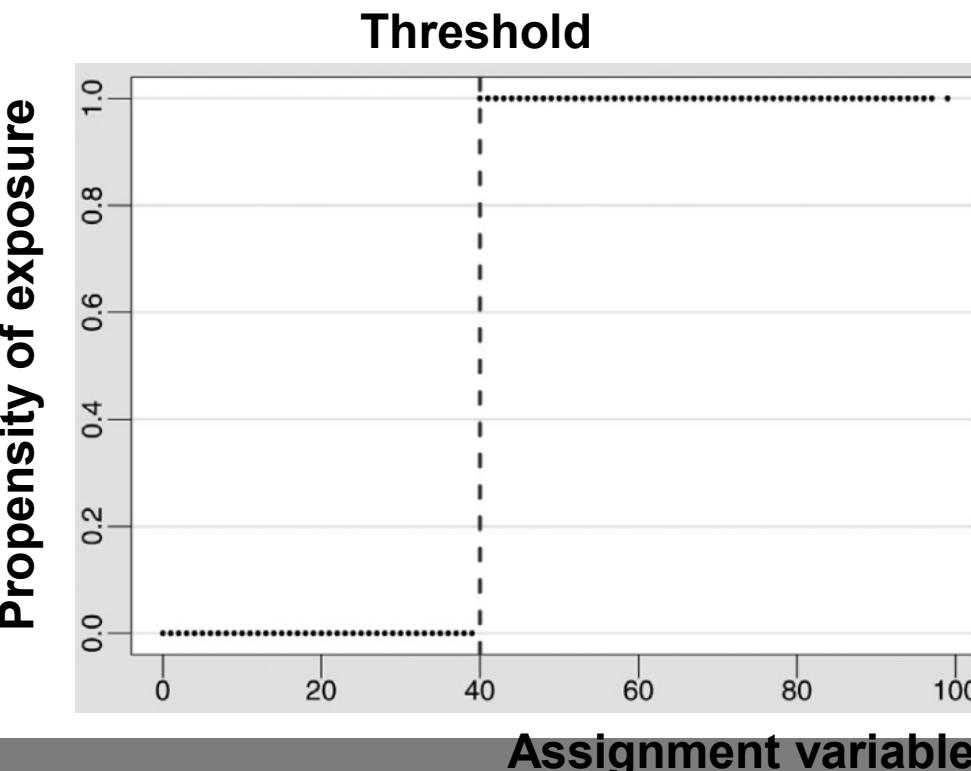


- If not recognised, this would dilute the apparent effect

PROBABILITY OF EXPOSURE

In a **sharp** regression discontinuity design, the propensity of exposure is either 0 (below the threshold) or 1 (above the threshold).

In a **fuzzy** regression discontinuity design, the relationship between the assignment variable and the propensity of exposure should be modelled



REGRESSION DISCONTINUITY DESIGN

In simple linear situation with:

- X = Propensity of exposure (e.g. probability of being treated)
- Z = ‘Assignment variable’ (e.g. variable that determines treatment),
- T = Threshold (i.e. value of Z where X is usually introduced)

$$Y = \beta_0 + \beta_1 X + \beta_2(1-X)(Z-T) + \beta_3 X(Z-T) + \varepsilon$$



Effect of
exposure
(e.g. statin)



Effect of
assignment
variable (e.g. BP)
before threshold



Effect of
assignment
variable (e.g. BP)
after threshold

RDD ASSUMPTIONS AND CAVEATS

Units above and below the threshold must be exchangeable, whether by true randomness or explainable factors

- Method fails with selection biases; e.g. doctors repeating borderline tests

Requires large sample to have sufficient data within bandwidth

Wider bandwidths require smaller samples, but this weakens exchangeability

- Should carry out sensitivity analyses with different bandwidths

Assumes ‘random’ exposure assignment after modelling assignment variable

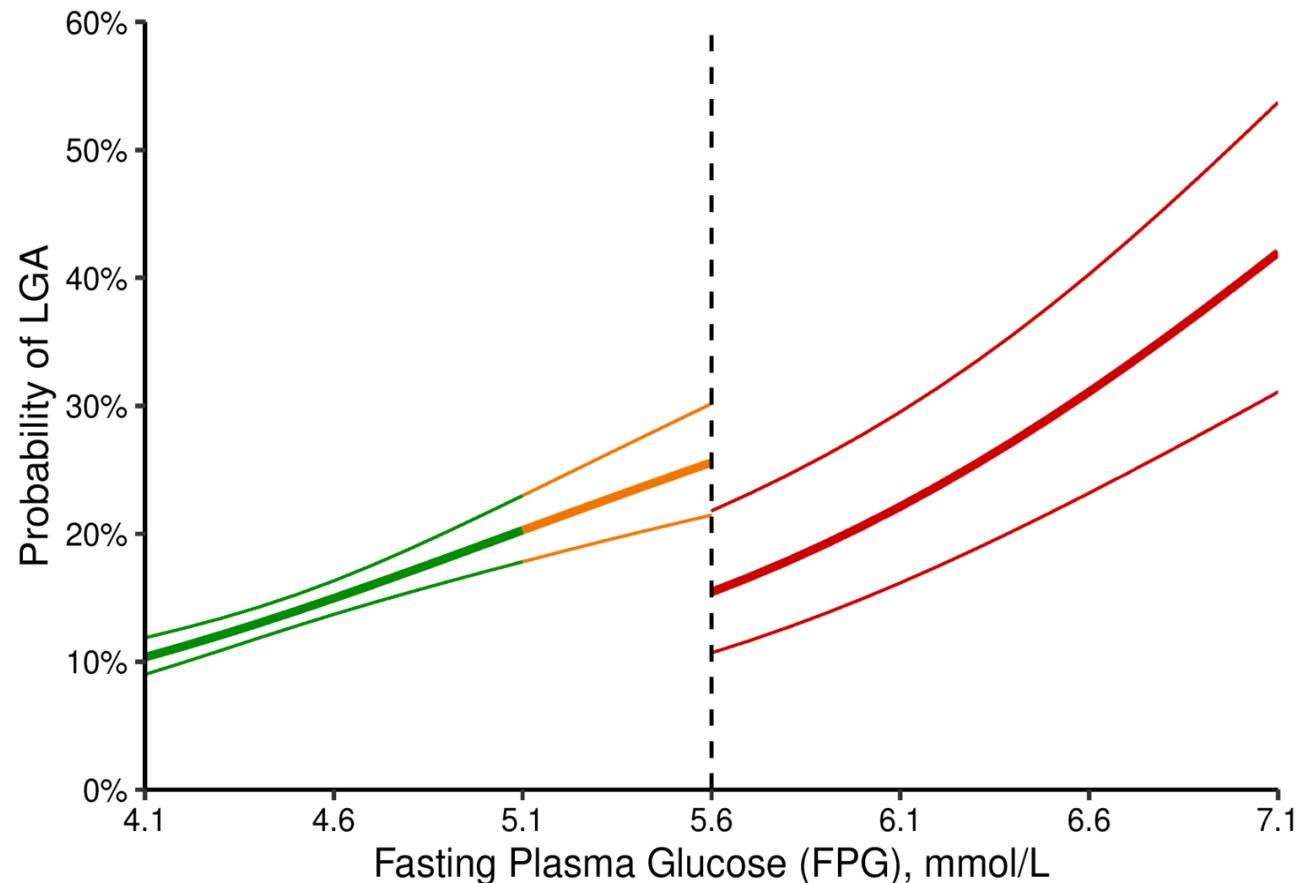
- Should explore other determinants of exposure propensity (i.e. confounders), and condition if necessary

Relies on accurately modelling exposure propensity and relationship between ‘assignment variable’ and outcome

REGRESSION DISCONTINUITY DESIGN

Example: Diagnosis of diabetes and risk of large for gestational age

- In the UK, gestational diabetes is diagnosed when the fasting glucose is $\geq 5.6\text{mmol/l}$
- Study found that receipt of diagnosis leads to half the risk of LGA – equivalent to around 1mmol/L drop in glucose



Tennant et al. *BJOG, in press.*

RECOMMENDED READING

- Craig P, Katikireddi SV, Leyland A, Popham F. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annual review of public health* 2017; **38**: 39–56.
- Baiocchia M, Chengb J, Smallc DS. Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference. .
- Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *The American journal of clinical nutrition* 2016; **103**: 965–78.
- Dimick JB, Ryan AM. Methods for evaluating changes in health care policy: the difference-in-differences approach. *Jama* 2014; **312**: 2401–2.
- Bernal JL, Cummins S, Gasparini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology* 2017; **46**: 348–55.
- Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology (Cambridge, Mass)* 2014; **25**: 729.

SUMMARY

Natural experiments provide an opportunity to examine and estimate causal effects in observational data

There are very few ‘true’ natural experiments available, where an interesting exposure has been assigned ‘as random’ to a representative sample

Various methods are available to exploit quasi-natural experiment scenarios to obtain better counterfactual estimates:

- instrumental variables
- difference-in-differences
- regression discontinuity designs
- interrupted time-series

These have long been popular in econometrics research, but remain relatively rare in health and medical research

4.1 - EXPOSURE REGIMES, CAUSAL MEDIATION, & 'INTERACTIONS'

GEORGIA



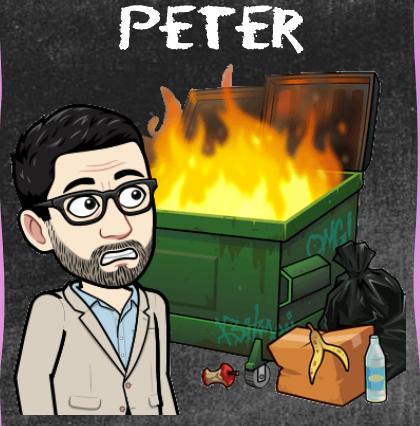
@GEORGIATOMOVA

MARK



@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

DAY 4

09:30-10:45 LECTURE 4.1

10:45-11:00 Q&A

11:00-11:30 TEA & COFFEE

11:30-12:15 LECTURE 4.2

12:15-13:00 ACTIVITY 4-A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 4.3

14:45-15:30 ACTIVITY 4-B

15:30-16:00 TEA & COFFEE

16:00-16:45 ACTIVITY 4-C

16:45-17:45 ACTIVITY 4-D

17:30-18:00 Q&A

LEARNING OBJECTIVES

By the end of this session, you will be able to:

- Explain why traditional **mediation analyses** are optimistic / naïve
- Explain how counterfactual methods can be extended to consider the joint and separate effects of **exposures**, **mediators**, or **time varying exposures**
- Define an **exposure regime**
- Explain how mediation methods can be used to estimate **theoretical interventions**
- Appreciate why the best solution is often to redefine the **mediator** as the **exposure**
- Explain why standard methods fail in the presence of **time-varying confounding**

HERE BE DRAGONS

Adjusting for mediators is a dangerous business

Blocks (part of) true causal effect

HERE BE DRAGONS

Adjusting for mediators is a dangerous business

Blocks (part of) true causal effect

Risks summoning **collider bias**

General advice is therefore:

- DO NOT adjust for **mediators!**



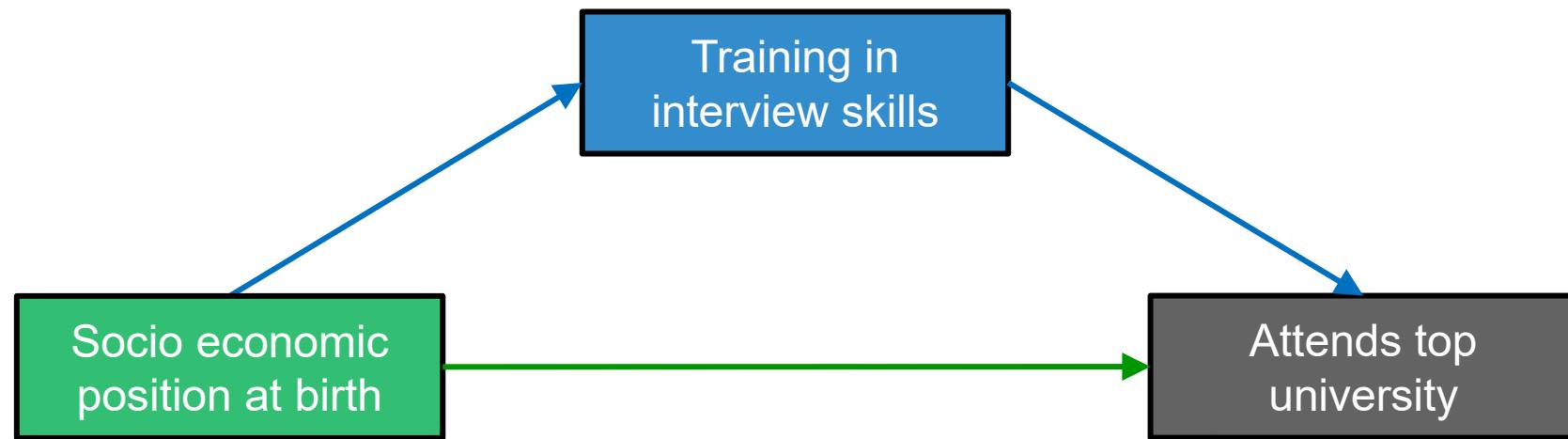
Collider bias (artist's impression)

BUT...

What if:

- exposure is not (directly) modifiable and ...
- we still want to change its effect on the outcome

e.g. the exposure is beyond our control

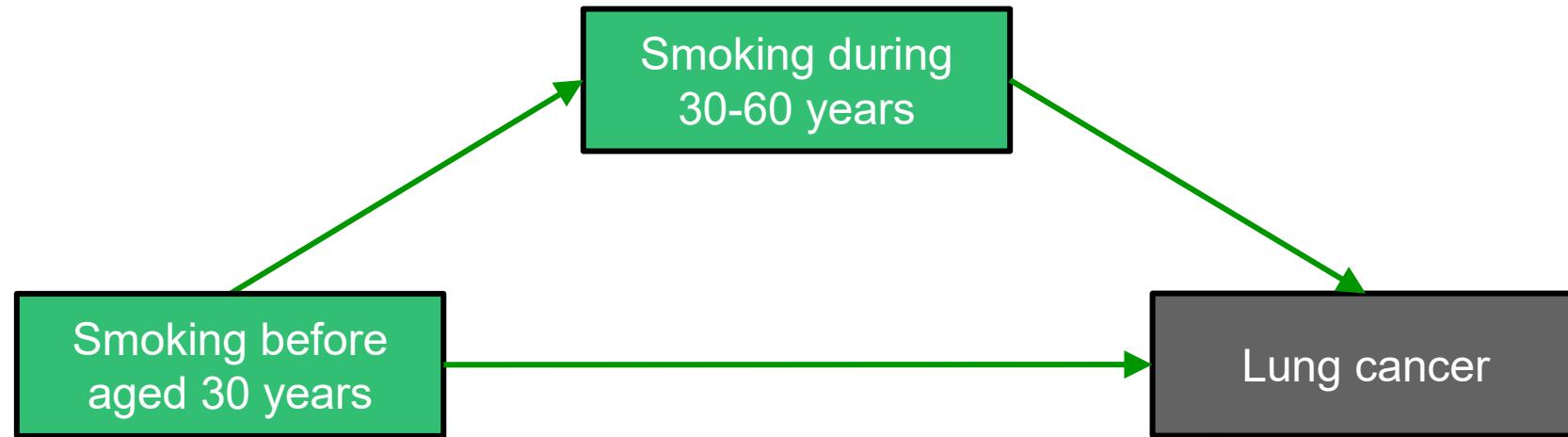


OR...

Want to know:

- how a **time-varying exposure** operates over time and ...
- how much of its effect can still be ‘changed’?

e.g. repeated measures

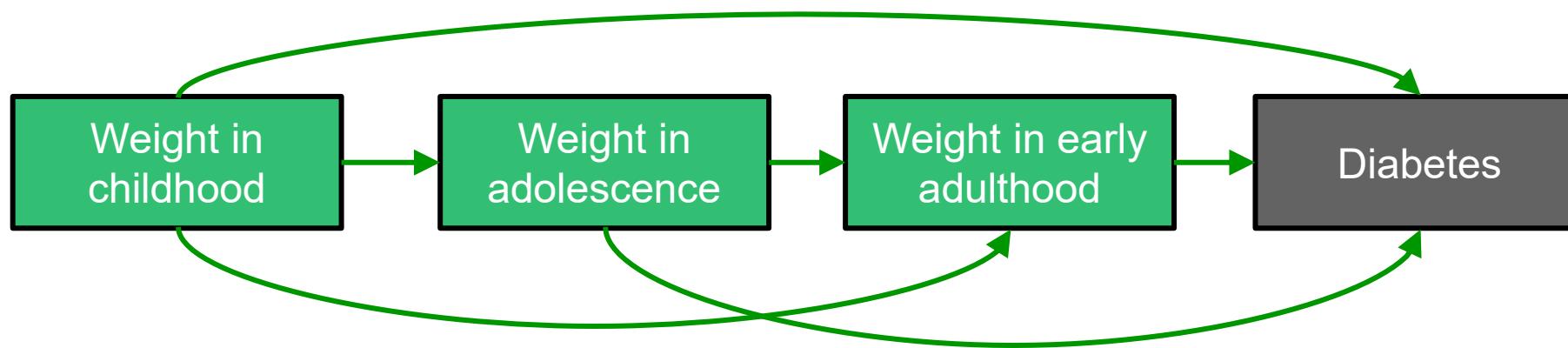


OR...

Want to know:

- how a **time-varying exposure** operates over time and ...
- how much of its effect can still be ‘changed’?

e.g. multiple repeated measures



TIME-INVARIANT AND TIME-VARYING

Time-invariant ('fixed')

- Occur only once
 - ✓ e.g. one-dose vaccine, birthweight
- Do not change over time
 - ✓ e.g. sex, (ab)normal BRAC1/BRAC2 genes, height (in middle age)
- Evolve over time in deterministic way
 - ✓ e.g. age, time since treatment

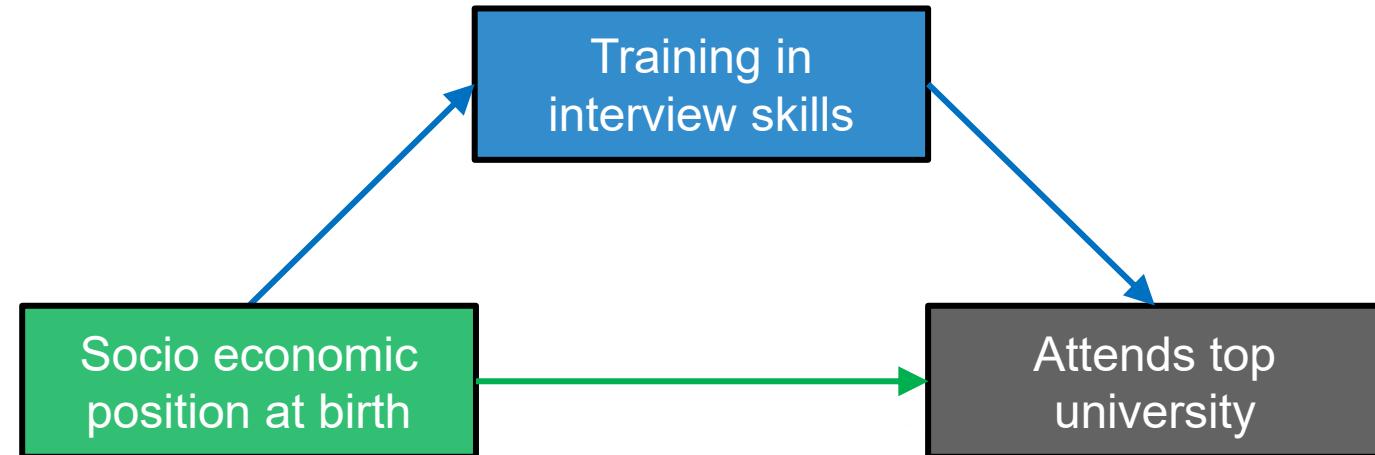
Time-varying ('changing')

- Occur multiple times
 - ✓ e.g. daily doses of medicine
- Change over time
 - ✓ e.g. smoking status, weight, height (in childhood), blood sugar

EASY SOLUTION

Consider:

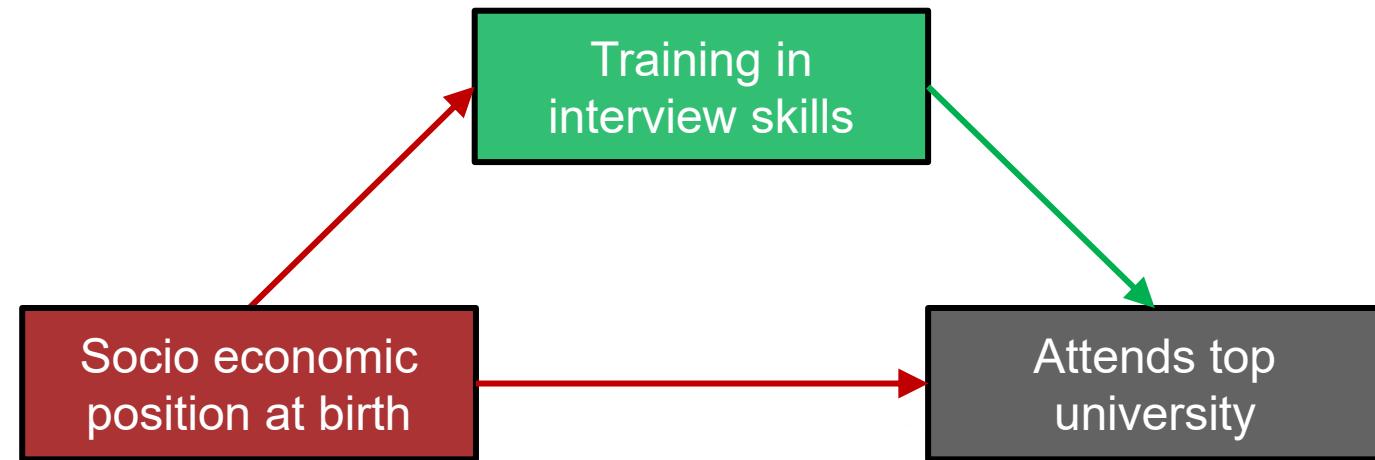
- are you *really* interested in the **joint effects** of the **exposure** and **mediator**?
- or is the '**mediator**' actually just your true **exposure** of interest?



EASY SOLUTION

Consider:

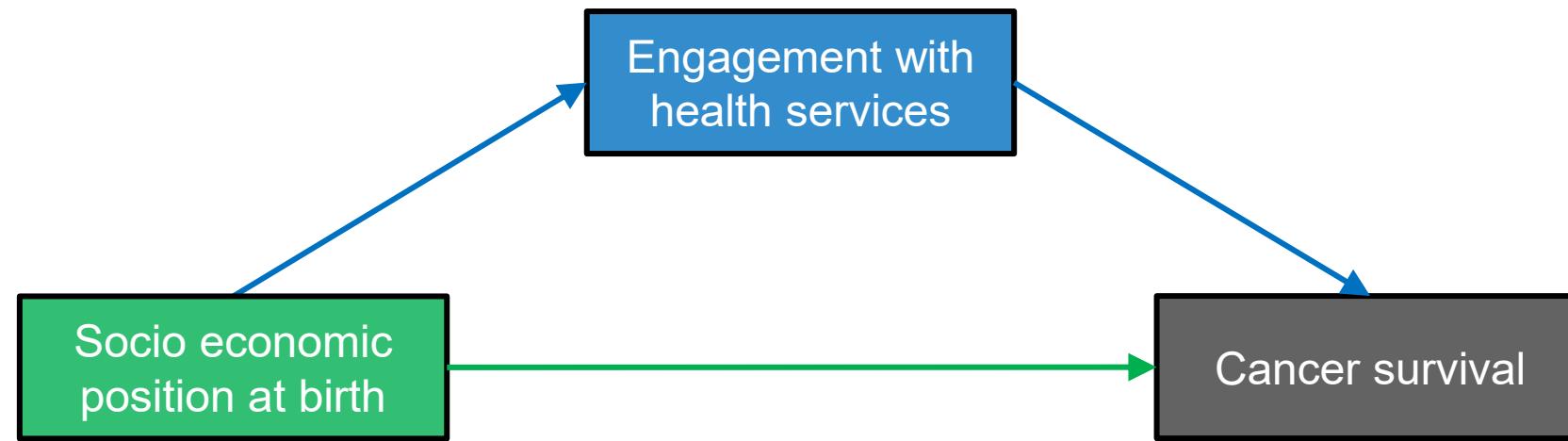
- are you *really* interested in the **joint effects** of the **exposure** and **mediator**?
- or is the '**mediator**' actually just your true **exposure** of interest?
- if so, focus on correctly estimating the effect of this true **exposure** on your outcome, treating the exposure as a **confounder** that is not to be interpreted



PANDORA'S BOX

Sometimes: we really want to know how one variable acts through another

- e.g. people from **more deprived socio-economic positions** have **poorer cancer survival prospects**



How much of this is because they are **less 'engaged' with health services** (i.e. attend later with more advanced / less treatable disease)?

TRADITIONAL MEDIATION ANALYSIS

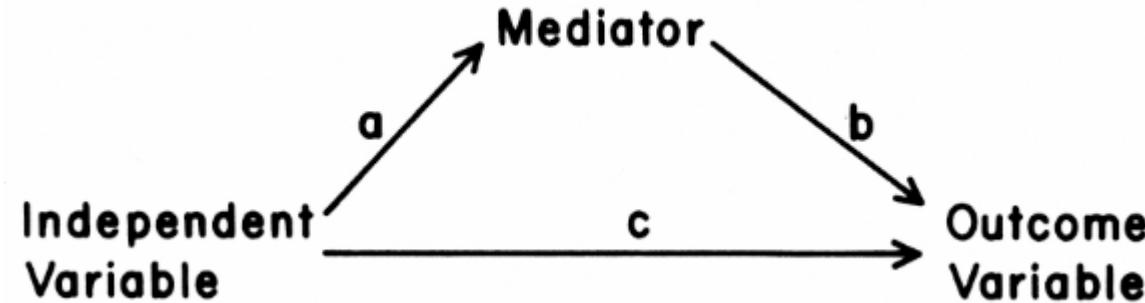
Baron and Kenny (1986)

Journal of Personality and Social Psychology
1986, Vol. 51, No. 6, 1173–1182

Copyright 1986 by the American Psychological Association, Inc.
0022-3514/86/\$00.75

The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations

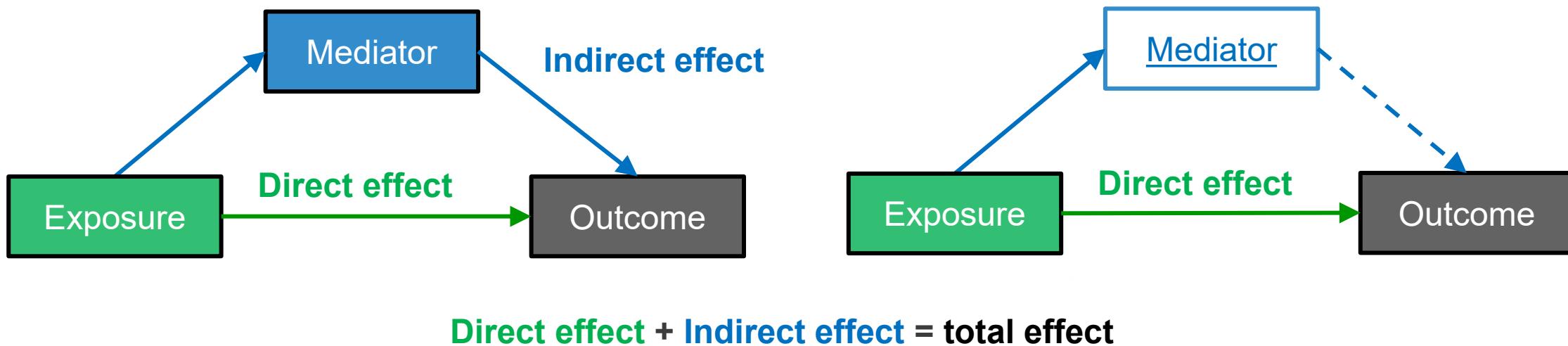
Reuben M. Baron and David A. Kenny
University of Connecticut



TRADITIONAL MEDIATION ANALYSIS

Baron and Kenny (1986)

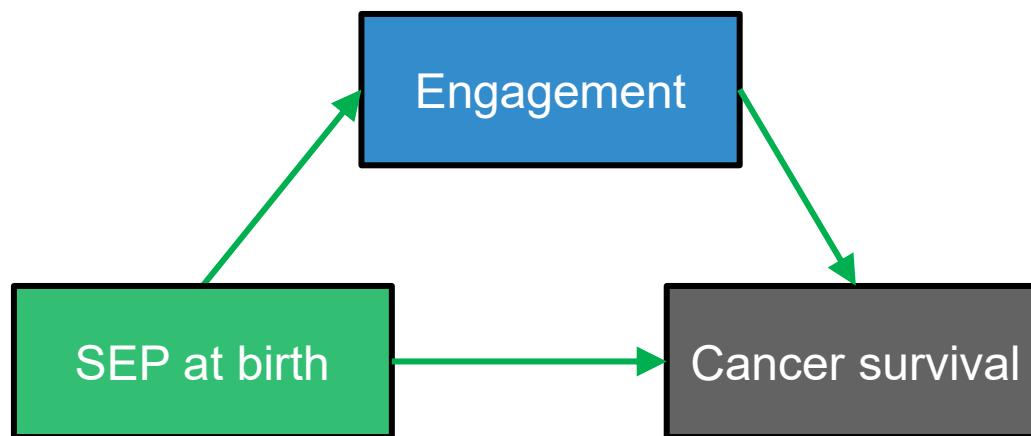
- Decompose ‘**total effect**’ by conditioning on **mediator**
- **Total effect** = effect **without** conditioning
- **Direct effect** = effect **with** conditioning
- **Indirect effect** = **total effect** – **direct effect**



TRADITIONAL MEDIATION ANALYSIS

This approach is encoded within '**structural equation modelling**'

Most SEM software & commands will automatically decompose effects this way



$$0.2243268 + 0.0865526 = 0.3108794$$

Direct effects						
	OIM Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
structural survival <- engagement sep	.1769575 .2243268	.0329057 .0326572	5.38 6.87	0.0000 0.0000	.1124634 .1603198	.2414515 .2883338
Indirect effects						
	OIM Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
structural survival <- sep	.0865526	.0168045	5.15	0.0000	.0536165	.1194888
Total effects						
	OIM Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
structural survival <- engagement sep	.1769575 .3108794	.0329057 .0288237	5.38 10.79	0.0000 0.0000	.1124634 .254386	.2414515 .3673729

TRADITIONAL MEDIATION ANALYSIS

Barron & Kenny's approach is rather optimistic (naïve) with its assumptions ...

Parametric assumptions:

- **Multivariate normality**
 - ✓ though can now be generalised ...
- **Linear relationships**
 - ✓ though non-linearity can be modelled ...
- **Additive effects**
 - ✓ i.e. no interactions !

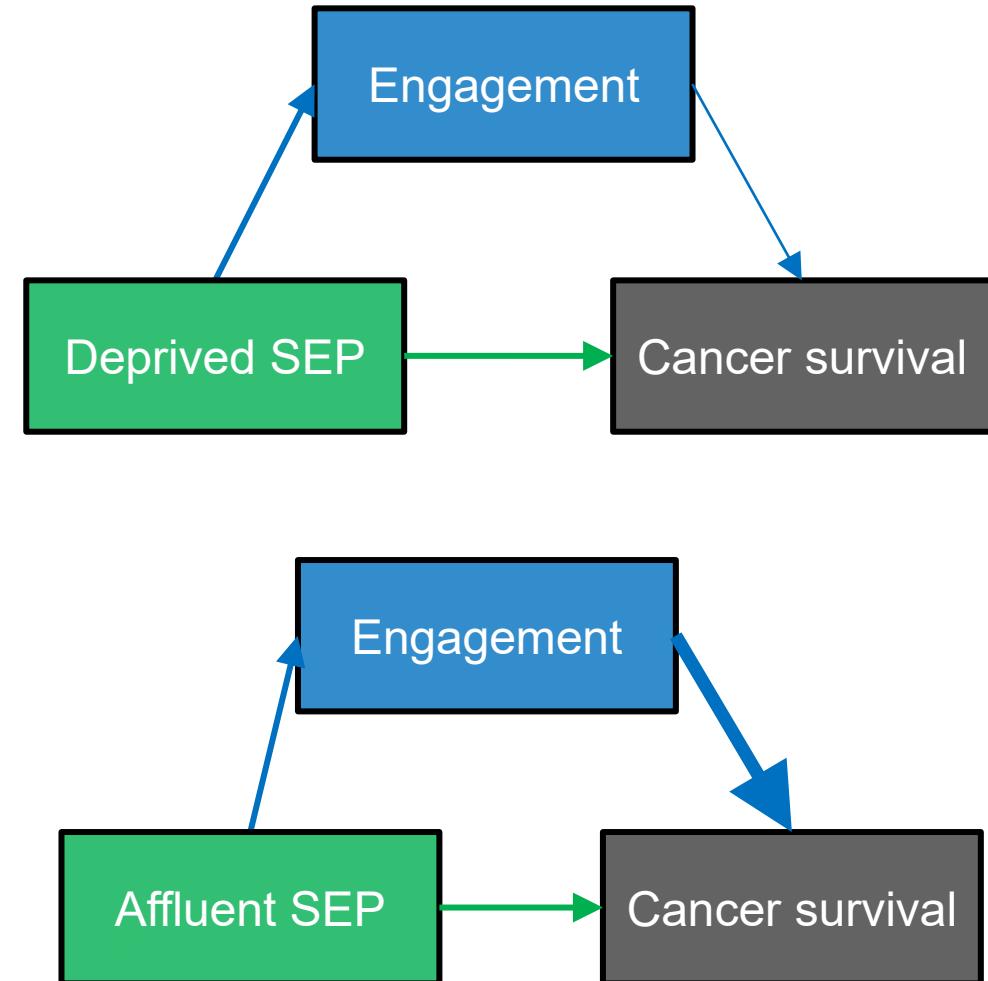


TRADITIONAL MEDIATION ANALYSIS

Interactions break this decomposition

Suppose effect of '**mediator**' varies by
exposure?

- e.g. people with **more deprived SEP** get dismissed by GPs, so effect of **engagement** is smaller?



TRADITIONAL MEDIATION ANALYSIS

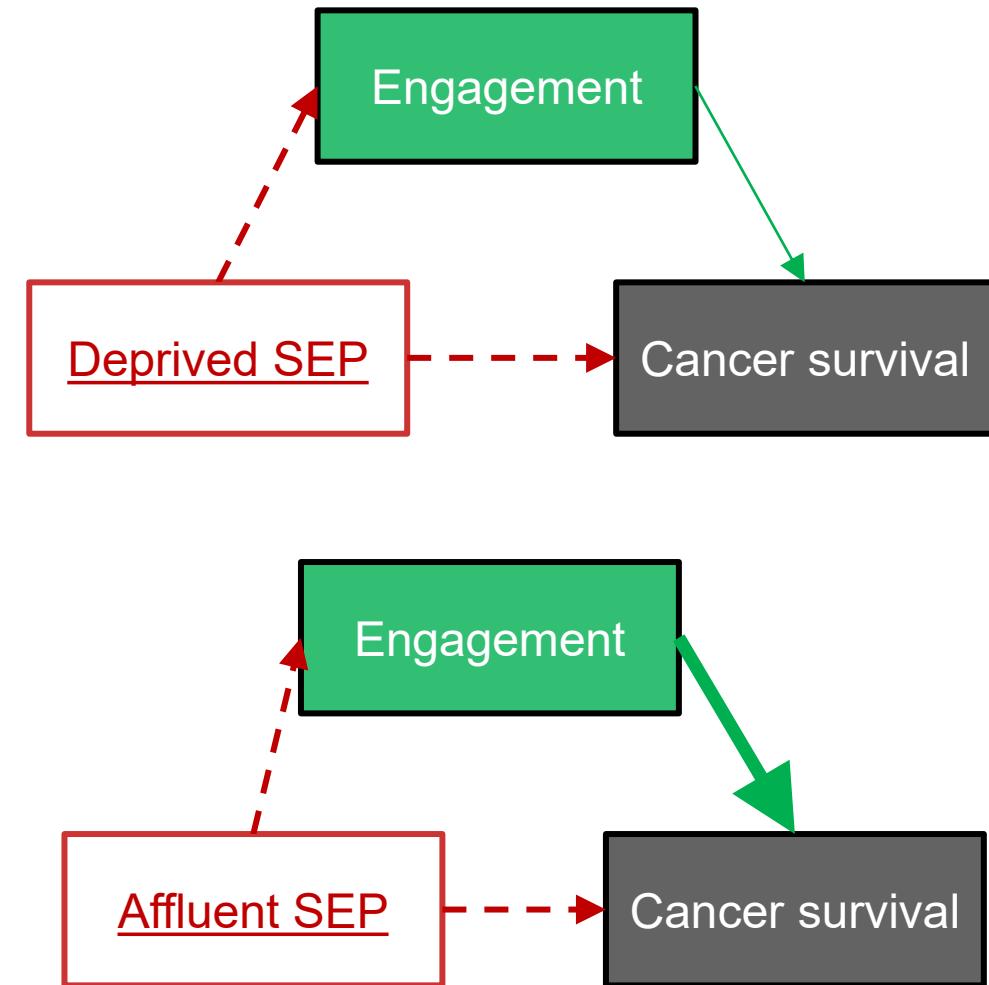
Could stratify over **exposure**, to describe unique effect of the '**mediator**' at each level

Note:

- you've actually just made the original **mediator** into your new **exposure**

Also:

- this doesn't really help you fully understand the **joint** effects – the '**exposure**' has become a nuisance



TRADITIONAL MEDIATION ANALYSIS

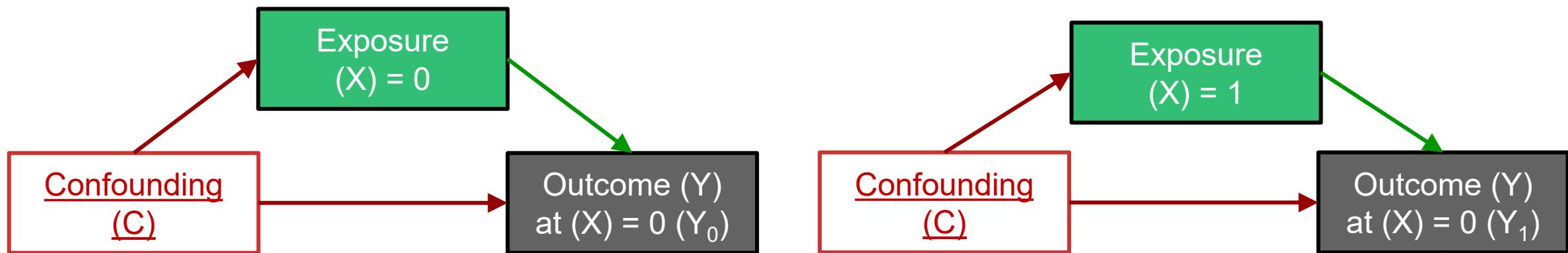
Could add an **interaction term**, but these are tricky to interpret – and certainly not obvious how you'd decompose into **direct**, **indirect**, and **total effects** ...



CAUSAL MEDIATION ANALYSES

A solution arises from taking a counterfactual (**potential outcome**) approach

Recall: To estimate the **total effect** of an **exposure** on an **outcome**, we **compare** the **potential outcomes** for **contrasting levels** of the **exposure**, conditioning for all **confounding**

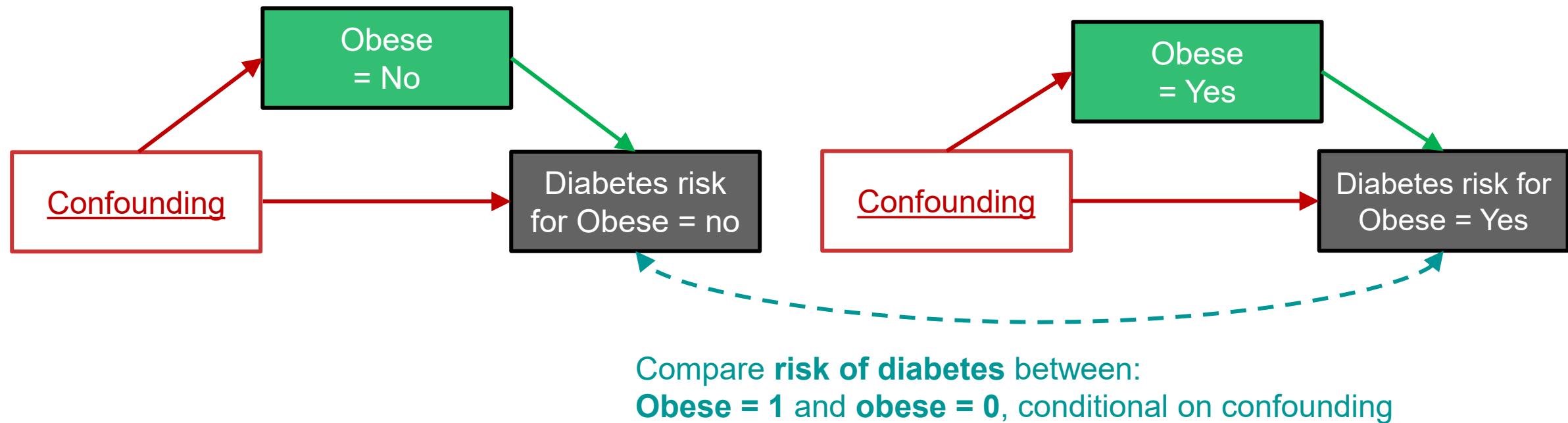


Compare levels of outcome (Y) between contrasting levels of exposure (X), conditional on confounding (C) to estimate total causal effect (e.g. $Y_1 - Y_0$)

CAUSAL MEDIATION ANALYSES

e.g. effect of **obesity** on risk of **diabetes**

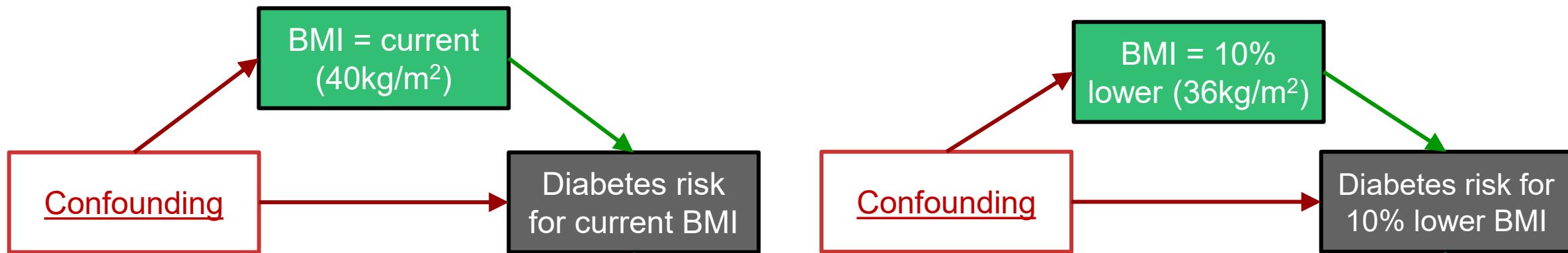
Contrast 1: **compare diabetes** between **obesity = 1** and **obesity = 0**



CAUSAL MEDIATION ANALYSES

e.g. effect of **BMI** on risk of **diabetes**

Contrast 2: **compare diabetes** between **BMI = current (40kg/m^2)** and **BMI = 10% lower (36kg/m^2)**

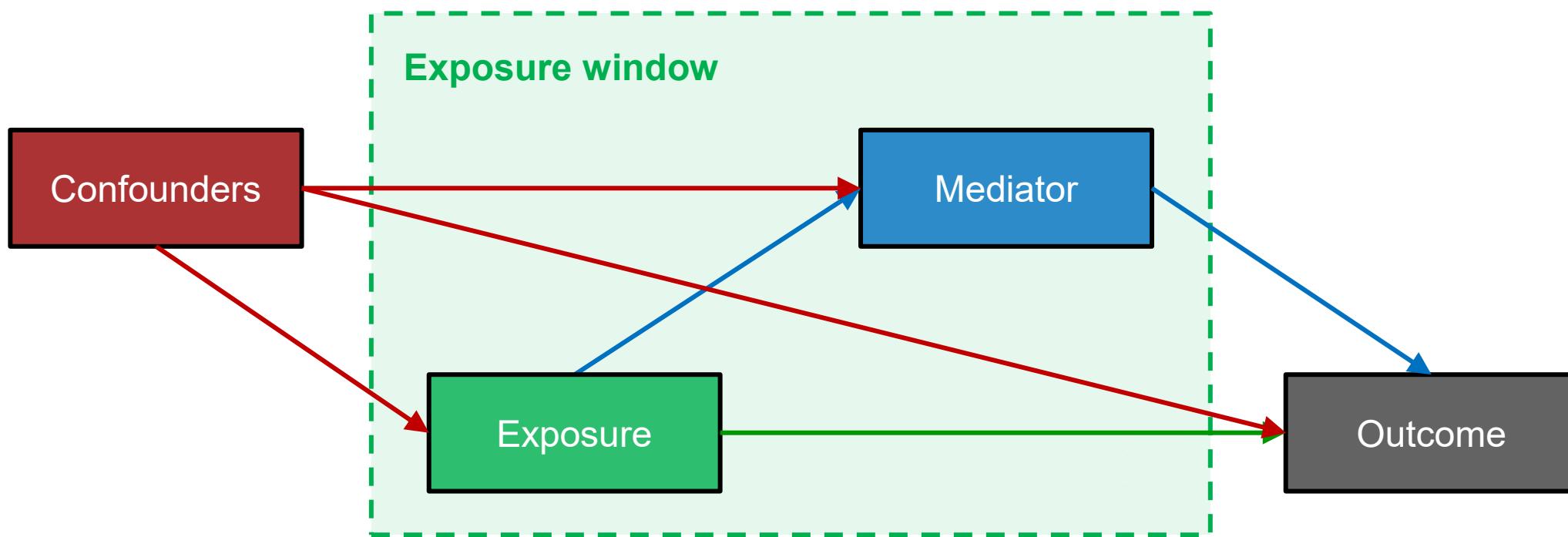


Compare **risk of diabetes** between:
BMI = current (40kg/m^2) and **BMI = 10% lower (36kg/m^2)**, conditional on confounding

CAUSAL MEDIATION ANALYSES

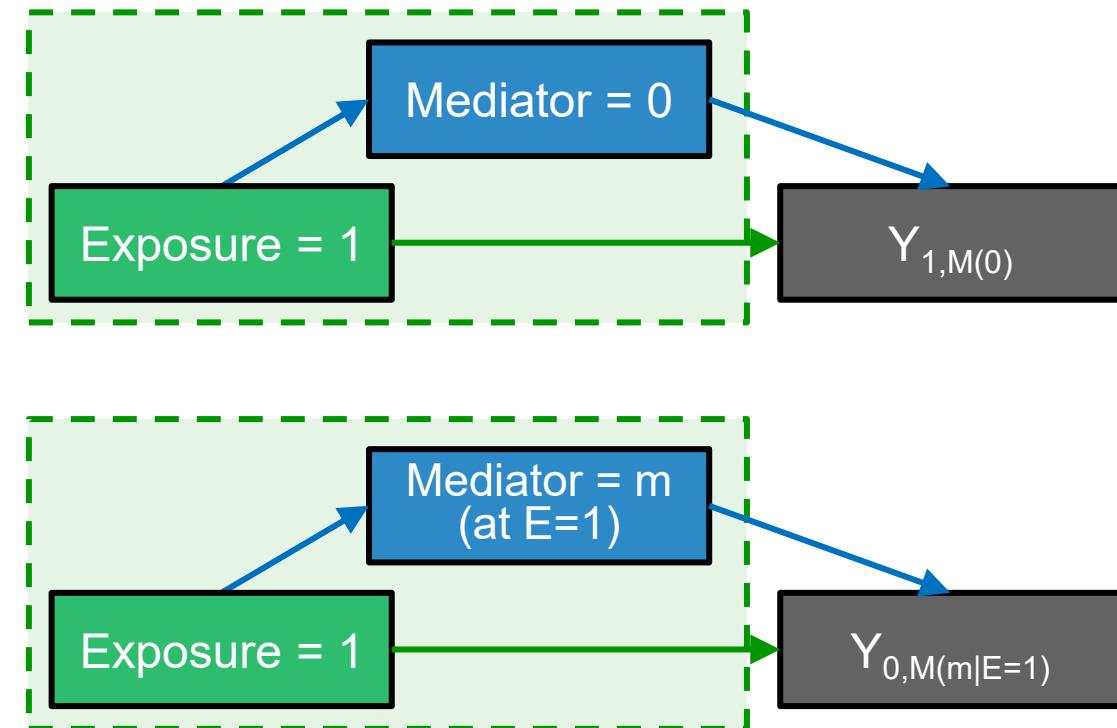
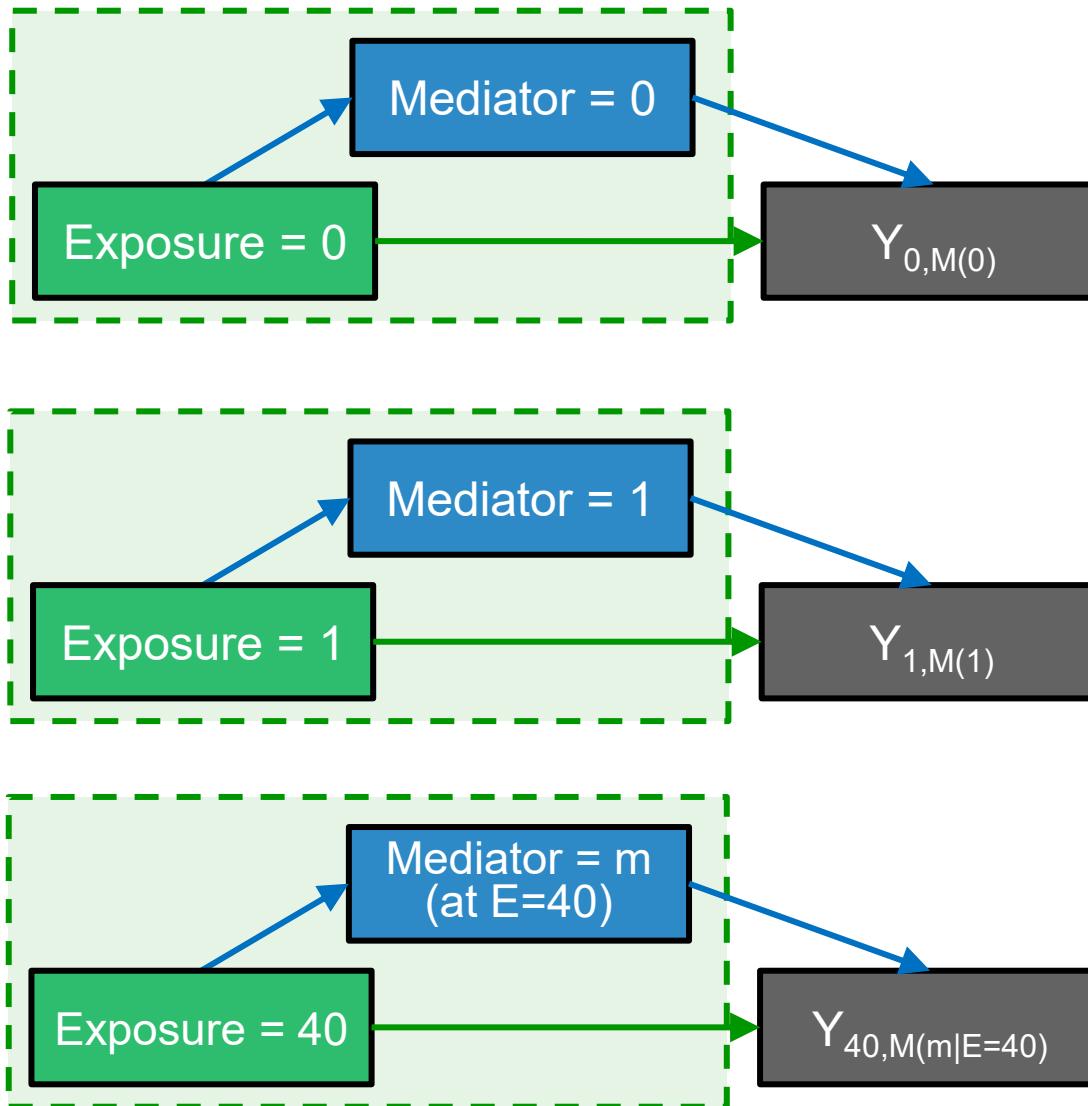
In counterfactual **mediation analysis**, we extend the ‘**exposure window**’ to include values of the **exposure** and **mediator**

This creates ‘**exposure regimes**’ – representing specific combinations of the exposure and mediator – for which we can estimate the **potential outcomes**



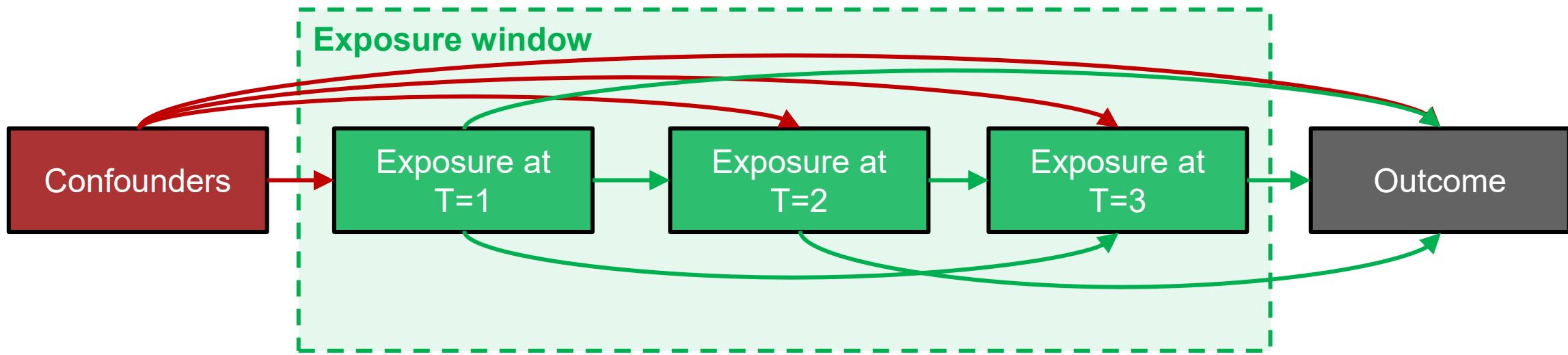
CAUSAL MEDIATION ANALYSES

Example regimes



MULTIPLE EXPOSURES OR MEDIATORS

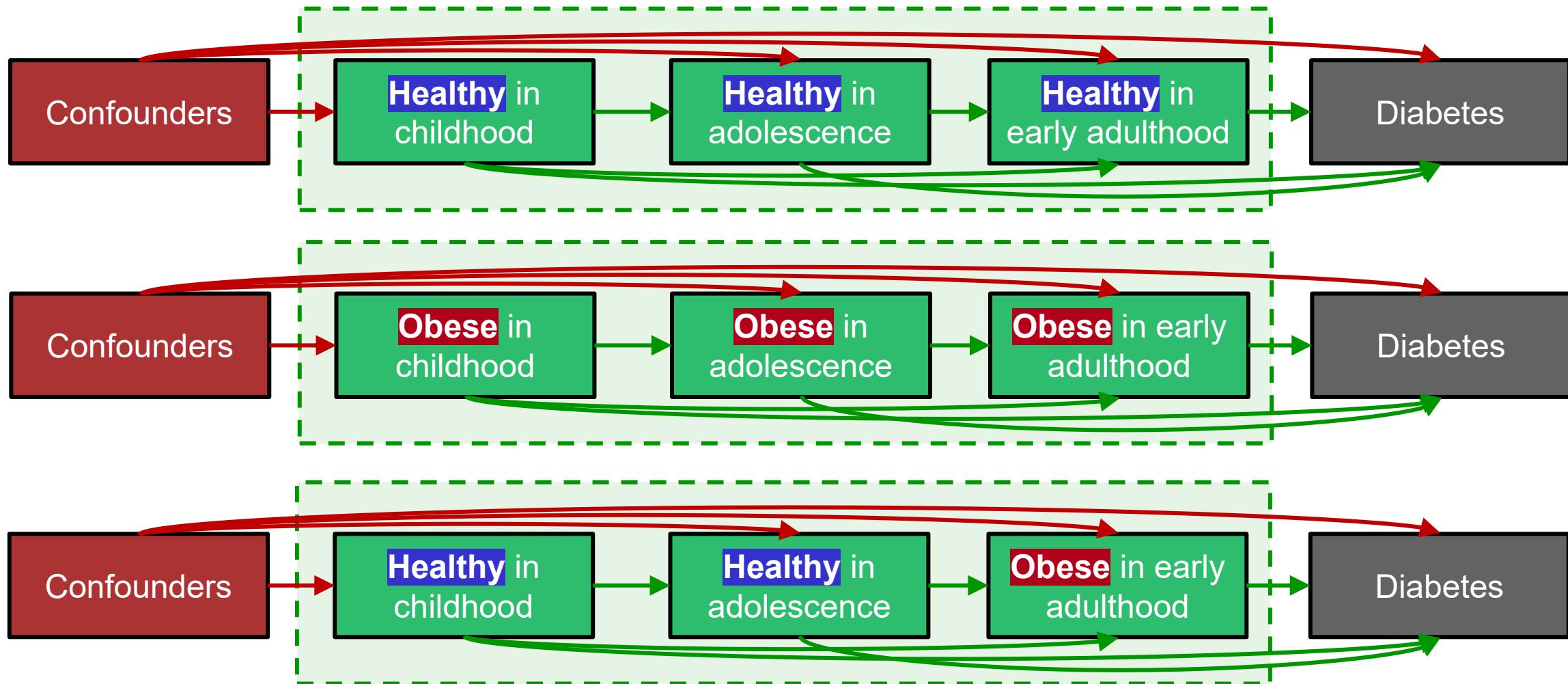
Can be easily extended for **multiple exposures** and/or **mediators**, provided all confounding occurs before the first exposure (**time-invariant confounding**)



Note: This is a special case where your **mediators** are repeated measures of your exposure known as a **time-varying exposure**

REPEATED EXPOSURES

Example regimes for time varying exposure



WHICH EFFECT?

- Estimating the effect of changing the exposure and/or mediator now becomes a matter of comparing the potential outcomes between regimes
- The format of your variables doesn't matter, nor whether your model includes non-linear relationships, interactions, etc.
- Unfortunately, with so many regimes, there are many possible comparisons, creating a dizzying array of potential estimands
- Counterfactual mediation effects therefore include many words like '**controlled**', '**natural**', '**pure**', '**mediated**', '**interactive**' '**marginal**', '**conditional**'
- e.g. **conditional pure natural indirect effect**



FOUR WAY DECOMPOSITION

VanderWeele's four way decomposition offers perhaps the best way to separate the **total causal effect** into constituent parts:

Controlled direct effect

- the effect of the exposure in the absence of the mediator

Pure indirect effect

- the effect of the exposure that acts purely through the mediator

Reference interaction

- the changing 'direct' effect of the exposure for changing values of the mediator

Mediated interaction

- the changing mediated effect of the exposure for changing values of the mediator

FOUR WAY DECOMPOSITION

Question: How useful is this decomposition?

Example: You find **20%** of the effect of the exposure is due to the **reference interaction** and **10%** due to the **mediating interaction**

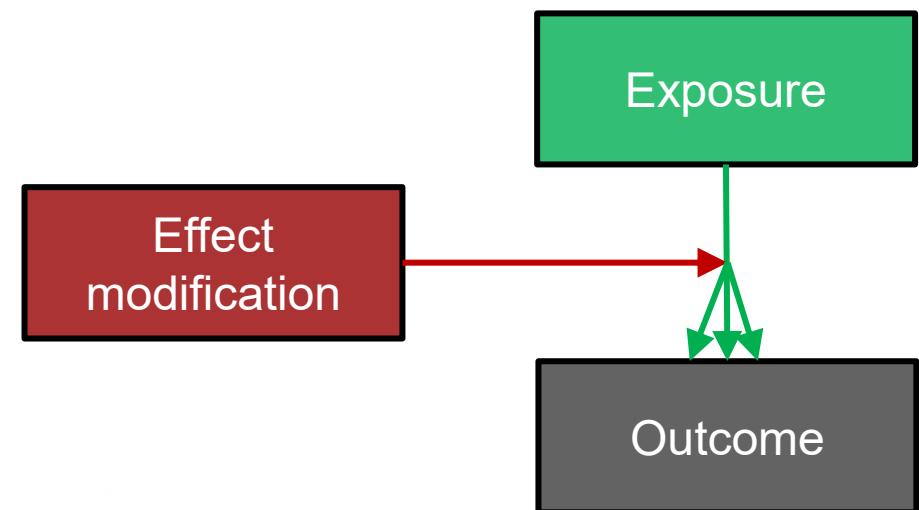
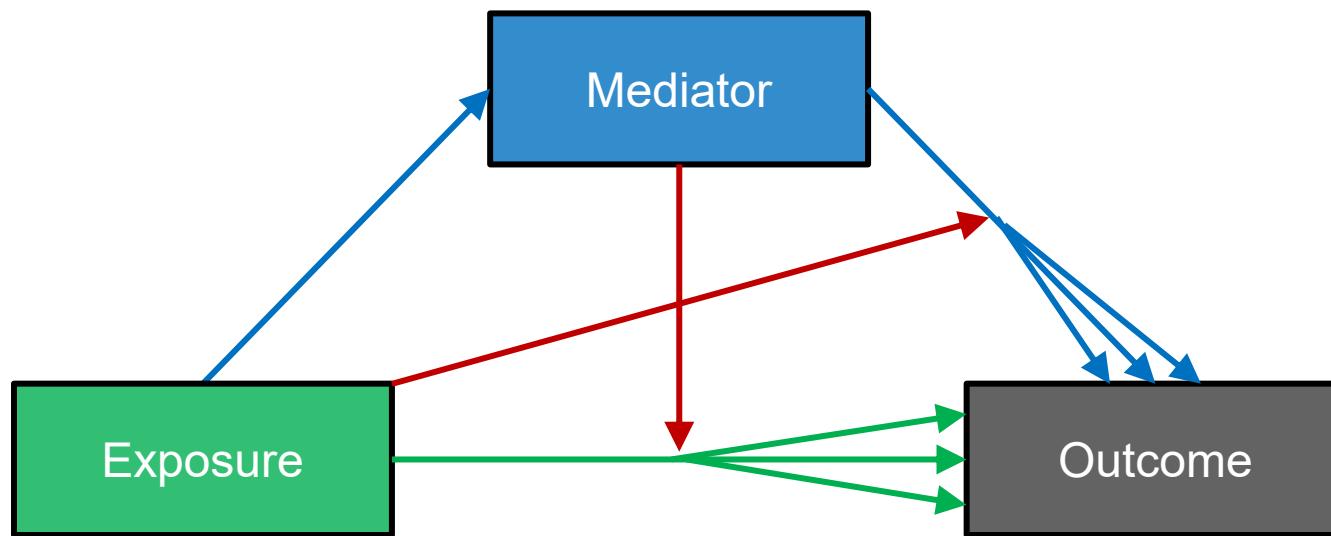


- Is this useful?
- What does it mean?
- Do we care?
- Direct effects are arbitrary – collecting and exploring more mediator variables would create smaller direct effects !

CAUSAL INTERACTIONS AND EFFECT MODIFICATION

Many researchers are interested in exploring whether **causal interactions** or **effect modification** is taking place

- **causal interaction** = the effect(s) of 2+ exposures depend on each other
- **effect modification** = the effect(s) of an exposure depends on a third variable
(AKA: **effect moderation**)



CAUSAL INTERACTIONS AND EFFECT MODIFICATION

Distinguishing **causal interactions** from **statistical interactions** is tricky!

Statistical interactions will often appear if you have unmodelled confounding or have not parameterised your model correctly

Statistical interactions are also scale dependent

- an interaction may be apparent / required for a linear outcome yet absent / not required for a binary version of the same outcome

Interpretation of statistical interactions should therefore be approached with extreme caution !



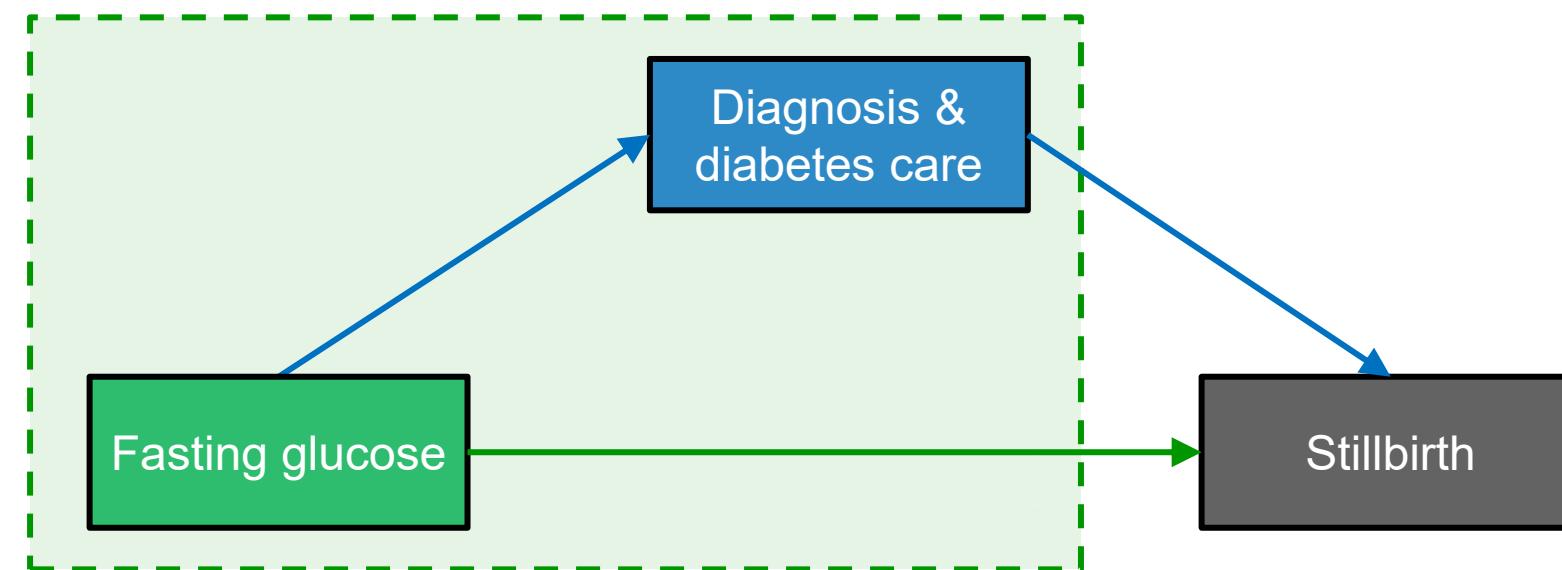
COMPARING REGIMES

More useful to compare **regimes** of interest, to estimate how different combinations of your **exposure** and/or **mediator** effect the outcome

- e.g. effect of **diabetes in pregnancy** on **risk of stillbirth**?

Diabetes in pregnancy has a very modest effect on **stillbirth**

Reason: underlying **biological harm** may be masked by **enhanced care**



Gestational diabetes and the risk of late stillbirth: a case-control study from England, UK

T Stacey,^a PWG Tennant,^{b,c,d} LME McCowan,^e EA Mitchell,^f J Budd,^{g,h} M Li,^e JMD Thompson,^{e,f} B Martin,ⁱ D Roberts,^j AEP Heazell^{g,h}

^a School of Healthcare, University of Leeds, Leeds, UK ^b Leeds Institute for Data Analytics, University of Leeds, Leeds, UK ^c School of Medicine, University of Leeds, Leeds, UK ^d The Alan Turing Institute, London, UK ^e Department of Obstetrics and Gynaecology, The University of Auckland, Auckland, New Zealand ^f Department of Paediatrics: Child and Youth Health, The University of Auckland, Auckland, New Zealand ^g Maternal and Fetal Health Research Centre, School of Medical Sciences, Faculty of Biological, Medical and Human Sciences, University of Manchester, Manchester, UK ^h Manchester Academic Health Science Centre, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, UK ⁱ Birmingham Women's Hospital NHS Foundation Trust, Birmingham, UK ^j Liverpool Women's Hospital NHS Foundation Trust, Liverpool, UK

Correspondence: T Stacey, School of Human and Health Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK.
Email: t.stacey@hud.ac.uk

Accepted 31 January 2019. Published Online 20 March 2019.

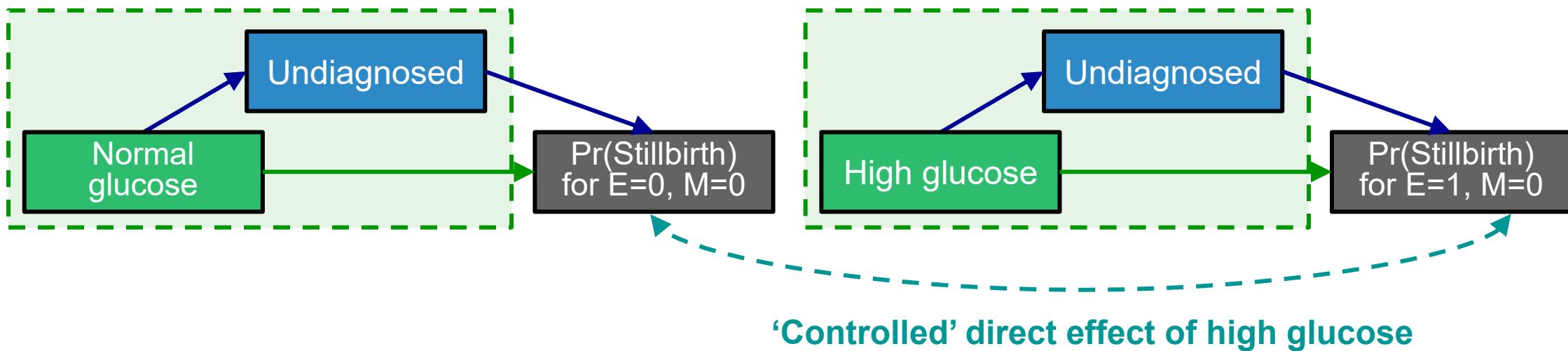
Stacey T et al 2019. Gestational diabetes and the risk of late stillbirth: a case-control study from England, UK.
BJOG. DOI: 10.1111/1471-0528.15659

COMPARING REGIMES

Due to poor adherence to guidelines, many women with **high fasting glucose** don't get **diagnosed or receive care**

- e.g. **controlled direct effect** – harmful effect of glucose in absence of specialist care

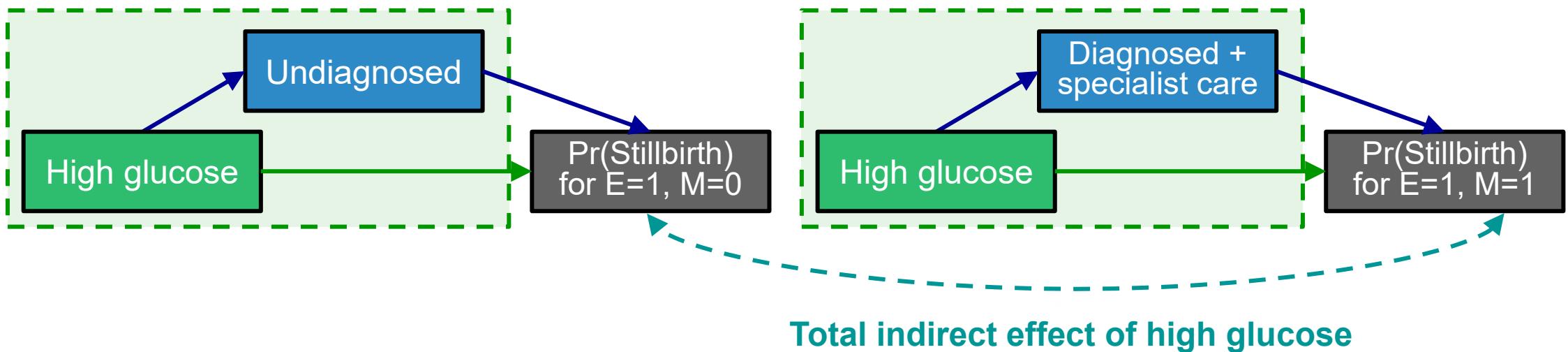
Stillbirth RR = **4.2** in **high glucose vs normal**



COMPARING REGIMES

e.g. **total indirect effect** – benefit of receiving specialist care in presence of exposure

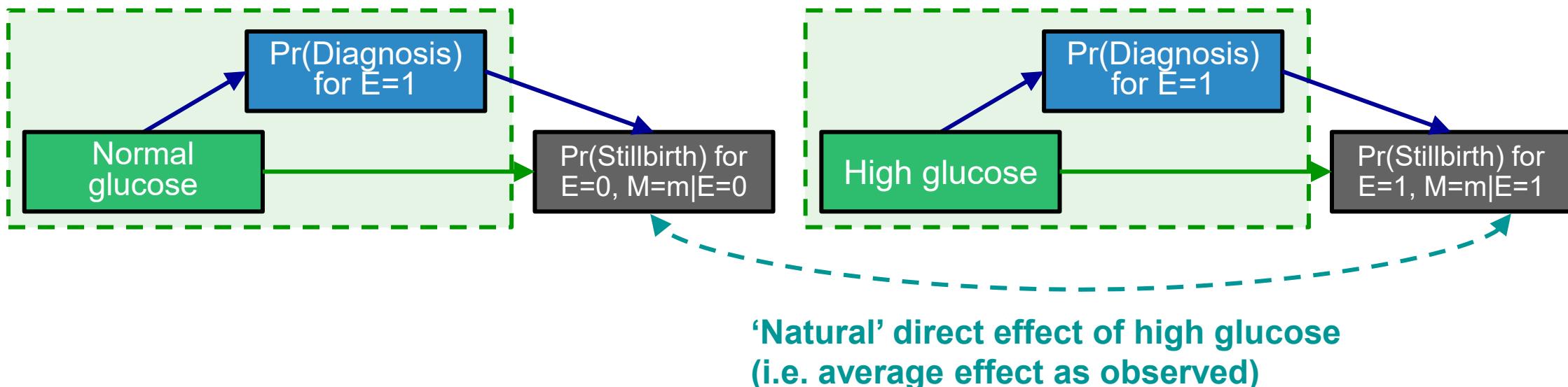
Stillbirth RR = **0.3** in **diagnosed** vs **undiagnosed**



COMPARING REGIMES

E.g. **natural direct effect** – average harmful effect of glucose in women with high glucose, **in presence of current care** (~70%)

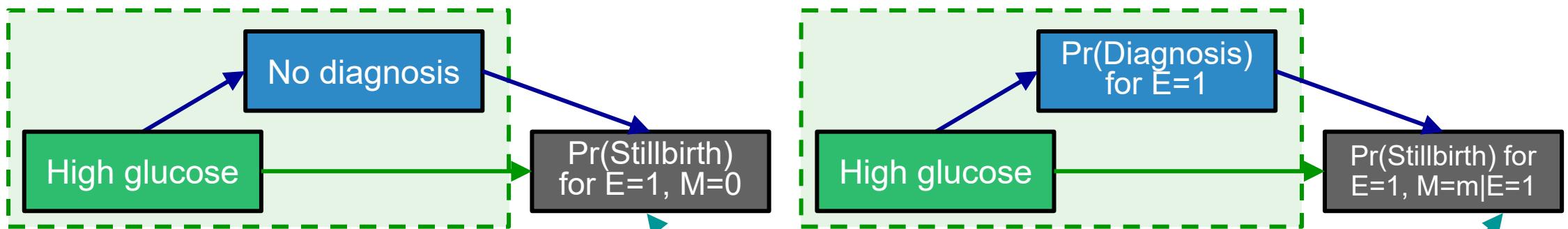
Stillbirth RR = **1.8** in **high glucose vs normal**



COMPARING REGIMES

e.g. **natural indirect effect** – average beneficial effect of current provision of care (70%)

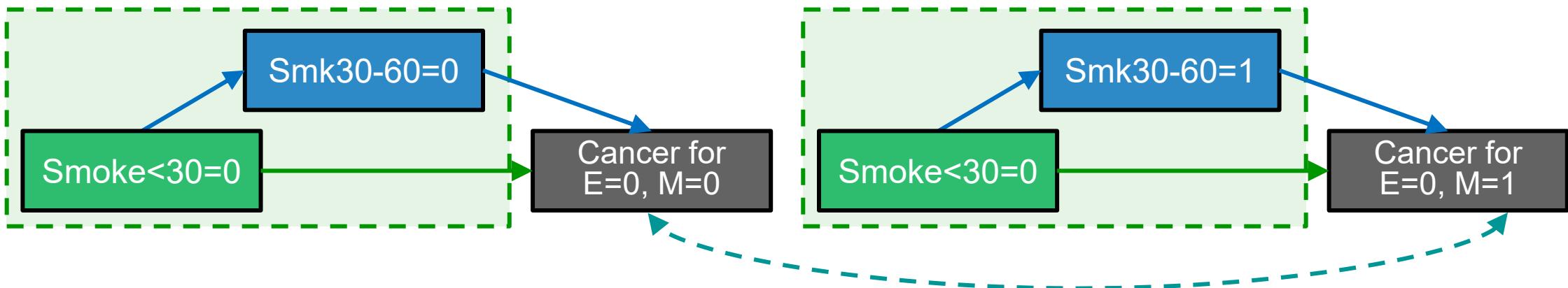
Stillbirth RR = **0.5** for **current diagnosis** vs **no diagnosis**



**'Natural' indirect effect of high glucose
(i.e. average impact as observed)**

POSITIVITY

The regimes you compare need to be plausible and possible to estimate from your data



This scenario tries to estimate effect of *starting* smoking in later life

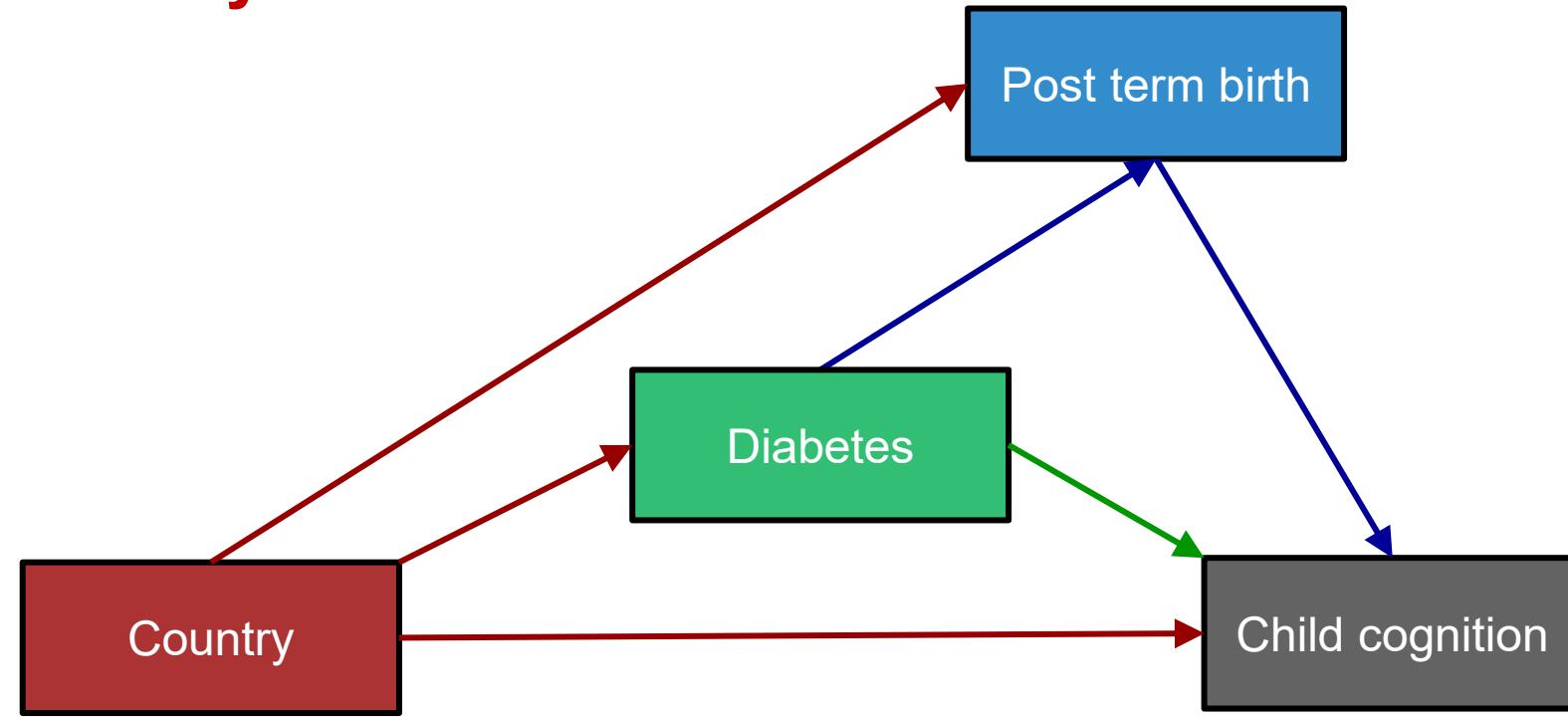
This is extremely rare and very unlikely to be observed !

POSITIVITY

- **Recall:** It must be *possible* for every participant subgroup to receive every exposure regime
- So all possible values of the **exposure** and **mediator** must be possible for all confounder combinations
- Likewise, all **mediator** values must be possible for all **exposure** values
- If certain **exposure** values mean you cannot have certain **mediator** values, then these impossible combinations cannot be estimated (and you have a positivity violation)

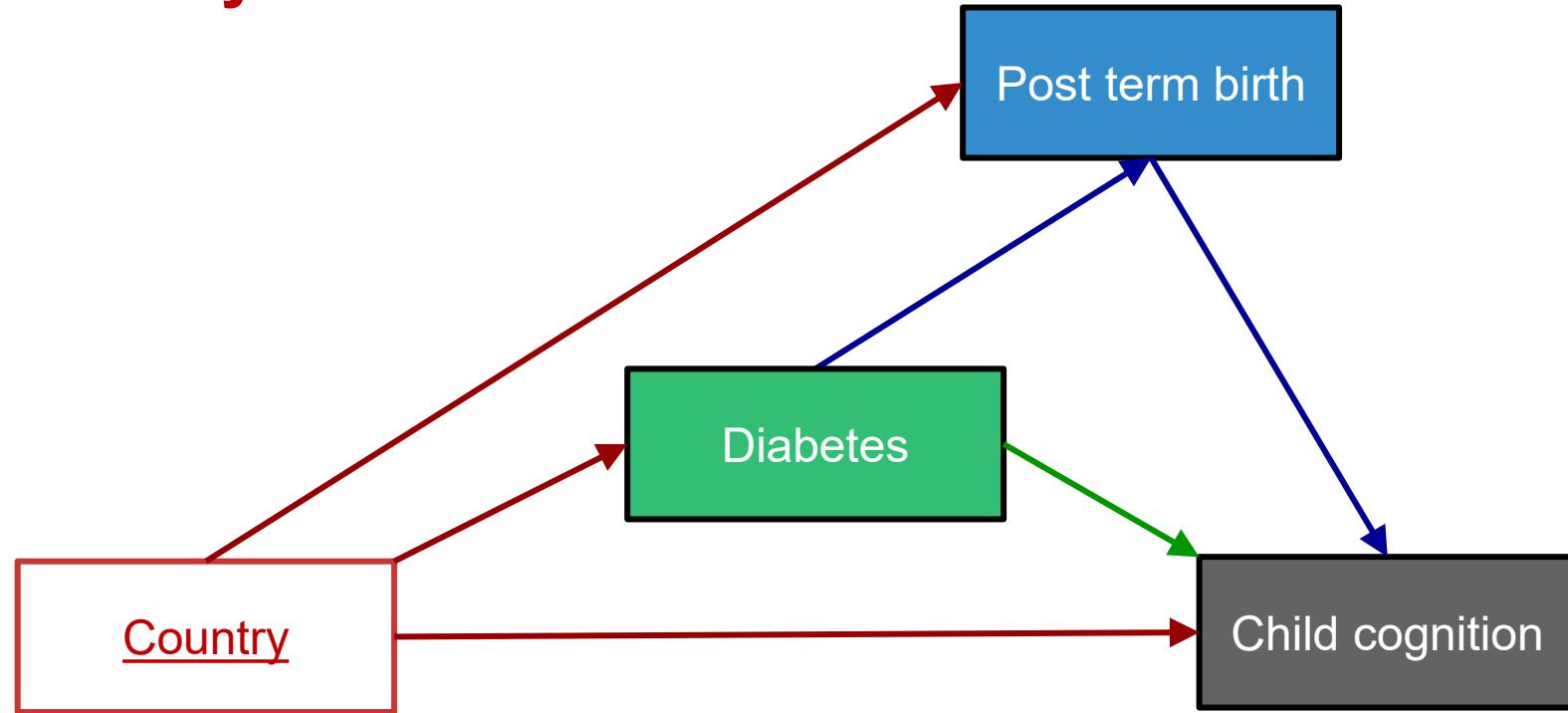
POSITIVITY

e.g. maternal diabetes, post term delivery, and cognitive outcomes confounded by country of birth



POSITIVITY

e.g. **maternal diabetes**, **post term delivery**, and **cognitive outcomes** confounded by **country of birth**



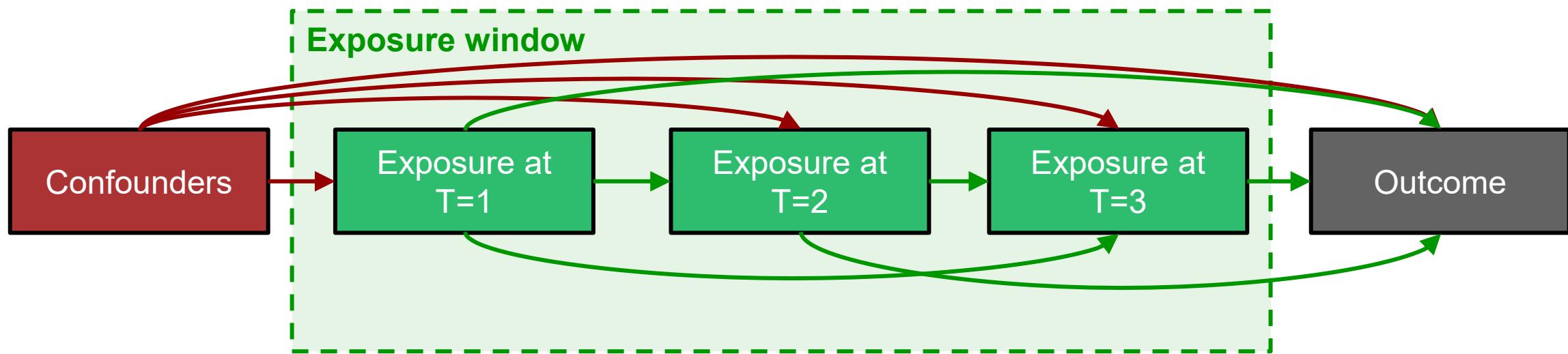
In **England**, pregnant women with pre-existing diabetes are **induced before full term**; we thus cannot observe **diabetes + post term birth**

EXCHANGEABILITY

Causal mediation analyses requires conditioning for all confounding

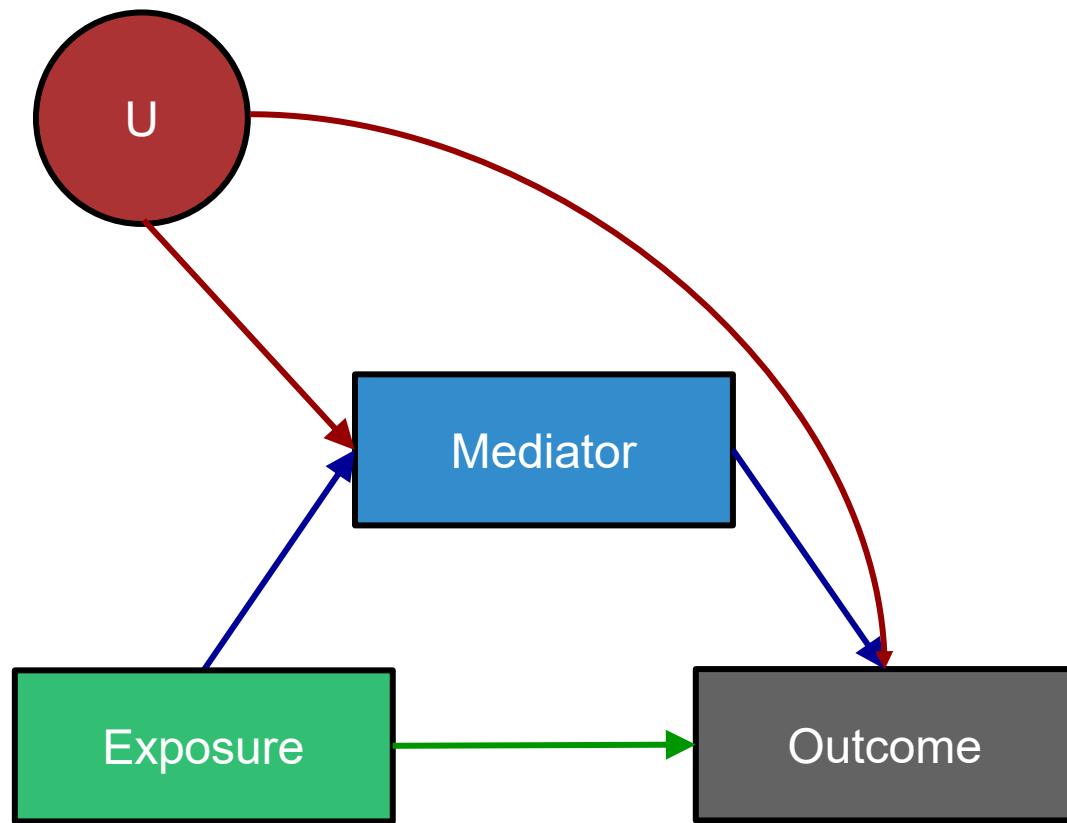
But this is only possible if all confounding is **time invariant confounding**

- i.e. all confounding variables **crystallised** before the first exposure of interest



MEDIATOR-OUTCOME CONFOUNDING

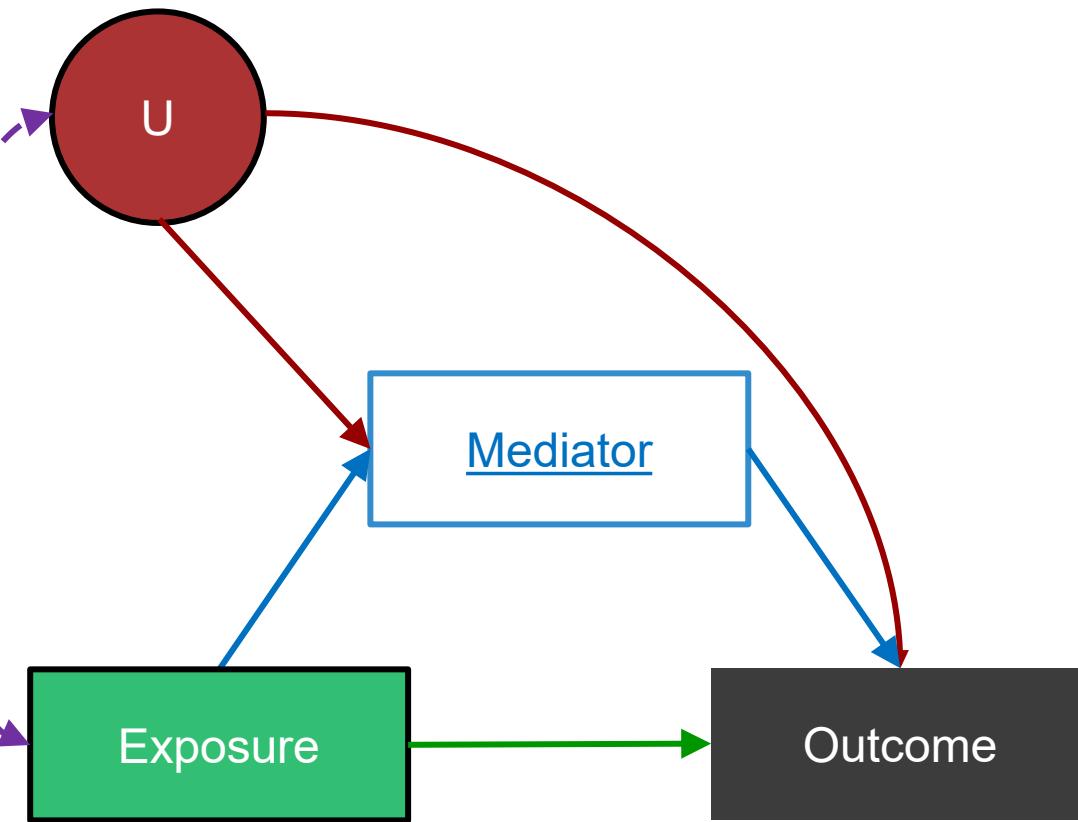
Main problem with mediation analyses is **mediator-outcome confounding**



MEDIATOR-OUTCOME CONFOUNDING

Main problem with mediation analyses is **mediator-outcome confounding**

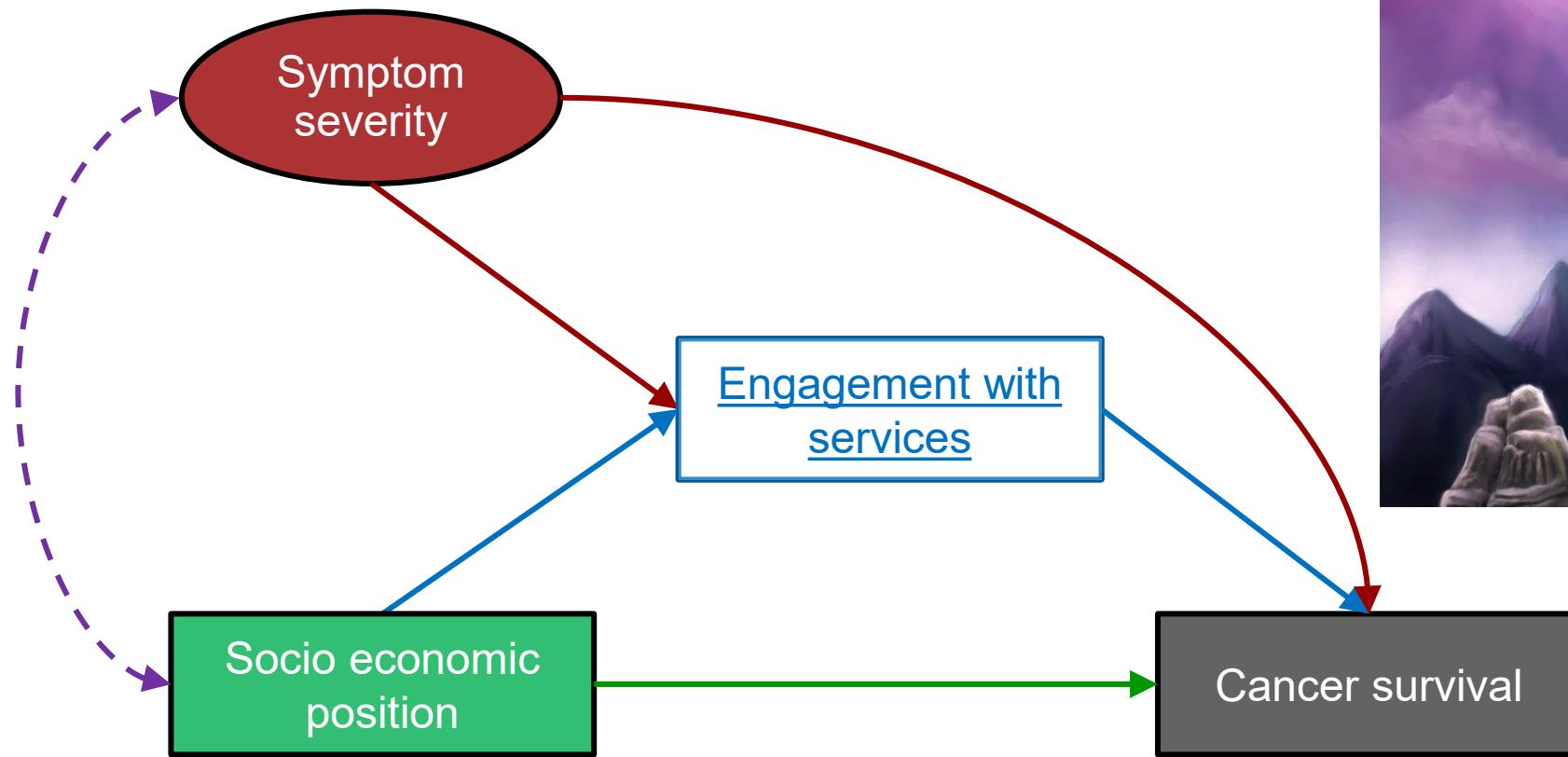
Mediator adjustment + **MO confounding** summons **collider bias!**



Collider bias (artist's impression)

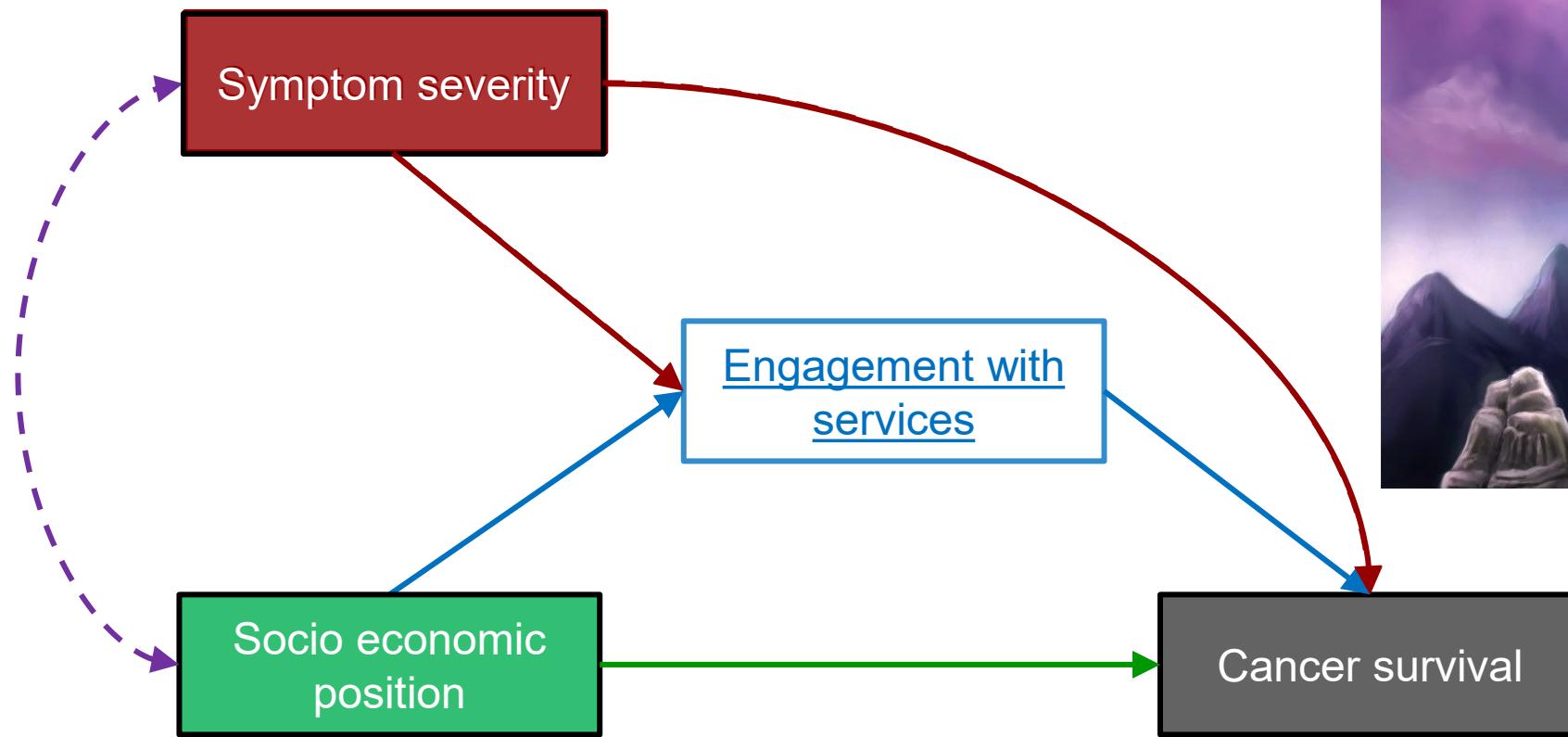
MEDIATOR-OUTCOME CONFOUNDING

e.g. Identifying effect of **socio-economic position** on **cancer survival** mediated through **engagement with healthcare services**



MEDIATOR-OUTCOME CONFOUNDING

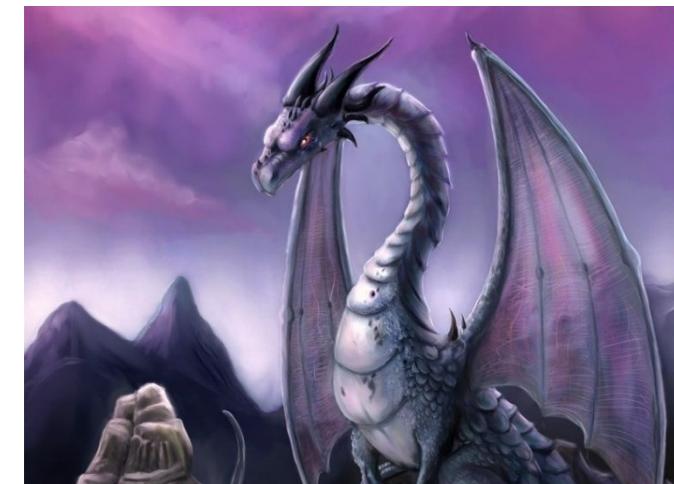
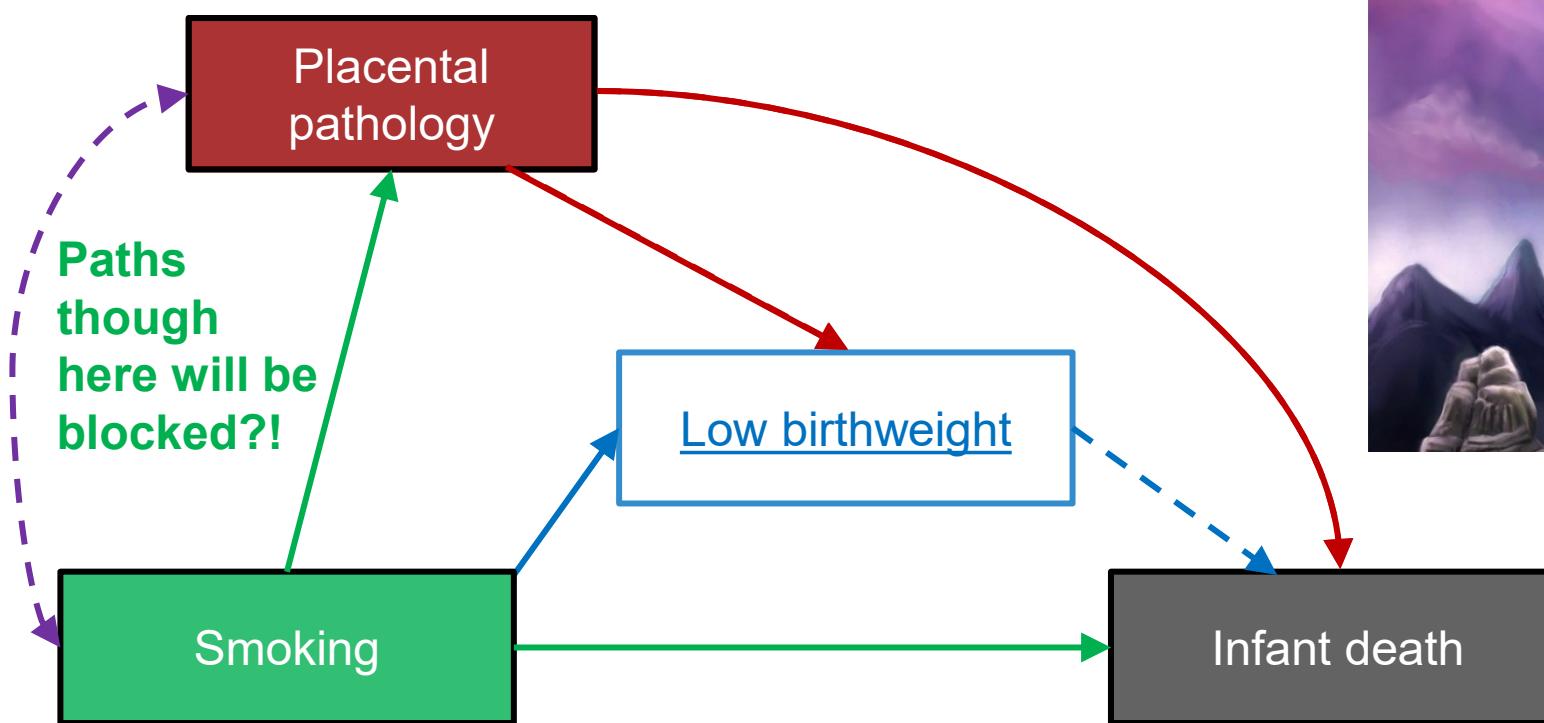
Solution: Measure **mediator-outcome confounders** & condition!?



INTERMEDIATE CONFOUNDING

Problem: This is not suitable if the **mediator-outcome confounder** is also a **mediator** for the **exposure**

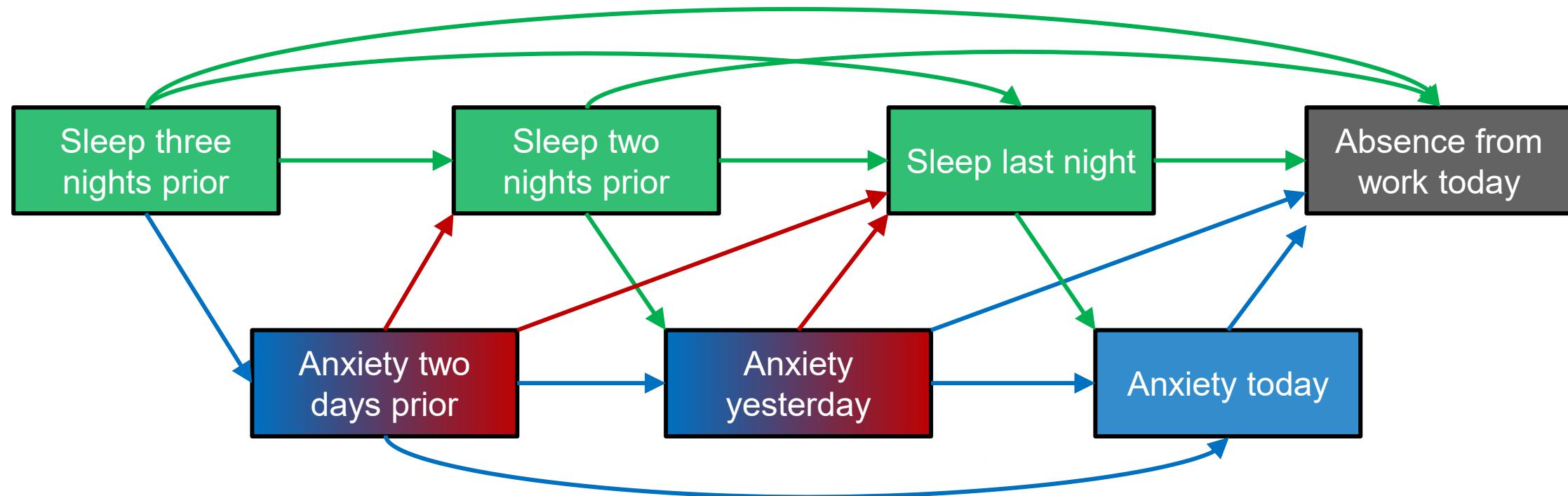
We cannot condition for these **intermediate confounders** without blocking part of the causal effect of the exposure



TIME-VARYING CONFOUNDING

Intermediate confounding occurs most commonly when studying **time-varying exposures**, where it is then known as **time-varying confounding**

Here, daytime **anxiety** is a **time-varying confounder** for night-time sleep, and simultaneously **confounds** and **mediates** the combined effect of '**sleep**'



TIME VARYING CONFOUNDING

Estimating individual causal effects in the presence of **intermediate / time-varying confounding** requires **g-methods**



REPORTING GUIDELINES

JAMA | Special Communication

A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies

The AGRema Statement

Hopin Lee, PhD; Aidan G. Cashin, PhD; Sarah E. Lamb, DPhil; Sally Hopewell, DPhil; Stijn Vansteelandt, PhD; Tyler J. VanderWeele, PhD; David P. MacKinnon, PhD; Gemma Mansell, PhD; Gary S. Collins, PhD; Robert M. Golub, MD; James H. McAuley, PhD; and the AGRema group

IMPORTANCE Mediation analyses of randomized trials and observational studies can generate evidence about the mechanisms by which interventions and exposures may influence health outcomes. Publications of mediation analyses are increasing, but the quality of their reporting is suboptimal.

OBJECTIVE To develop international, consensus-based guidance for the reporting of mediation analyses of randomized trials and observational studies (A Guideline for Reporting Mediation Analyses; AGRema).

DESIGN, SETTING, AND PARTICIPANTS The AGRema statement was developed using the Enhancing Quality and Transparency of Health Research (EQUATOR) methodological framework for developing reporting guidelines. The guideline development process included (1) an overview of systematic reviews to assess the need for a reporting guideline; (2) review of systematic reviews of relevant evidence on reporting mediation analyses; (3) conducting a Delphi survey with panel members that included methodologists, statisticians, clinical trialists, epidemiologists, psychologists, applied clinical researchers, clinicians, implementation scientists, evidence synthesis experts, representatives from the EQUATOR Network, and journal editors ($n = 19$; June–November 2019); (4) having a consensus meeting ($n = 15$; April 28–29, 2020); and (5) conducting a 4-week external review and pilot test that included methodologists and potential users of AGRema ($n = 21$; November 2020).

RESULTS A previously reported overview of 54 systematic reviews of mediation studies demonstrated the need for a reporting guideline. Thirty-three potential reporting items were identified from 3 systematic reviews of mediation studies. Over 3 rounds, the Delphi panelists ranked the importance of these items, provided 60 qualitative comments for item refinement and prioritization, and suggested new items for consideration. All items were reviewed during a 2-day consensus meeting and participants agreed on a 25-item AGRema statement for studies in which mediation analyses are the primary focus and a 9-item short-form AGRema statement for studies in which mediation analyses are a secondary focus. These checklists were externally reviewed and pilot tested by 21 expert methodologists and potential users, which led to minor adjustments and consolidation of the checklists.

CONCLUSIONS AND RELEVANCE The AGRema statement provides recommendations for reporting primary and secondary mediation analyses of randomized trials and observational studies. Improved reporting of studies that use mediation analyses could facilitate peer review and help produce publications that are complete, accurate, transparent, and reproducible.

- ➡ Editorial page 1011 and Editor's Note page 1057
- ➡ Supplemental content
- ➡ CME Quiz at [jamacmlookup.com](#) and CME Questions page 1062

Author Affiliations: Author affiliations are listed at the end of this article.
Grant Information: The AGRema

Table 1. A Guideline for Reporting Mediation Analyses (AGRema) Long-Form Checklist^a

Section and topic	Item No.	Item description
Title and abstract		
Title	1	<ul style="list-style-type: none">▪ Identify that the study uses mediation analyses
Abstract	2	<ul style="list-style-type: none">▪ Provide a structured summary of the objectives, methods, results, and conclusions specific to mediation analyses
Introduction		
Background and rationale	3	<ul style="list-style-type: none">▪ Describe the study background and theoretical rationale for investigating the mechanisms of interest▪ Include supporting evidence or theoretical rationale for why the intervention or exposure might have a causal relationship with the proposed mediators▪ Include supporting evidence or theoretical rationale for why the mediators might have a causal relationship with the outcomes
Objectives	4	<ul style="list-style-type: none">▪ State the objectives of the study specific to the mechanisms of interest▪ The objectives should specify whether the study aims to test or estimate the mechanistic effects
Methods		
Study registration	5	<ul style="list-style-type: none">▪ If applicable, provide references to any protocols or study registrations specific to mediation analyses and highlight any deviations from the planned protocol
Study design and source of data	6	<ul style="list-style-type: none">▪ Specify the design of the original study that was used in mediation analyses and where the details can be accessed, supported by a reference▪ If applicable, describe study design features that are relevant to mediation analyses
Participants	7	<ul style="list-style-type: none">▪ Describe the target population, eligibility criteria specific to mediation analyses, study locations, and study dates (start of participant enrollment and end of follow-up)
Sample size	8	<ul style="list-style-type: none">▪ State whether a sample size calculation was conducted for mediation analyses▪ If so, explain how it was calculated
Effects of interest	9	<ul style="list-style-type: none">▪ Specify the effects of interest
Assumed causal model	10	<ul style="list-style-type: none">▪ Include a graphic representation of the assumed causal model including the exposure, mediator, outcome, and possible confounders
Causal assumptions	11	<ul style="list-style-type: none">▪ Specify assumptions about the causal model
Measurement	12	<ul style="list-style-type: none">▪ Clearly describe the interventions or exposures, mediators, outcomes, confounders, and moderators that were used in the analysis▪ Specify how and when they were measured, the measurement properties, and whether blinded assessment was used
Measurement levels	13	<ul style="list-style-type: none">▪ If relevant, describe the levels at which the exposure, mediator, and outcome were measured
Statistical methods	14	<ul style="list-style-type: none">▪ Describe the statistical methods used to estimate the causal relationships of interest▪ This description should specify analytic strategies used to reduce confounding, model building procedures, justification for the inclusion or exclusion of possible interaction terms, modeling assumptions, and methods used to handle missing data▪ Provide a reference to the statistical software and package used
Sensitivity analyses	15	<ul style="list-style-type: none">▪ Describe any sensitivity analyses that were used to explore causal or statistical assumptions and the influence of missing data
Ethical approval	16	<ul style="list-style-type: none">▪ Name the institutional research board or ethics committee that approved the study▪ Provide a description of participant informed consent or ethics committee waiver of informed consent
Results		
Participants	17	<ul style="list-style-type: none">▪ Describe baseline characteristics of participants included in mediation analyses▪ Report the total sample size and number of participants lost during follow-up or with missing data
Outcomes and estimates	18	<ul style="list-style-type: none">▪ Report point estimates and uncertainty estimates for the exposure-mediator and mediator-outcome relationships▪ If inference concerning the causal relationship of interest is considered feasible given the causal assumptions, report the point estimate and uncertainty estimate
Sensitivity parameters	19	<ul style="list-style-type: none">▪ Report the results from any sensitivity analyses used to assess robustness of the causal or statistical assumptions and the influence of missing data
Discussion		
Limitations	20	<ul style="list-style-type: none">▪ Discuss the limitations of the study including potential sources of bias
Interpretation	21	<ul style="list-style-type: none">▪ Interpret the estimated effects considering the study's magnitude and uncertainty, plausibility of the causal assumptions, limitations, generalizability of the findings, and results from relevant studies
Implications	22	<ul style="list-style-type: none">▪ Discuss the implications of the overall results for clinical practice, policy, and science
Other information		
Funding and role of sponsor	23	<ul style="list-style-type: none">▪ List all sources of funding or sponsorship for mediation analyses and the role of the funders/sponsors in the conduct of the study, writing of the manuscript, and decision to submit the manuscript for publication
Conflicts of interest and financial disclosures	24	<ul style="list-style-type: none">▪ State any conflicts of interest and financial disclosures for all authors
Data and code	25	<ul style="list-style-type: none">▪ Authors are encouraged to provide a statement for sharing data and code for mediation analyses

^a Designed for articles that report primary mediation analyses of randomized trials or observational studies or those that report secondary mediation analyses as the primary focus of an article. Republished with permission from the AGRema group. This checklist is copyrighted by the AGRema group under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0) license.

RECOMMENDED READING

Books

- VanderWeele TJ. Explanation in Causal Inference: Methods for Mediation and Interaction. New York: Oxford University Press, 2015.

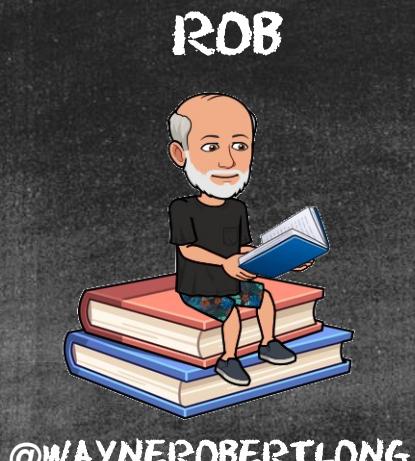
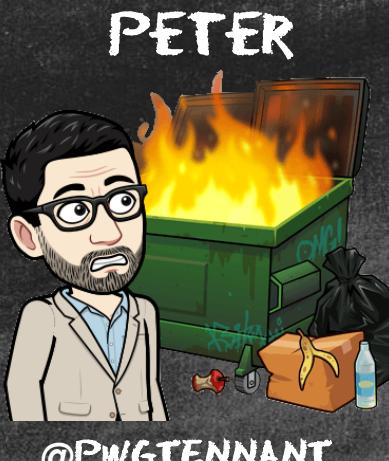
Papers

- VanderWeele TJ. A unification of mediation and interaction: a four-way decomposition. *Epidemiology (Cambridge, Mass)* 2014; 25: 749.
- VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annual review of public health* 2016; 37: 17–32.
- Lange, T., Hansen, K.W., Sørensen, R. and Galatius, S., 2017. Applied mediation analyses: a review and tutorial. *Epidemiology and health*, 39.
- Lee, H., Cashin, A.G., Lamb, S.E., Hopewell, S., Vansteelandt, S., VanderWeele, T.J., MacKinnon, D.P., Mansell, G., Collins, G.S., Golub, R.M. and McAuley, J.H., 2021. A guideline for reporting mediation analyses of randomized trials and observational studies: The AGReMA Statement. *JAMA*, 326(11), pp.1045-1056.

SUMMARY

- Traditional methods for **mediation analyses** make naïve assumptions about lack of interactions
- Counterfactual methods suggest extending our ‘**exposure window**’ to cover all our exposures (and mediators) of interest
- If you wish to break down your total effect into different parts, **VanderWeele's four-way decomposition** is probably your best option
- Mediation methods fail in the presence of **mediator-outcome confounding**
- Conditioning is possible if mediator-outcome confounder is not caused by exposure (i.e. is not an **intermediate confounder**)
- With **intermediate confounding** and **time-varying confounding** you will need to use **g-methods**

4.2 - INTRODUCTION TO G-METHODS



09:30-10:45 LECTURE 4.1

10:45-11:00 Q&A

11:00-11:30 TEA & COFFEE

11:30-12:15 LECTURE 4.2

12:15-13:00 ACTIVITY 4-A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 4.3

14:45-15:30 ACTIVITY 4-B

15:30-16:00 TEA & COFFEE

16:00-16:45 ACTIVITY 4-C

16:45-17:45 ACTIVITY 4-D

17:30-18:00 Q&A

LEARNING OBJECTIVES

- Understand the types of situations in which the ‘g-methods’ are required
- Define the total causal effect (TCE) of a time-varying exposure
- Identify the 3 different g-methods
 - ✓ basic principles
 - ✓ caveats and limitations

Important concepts

- what features of longitudinal data are we interested in?
- time-varying exposures and time-dependent confounding
- exposure/treatment regimes
- total causal effects (TCEs)
- the identifiability conditions

G-methods

- the g-formula
- inverse probability of treatment weighting (IPTW) of marginal structural models (MSMs)
- g-estimation of structural nested models (SNMs)

LONGITUDINAL DATA

Many methods focus on longitudinal data

- spawned the field of ‘lifecourse epidemiology’

Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in lifecourse research

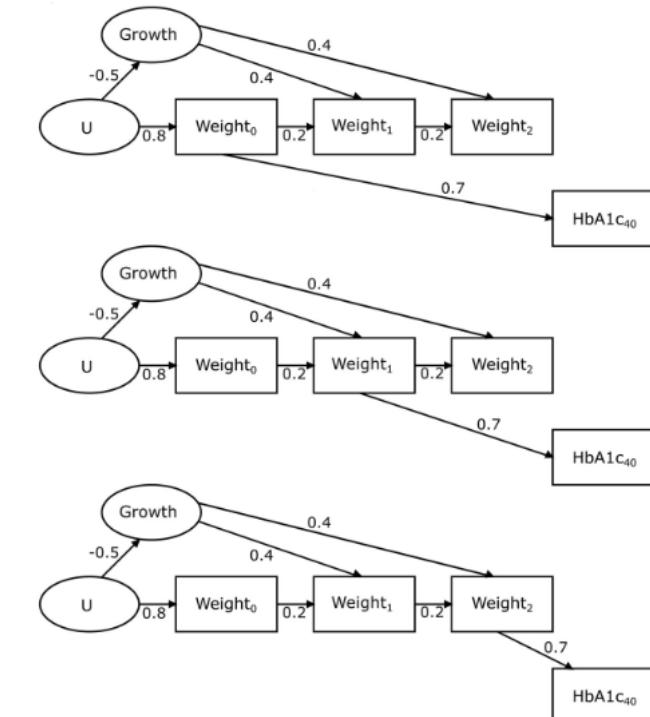
Sarah C. Gadd^{1,2*}, Peter W. G. Tennant^{1,3,4}, Alison J. Heppenstall^{1,2,4}, Jan R. Boehnke⁵,
Mark S. Gilthorpe^{1,3,4}

But what are we asking of the data?

If interested in body weight, what matters most ...

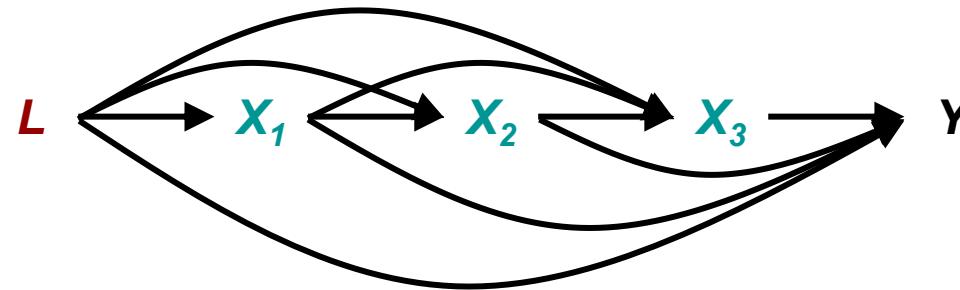
- is it weight at specific times?
- is it the accumulation of weight over time?
- is it the ‘pattern’ (often called ‘trajectory’) over time?

How do we unpick this?

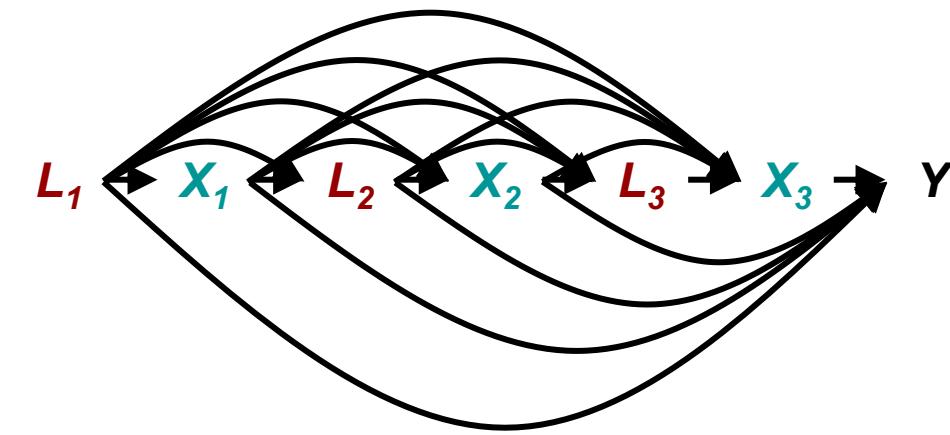


TIME-VARYING VARIABLES

Time-varying exposures:



Time-dependent confounding:



Possible methods of analysis:

- Estimate the total causal effect of each exposure measurement on the outcome *individually* using...
 - ✓ separate standard regression models
 - ✓ ‘unexplained residuals’ (UR) model
- Estimate the *joint* total causal effect of all measurements of the exposure on the outcome by considering them together as ‘exposure regimes’ using...
 - ✓ **g-methods!**

EXPOSURE REGIMES

Exposure regimes describe a sequence of exposures / treatments

For a **binary exposure** measured at **two time points**, there are **four possible exposure regimes**:

- $\bar{a} = (0,0)$ \Rightarrow 'never treat'
- $\bar{a} = (0,1)$ \Rightarrow 'treat late'
- $\bar{a} = (1,0)$ \Rightarrow 'treat early'
- $\bar{a} = (1,1)$ \Rightarrow 'always treat'

TOTAL CAUSAL EFFECT (TCE) OF A TIME-VARYING EXPOSURE

The TCE represents a counterfactual contrast between the 'always treat' ($\bar{a} = (\mathbf{1}, \mathbf{1})$) and 'never treat' ($\bar{a} = (\mathbf{0}, \mathbf{0})$) exposure regimes

- *What would be the average difference in outcomes if everyone were **always treated** compared to if they were **never treated**?*
- $E(Y^{\mathbf{1},\mathbf{1}} - Y^{\mathbf{0},\mathbf{0}}) = \hat{Y}^{\mathbf{1},\mathbf{1}} - \hat{Y}^{\mathbf{0},\mathbf{0}}$

THE IDENTIFIABILITY CONDITIONS

1. Consistency:

- The **counterfactual outcome** $Y^{\bar{a}}$ for an individual who received the exposure regime \bar{a} is equal to their **observed outcome** Y

2. (Sequential) conditional exchangeability:

- The counterfactual outcome $Y^{\bar{a}}$ associated with the particular exposure regime \bar{a} is independent of the exposure regime that was actually observed
- ‘Assumption of **no unmeasured confounding**’

3. Positivity:

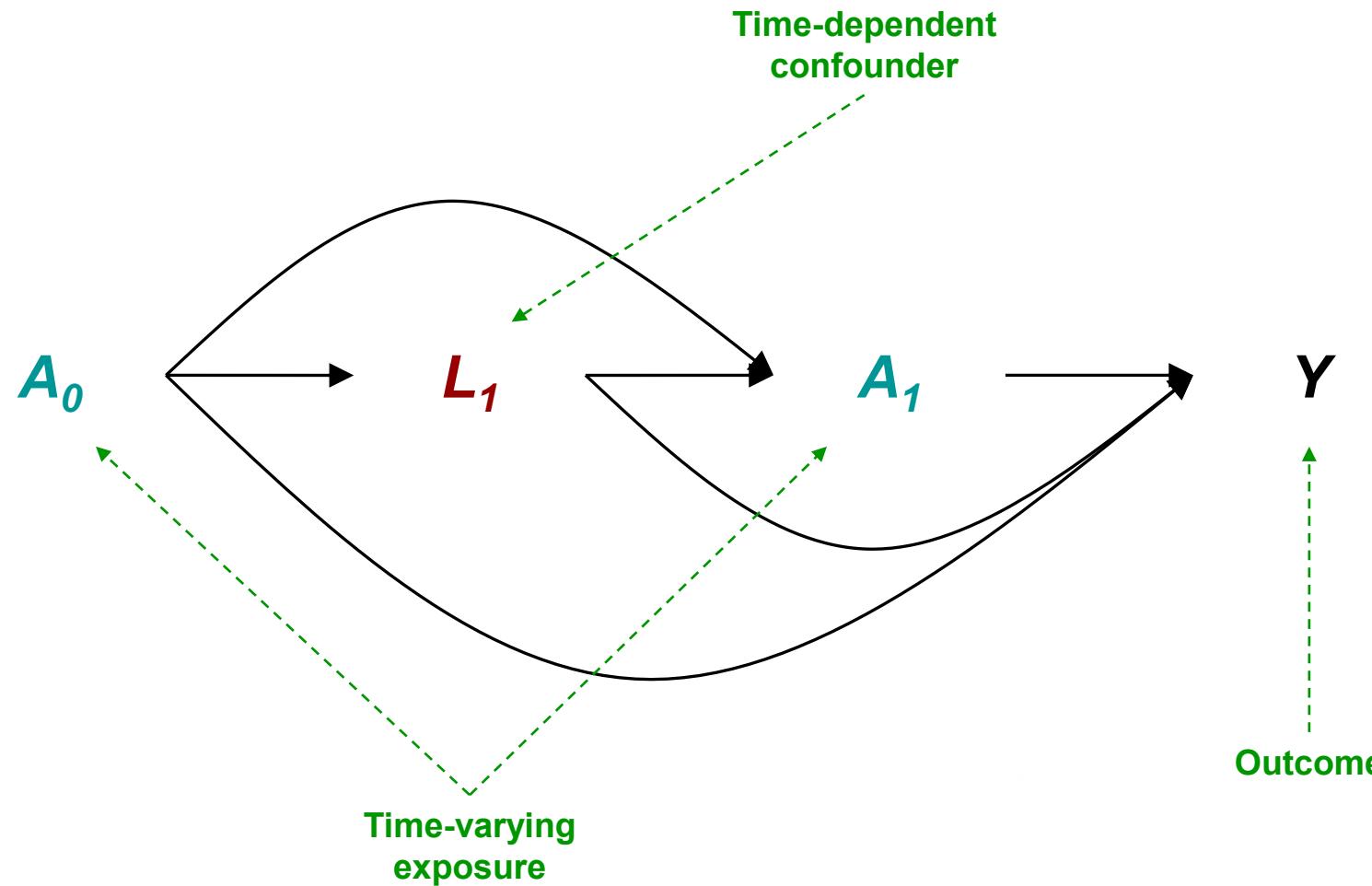
- Every individual has a **non-zero chance** of receiving the exposure at each time point, regardless of prior exposure and confounder history

The g-methods include...

- the g-formula
- inverse probability of treatment weighting (IPTW) of marginal structural models (MSMs)
- g-estimation of structural nested models (SNMs)

These methods allow for the estimation of the total causal effect of a time-varying exposure, even *in the presence of **time-dependent confounding***

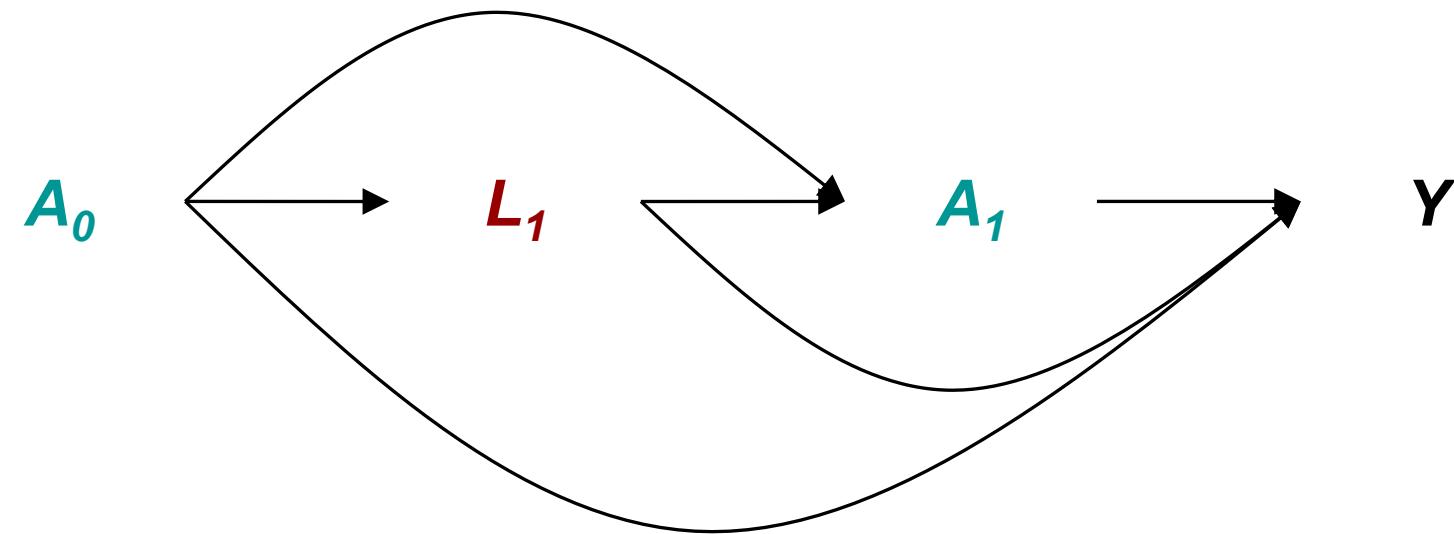
EXAMPLE SCENARIO



THE G-FORMULA

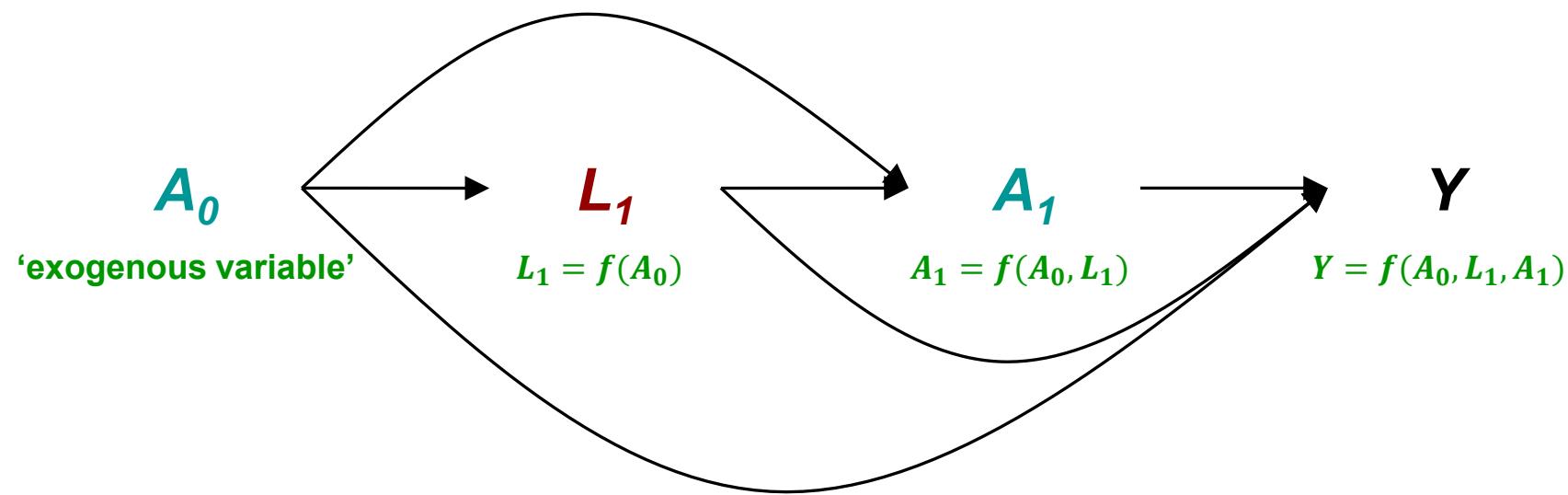
2-step process:

1. ‘parameterise’ the DAG
2. estimate the target counterfactual quantities to compute the TCE



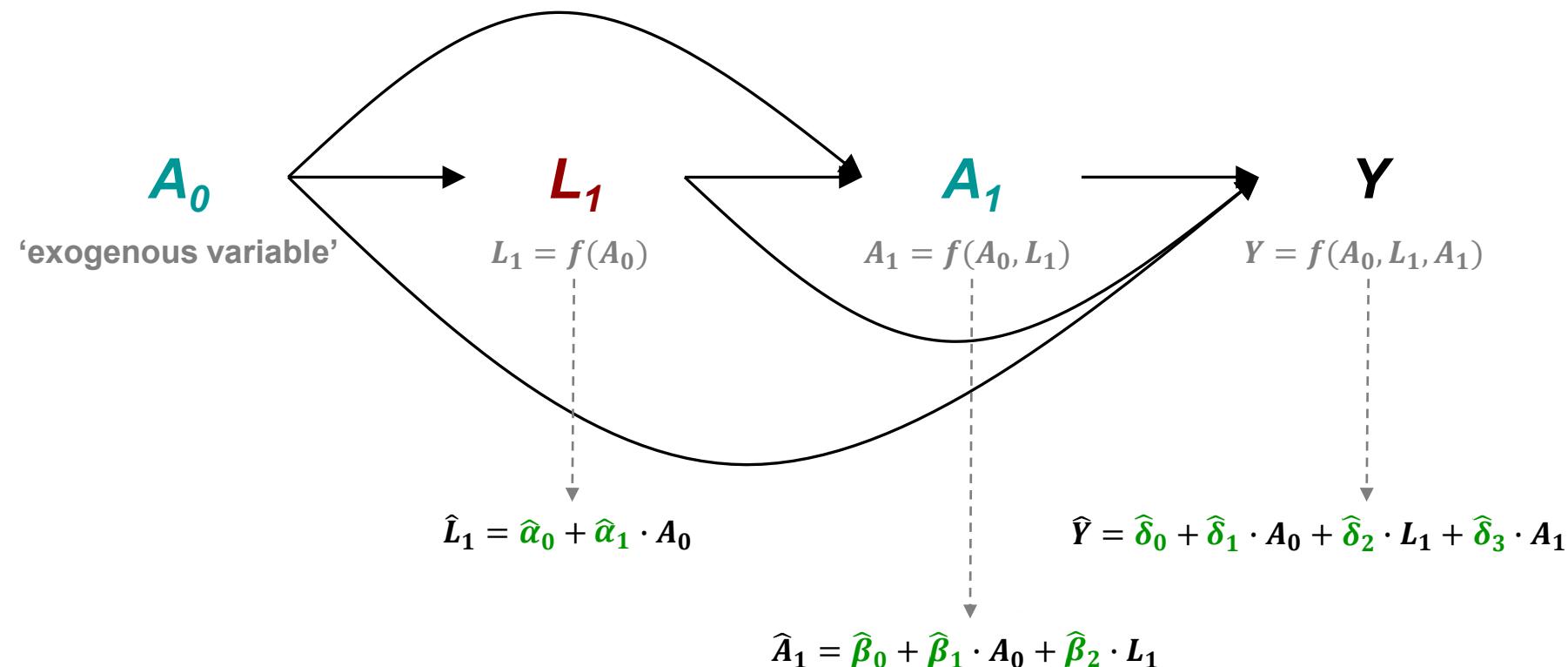
1. ‘PARAMETERISE’ THE DAG

A DAG represents the **presumed data-generating process** – the value for each variable is determined by the variables that causally precede it



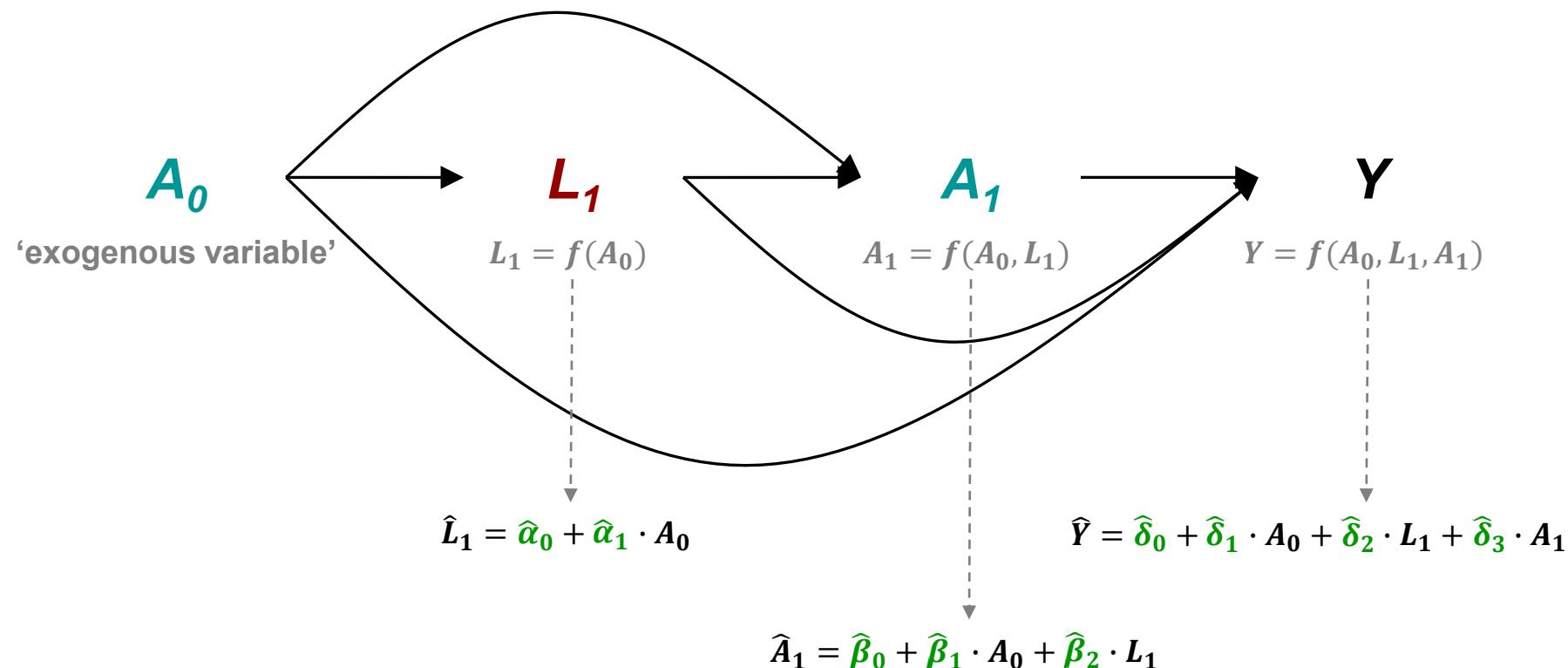
1. 'PARAMETERISE' THE DAG

Assuming linearity...



1. 'PARAMETERISE' THE DAG

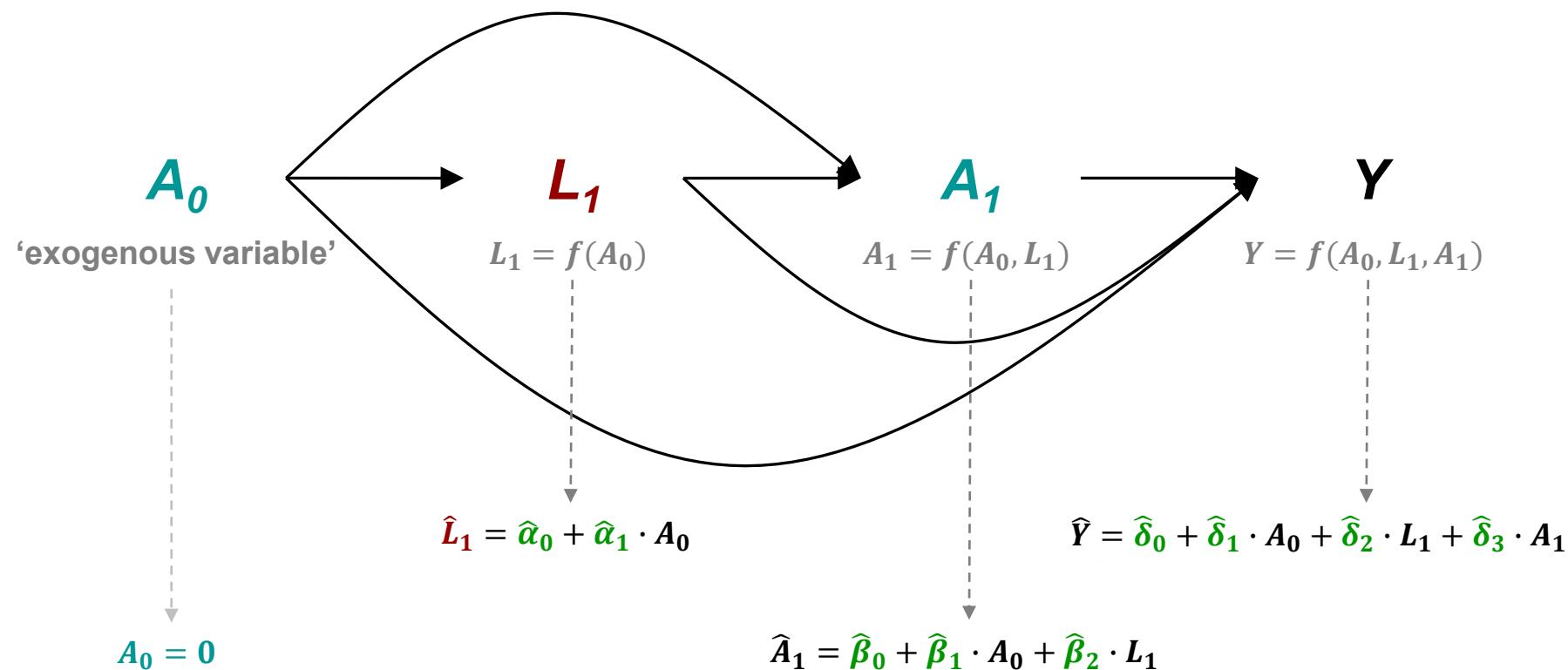
Assuming linearity...



2. ESTIMATE THE TARGET COUNTERFACTUALS

For example...

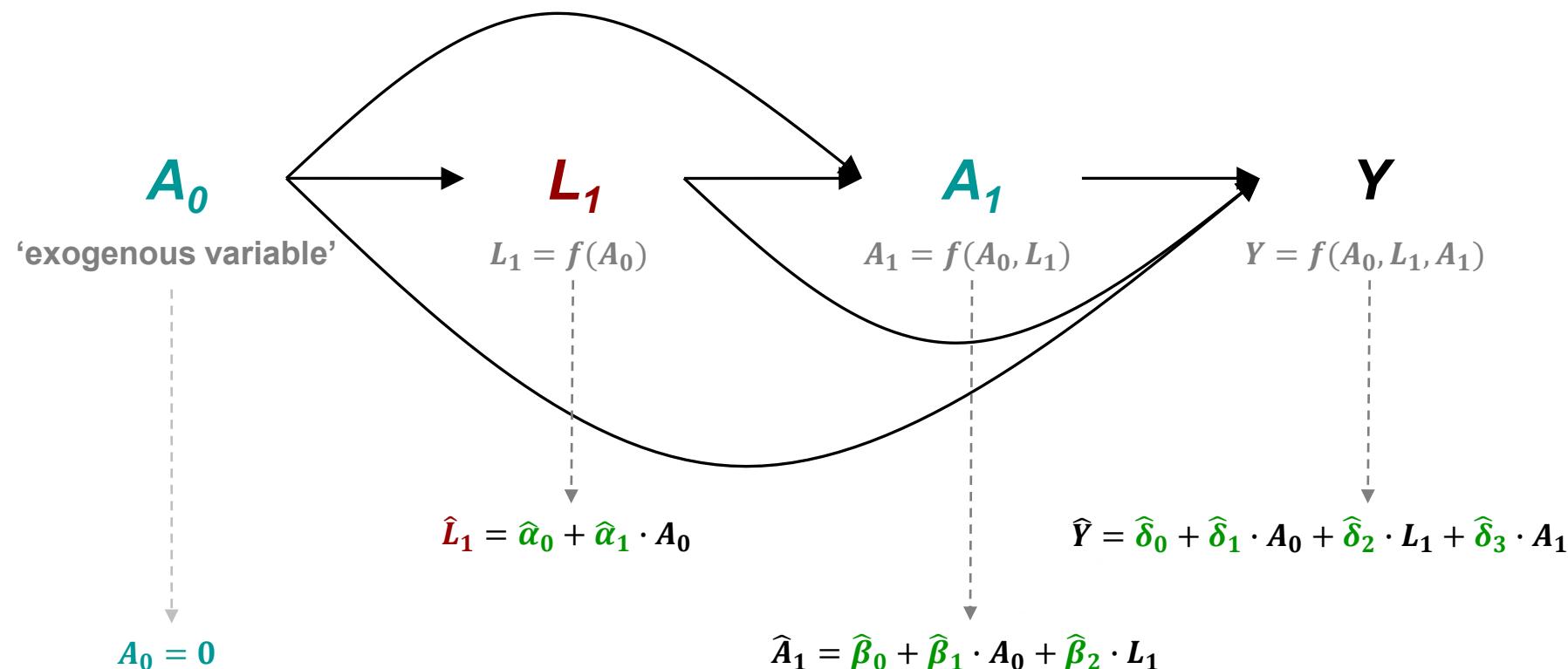
- what is the expected value of Y if no one were treated (i.e. $A_0 = 0$ and $A_1 = 0$)?



2. ESTIMATE THE TARGET COUNTERFACTUALS

For example...

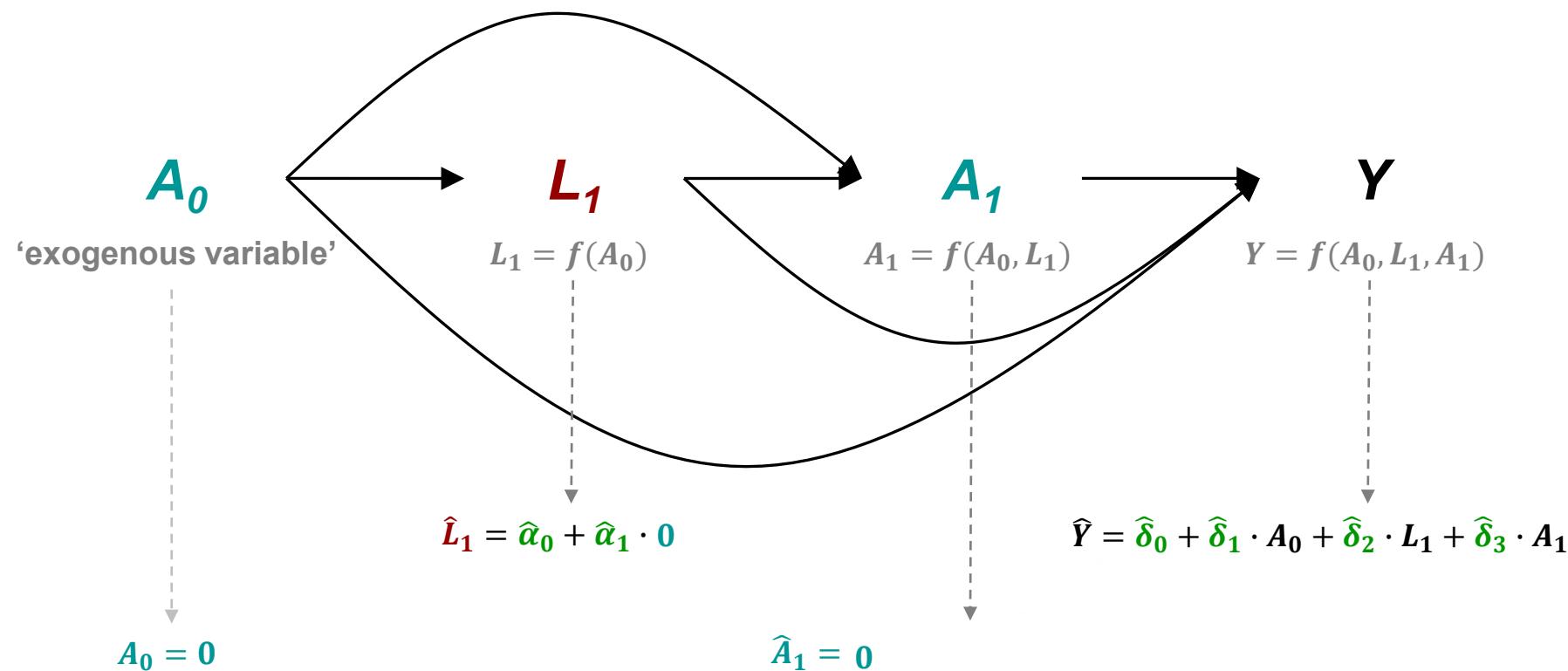
- What is the expected value of Y if no one were treated (i.e. $A_0 = 0$ and $A_1 = 0$)?



2. ESTIMATE THE TARGET COUNTERFACTUALS

For example...

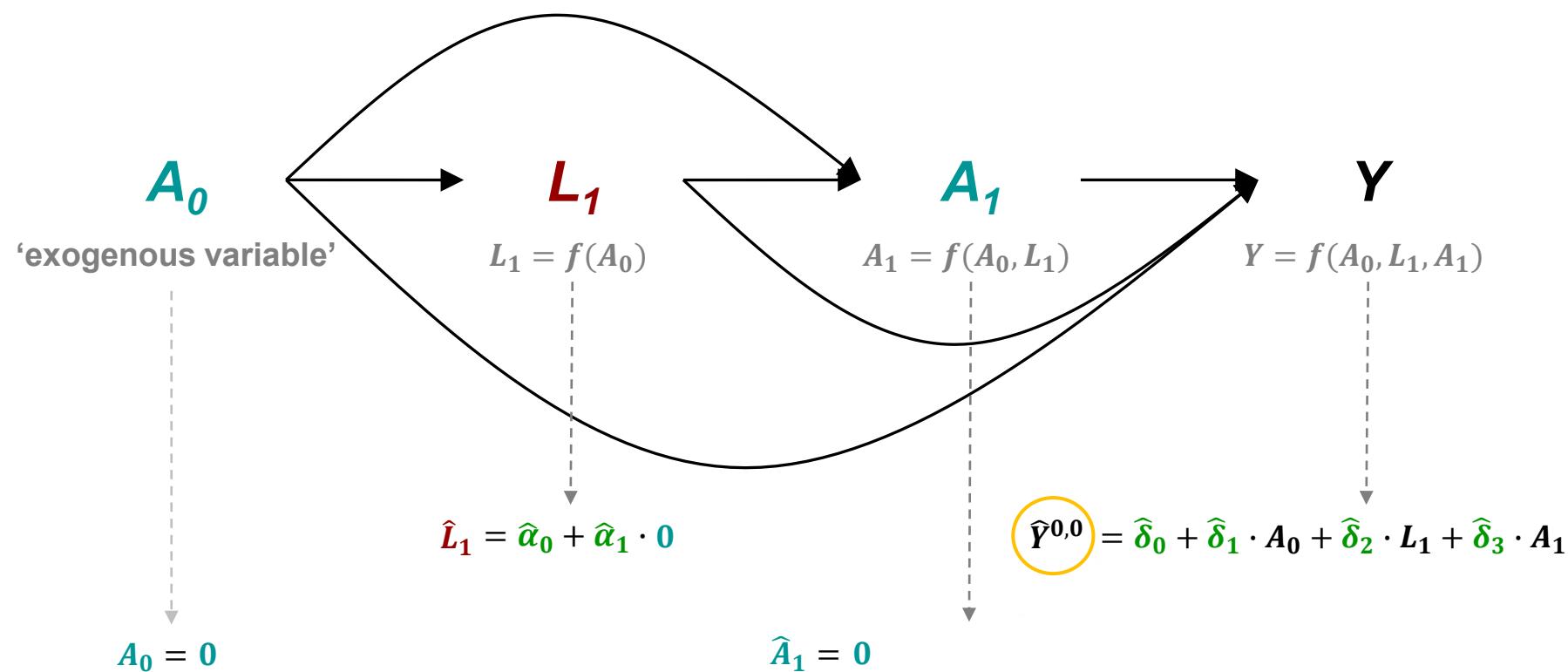
- What is the expected value of Y if no one were treated (i.e. $A_0 = 0$ and $A_1 = 0$)?



2. ESTIMATE THE TARGET COUNTERFACTUALS

For example...

- What is the expected value of Y if no one were treated (i.e. $A_0 = 0$ and $A_1 = 0$)?



THE G-FORMULA (IN A NUTSHELL)

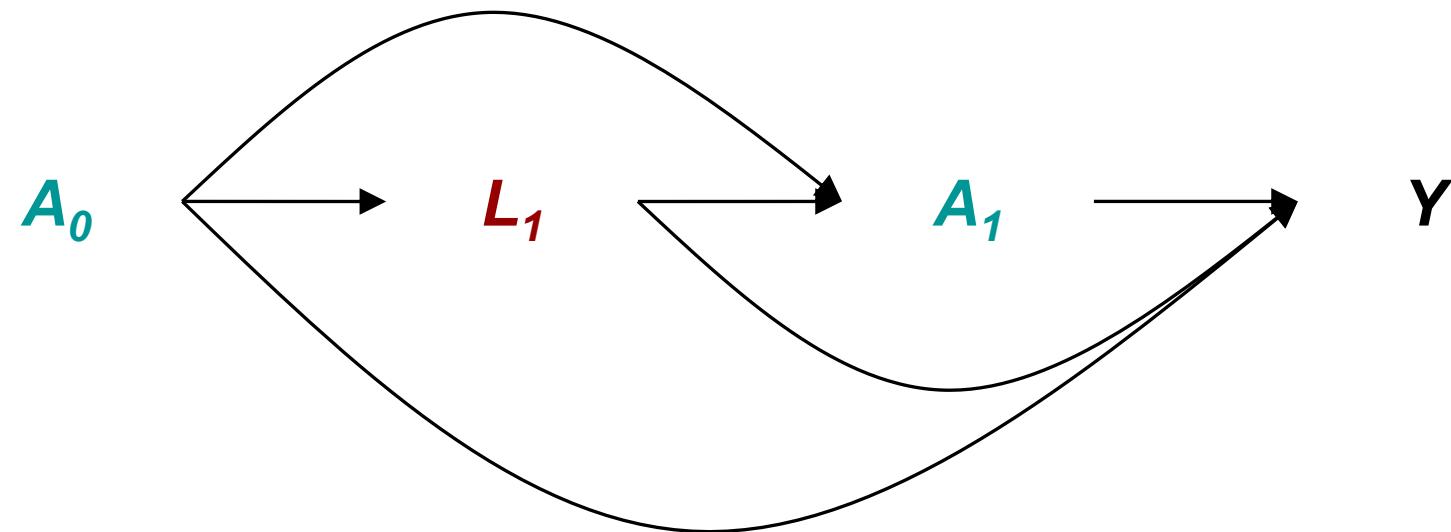
Simulates the joint distribution of the variables that **would have been observed** under a **hypothetical intervention** targeting the exposure – based on the joint distribution that **was actually observed**

Note:

- can lead to bias if any parts of the joint distribution are not correctly parameterised
- computationally intensive (requires Monte Carlo simulation & bootstrapping)

2-step process:

1. create a 'pseudo population' using IPTW
2. estimate MSM parameters



1. CREATE A ‘PSEUDO POPULATION’

Weight each individual by the inverse of the conditional probability of receiving the treatment that they actually received (this is based on their **propensity score!**)

Unstabilised weights:

$$w = \prod_{t=0}^{t=1} \frac{1}{P(A_t | \bar{A}_{t-1}, \bar{L}_t)} = \frac{1}{P(A_0)} \cdot \frac{1}{P(A_1 | A_0, L_1)}$$

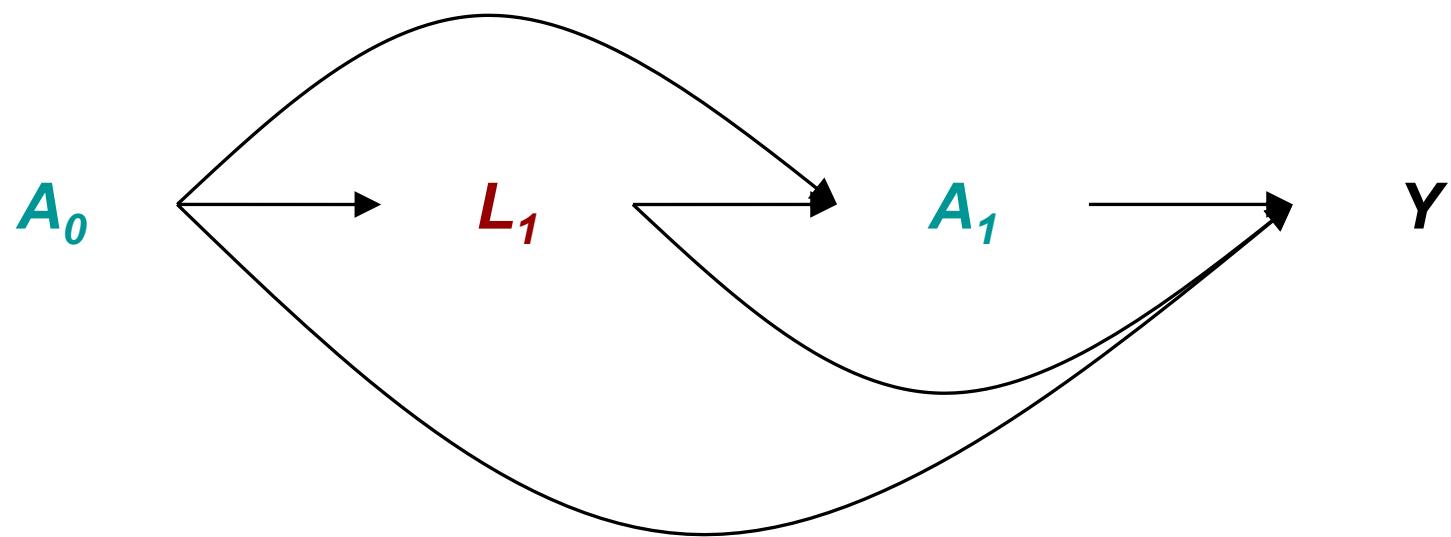
- All individuals up-weighted
- Pseudo-population is 4x the size of original population

Stabilised weights:

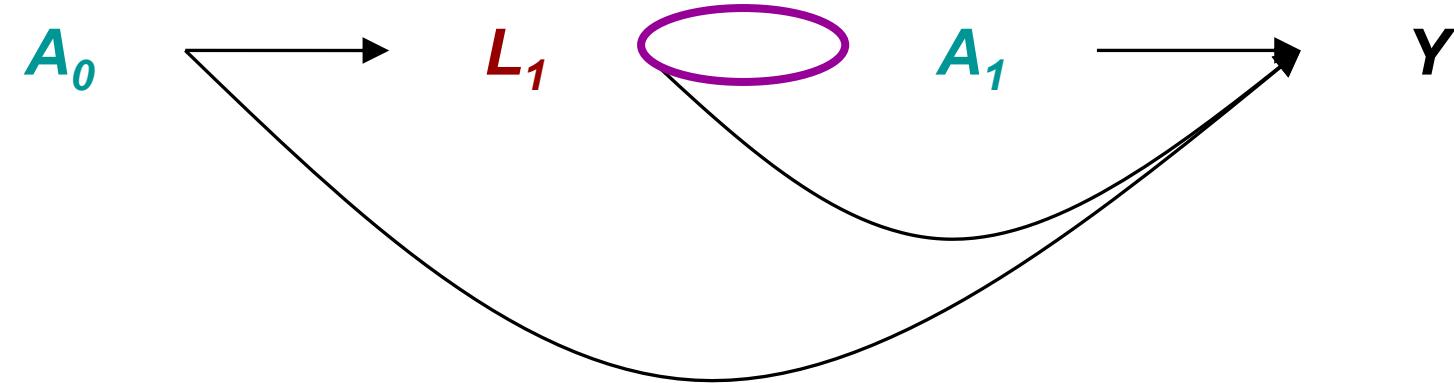
$$sw = \prod_{t=0}^{t=1} \frac{P(A_t | \bar{A}_{t-1})}{P(A_t | A_{t-1}, \bar{L}_t)} = \frac{P(A_0)}{P(A_0)} \cdot \frac{P(A_1 | A_0)}{P(A_1 | A_0, L_1)}$$

- Some individuals up-weighted, others down-weighted
- Pseudo-population is the same size as original population

Original DAG (unweighted population):



Modified DAG (weighted population):



2. ESTIMATE MSM PARAMETERS

Marginal structural model:

$$\hat{Y}^{a_0, a_1} = \alpha + \beta_0 \cdot a_0 + \beta_1 \cdot a_1 + \beta_2 \cdot a_0 a_1$$

Average effect of a_0

Average effect of a_1

Average additional joint
effect of a_0 and a_1

$$TCE = \hat{Y}^{1,1} - \hat{Y}^{0,0}$$

$$= [\alpha + \beta_0 \cdot 1 + \beta_1 \cdot 1 + \beta_2 \cdot 1 \cdot 1] - [\alpha + \beta_0 \cdot 0 + \beta_1 \cdot 0 + \beta_2 \cdot 0 \cdot 0]$$

$$= \beta_0 + \beta_1 + \beta_2$$

IPTW OF MSMS (IN A NUTSHELL)

IPTW creates a '**pseudo-population**' in which there exists ***no time-dependent confounding***, and MSMs are used to estimate the causal parameters of interest in this population

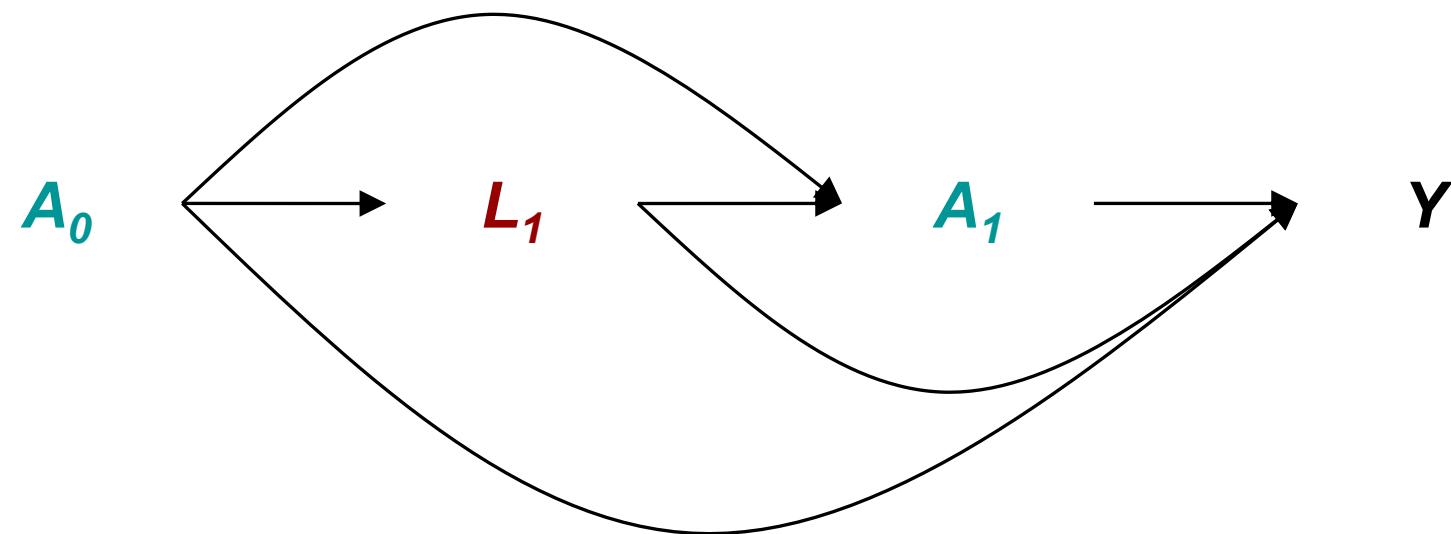
Note:

- extreme up-weighting can occur for individuals with rare exposure/treatment patterns (aka issues with propensity scores)
- not well-suited to continuous exposures
- assumes effect is the same for all levels of the time-dependent confounder (i.e. does not accommodate interactions)

G-ESTIMATION OF SNMS

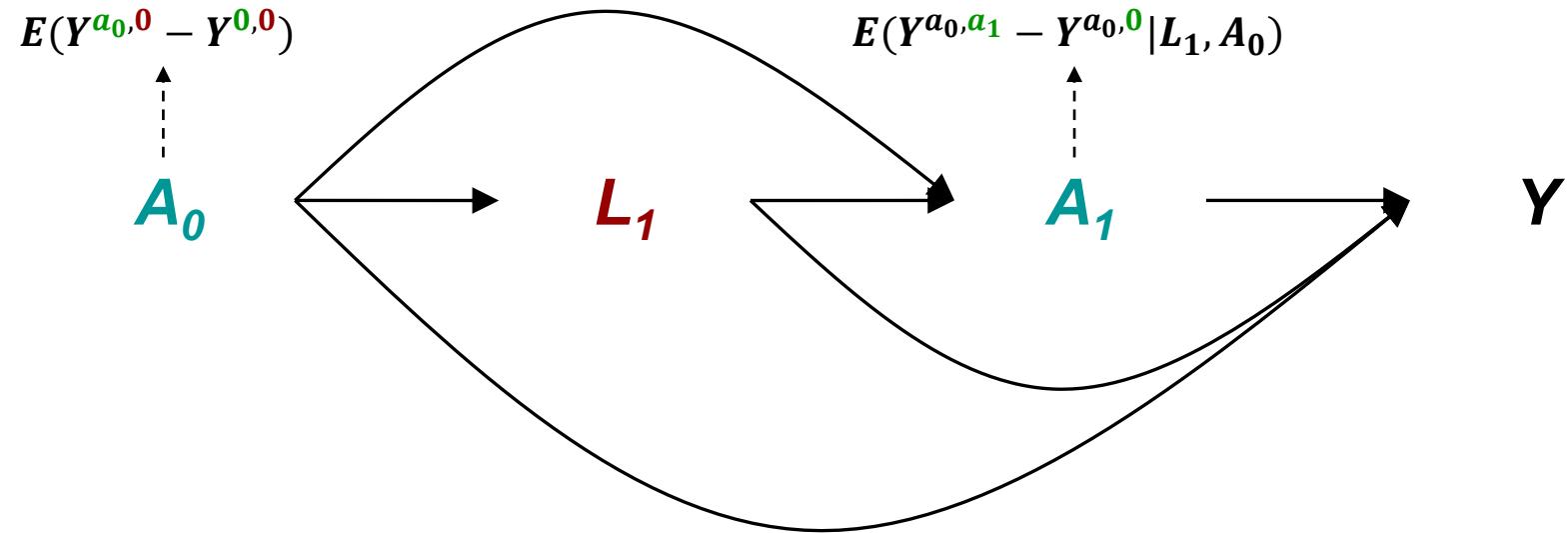
2-step process:

1. define structural nested model (SNM)
2. estimate SNM parameters using g-estimation



1. DEFINE SNM

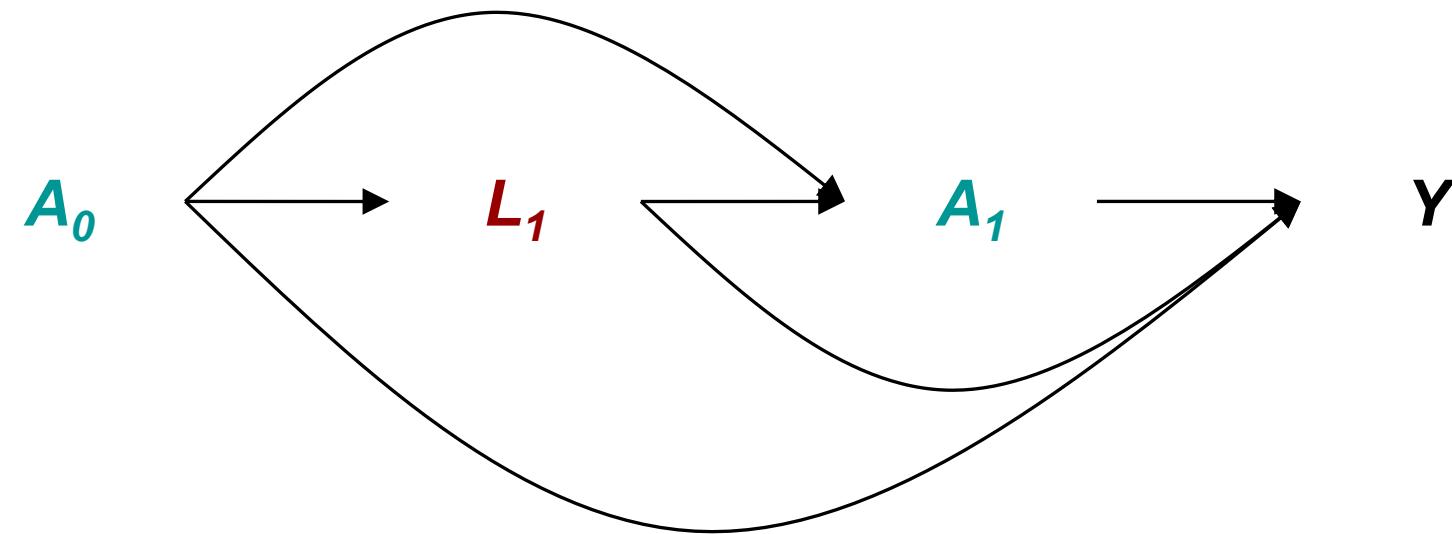
An SNM describes the **average effect of exposure** at each time point, assuming **no future exposure**



STRUCTURAL NESTED MODEL

$$E(Y^{a_0,0} - Y^{0,0}) = \beta_0 \cdot a_0$$

$$E(Y^{a_0,a_1} - Y^{a_0,0} | L_1, A_0) = \delta_0 \cdot a_1 + \delta_1 \cdot a_1 \cdot l_1 + \delta_2 \cdot a_1 \cdot a_0 + \delta_3 \cdot a_1 \cdot l_1 \cdot a_0$$



2. ESTIMATE SNM PARAMETERS

g-estimation is a process which exploits the (sequential) conditional exchangeability assumption to produce estimates of $\beta_0, \delta_0, \delta_1, \delta_2, \delta_3$:

$$Y^{a_0, a_1} \coprod A_0$$

and

$$Y^{a_0, a_1} \coprod A_1 | L_1, A_0$$

G-ESTIMATION OF SNMS (IN A NUTSHELL)

Estimates the effect of exposure at each time point by solving equations that directly result from (sequential) conditional exchangeability

Note:

- more **robust** than the g-formula, and more **flexible** than IPTW of MSMs
- does not require positivity condition
- most **conceptually and practically challenging** to implement

RECOMMENDED READING

- Naimi, A.I., Cole, S.R. and Kennedy, E.H., 2017. An introduction to g methods. *International journal of epidemiology*, 46(2), pp.756-762.
- Mansournia, M.A., Etminan, M., Danaei, G., Kaufman, J.S. and Collins, G., 2017. Handling time varying confounding in observational research. *bmj*, 359.
- Bodnar, L.M., Davidian, M., Siega-Riz, A.M. and Tsiatis, A.A., 2004. Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. *American Journal of Epidemiology*, 159(10), pp.926-934.

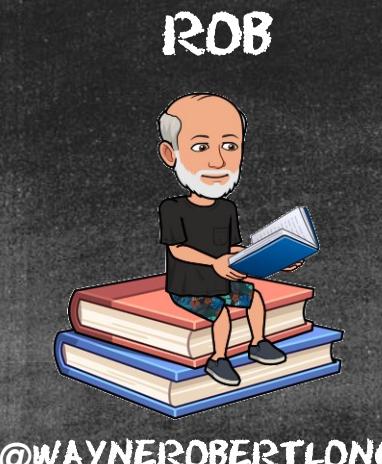
SUMMARY

The three g-methods may be used to estimate causal effects for time-varying exposures, even when there exists time-dependent confounding

All three methods...

- are more conceptually and practically challenging to implement than standard regression analyses
- give identical estimates of the total causal effect under ‘ideal conditions’
- but have different benefits and drawbacks for their implementation

4.3 - DETERMINISTIC RELATIONSHIPS AND TAUTOLOGICAL ASSOCIATIONS



DAY 4

09:30-10:45 LECTURE 4.1

10:45-11:00 Q&A

11:00-11:30 TEA & COFFEE

11:30-12:15 LECTURE 4.2

12:15-13:00 ACTIVITY 4-A

13:00-14:00 LUNCH

14:00-14:45 LECTURE 4.3

14:45-15:30 ACTIVITY 4-B

15:30-16:00 TEA & COFFEE

16:00-16:45 ACTIVITY 4-C

16:45-17:45 ACTIVITY 4-D

17:30-18:00 Q&A

LEARNING OBJECTIVES

By the end of this lecture, you will be able to:

- Define **derived (simple and complex)** and **compositional variables** as the two forms of **deterministic variables**
- Depict **deterministic variables** within DAGs
- Discuss how **tautological associations** can arise from naïve analyses of deterministic variables
- Explain how tautological associations can create misleading results in analyses of **change-scores** and **ratio variables**
- Suggest alternative – more appropriate – analytical approaches to avoid tautological associations

DETERMINISTIC VARIABLES

DAGs are usually used to depict **probabilistic** relationships and **probabilistic** variables

- i.e. variables whose values cannot be fully known

In health and social science research, some relationships and variables are **deterministic**

- i.e. their values can be completely known from other variables

This typically occurs when we create variables ourselves, such as:

- **simple derived variables**
- **complex derived (composite) variables**

It can also occur in other situations, such as:

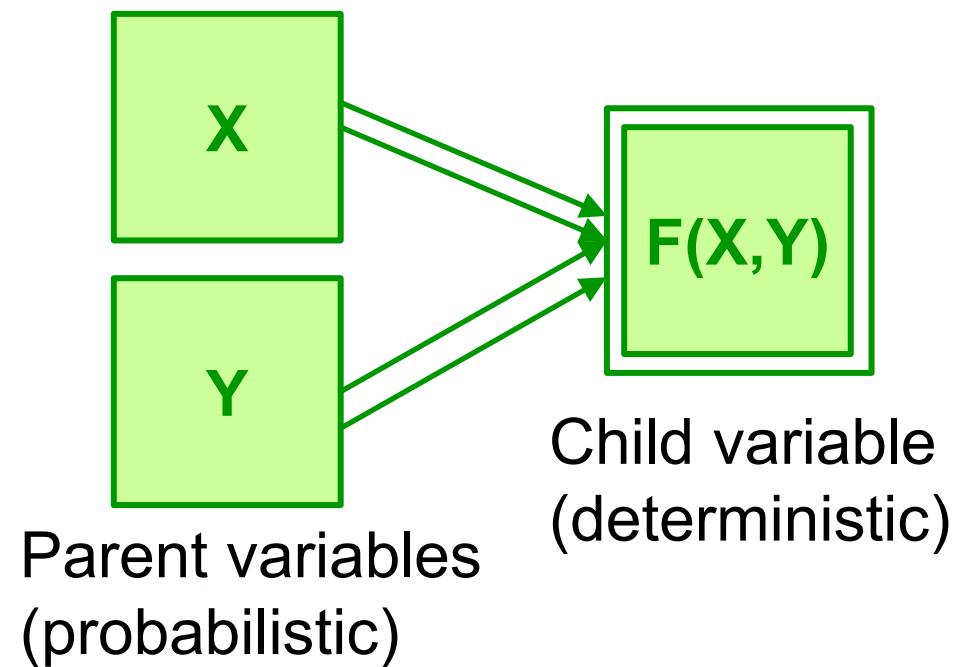
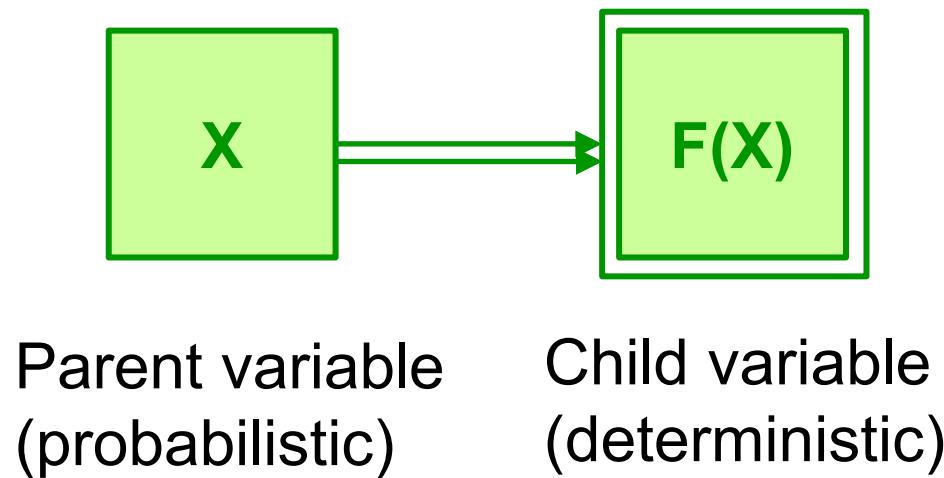
- **compositional data**

DETERMINISTIC VARIABLES IN DAGS

Deterministic variables need to be considered **very** carefully

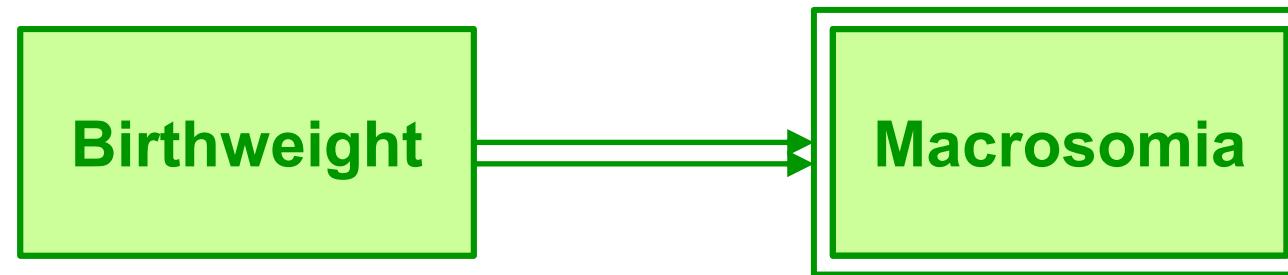
We therefore depict deterministic variables differently to probabilistic variables; by adding a **double outline and double arrows**

We describe a deterministic variable as a **child** of its probabilistic **parent(s)**



SIMLPE DERIVED VARIABLES

Simple derived variables are deterministic variables that are functionally created from – and **fully explained by** – a **single parent variable**



Macrosomia = 1 if **birthweight** \geq 4kg

Macrosomia = 0 if **birthweight** < 4kg

SIMPLE DERIVED VARIABLES

Simple derived variables include any ‘disease’ classification or other characteristic that is defined by values of single parent variable:

- Hypertension = $f(\text{Blood pressure})$
- Diabetes = $f(\text{Blood glucose concentration})$
- Binge drinking = $f(\text{Units of alcohol consumed in a single session})$
- Absolute poverty = $f(\text{Total daily income})$

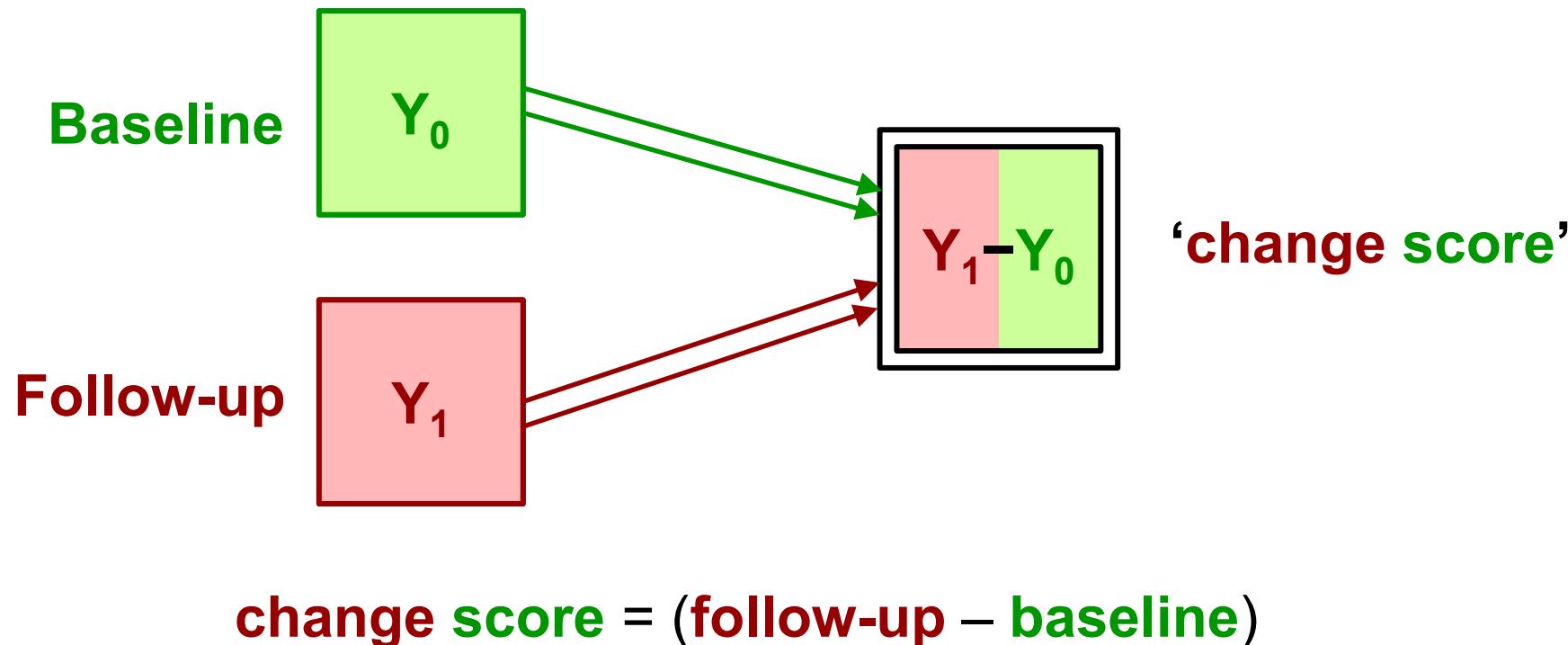
They also include any functional transformation:

- Log income = $f(\text{income})$

COMPLEX DERIVED (COMPOSITE) VARIABLES

Complex derived (composite) variables are deterministic variables that are functionally created from, and **fully explained by**, two or more parent variables

- e.g. **change score** (or difference-from-baseline variable)



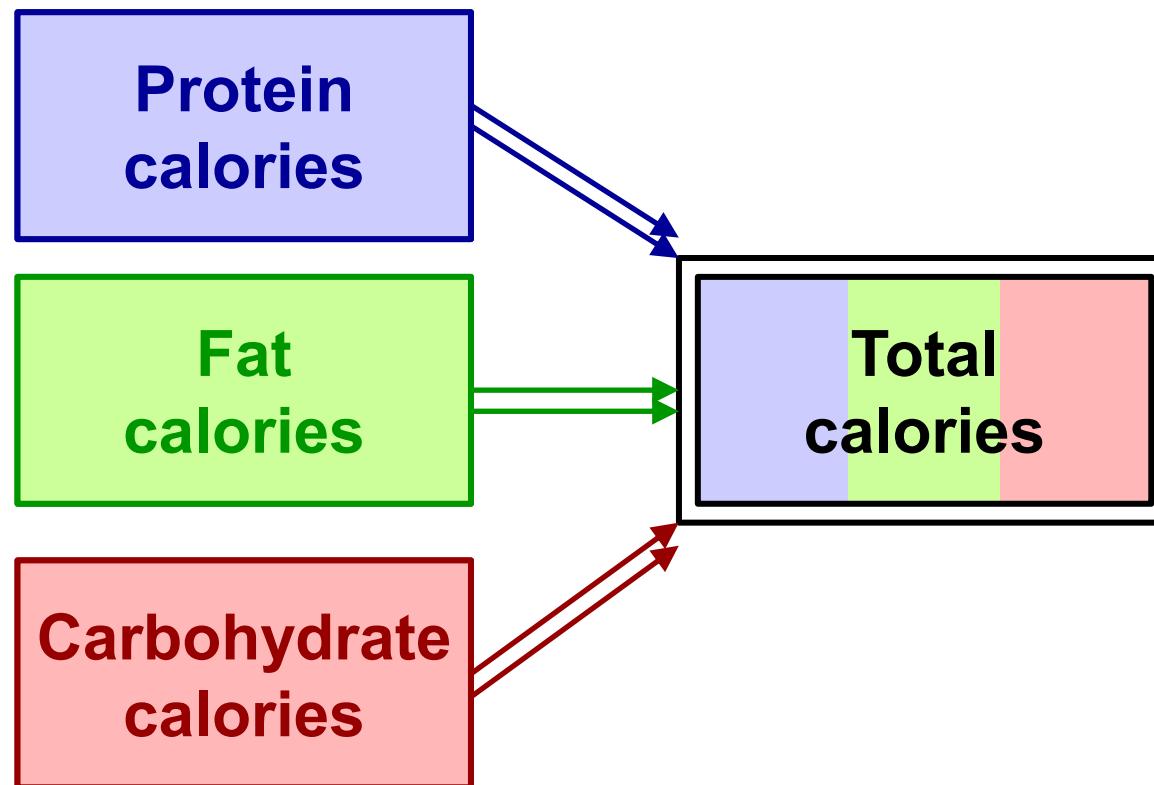
COMPLEX DERIVED (COMPOSITE) VARIABLES

Other examples of complex derived (**composite**) variables:

- **Ratio variables** (e.g. waist-to-hip ratio)
 - ✓ WHR = Waist circumference/hip circumference
- **Functional variables** (e.g. body mass index)
 - ✓ BMI = Weight/Height²
- **risk / symptom / disease scores** (e.g. Metabolic Syndrome)
 - ✓ MetS = f(Waist Circumference, BP, Blood glucose, Triglycerides, HDL)

COMPOSITIONAL DATA

Deterministic variables also occur in **compositional data**, i.e. where two or more '**part**' variables make up a '**whole**' variable



Compositional 'totals' **may** also be derived if we measure / observe the constituent components first

BUT if we measure the total first, the components **cannot** be derived

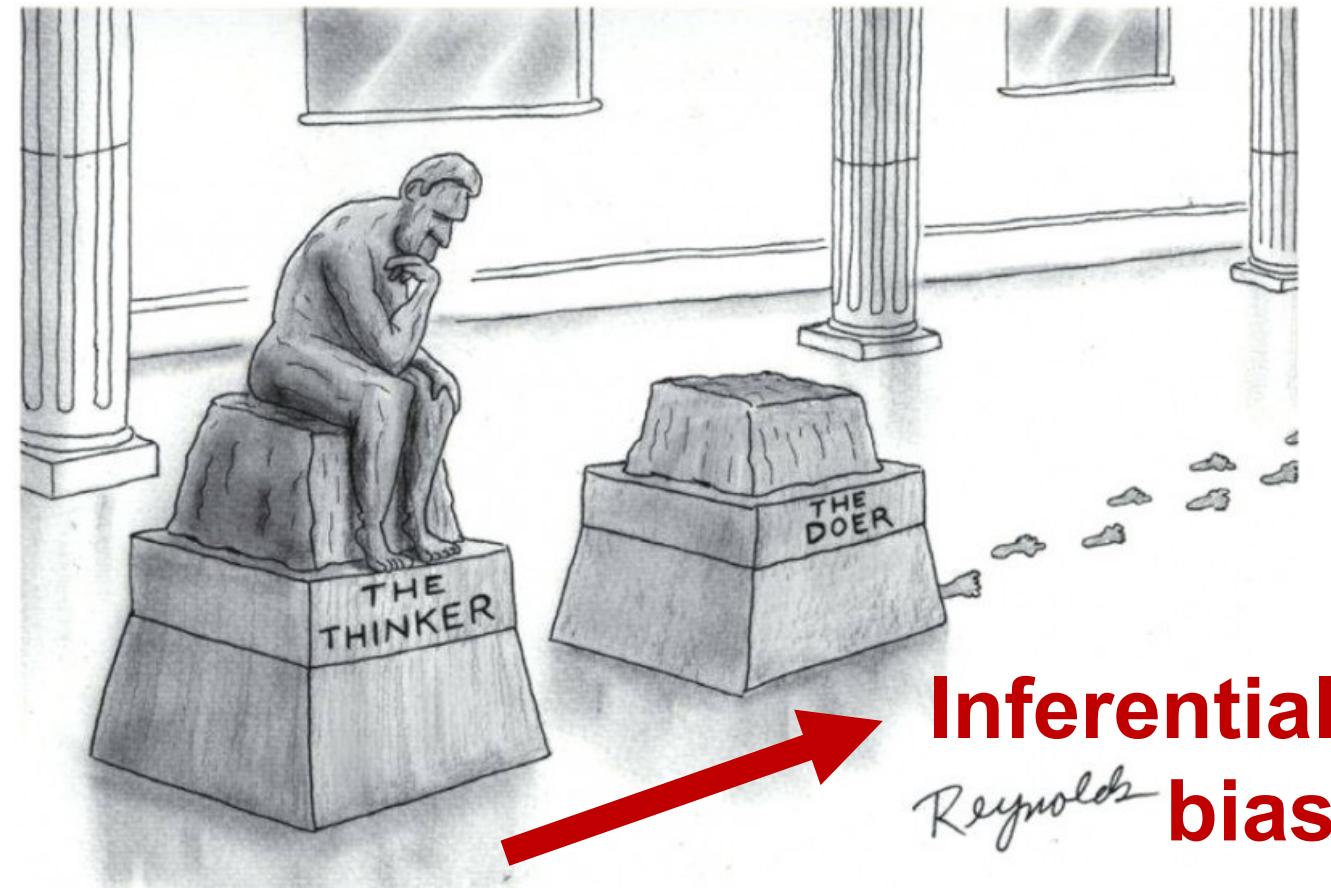
$$\text{Total calories} = \text{Protein calories} + \text{Fat calories} + \text{Carb calories}$$

INFERENTIAL BIAS

Deterministic variables must be treated carefully so that the underlying **deterministic relationships** are recognised and respected

Deterministic relationships can be very strong and can overwhelm other causal relationships

If they are not recognised, they can thus lead to serious **inferential bias**



TAUTOLOGICAL ASSOCIATIONS

The simplest form of **inferential bias** occurs ...

WHEN:

- a **derived variable** is analysed in direct relation to (one of) its parent(s)

OR

- two (sibling) **derived variables** that share one or more common algebraic parent(s) are analysed in relation to one another

AND:

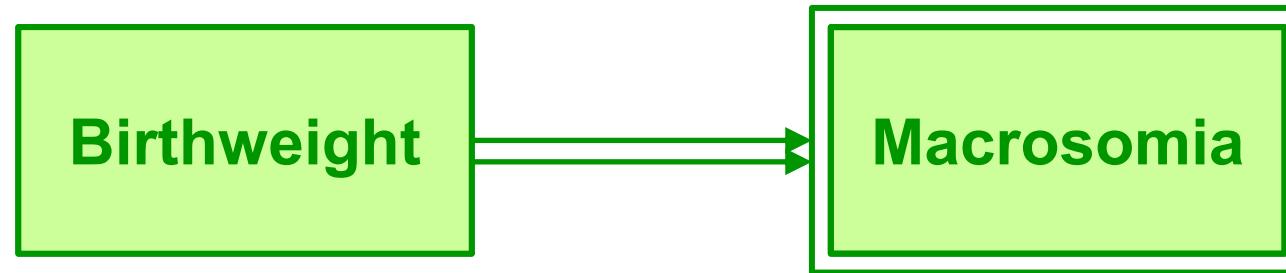
- we mistake the resulting **tautological association** as causally meaningful

Tautological associations are the **self-fulfilling associations** that occur between variables that are related, by their very definition

EXAMPLE: TAUTOLOGICAL ASSOCIATIONS

A **simple derived variable** will (obviously?!) share a **tautological association** with the **parent variable** from which it was defined ...

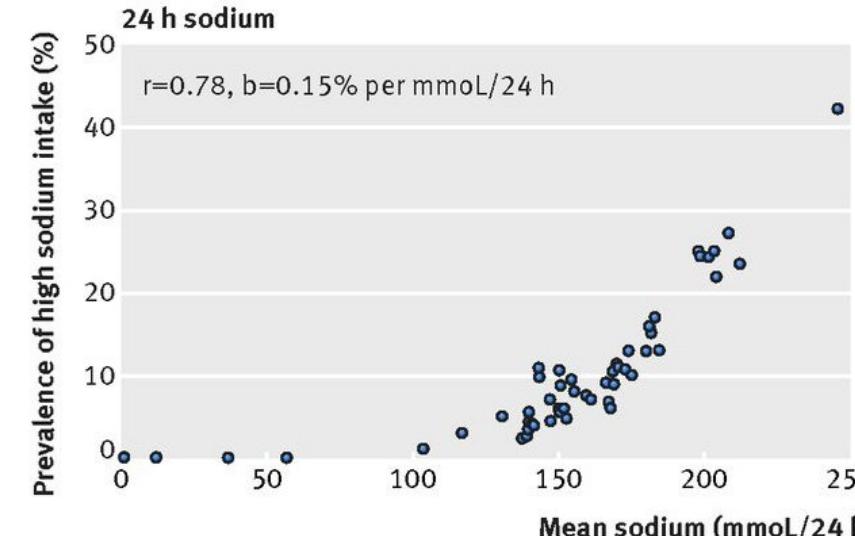
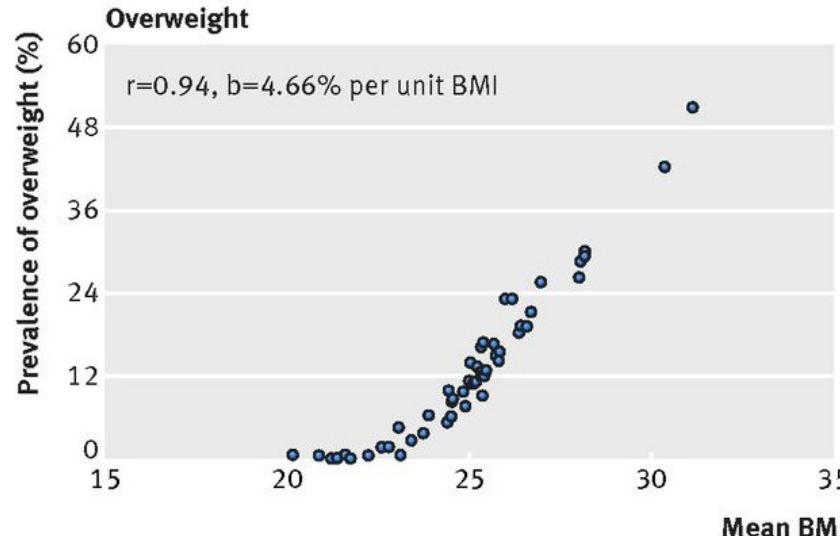
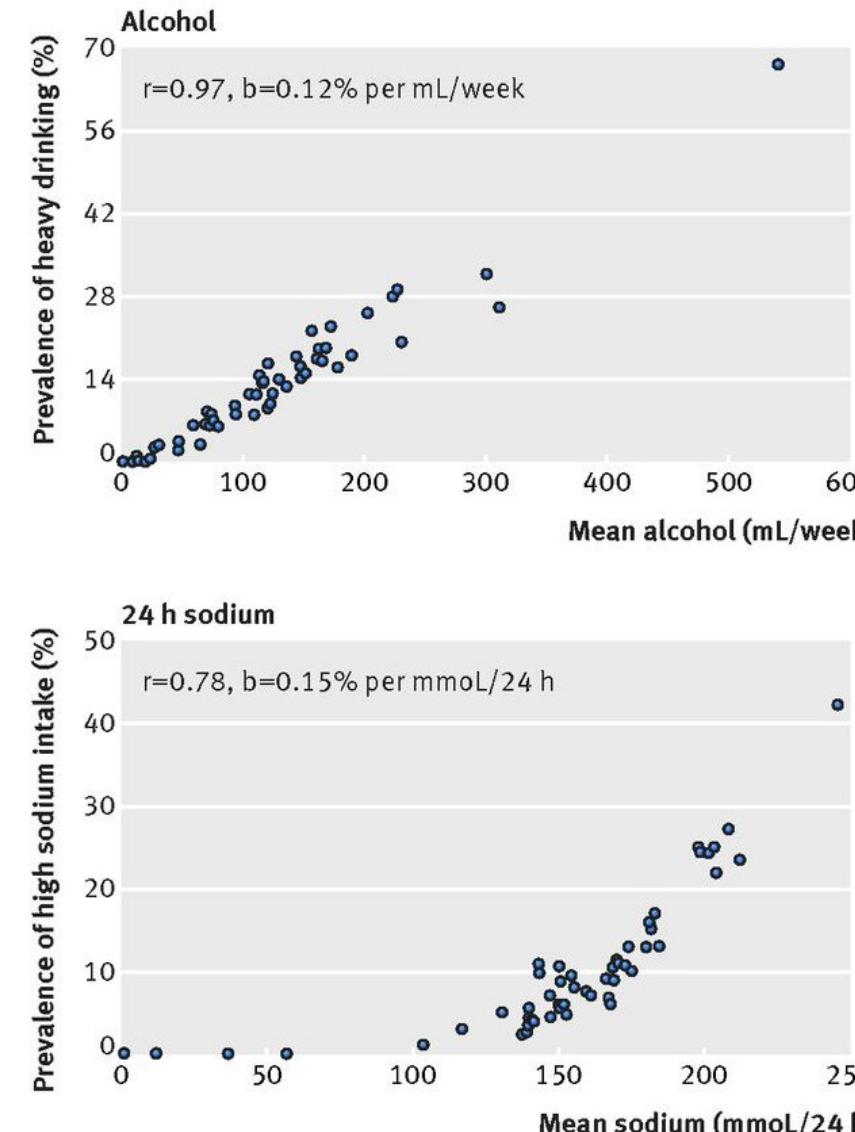
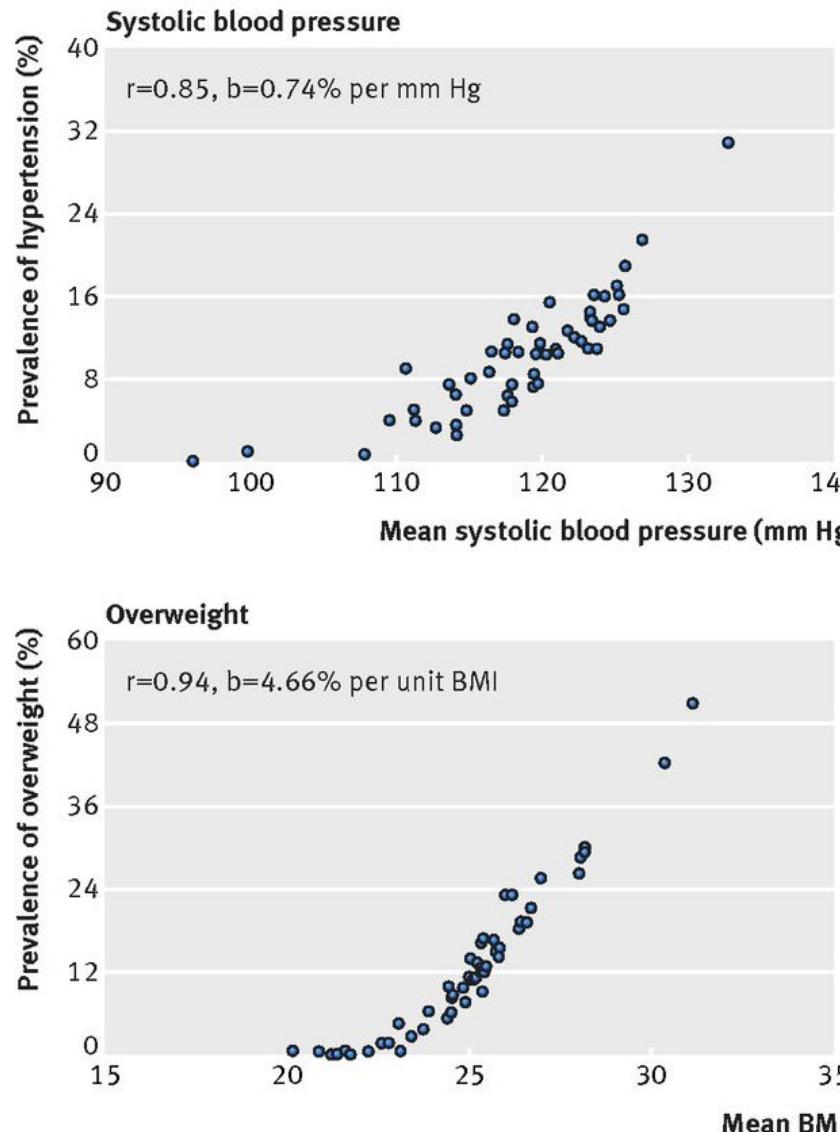
- e.g. **Birthweight** and **Macrosomia**



It's true that macrosomia is caused by birthweight, but the causation is tautological; macrosomia is birthweight because we define it as such !

It's hard to imagine any tautological association being useful / interesting

EXAMPLE: TAUPOLOGICAL ASSOCIATIONS



Razak F et al. BMJ.
2018, 3;362:k3147.

TAUTOLOGICAL ASSOCIATIONS

Inferential bias involving **simple derived variables** is very rare, presumably because the tautology is so obvious and the associations so large

Unfortunately, the same cannot be said about **complex derived variables**, where examples of inferential bias are common and widespread, particularly involving:

- **change score variables**
- **ratio variables**

This is despite a long and esteemed literature on the issue ...

THE ORIGINAL SPURIOUS CORRELATION

Karl Pearson wrote about tautological associations in 1897 !

He, however, coined a different term:



Karl Pearson, father of statistics

*'I term this a spurious organic correlation, or simply a **spurious correlation**'*

"Mathematical Contributions to the Theory of Evolution.—On a Form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs." By KARL PEARSON, F.R.S., University College, London. Received December 29, 1896,—Read February 18, 1897.

(1) If the ratio of two absolute measurements on the same or different organs be taken it is convenient to term this ratio an *index*.

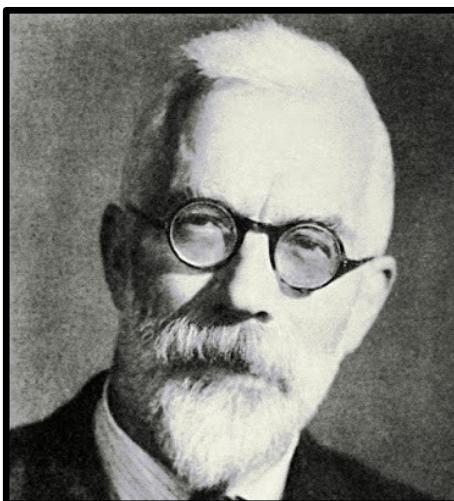
If $u = f_1(x, y)$ and $v = f_2(z, y)$ be two functions of the three variables x, y, z , and these variables be selected at random so that there exists no correlation between x, y , y, z , or z, x , there will still be found to

THE ORIGINAL SPURIOUS CORRELATION

Ronald Fisher (1947) and Jerzy Neyman (1952) later made the same observations (and repeated similar warnings) !

THE ANALYSIS OF COVARIANCE METHOD FOR THE
RELATION BETWEEN A PART AND THE WHOLE

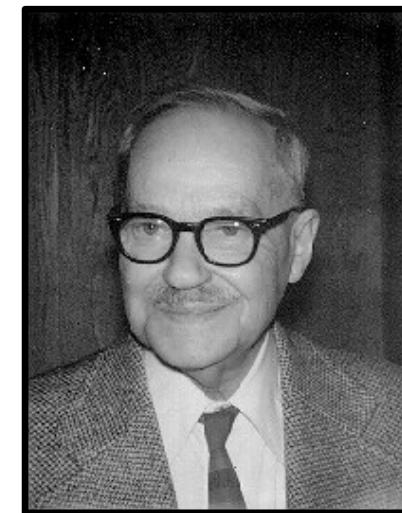
R. A. FISHER
University of Cambridge



Ronald Fisher – Pearson's pupil
– another father of statistics

Part 3. On a Most Powerful Method of Discovering
Statistical Regularities

(This section is based on a talk given before the members of Sigma Xi at a meeting of the Society held in Berkeley, California, April 9, 1947.)



Jerzy Neyman – another
father of statistics

CHANGE SCORES AND TAUTOLOGICAL ASSOCIATIONS

In 1962, Peter Oldham wrote in Journal of Chronic Disease:

"When two measurements have been made, it is natural to think of analysing the first, and the difference between second and first..."

(But) these... not independent, however, and to treat them as independent is to introduce spurious correlations"

A NOTE ON THE ANALYSIS OF REPEATED MEASUREMENTS OF THE SAME SUBJECTS

P. D. OLDHAM, M.A.(Oxon.), F.S.S.

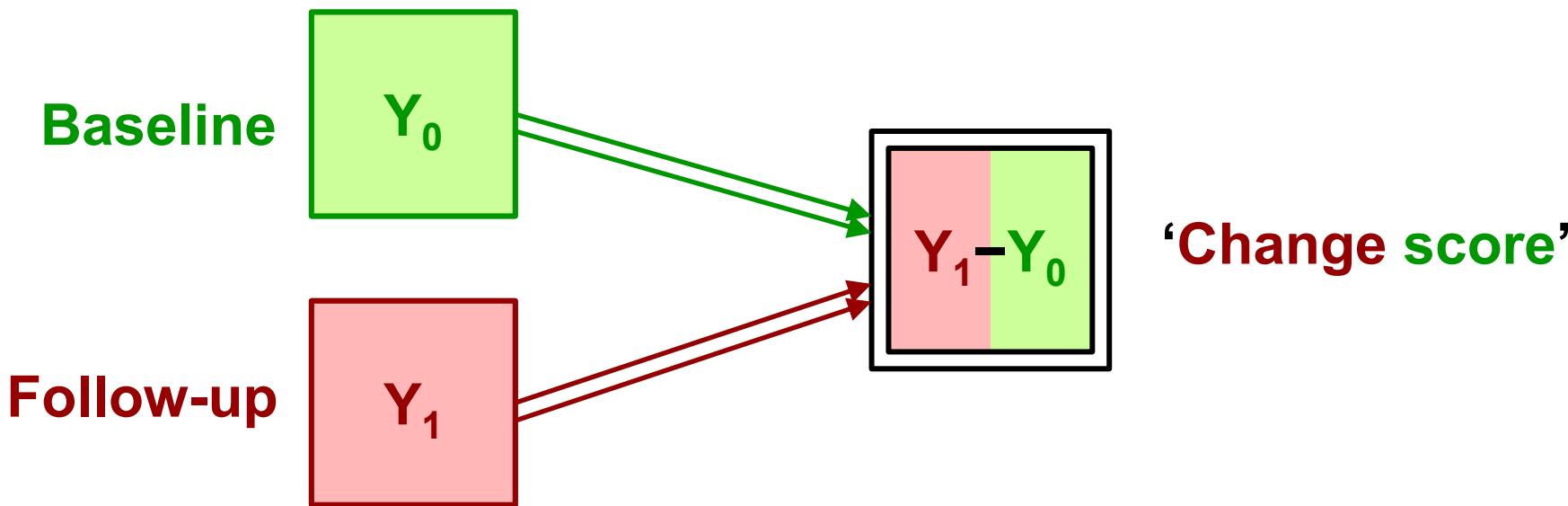
Pneumoconiosis Research Unit, Llandough Hospital, Penarth, Glamorgan



Peter Oldham – doctor and statistician, largely forgotten

CHANGE SCORES AND TAUTOLOGICAL ASSOCIATIONS

'Change-scores' are **complex derived variables** constructed from repeated measures of a single parent variable by subtracting a subsequent measure (Y_1 : '**follow-up**') from a preceding measure (Y_0 : '**baseline**'')

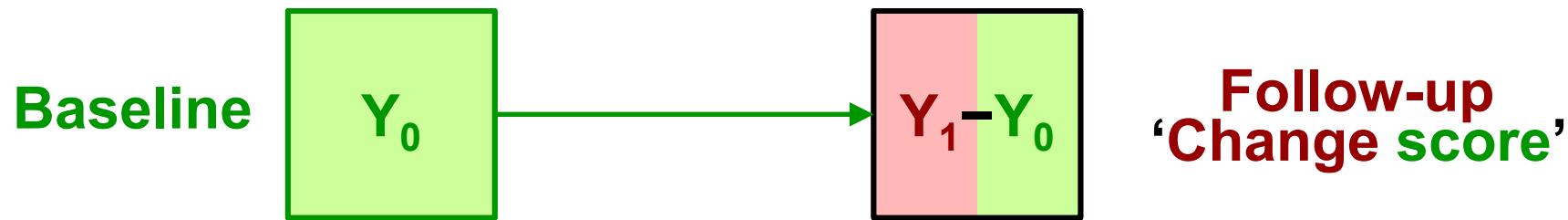


The intention is perhaps to '**standardise**' against starting levels of Y , and to 'isolate' the part of Y that changed between the two timepoints ...

CHANGE SCORES AND TAUTOLOGICAL ASSOCIATIONS

The composite **change score** variable will share a **tautological association** with either parent variable if analysed by correlation or regression

- e.g. $(Y_1 - Y_0) = \beta_0 + \beta_1 Y_0 + \dots + \varepsilon$



Composite variable bias occurs when we misinterpret this tautological association as meaningful

CHANGE SCORES AND TAUTOLOGICAL ASSOCIATIONS

e.g. **Law of initial value:**

“the direction of response of a body function to any agent depends to a large degree on the initial level of that function”

STIMULUS AND RESPONSE

THE LAW OF INITIAL VALUE

BY

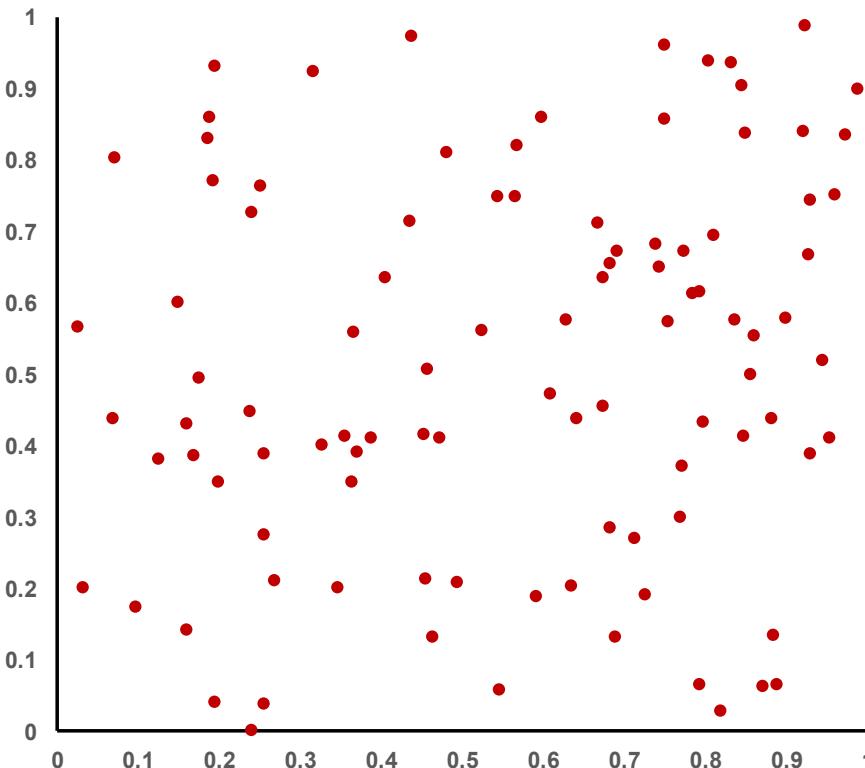
JOSEPH WILDER, M.D.

Clinical Professor of Neurology, New York Medical College

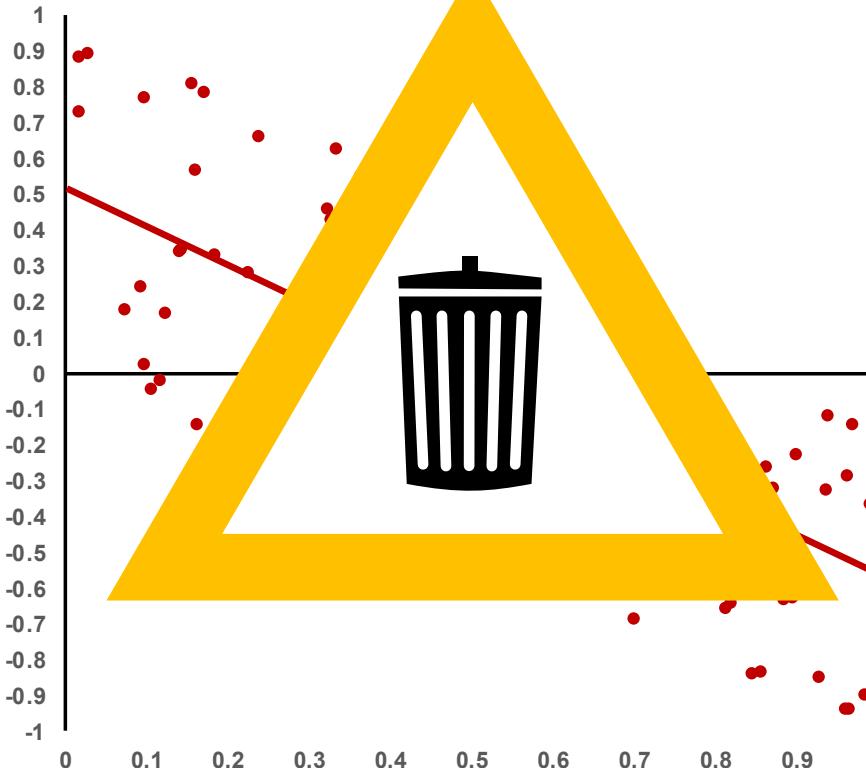
In fact, this is a consequence of the tautological association between Y_0 (**baseline**) and the negative $-Y_0$ in the change-score

CHANGE SCORES AND TAUTOLOGICAL ASSOCIATIONS

$Y \sim X$



$(Y-X) \sim X$



CORRELATION (r) $\approx \pm 0.71$

EXAMPLE: CHANGE SCORES



CHEST

Original Research

LUNG FUNCTION

Childhood Respiratory Illness and Lung Function at Ages 14 and 50 Years

Childhood Respiratory Illness and Lung Function

Peter W. G. Tennant, MSc; G. John Gibson, MD; Louise Parker, PhD
and Mark S. Pearce, PhD

Background: Although childhood respiratory tract infections and low birth weight were associated with reduced adult lung function, little is known about the timing during life. We used data from the Newcastle Thousand Families Study to examine other factors influenced FEV₁ at age 14 years and between 14 and 49 to 51 years.

Methods: Detailed information was collected prospectively during childhood. Members of the cohort were recruited into a case-control study of respiratory illness. Data included measurement of FEV₁. One hundred twenty-two of these were measured at age 14 years and 111 at age 49 to 51 years. Linear regression models were used to examine cross-sectional influences on FEV₁.

Results: Lower height ($P < .001$), lower BMI ($P < .001$), being breast fed for longer than 6 months ($P = .028$), childhood history of severe respiratory illness ($P = .014$), childhood history of LRTI ($P = .004$), childhood history of TB ($P = .023$), and birth into a lower social class ($P = .047$) were significant independent predictors of lower FEV₁ at 14 years of age. Correspondingly, higher height ($P < .001$), lower birth weight ($P = .025$), having a higher FEV₁ at age 14 years ($P < .001$), a lower lifetime number of cigarettes smoked ($P = .007$), and having a greater lifetime history of severe respiratory illness ($P = .047$) were all independently associated with a greater increase (or a smaller increase) in FEV₁ between age 14 and 49 to 51 years.

Conclusions: This study suggests that the change in FEV₁ between youth and middle age depends on several factors acting throughout life, including FEV₁ in adolescence, sex, cigarette smoking, birth weight, and childhood respiratory health.

CHEST 2010; 137(1):146–155

Abbreviations: LRTI = lower respiratory tract infection; URTI = upper respiratory tract infection

Two results are striking for their magnitude and/or direction: the relative decline in women compared with men, and the negative association between FEV₁ at 14 years and change thereafter. Although both may represent regression to the mean, other explanations are plausible. On average, women reach their maximum FEV₁ at a younger age than men.³⁹ Figure 1 (constructed using established reference equations)⁴⁰

Tennant et al 2010
Chest; 137(1):146-155

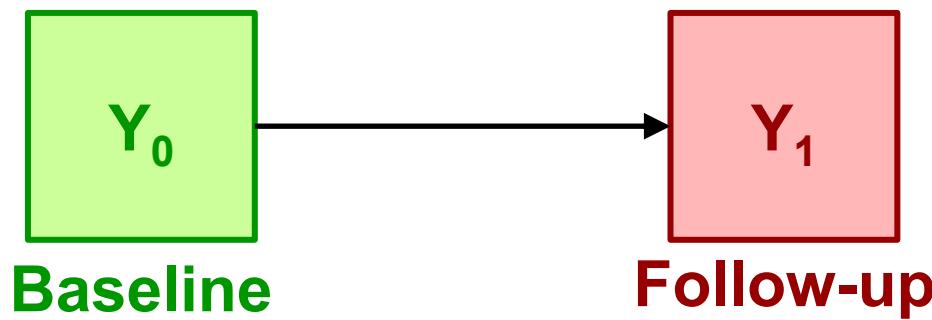
THE CORRECT APPROACH

If we can't analyse change on baseline, what should we do?



What's your **estimand**!?

If you're interested in the extent that a variable (Y) is auto-correlated over time ($Y_0 \rightarrow Y_1$), this is (relatively) simple:



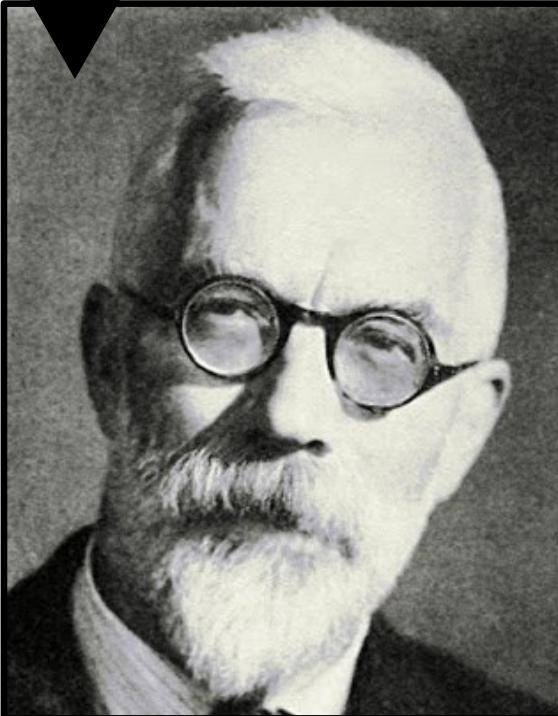
Although, with just two measures, the estimate will be diluted by error and uncertainty in Y_0 and Y_1

RATIO VARIABLES

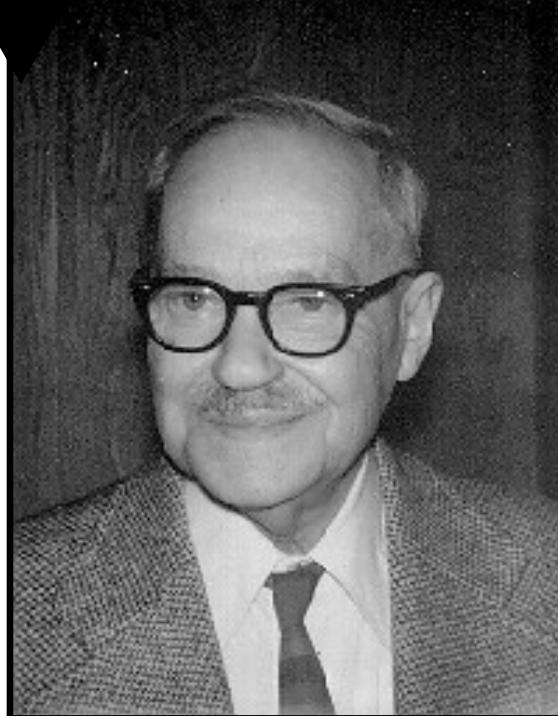
Who cares about change scores?! Tell them about ratio variables!



Pearson – old man 1



Fisher – old man 2



Neyman – old man 3

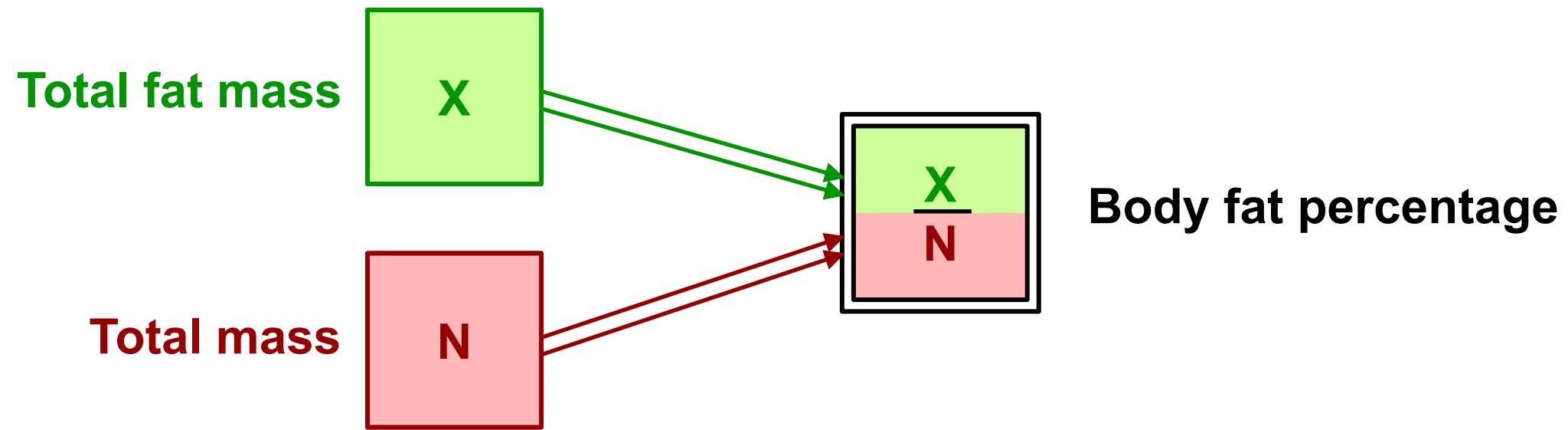


Gilthorpe – old man 4

RATIO VARIABLES

Ratio variables (AKA **ratio-index variables** or **proportion variables**) are composite variables made by dividing one variable by another

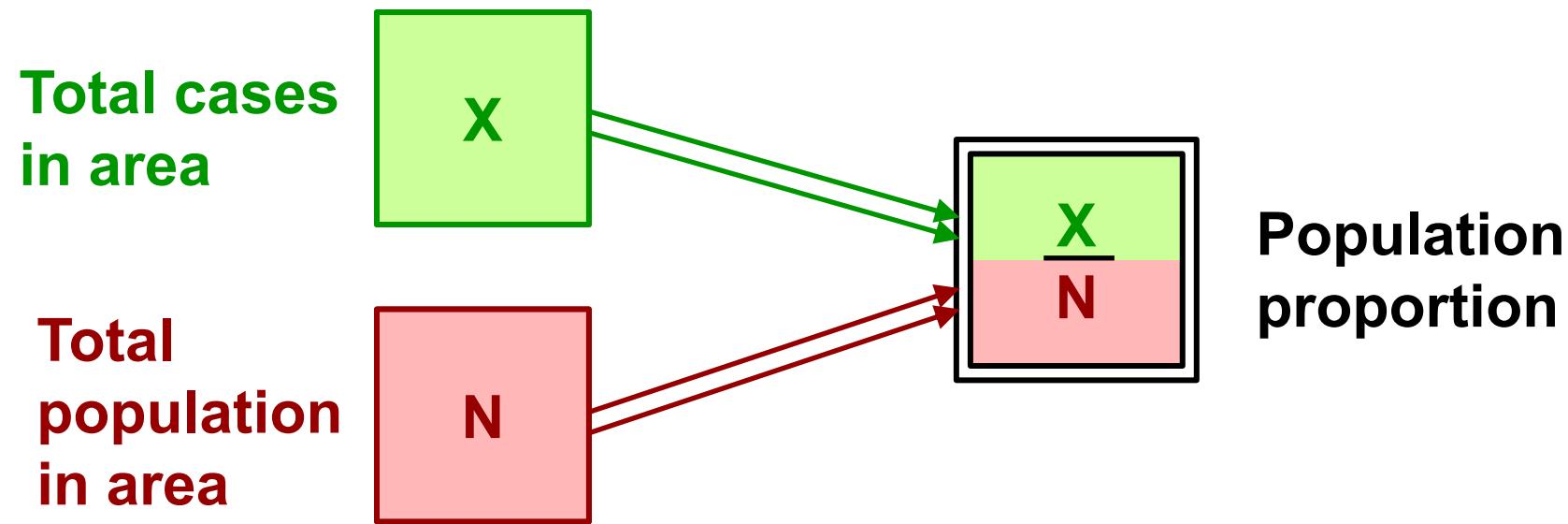
- e.g. body fat percentage



RATIO VARIABLES

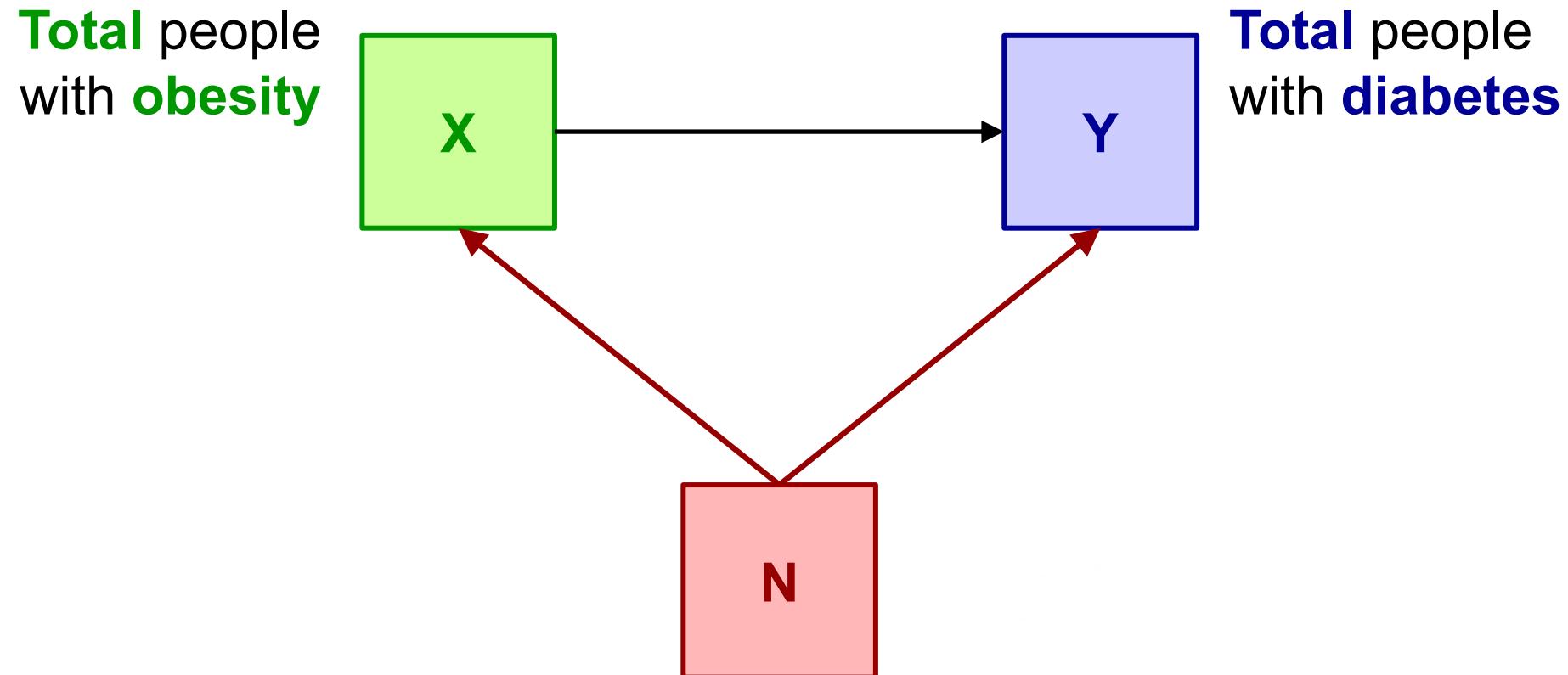
Ratio variables are extremely common in geographical research, where proportions are often calculated and compared between geographical areas

- e.g. prevalence or incidence proportions



THE PROBLEM WITH N

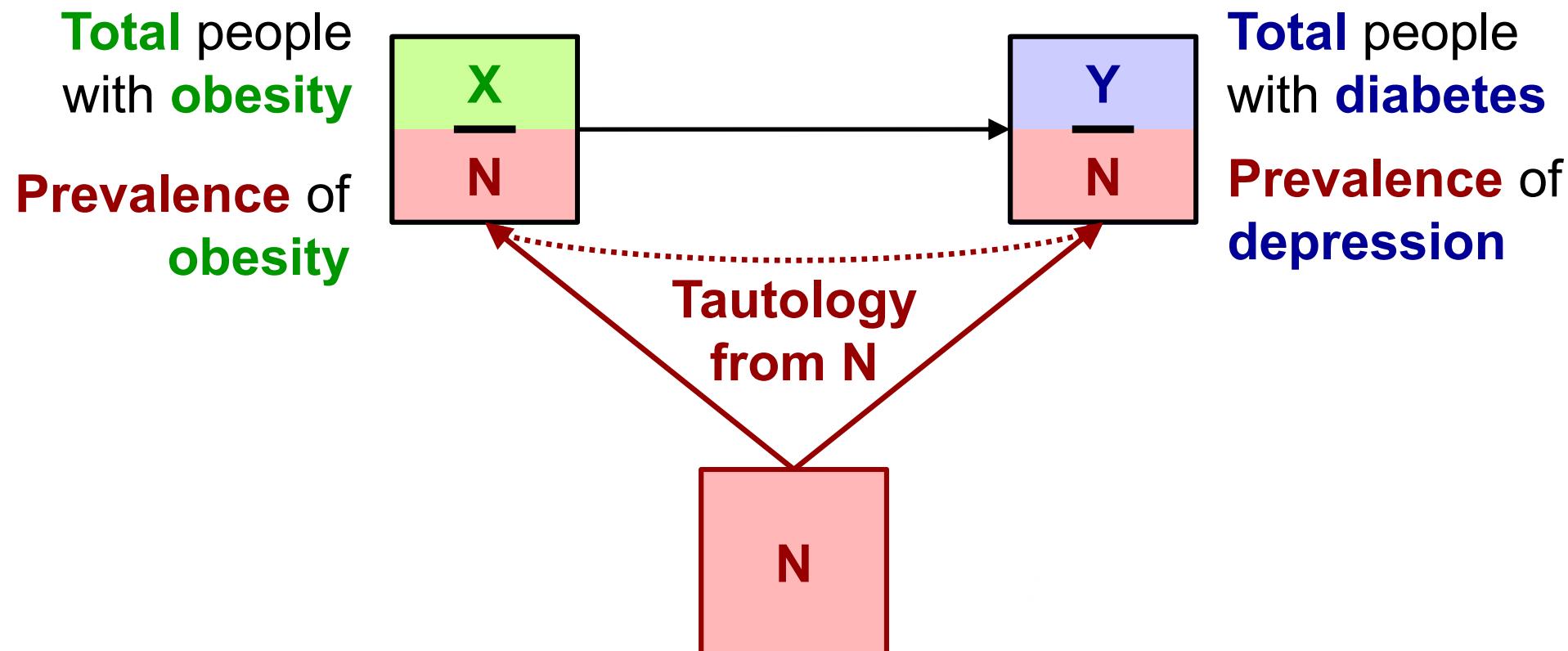
When the **number of events** is caused by the **population size (N)**, the relationship between different events (e.g. **X** and **Y**) will be **confounded**:



MATHEMATICAL COUPLING BIAS AND RATIO VARIABLES

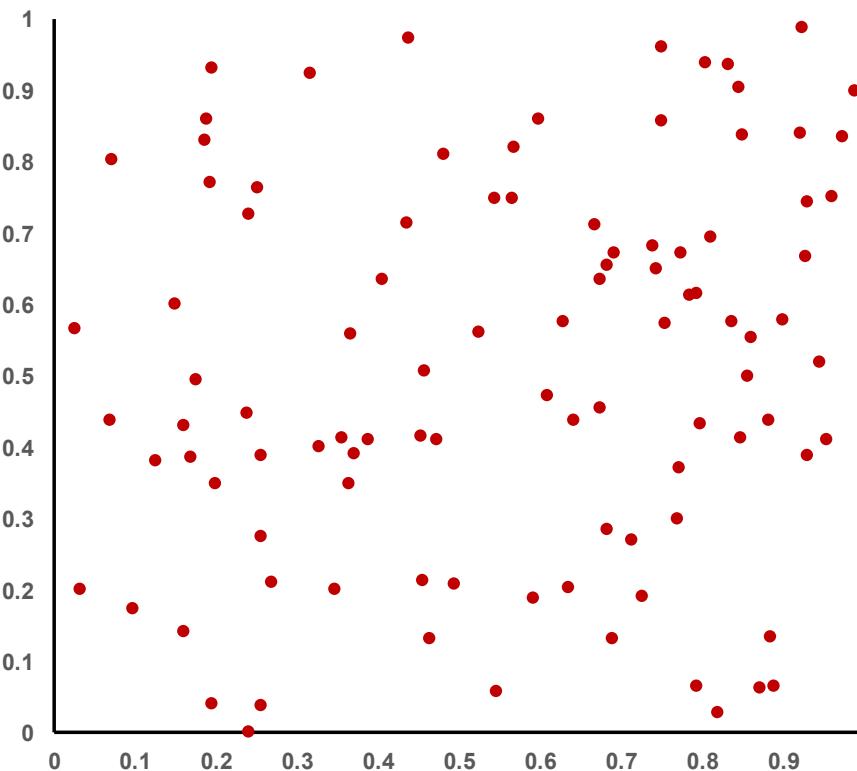
To attempt to resolve this, analysts create **ratio variables** with **common denominators**, inadvertently introducing a **tautological association**:

- i.e. $(Y/N) = \beta_0 + \beta_1(X/N) + \dots + \varepsilon$



MATHEMATICAL COUPLING BIAS AND RATIO VARIABLES

$Y \sim X$



$Y/N \sim X/N$



CORRELATION (r) ≈ 0.5

EXAMPLE: RATIO VARIABLES

To demonstrate the problem, **Neyman** described how a ‘friend’ had collected data from 54 countries on

- Number women of childbearing age
- Number of babies
- Number of storks

The question his friend allegedly sought to answer was:

‘Do storks bring babies?’



EXAMPLE: RATIO VARIABLES

Neyman noted a large correlation between the **crude number of storks** and the **crude number of babies** that was attributed to underlying country size

He then analysed the relationship between stork density (per 1000 women) and birth rate (per 1000 women)

My friend's conclusion was that, although there is no evidence of storks actually bringing babies, there is overwhelming evidence that, by some mysterious process, they influence the birth rate! I know that some of

There is nothing spurious in the correlation between X and Y . When $A_0 \neq 0$ and $C_0 \neq 0$, the correlation between these two variables is quite real. Therefore the term "spurious correlation" seems to miss the point. The real point of the discussion is that the computation of the quotients X and Y is undertaken in order to study the correlation, not between these variables themselves, but between the social, economic or biological factors that these quotients are supposed to represent. It is the method of study that is faulty

EXAMPLE: RATIO VARIABLES



OPEN ACCESS

'First, do no harm': are disability assessments associated with adverse trends in mental health? A longitudinal ecological study

B Barr,¹ D Taylor-Robinson,¹ D Stuckler,² R Loopstra,² A Reeves,² M Whitehead¹

ABSTRACT

Background In England between 2010 and 2013, just over one million recipients of the main out-of-work disability benefit had their eligibility reassessed using a new functional checklist—the Work Capability Assessment. Doctors and disability rights organisations have raised concerns that this has had an adverse effect on the mental health of claimants, but there are no population level studies exploring the health effects of this or similar policies.

Method We used multivariable regression to investigate whether variation in the trend in reassessments in each of 149 local authorities in England was associated with differences in local trends in suicides, self-reported mental health problems and antidepressant prescribing rates, while adjusting for baseline conditions and trends in other factors known to influence mental ill-health.

Results Each additional 10 000 people reassessed in each area was associated with an additional 6 suicides (95% CI 2 to 9), 2700 cases of reported mental health problems (95% CI 548 to 4840), and the prescribing of an additional 7020 antidepressant items (95% CI 3930 to 10100). The reassessment process was associated with the greatest increases in these adverse mental health outcomes in the most deprived areas of the country, widening health inequalities.

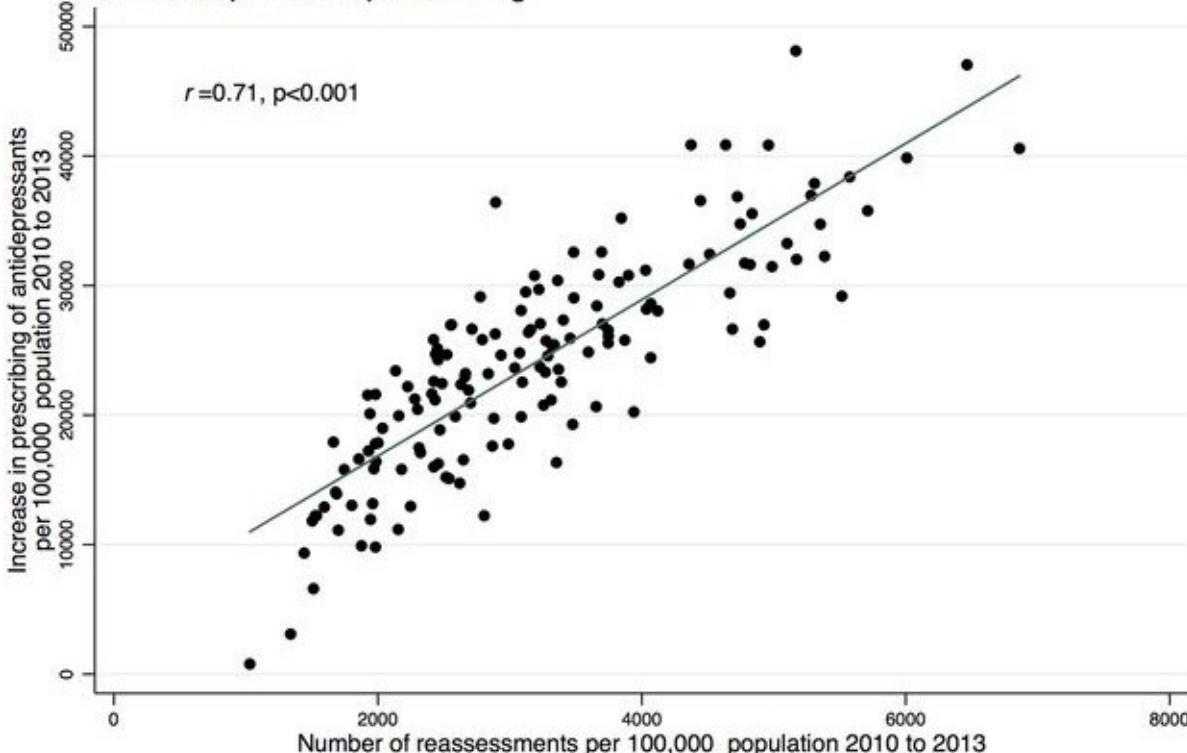
Conclusions The programme of reassessing people on disability benefits using the Work Capability Assessment was independently associated with an increase in suicides, self-reported mental health problems and antidepressant prescribing. This policy may have had serious adverse consequences for mental health in England, which could outweigh any benefits that arise

Disability

The provision of cash benefits to people who are unable to work because of disability is an essential component of health and welfare systems that aim to promote the social inclusion of people with disabilities.⁸ Over recent years many countries, including the UK, the Netherlands and Australia, have introduced more stringent functional assessment checklists to reduce the growing number of people receiving disability benefits.^{9, 10} While in most countries these more stringent criteria have only been applied to new benefit claimants, the UK and the Netherlands have gone further—reassessing their entire caseloads.⁸ In the UK this process started in 2010 when the government initiated a programme to reassess all existing claimants of out-of-work disability benefits using the WCA. Following reassessment the claimants were either moved off disability benefits, if found to be fit for work, or otherwise were transferred to a new disability benefit scheme called Employment Support Allowance.

The WCA has been the subject of a great deal of controversy. Nearly 40% of those who have appealed against the initial assessment decision have had this decision overturned,¹¹ and five independent reviews have raised concerns about the fairness and effectiveness of the process. In particular the reviews indicated that the process was impersonal and mechanistic and did not adequately capture the impact of many chronic health conditions.¹² The government has however accepted many of the recommendations of these reviews and changed the WCA over time. Many of these changes have particularly focused on the assess-

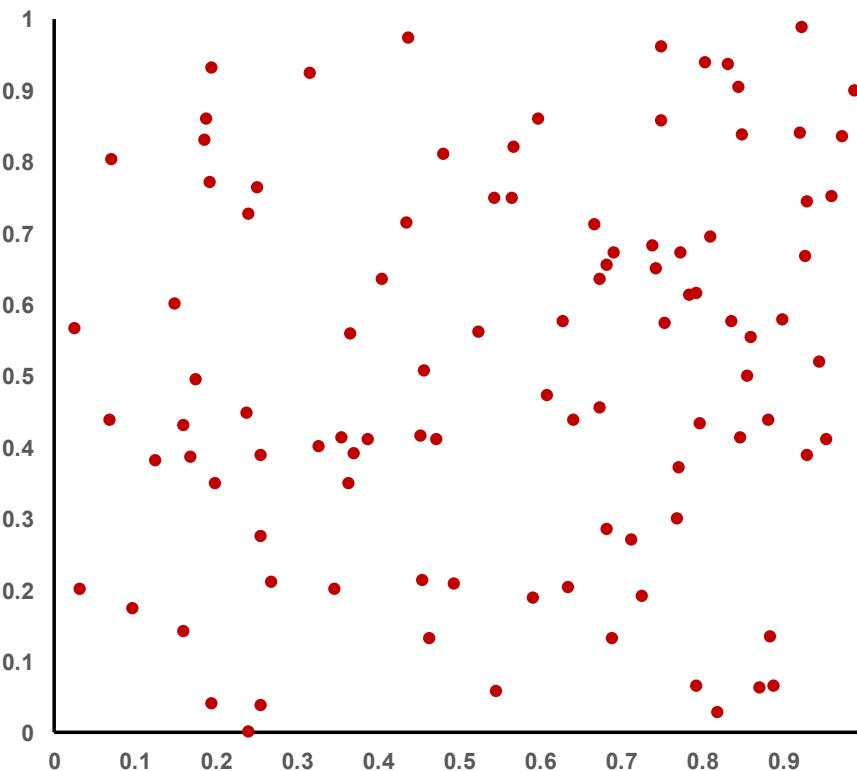
3. Antidepressant prescribing



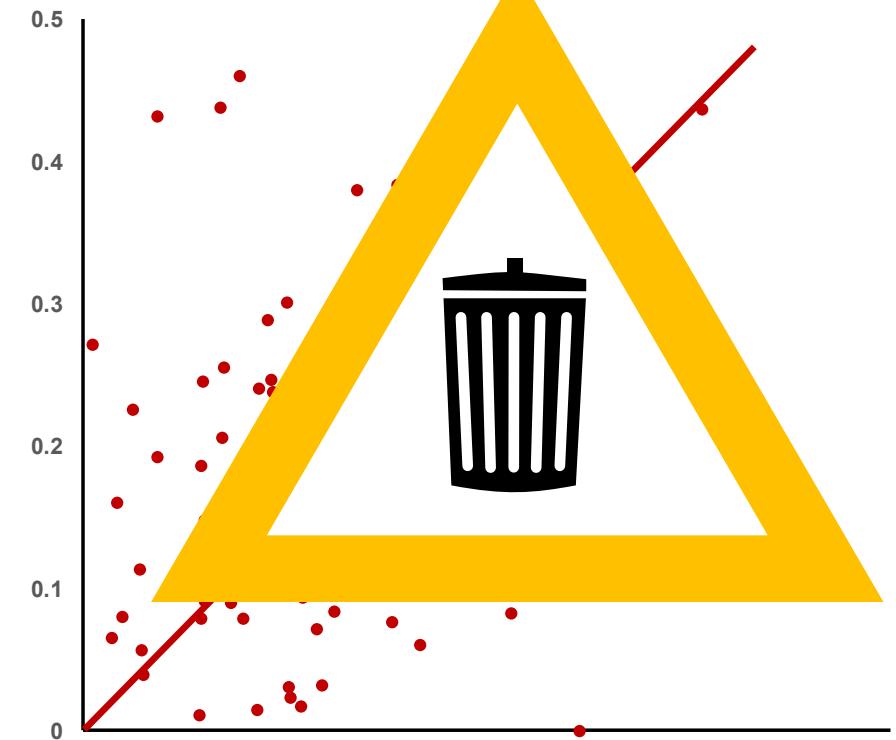
Barr et al 2016
JECH; 70:339-345

MATHEMATICAL COUPLING BIAS AND RATIO VARIABLES

$Y \sim X$



$Y/N \sim X/N$



CORRELATION (r) ≈ 0.5

HOW SPURIOUS?

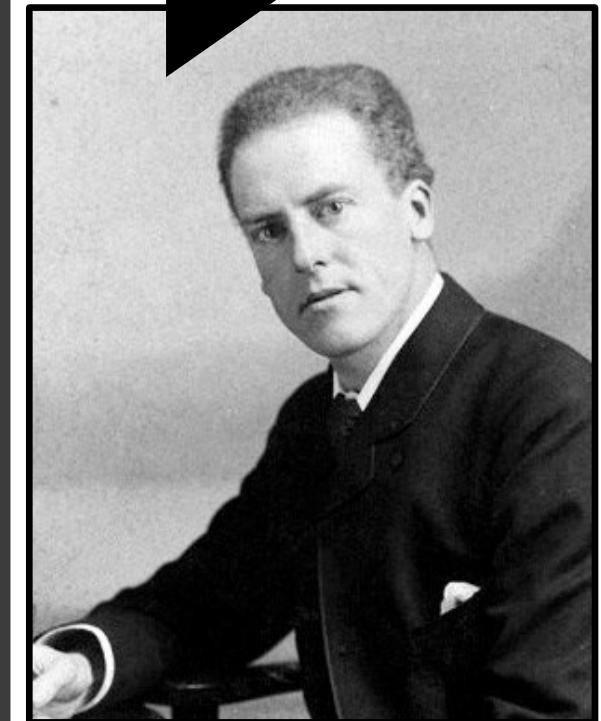
A spurious correlation of $\rho = 0.5$ is very unlikely, because **X**, **Y**, and **N** will not be completely random

The degree of the bias will largely depend on the variation due to N:

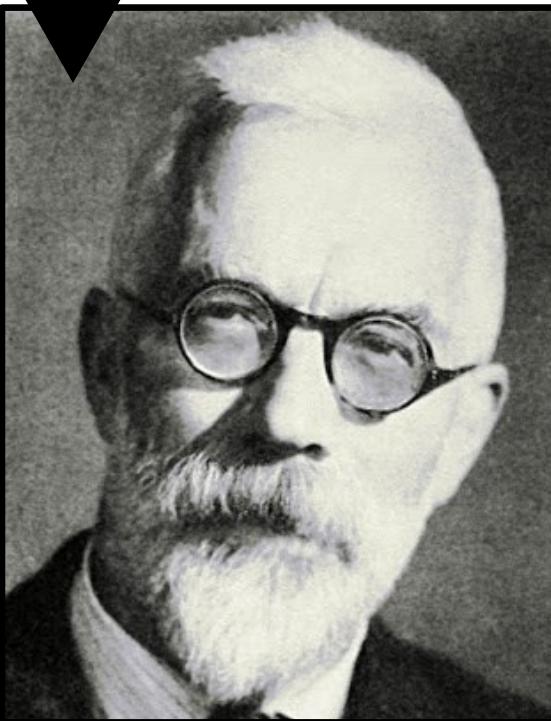
- the more variation in N, the more variation in X/N and Y/N is explained by N
- if N is very different between areas (e.g. countries), the bias will be larger
- if N is very similar between areas (e.g. lower super-output areas), the bias will be smaller

A SOLUTION

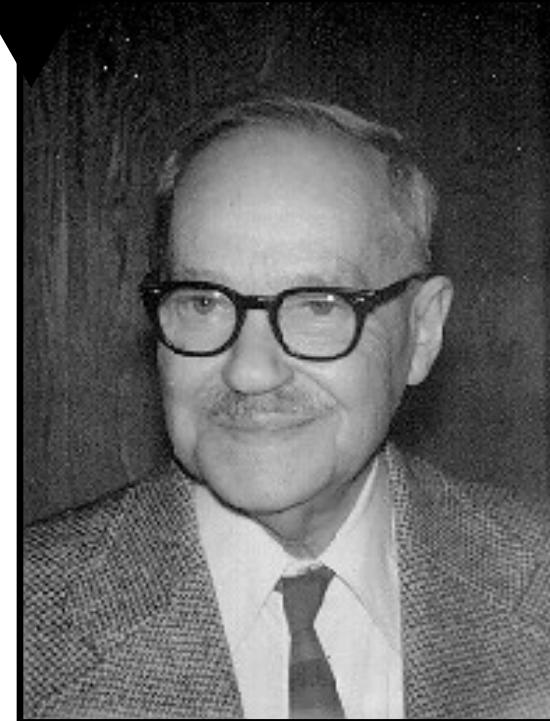
Tell them our solution!



Pearson – old man 1



Fisher – old man 2



Neyman – old man 3



Gilthorpe – old man 4

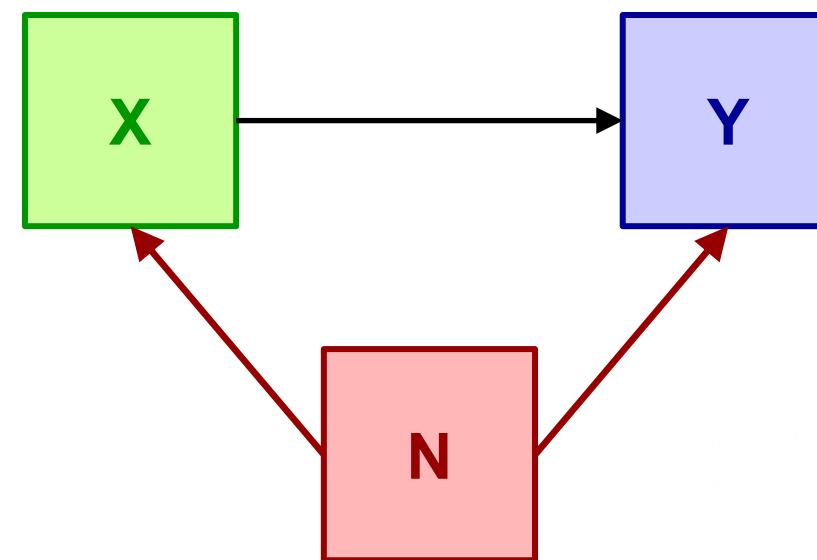
A SOLUTION

Pearson (1897), Fisher (1947), Neyman (1952) and Gilthorpe (2010) all offered the same solution:

Adjust for the **common denominator** as a covariate within linear regression!

$$Y = \beta_0 + \beta_1 X + \beta_2 N + \dots + \varepsilon$$

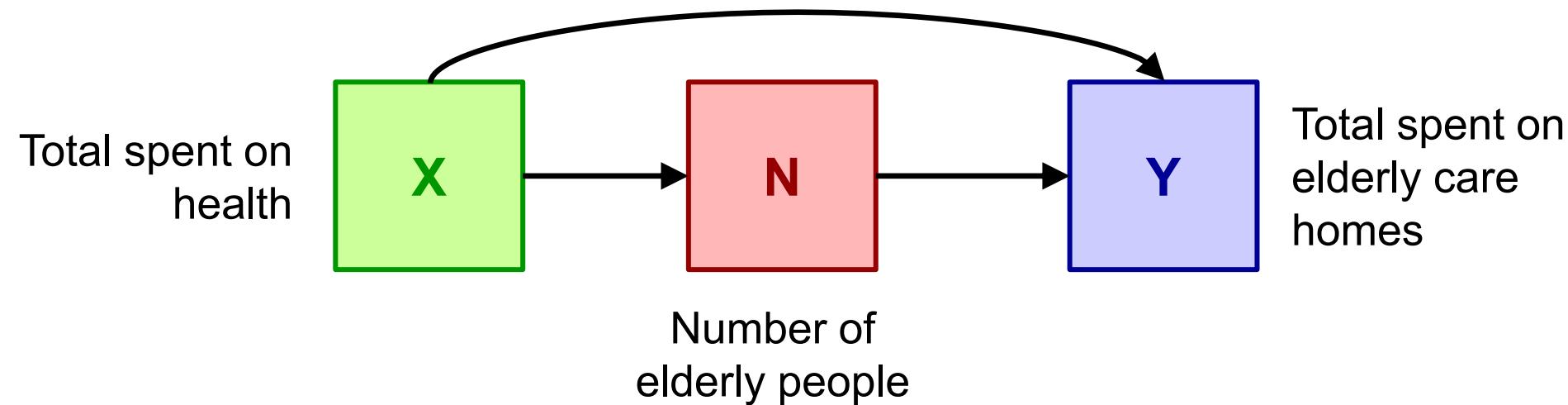
When the **denominator** is a **confounder**, this seems rather obvious from a causal diagram:



NOTE OF WARNING

Adjusting for the common denominator is **only** reasonable if it is a true confounder...

e.g.



Adjusting for this N would not provide the desired estimand

Must always think about our question and the causal relationships operating

Think: **estimand and covariate roles !**

SUMMARY

- **Deterministic variables** like **derived** and **compositional variables** are extremely common in applied health and social science research
- **Tautological associations** occur when deterministic variables are analysed in direct relation to themselves, parent, or sibling variables
- **Inferential bias** occurs when this tautology is not recognised and is mistaken as causally meaningful
- This is particularly common in analyses of **change scores** and **ratio variables**
- Robust analysis requires that the appropriate estimand is identified beforehand and that deterministic variables are very carefully considered (e.g. by drawing double-outlined nodes within your DAG)

Leeds Spring School in Causal Inference with Observational Data

09:30-10:15 LECTURE 5.1

10:15-11:00 ACTIVITY 5-A

11:00-11:30 TEA & COFFEE

11:30-12:30 LECTURE 5.2

12:30-13:00 Q&A

13:00-13:30 LUNCH

13:30-14:30 LECTURE 5.3

14:30-15:00 Q&A

15:00 CLOSE

5.1 - COMPOSITIONAL DATA

GEORGIA



@GEORGIATOMOVA

MARK



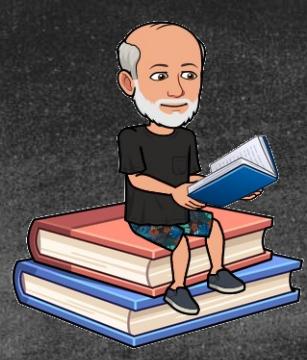
@STATSMETHODS

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

LEARNING OBJECTIVES

- Define and identify examples of compositional data
- Draw DAGs for compositional data
- Identify the two distinct types of compositional data
- Identify the two distinct estimands that exist for compositional data, and the strategies for estimating these

COMPOSITIONAL DATA

Data that comprise **parts of some whole**, for which all parts sum to that whole

- **Note:** The whole may *vary* across individual units of analysis, or remain *fixed*

Examples of compositional data include:

- *Leg length + Trunk length = Total height*
- *Brown fat mass + White fat mass = Total fat mass*
- *0-18 year-olds + 19-65 year-olds + 65+ year-olds = Total population*
- *Time spent sleeping + time spent awake = 24 hours*

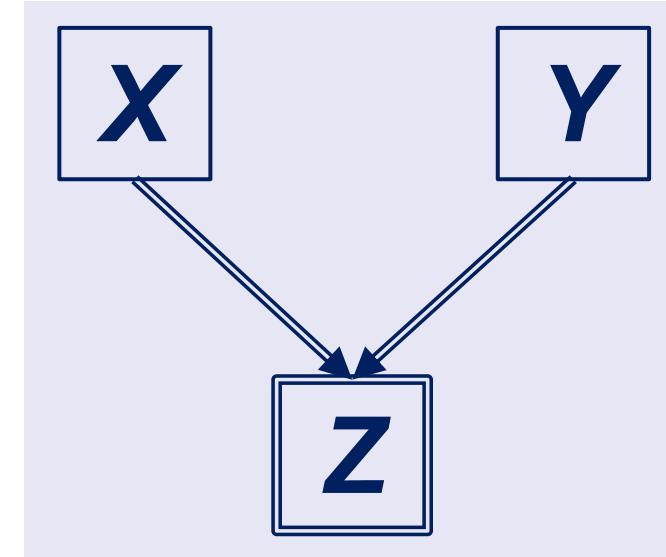
We often explicitly conceptualise data as being compositional when there is interest in one or more of the components in relation to the whole

DAGS FOR COMPOSITIONAL DATA

Variables X , Y , and Z , for which:

$$X + Y = Z$$

Z is a **collider** on the path between X and Y



DAGS FOR COMPOSITIONAL DATA

Variables X , Y , and Z , for which:

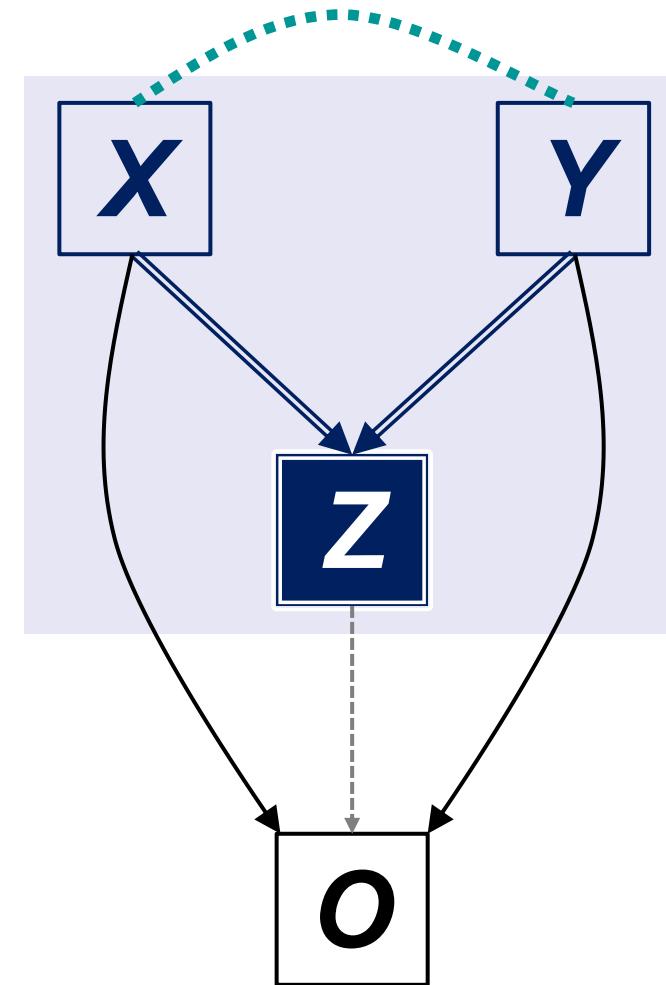
$$X + Y = Z$$

Z is a **collider** on the path between X and Y

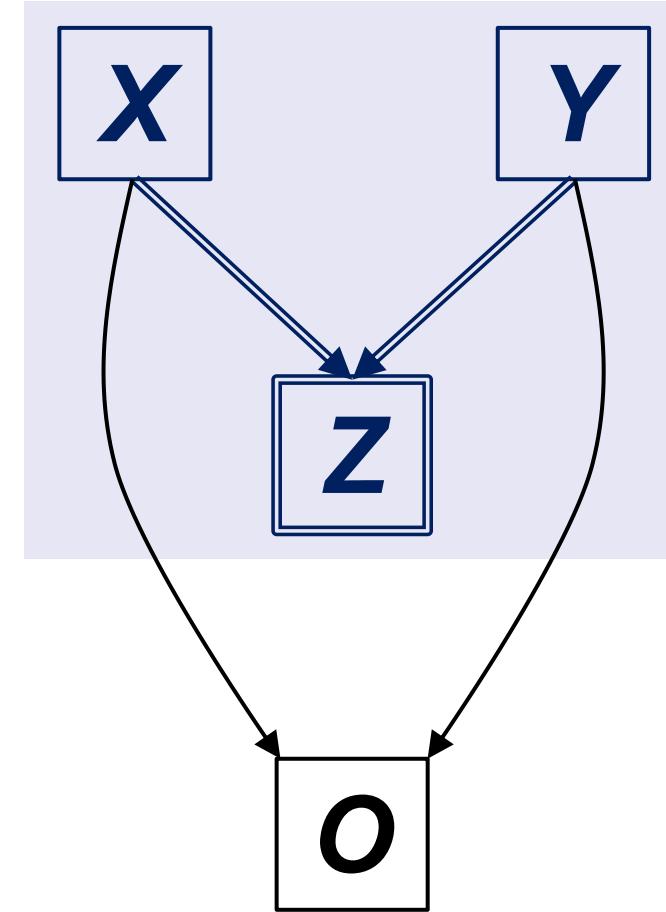
Conditioning on Z creates a **dependency** between X and Y



Collider 'bias'?

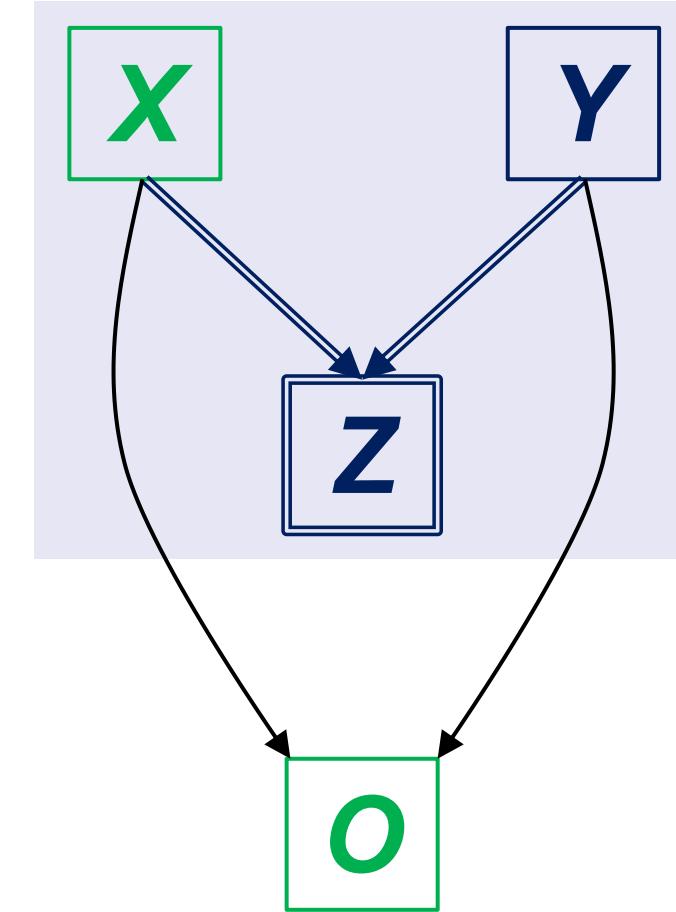


CAUSAL ESTIMANDS



CAUSAL ESTIMANDS

There exist *two distinct effects* of a particular component (e.g. **X**) on the outcome **O**

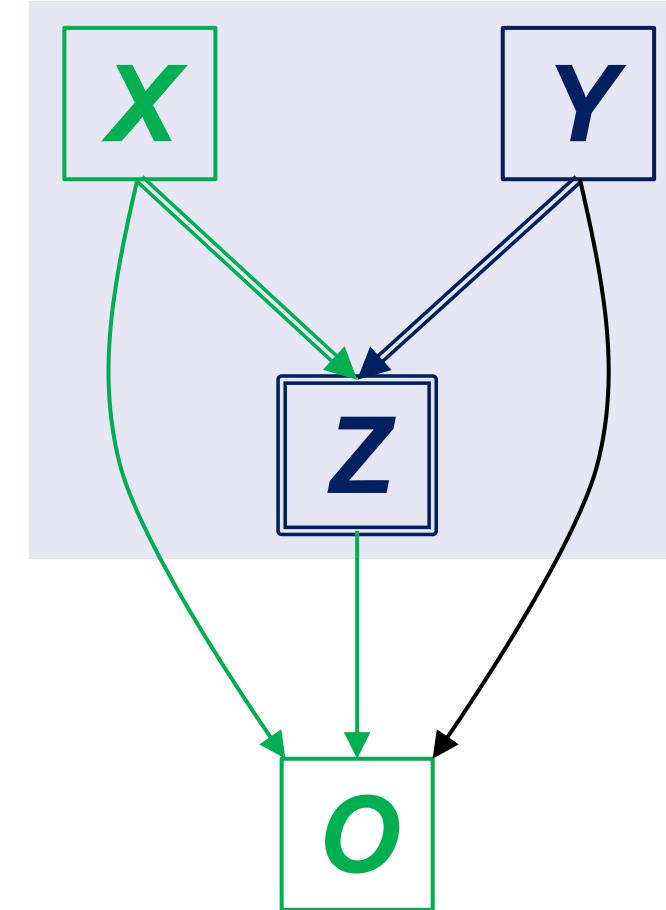


CAUSAL ESTIMANDS

There exist *two distinct effects* of a particular component (e.g. **X**) on the outcome **O**

1. Total (unconditional) effect:

- The effect on **O** of **increasing X** (and thereby increasing **Z**), regardless of **Y**



CAUSAL ESTIMANDS

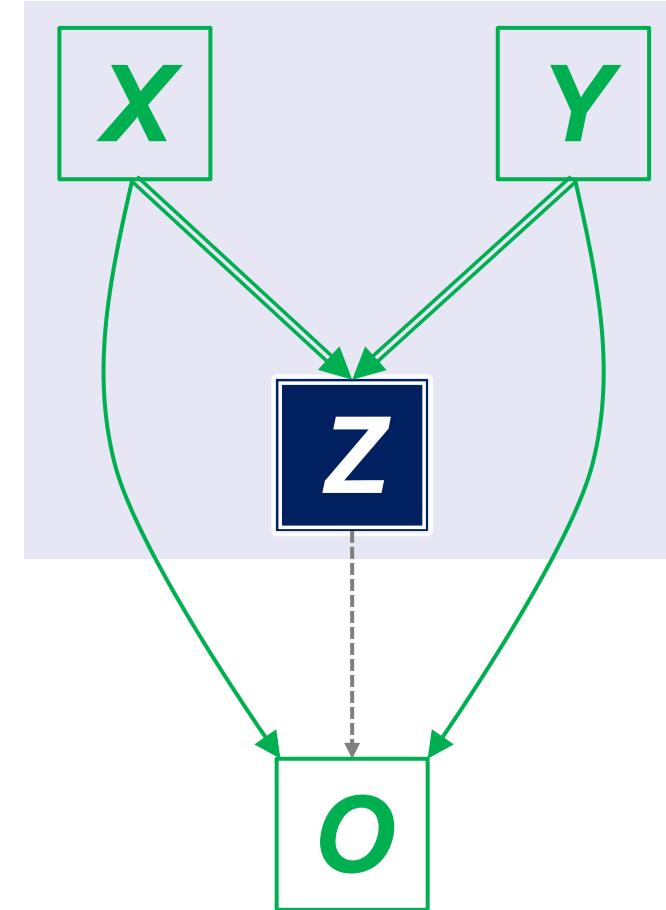
There exist *two distinct effects* of a particular component (e.g. X) on the outcome O

1. Total (unconditional) effect:

- The effect on O of **increasing X** (and thereby increasing Z), regardless of Y

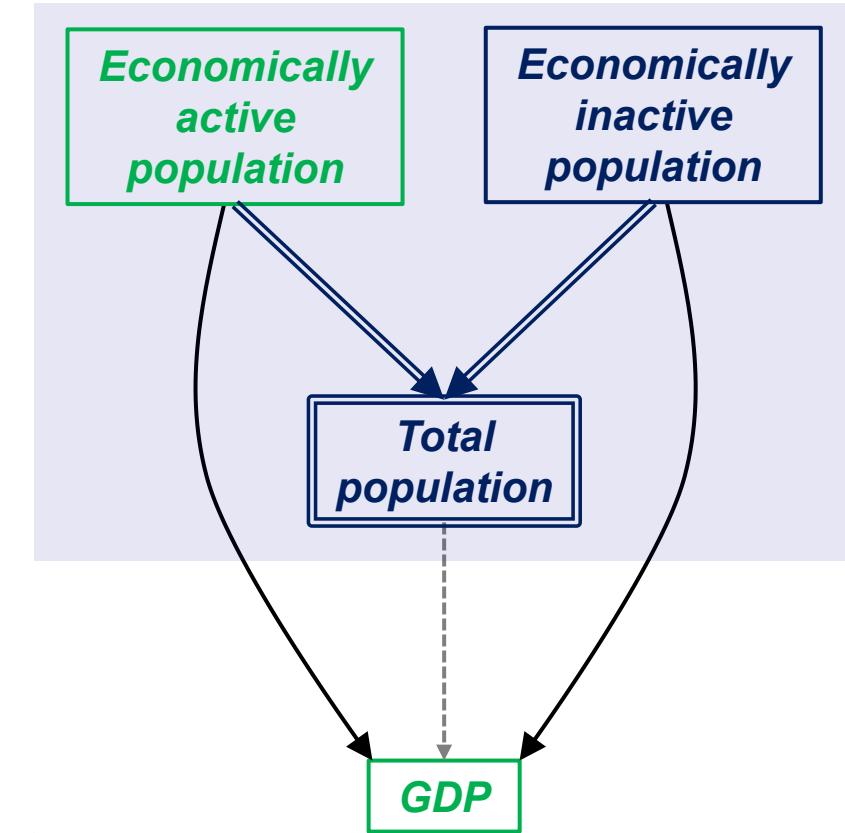
2. Relative (conditional) effect:

- The effect on O of **increasing X** while simultaneously **decreasing Y**



EXAMPLE 1

Economically active population → area-level GDP



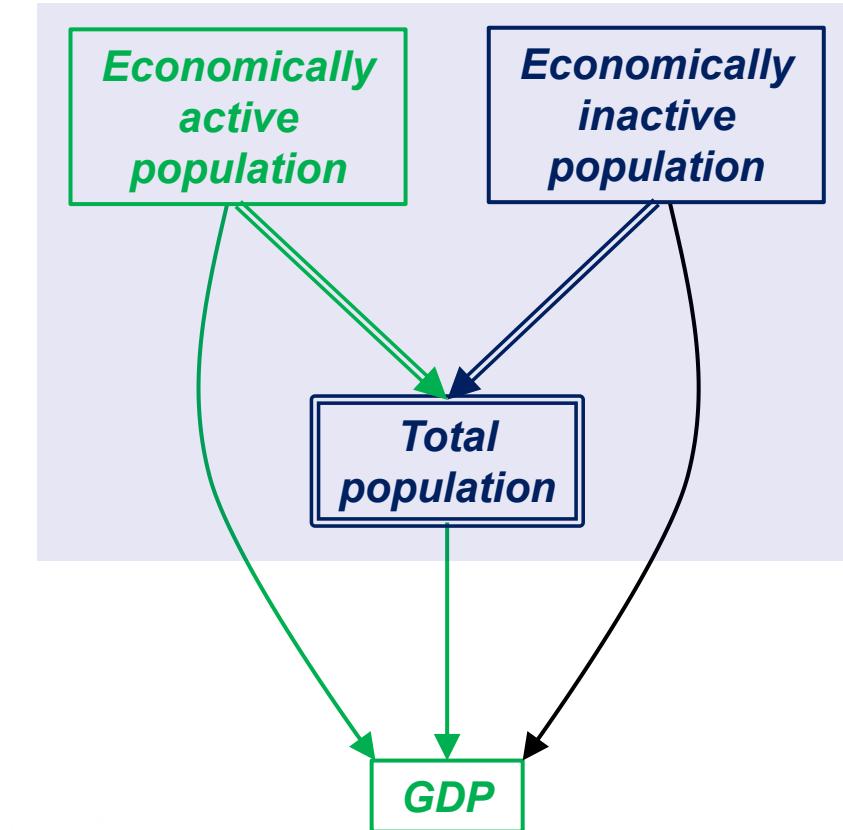
EXAMPLE 1

Economically active population → area-level GDP

1. Total (unconditional) effect:

- The average change in GDP due to adding economically active individuals to the area

↑ *Total population*



EXAMPLE 1

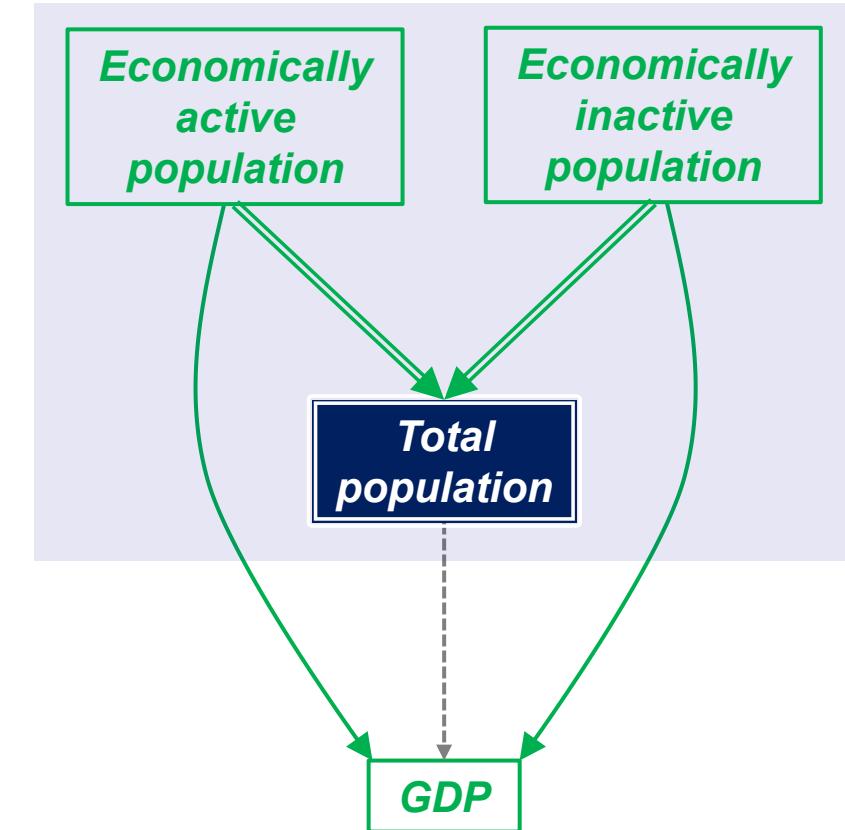
Economically active population → area-level GDP

1. Total (unconditional) effect:

- The average change in GDP due to adding economically active individuals to the area
↑ *Total population*

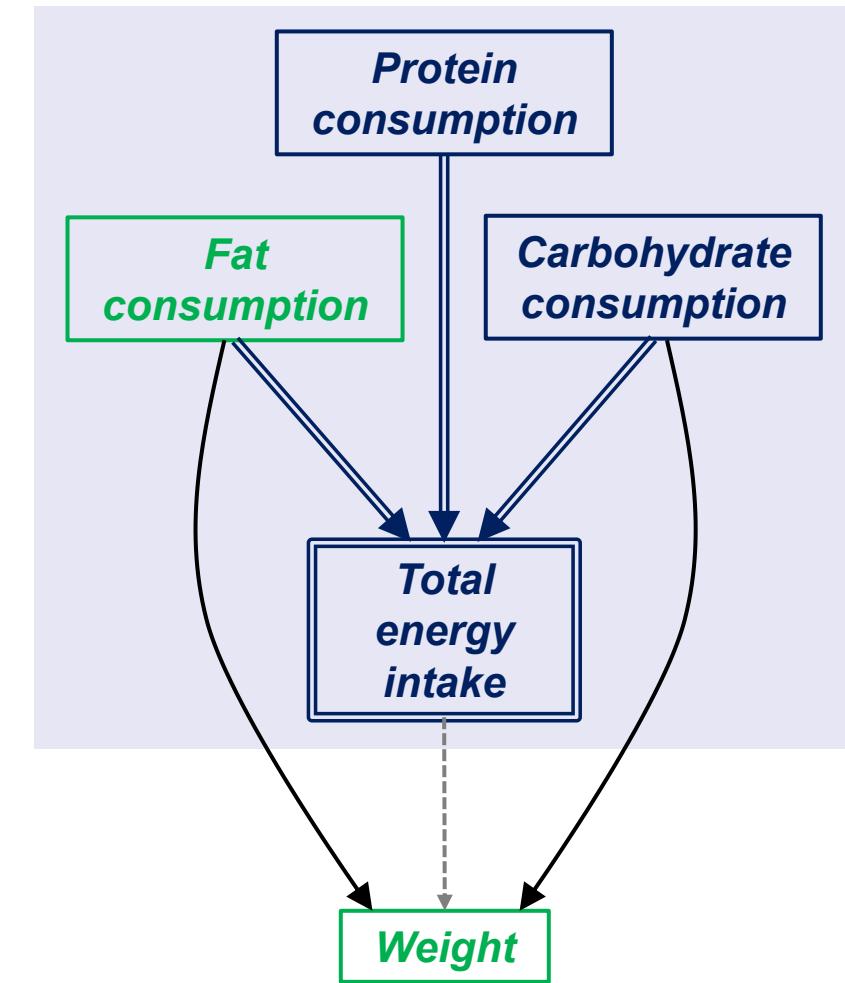
2. Relative (conditional) effect:

- The average change in GDP due to swapping economically inactive individuals for economically active ones
= *Total population*



EXAMPLE 2

Fat consumption → weight

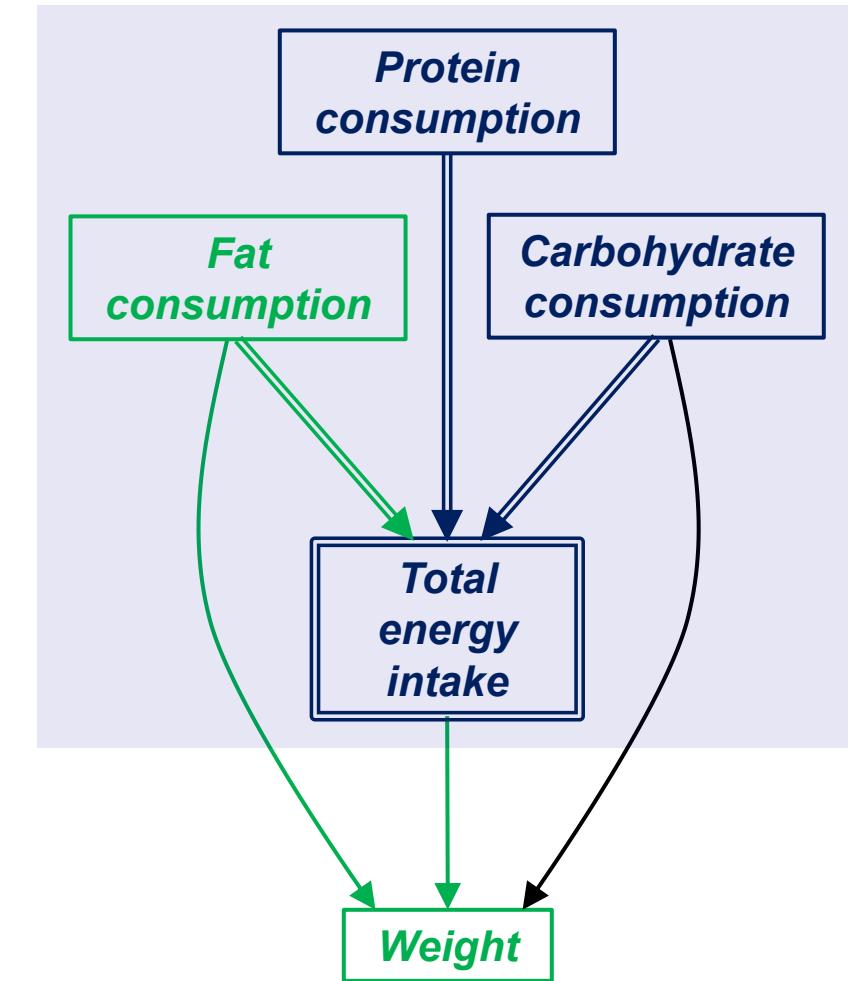


EXAMPLE 2

Fat consumption → weight

1. Total (unconditional) effect:

- The average change in weight due to adding fat to an individual's diet
 - ↑ *Total energy intake*



EXAMPLE 2

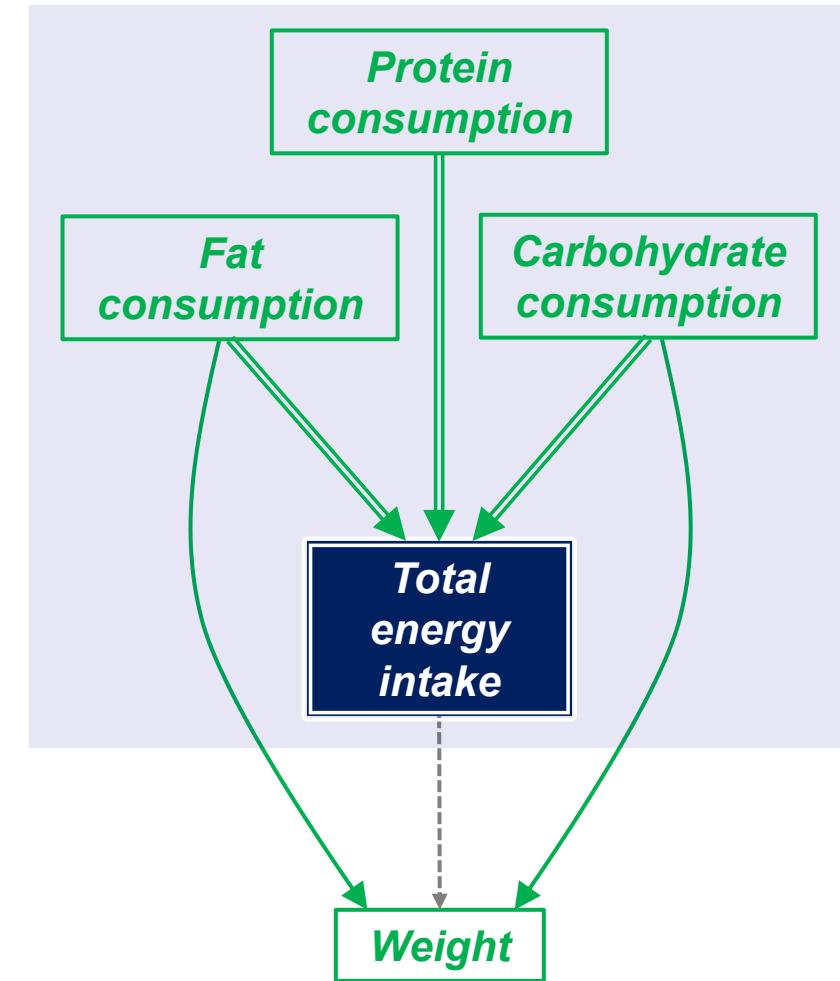
Fat consumption → weight

1. Total (unconditional) effect:

- The average change in weight due to adding fat to an individual's diet
 - ↑ *Total energy intake*

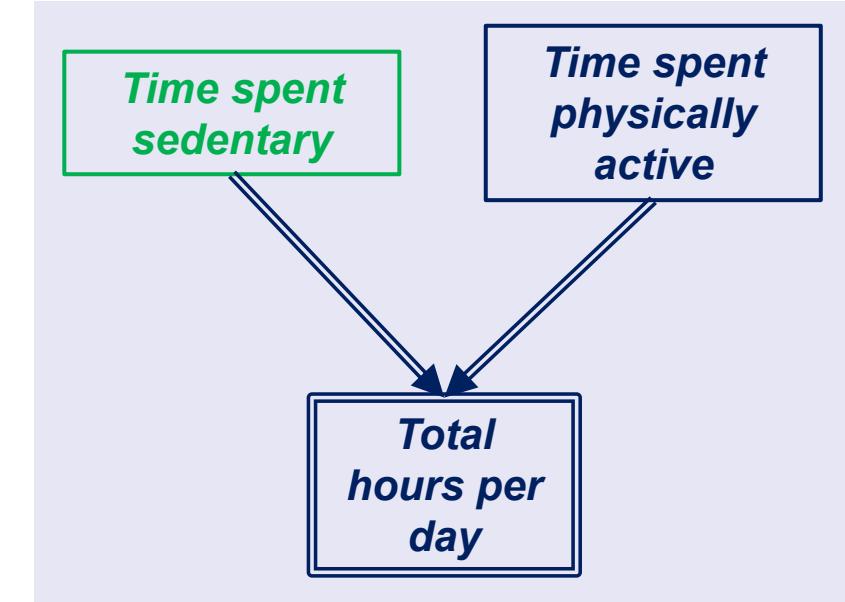
2. Relative (conditional) effect:

- The average change in weight due to swapping 'other' consumption for fat consumption
 - = *Total energy intake*
- **Note:** More specific contrasts can be made by conditioning on additional components



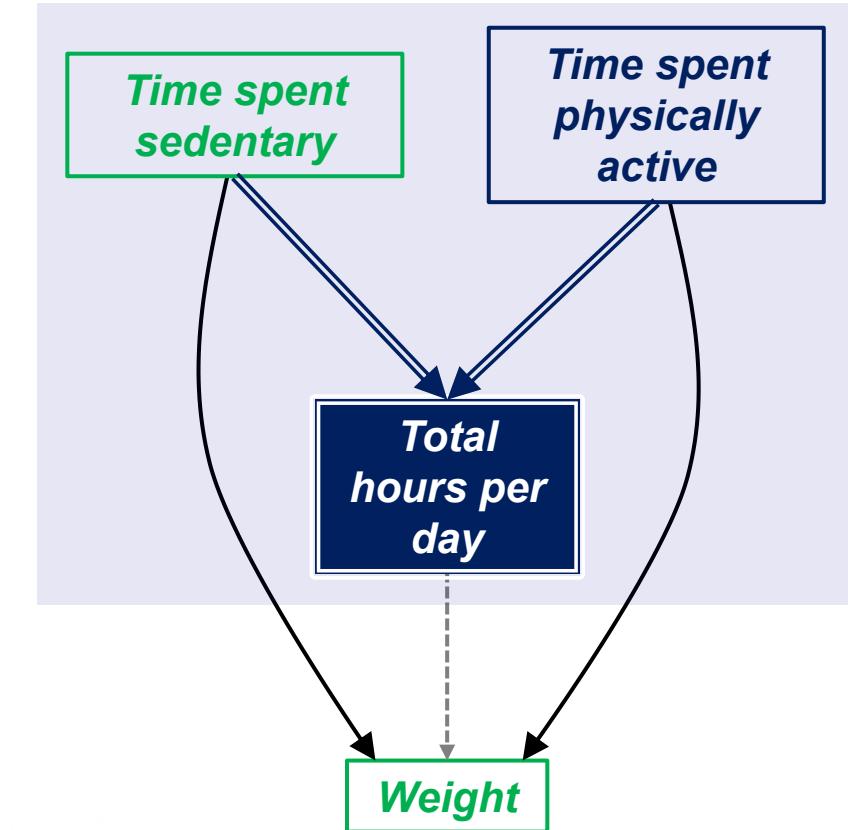
EXAMPLE 3

Time spent sedentary → weight



EXAMPLE 3

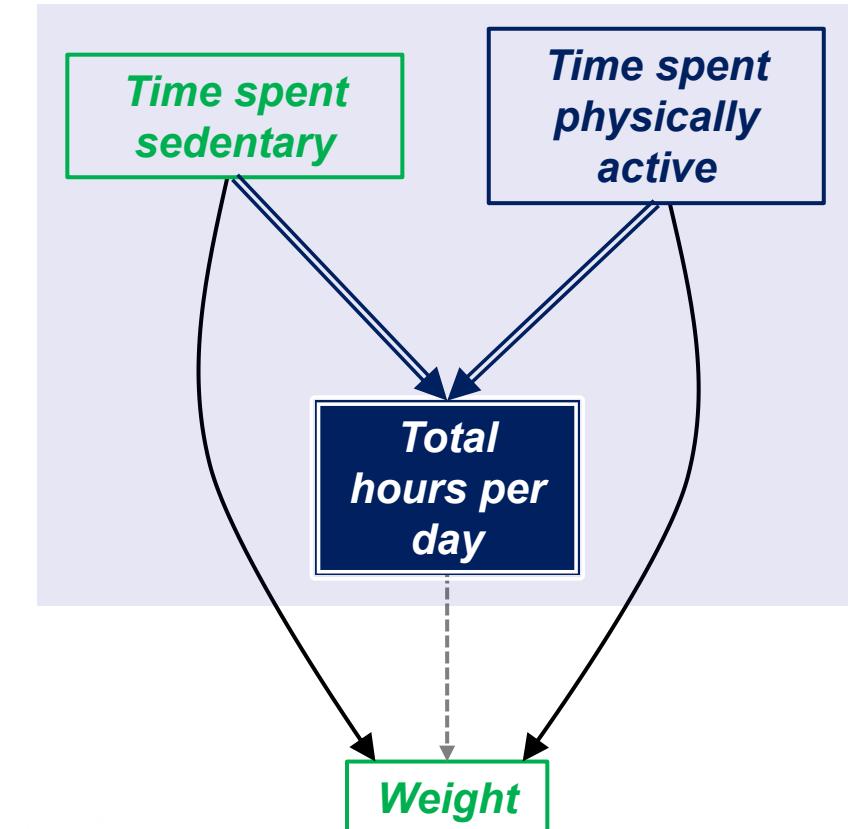
Time spent sedentary → weight



EXAMPLE 3

Time spent sedentary → weight

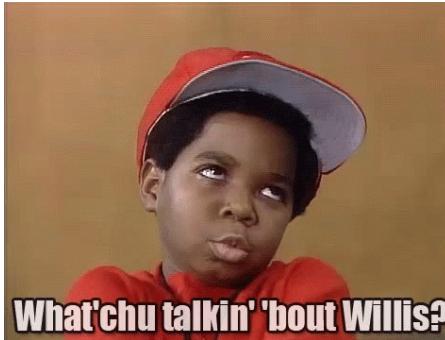
1. Total (unconditional) effect:



EXAMPLE 3

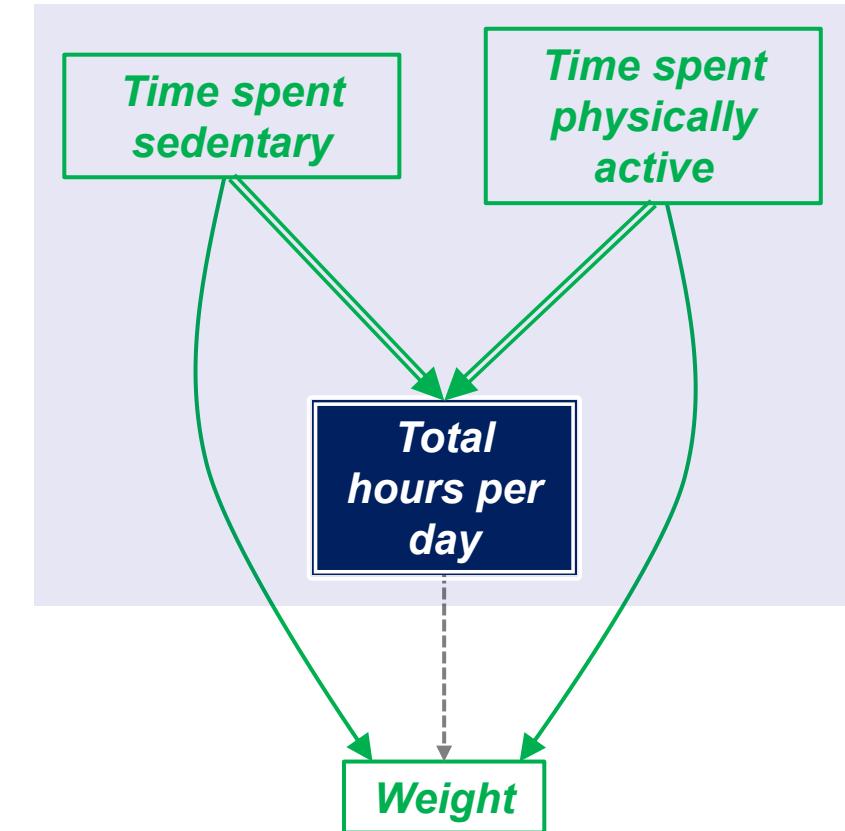
Time spent sedentary → weight

1. Total (unconditional) effect:



2. Relative (conditional) effect:

- The average change in weight due to swapping time spent sedentary for time spent physically active
= *Total hours*



VARIABLE VS FIXED TOTALS

Compositional data with variable totals:

- both the **total** and **relative** causal effects of one component on an outcome are identifiable
- either effect may be of interest, depending on context

Compositional data with fixed totals:

- only the **relative** causal effect of one component on an outcome is identifiable

ESTIMATING THE TOTAL (UNCONDITIONAL) EFFECT

Same variable selection rules apply!

DO

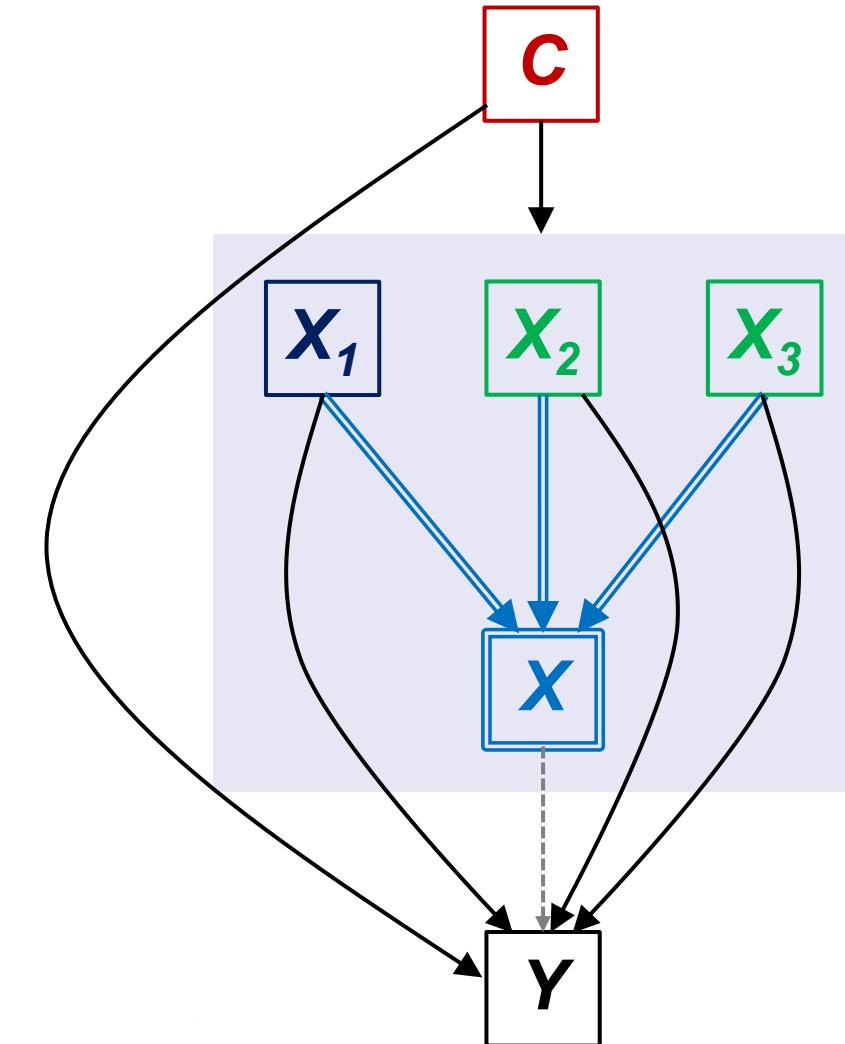
- include **confounders** (i.e. **C**) to block confounding paths

DO NOT

- include **mediators** (i.e. the total **X**), as this would block true causal paths (and cause collider ‘bias’)

DO (OPTIONALLY)

- include **competing exposures** (i.e. **X₂**, **X₃**), as this *improves* precision

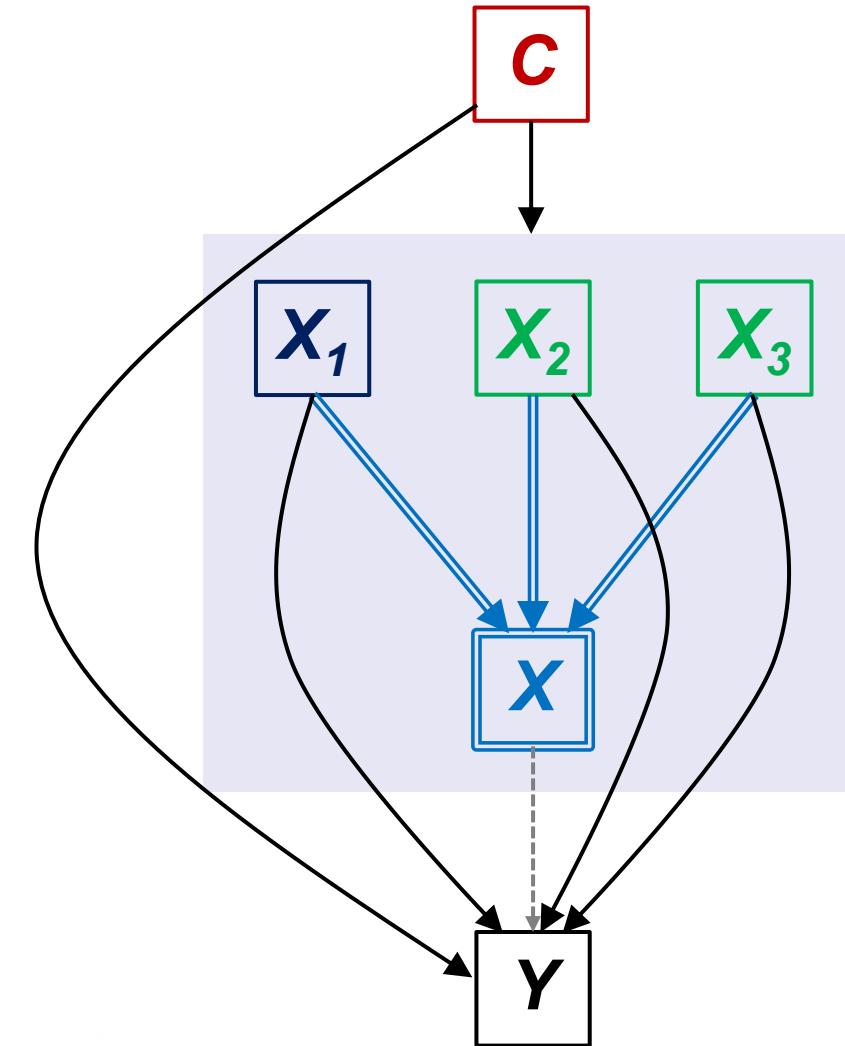


ESTIMATING THE RELATIVE (CONDITIONAL) EFFECT

Estimating the relative effect involves
conditioning on the total

However, the total (X) is a **composite variable** and thus an *imperfect summary* of its components (X_1, X_2, X_3)

Adjusting for the total will lead to
composite variable bias if the individual components have unique effects on the outcome



COMPOSITE VARIABLE BIAS

- **Recall:** **Composite variables** are variables that are perfectly determined by two or more parent variables
- A ‘total’ variable is a composite variable, because it summarises information from all parent ‘parts’ variables
- Although a ‘total’ variable has a clear physical and theoretical foundation, it contains far less information than the individual ‘parts’
- Where the ‘parts’ have different causal effects, the total will provide an imperfect average of those effects
- Where the total is used to account for confounding by the parts, this will lead to residual confounding

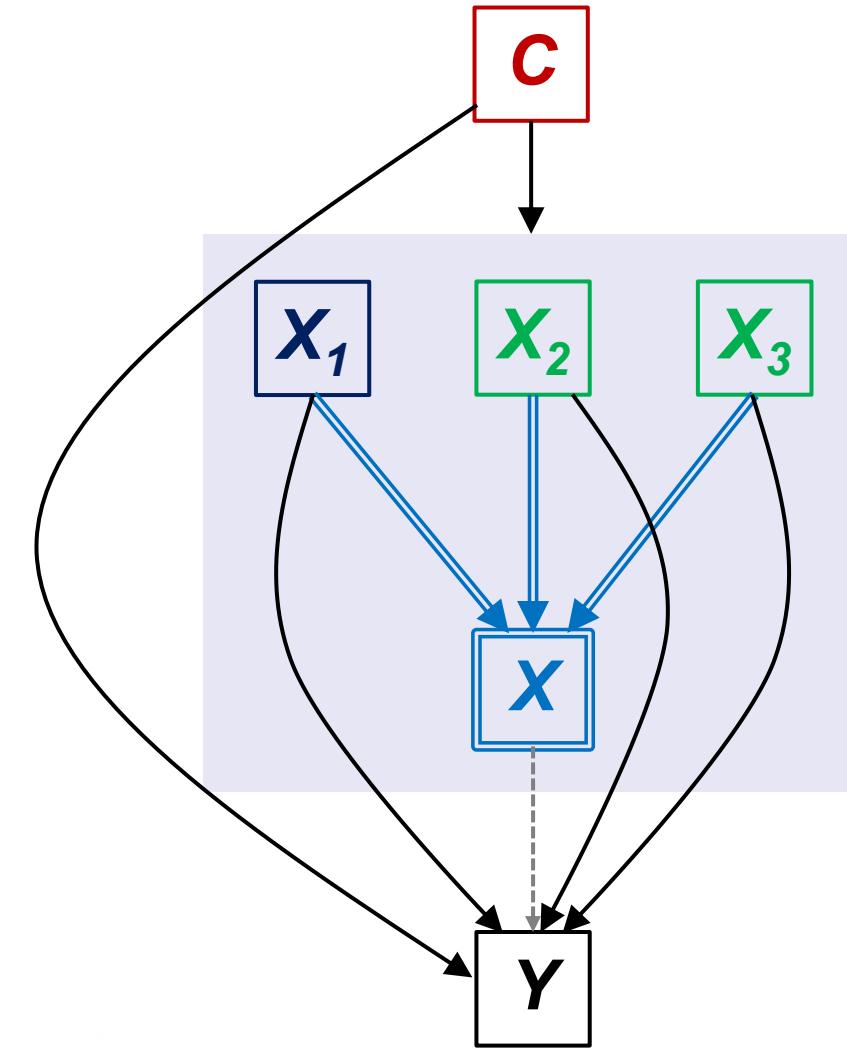
'ALL-COMPONENTS' MODEL: TOTAL ENERGY EXAMPLE

An '**all-components**' model can be used to adjust for all competing sources of energy other than the exposure (i.e. all 'other' energy intake):

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 C$$

Each coefficient represents the **total effect** of its component on the outcome

Relative effects can be obtained through simple algebra



RECOMMENDED READING

- Arnold, K.F., Berrie, L., Tennant, P.W.G., Gilthorpe, M.S. (2020). A causal inference perspective on the analysis of compositional data. *International Journal of Epidemiology*, 49(4), 1307-1313.
<https://doi.org/10.1093/ije/dyaa021>
- Tomova, G.D., Arnold, K.F., Gilthorpe, M.S., Tennant, P.W.G. (2022). Adjustment for energy intake in nutritional research: a causal inference perspective. *The American Journal of Clinical Nutrition*. 115(1), 189-198 <https://doi.org/10.1093/ajcn/nqab266>

SUMMARY

DAGs offer a useful conceptual framework for considering compositional data

For causal questions involving compositional data, two distinct effects may be of interest, depending on context

- Total (unconditional) effect
- Relative (conditional) effect

An ‘all-components’ model may be used to robustly estimate both total and relative effects for compositional data

Leeds Spring School in Causal Inference with Observational Data

5.2 - COMPOSITE VARIABLE BIAS

GEORGIA



MARK



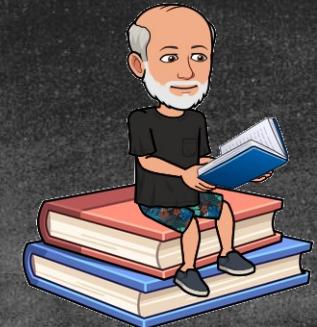
@GEORGIATOMOVA

PETER



@PWTENNANT

ROB



@WAYNEROBERTLONG

09:30-10:15 LECTURE 5.1

10:15-11:00 ACTIVITY 5-A

11:00-11:30 TEA & COFFEE

11:30-12:30 LECTURE 5.2

12:30-13:00 Q&A

13:00-13:30 LUNCH

13:30-14:30 LECTURE 5.3

14:30-15:00 Q&A

15:00 CLOSE

LEARNING OBJECTIVES

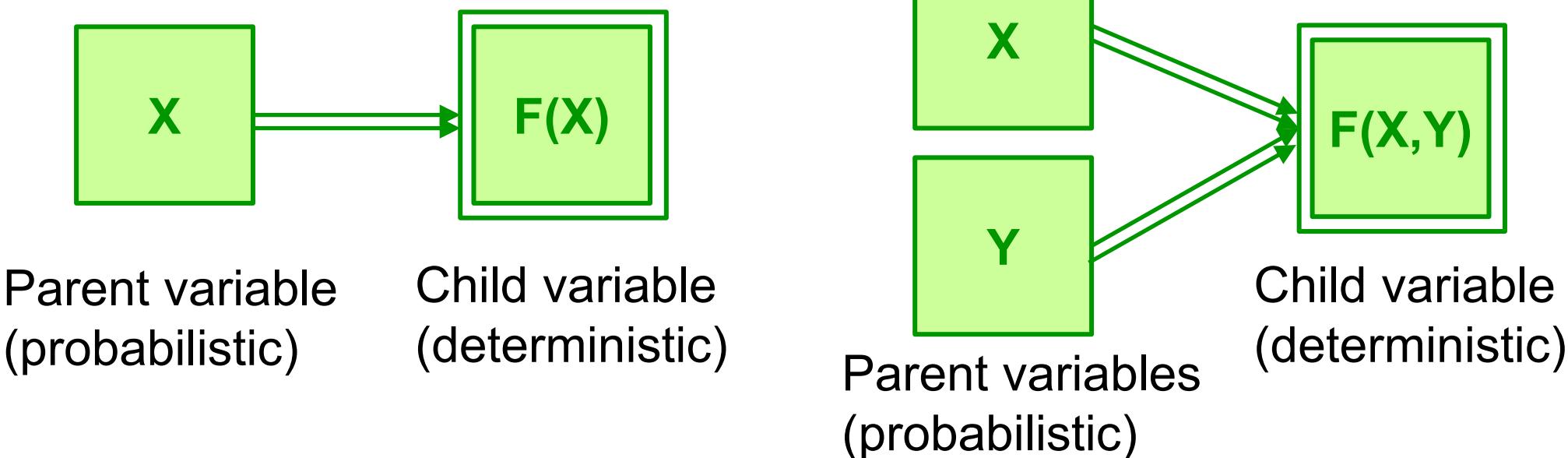
- By the end of this lecture, you will be able to:
 - Define **composite variable bias**
 - Explain how the analysis of composite variables for causal inference may be complicated by :
 - **Loss of information**
 - **Estimand confusion**
 - **Consistency violations**
 - **Exchangeability violations**

RECAP: DETERMINISTIC VARIABLES

- In health and social science research, some relationships and variables are **deterministic**
 - i.e. their values can be *completely* known from other variables
- This typically occurs when we create variables ourselves, such as:
 - **simple derived variables**
 - **complex derived (composite) variables**

RECAP: DETERMINISTIC VARIABLES IN DAGS

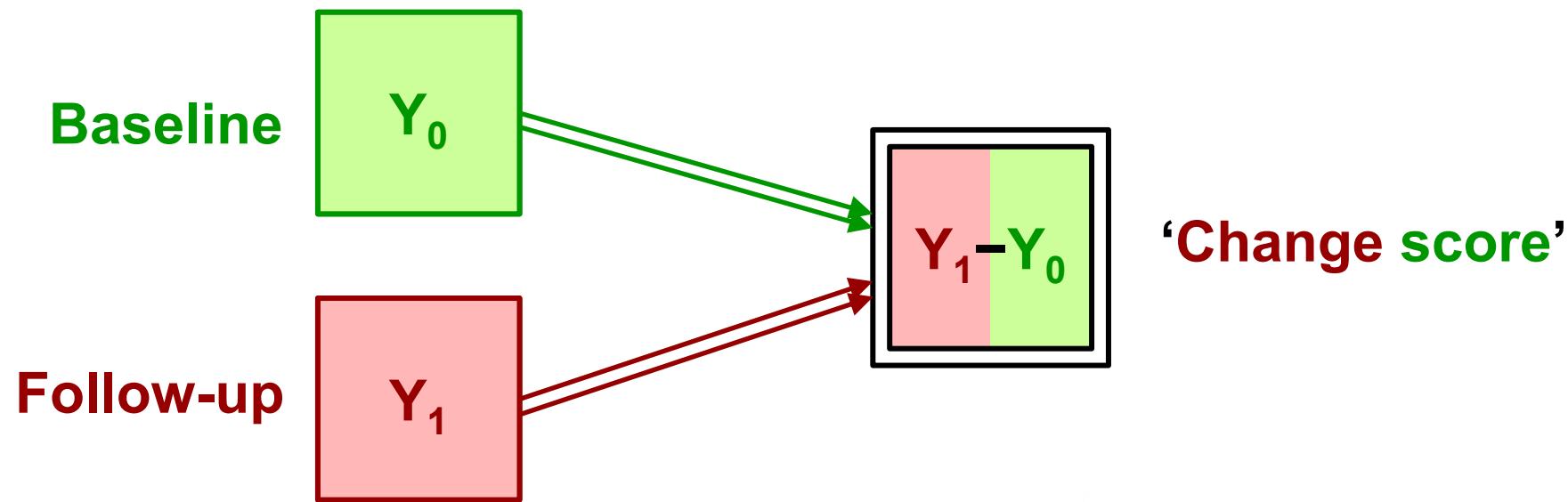
- **Deterministic variables** can be depicted in DAGs with **double-outline nodes and double arrows**
- We can describe deterministic variables as the **child** (or **descendent**) of its probabilistic **parent(s)** (or **ancestors**)



RECAP: TAUTOLOGICAL ASSOCIATIONS

DAGs with **deterministic nodes** can help understand and avoid **tautological associations**

- e.g. the analysis of change-scores with respect to the baseline:



WHAT'S ALL THE FUSS ABOUT?

The benefits of DAGs for recognising and understanding **tautological associations** are however fairly modest

Arguably, these are actually easier to detect from the structural equation alone:

Tautological association in **change-score** analysis:

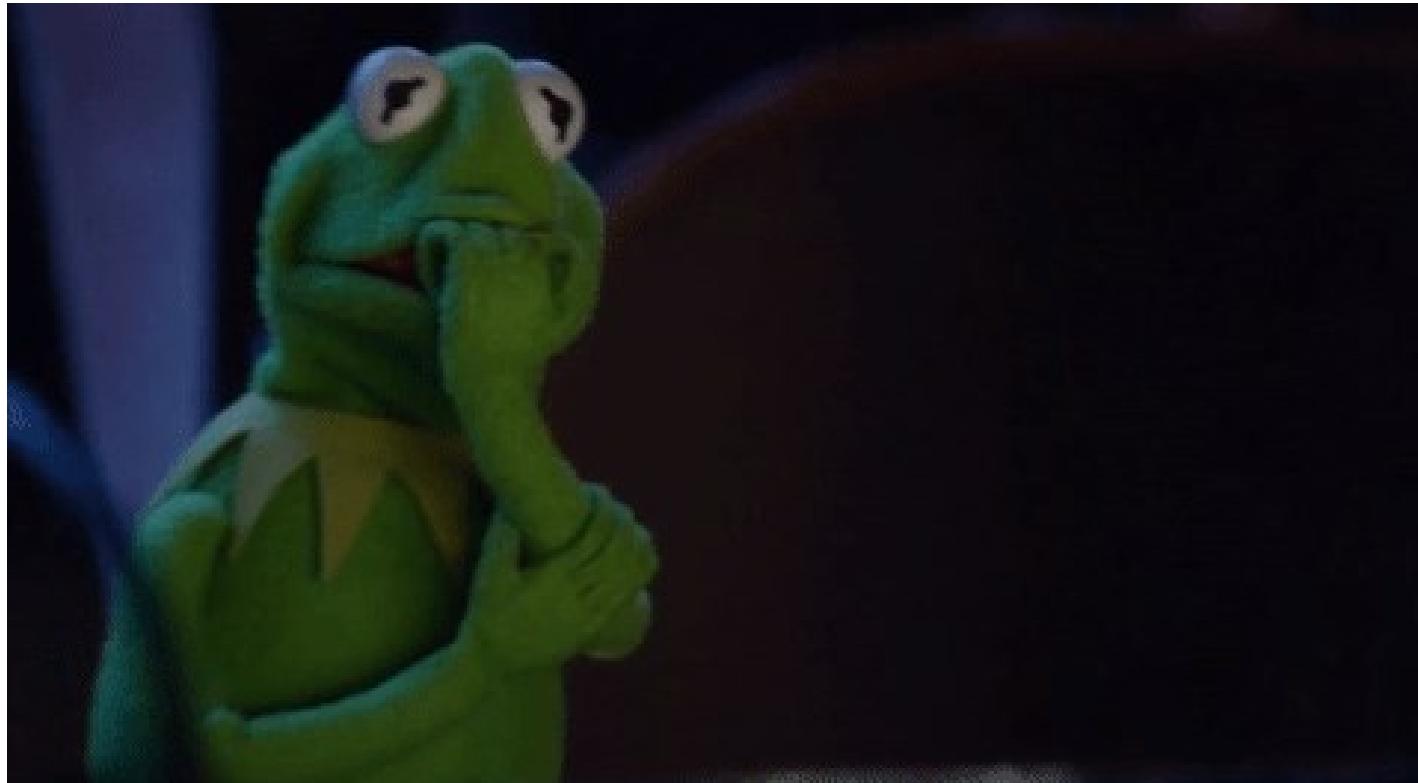
$$(Y_1 - Y_0) = \beta_0 + \beta_1 Y_0 + \dots + \varepsilon$$

Tautological association in **ratio-variable** analysis:

$$(Y/N) = \beta_0 + \beta_1 (X/N) + \dots + \varepsilon$$

WHAT'S ALL THE FUSS ABOUT?

Tautological associations are merely the tip of the iceberg when it comes to the (potential) problems with analysing **deterministic variables**



COMPOSITE VARIABLE BIAS

Composite variable bias is a variety of **inferential bias** that complicates the analyses of **composite variables**

- Need not involve **tautological associations**
- Has received very little research and is fairly poorly understood
- Although broad concept(s) may be relatively easy to understand, even the simplest examples can become brain-meltingly complex

COMPOSITE VARIABLE BIAS

One-line summary

- When analysing composite variables, you probably have no idea what you're estimating; never mind what it means!



COMPOSITE VARIABLE BIAS

- Estimating causal effects for composite variables faces four issues:

Information loss

- By combining multiple variables, you distort their effects to create an imperfect average

Consistency violations

- Composite variables *fundamentally* violate consistency

Exchangeability violations

- For composites with large temporal footprints, components may have different relationships with other variables, making it impossible to adjust for confounding

Estimand confusion

- Using composite variables may not give you the estimand you seek
- Depending on how the variables have been combined, your estimates may be misleading

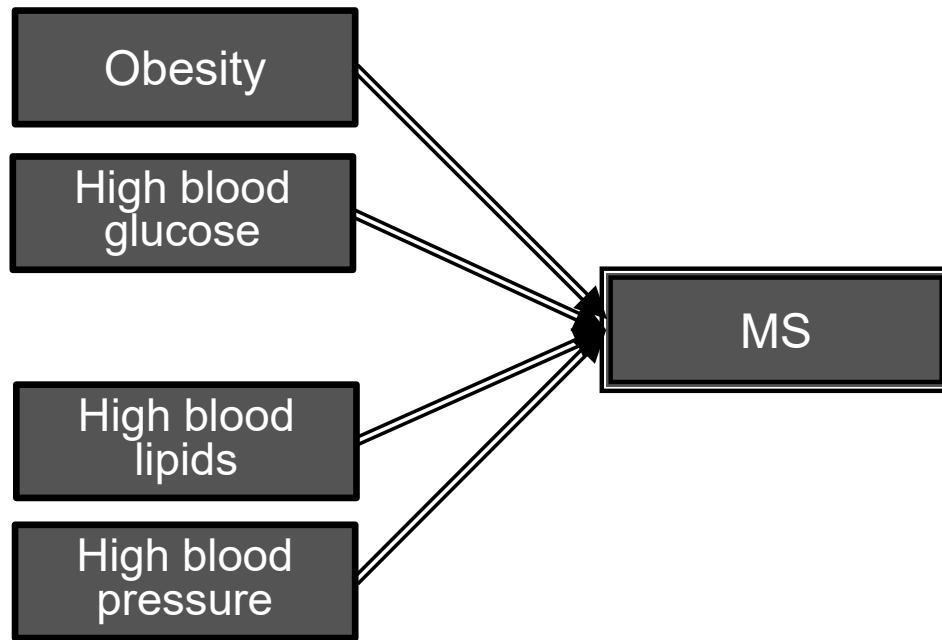
Information Loss

- **Composite variables** contain less information than their parent variables, meaning they return an *imperfect average* effect of each part
- The summary causal effect is determined by the relative *variance* of each variable
- Components with the largest variance will contribute more to the average effect

EXAMPLE: METABOLIC SYNDROME

Metabolic Syndrome

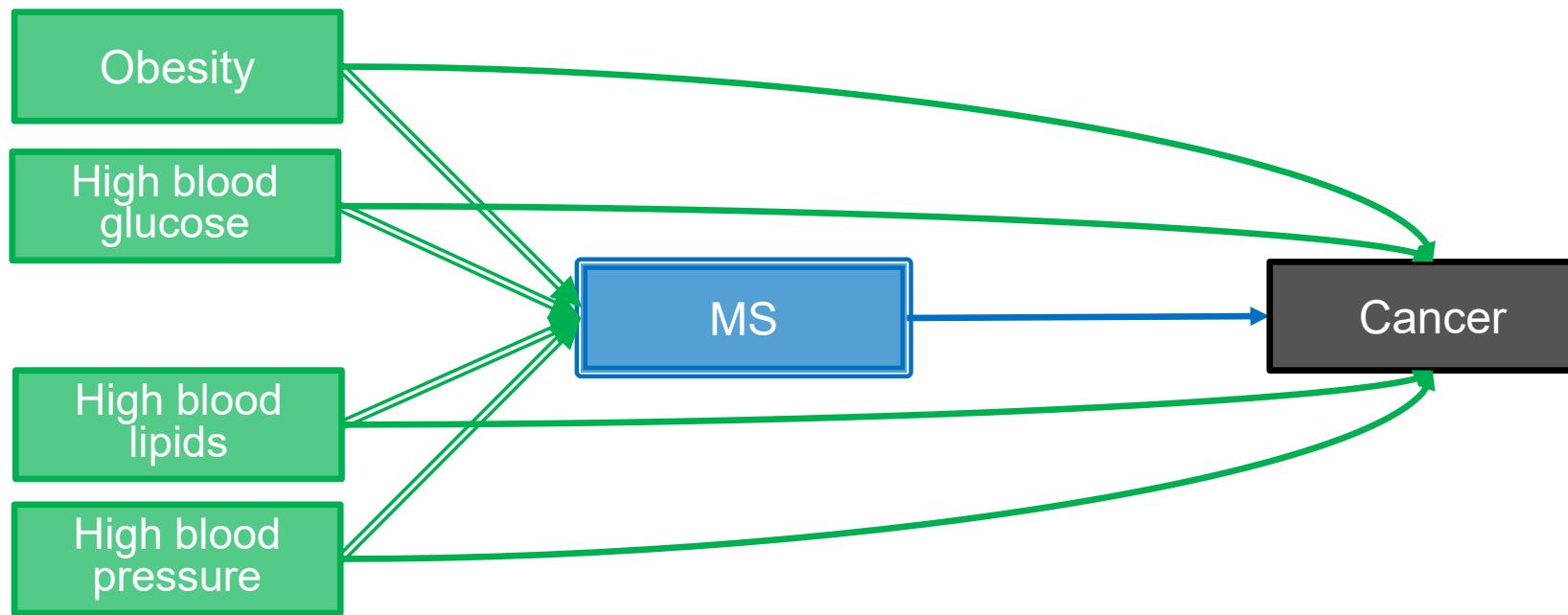
- A composite variable made from 4-5 parent metabolic variables



EXAMPLE: METABOLIC SYNDROME

Metabolic Syndrome

- Suppose you were interested in the **summary effect of MS on cancer**



- The effect would be an ***imperfect average*** of the effects of **obesity**, **high blood glucose**, **high blood lipids**, and **high blood pressure**; skewed towards whichever was most variable in your sample

CONSISTENCY VIOLATIONS

Recall:

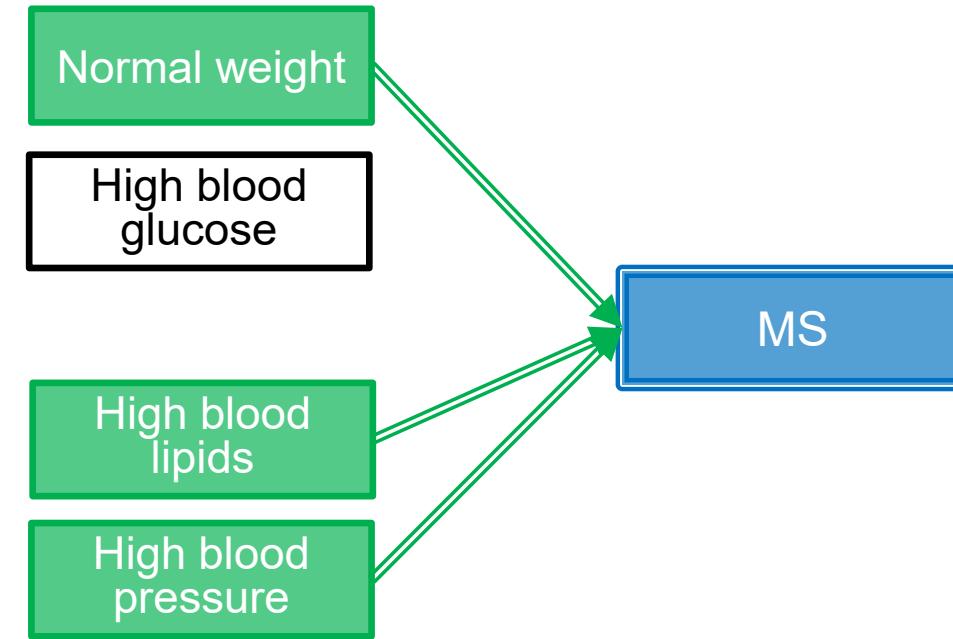
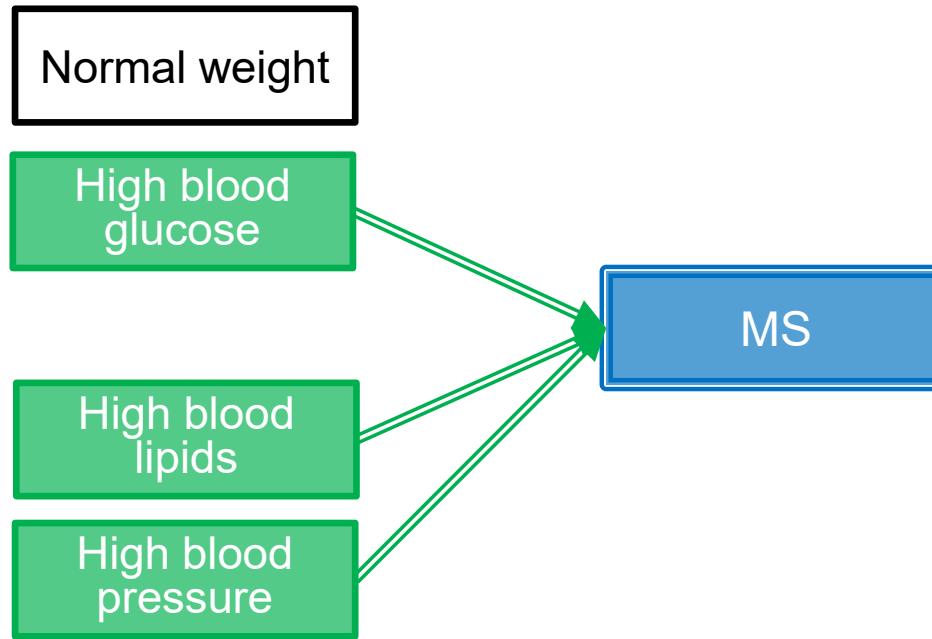
- **Consistency:** The effect of the exposure must be the same whether observed (i.e. in the factual universe) as if given by intervention (i.e. in the counterfactual universe)
- “*There are no two “flavors” or versions of treatment*”

Composite variables fundamentally violate consistency

- With composite variables, the same value can be obtained from multiple combinations of the parent variables; leading to multiple flavours

EXAMPLE: METABOLIC SYNDROME

- These two people with MS are not exchangeable, because they have different flavours of MS



- We cannot be sure that the units we are using as estimates of the counterfactual have the same 'flavour' as those units *would have had* if we intervened

CONSISTENCY VIOLATIONS

Recall:

- **Consistency:** The effect of the exposure must be the same whether observed (i.e. in the factual universe) as if given by intervention (i.e. in the counterfactual universe)
- “*There are no two “flavors” or versions of treatment*”

Composite variables fundamentally violate consistency

- With composite variables, the same value can be obtained from multiple combinations of the parent variables; leading to multiple flavours

Example 2: Body mass index = **weight / height²**

- With BMI, we have no idea whether the observed effect is due to **weight**, **height**, or **both**. The same value can be obtained from an infinite number of combinations of **weight** and **height**; a ‘one-unit change in BMI’ has no consistent meaning

EXCHANGEABILITY VIOLATIONS

Exchangeability violation

- Accurately identifying a causal effect requires that all relevant variables can be temporally situated within the DAG
- This is not possible for composite variables, because the parent variables may **crystallise** at different points in time!
- This makes it impossible to correctly identify and adjust for confounding

At best

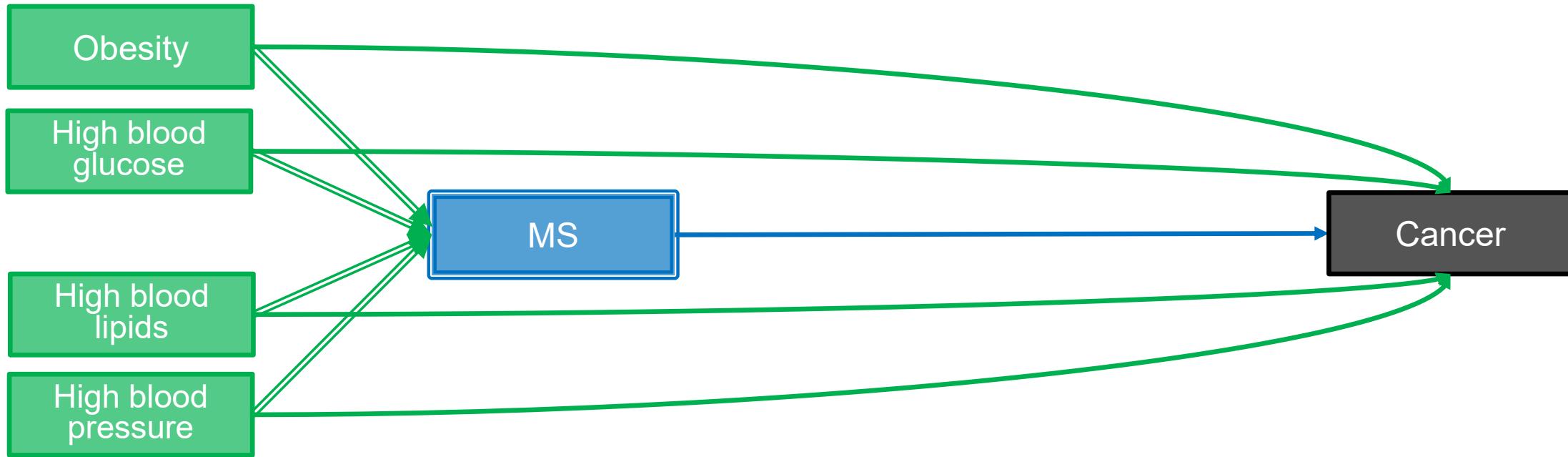
- The different components may have different relationships with each confounder

At worst

- Some variables may confound the effect of some components while mediating the effect of others !

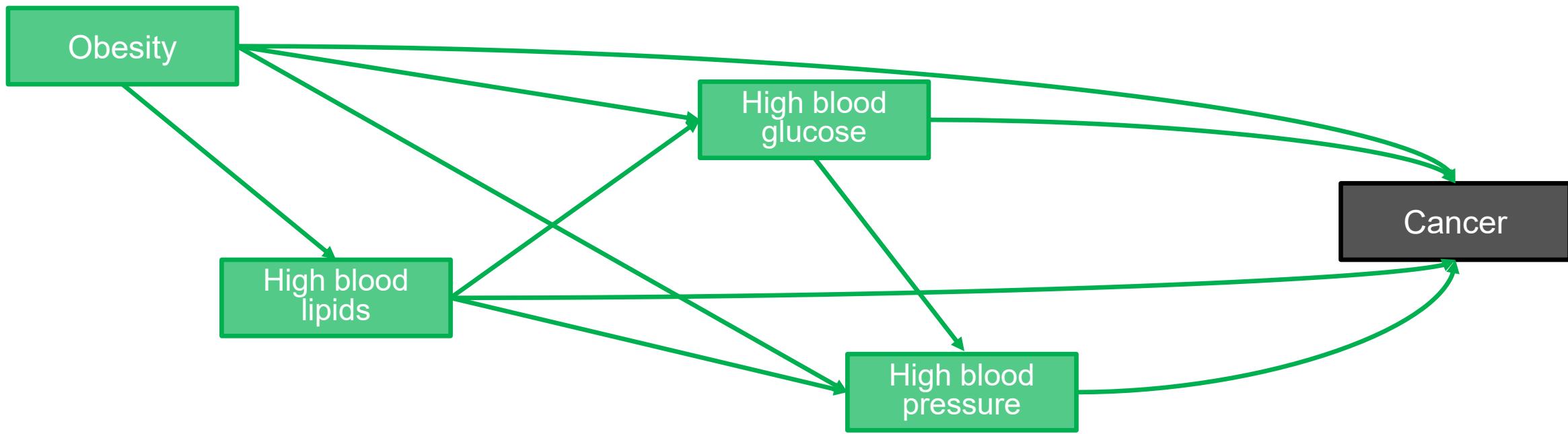
EXAMPLE: EXCHANGEABILITY VIOLATIONS

- Suppose again we were interested in the **summary effect of MS** on cancer
- We have drawn the parent variables as if they all occur together



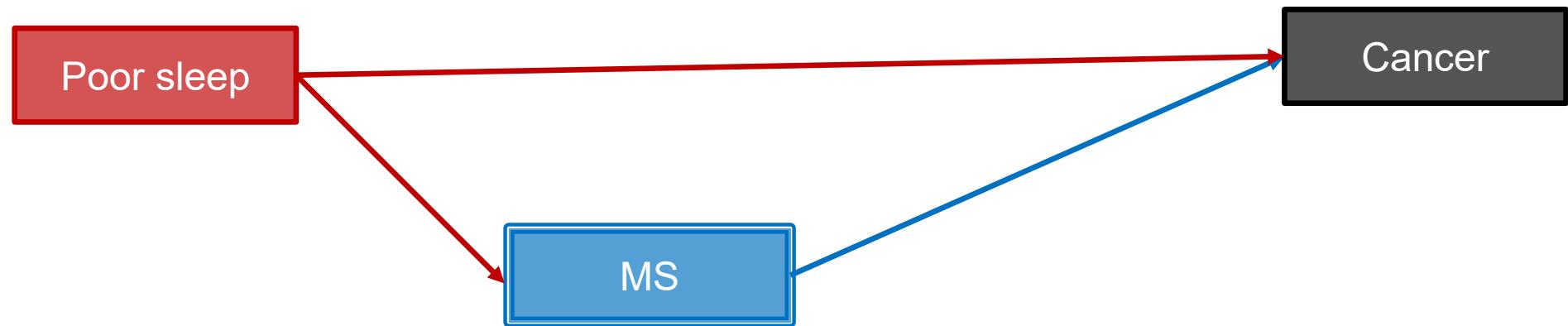
EXAMPLE: EXCHANGEABILITY VIOLATIONS

- Suppose again we were interested in the **summary effect of MS** on cancer
- We have drawn the parent variables as if they all occur together
- In fact, these variables are spread out over time...



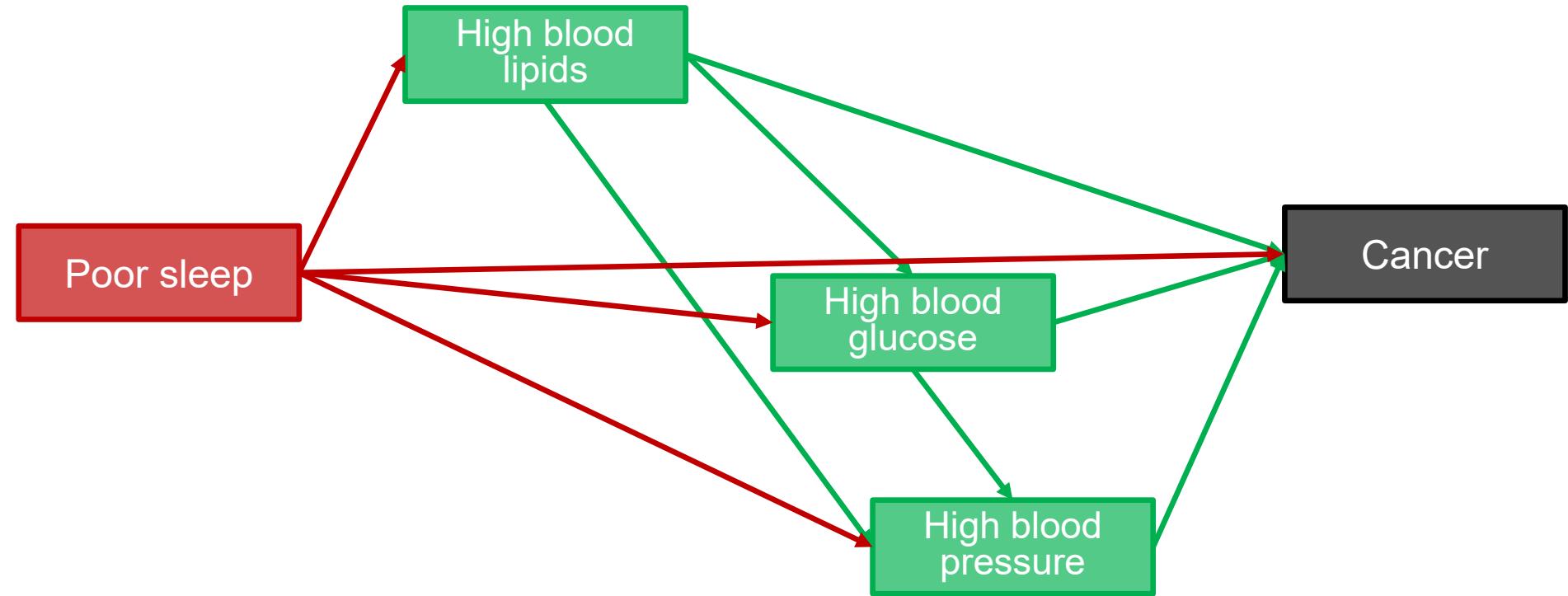
EXAMPLE: EXCHANGEABILITY VIOLATIONS

- Question: Is **poor sleep** a **confounder** for the effect of **MS** on cancer?



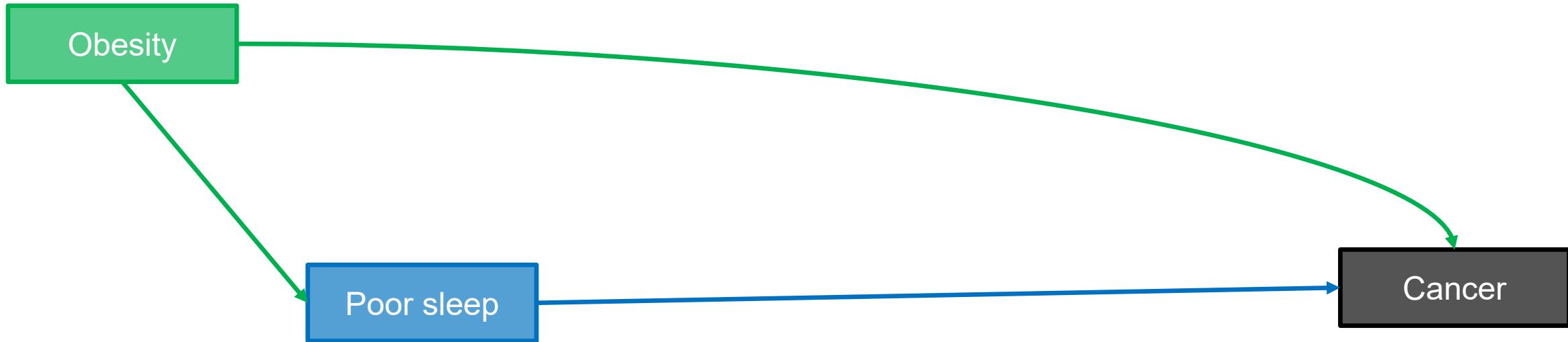
EXAMPLE: EXCHANGEABILITY VIOLATIONS

- Poor sleep a **confounder** for the effects of **lipids**, **glucose** and **high BP**



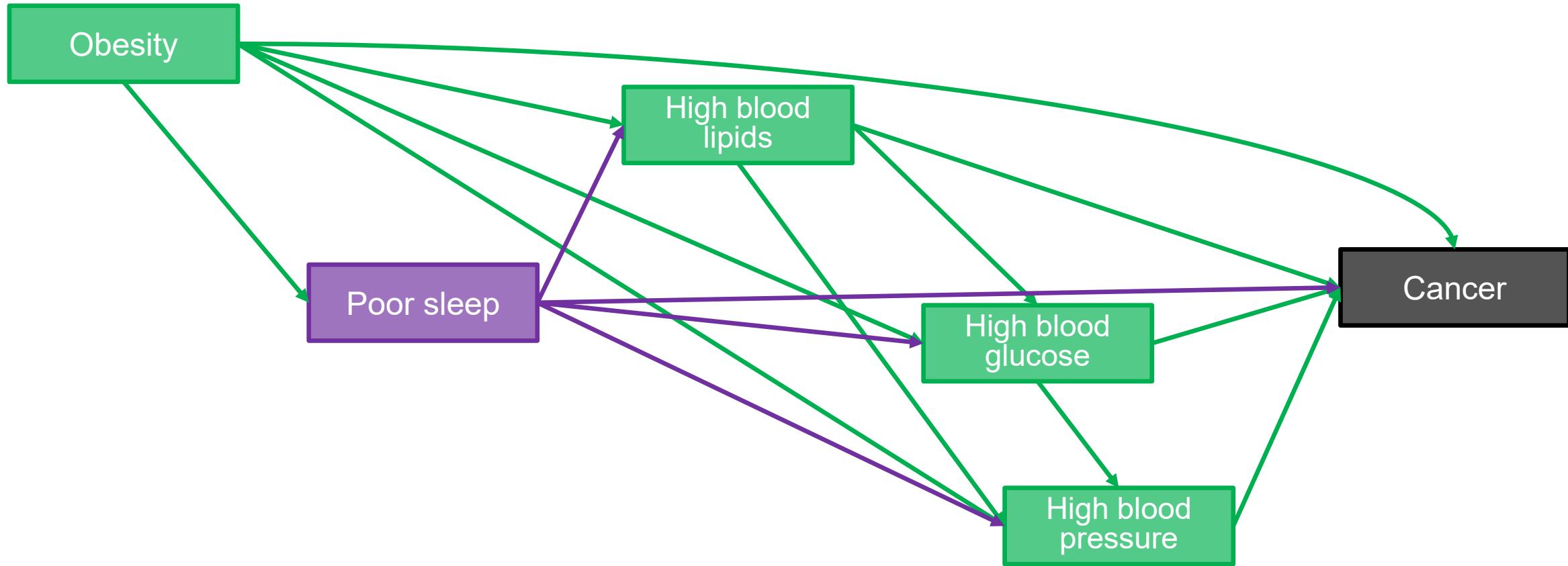
EXAMPLE: EXCHANGEABILITY VIOLATIONS

- But: Poor sleep a mediator for the effect of obesity



EXAMPLE: EXCHANGEABILITY VIOLATIONS

- Poor sleep is both **confounder** and **mediator**!

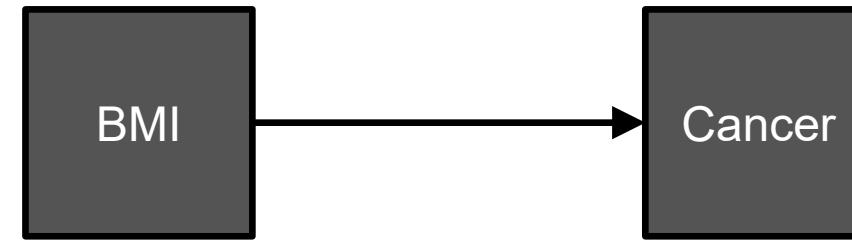


ESTIMAND CONFUSION

- Composite variables are typically made for two reasons:
 1. To provide useful summaries two or more variables
 2. To ‘standardise’ one variable by a nuisance variable or concept
- Common approaches to standardisation include:
 - Dividing by the nuisance variable (i.e. **ratio variables**)
 - Subtracting the nuisance variable (i.e. **change variables**)
- These variables are very unlikely to return the true estimand of interest!

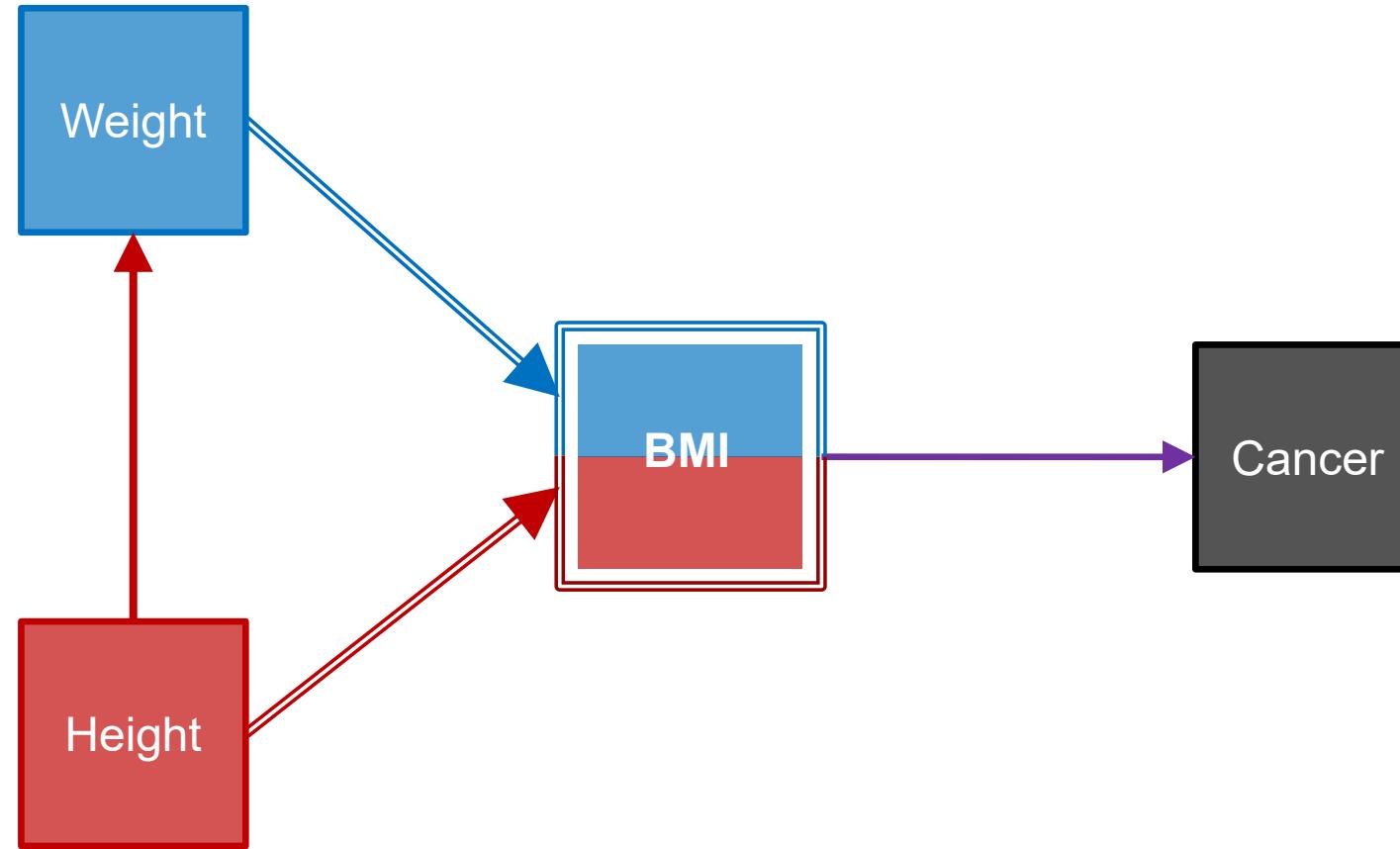
ESTIMAND CONFUSION

Suppose we are interested in the **total causal effect of BMI on cancer**



ESTIMAND CONFUSION

BMI is actually a composite of **weight** and **height**

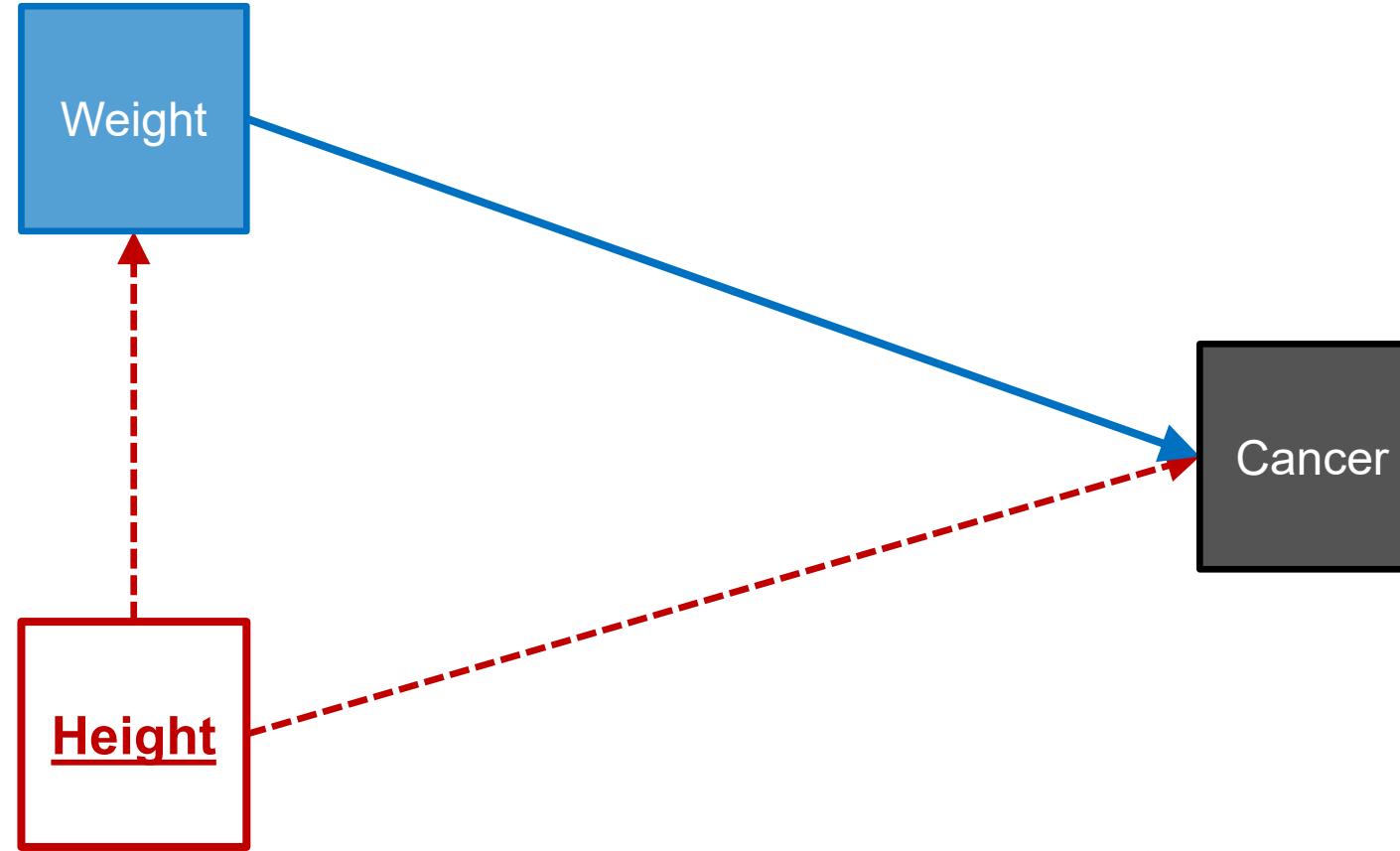


Wrong estimand:

- People typically interpret the ‘effect’ of **BMI** as the ‘effect’ of **adiposity**
- This is incorrect; the ‘effect’ of BMI is the **joint effect** of **weight** and **height²**
- It is not clear why we would want this, rather than **weight** conditional on **height**
- If **height** is a **confounder**, we should condition on it, not divide by it
- Dividing by **height²** will **not** remove the effect of height, it will simply transform it!

ESTIMAND CONFUSION

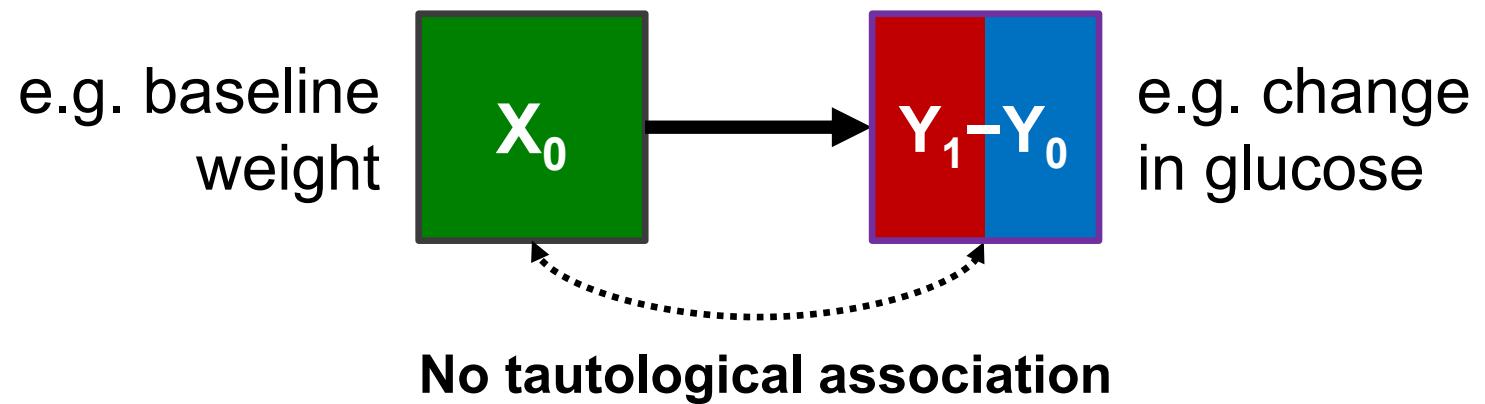
Why **BMI**? Why not **Weight** conditional on **Height**?



KEY EXAMPLE: CHANGE SCORES

Let's return to the analysis of **change scores**; this time considering when the **exposure (X)** and **outcome (Y)** do *not* share a **tautological association**?

- e.g. $[Y_1 - Y_0] = \beta_0 + \beta_1 X_0 + \dots + \varepsilon$



Most '**analyses of change**' involve this scenario: they seek to understand effect of a (baseline) exposure on 'change' in an outcome

EXPERIMENTAL DATA

- In randomised experimental data: **Change-score analyses** provide similar results to **follow-up adjusted for baseline analyses** – but are less efficient

Senn 2006¹²

STATISTICS IN MEDICINE
Statist. Med. 2006; 25:4334–4344
Published online 21 August 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2682



Change from baseline and analysis of covariance revisited

Stephen Senn*,†

Department of Statistics, 15 University Gardens, University of Glasgow, Glasgow G12 8QQ, U.K.

SUMMARY

The case for preferring analysis of covariance (ANCOVA) to the simple analysis of change scores (SACS) has often been made. Nevertheless, claims continue to be made that analysis of covariance is biased if the groups are not equal at baseline. If the required equality were in expectation only, this would permit the use of ANCOVA in randomized clinical trials but not in observational studies. The discussion is related to Lord's paradox. In this note, it is shown, however that it is not a necessary condition for groups to be equal at baseline, not even in expectation, for ANCOVA to provide unbiased estimates of treatment effects. It is also shown that although many situations can be envisaged where ANCOVA is biased it is very difficult to imagine circumstances under which SACS would then be unbiased and a causal interpretation could be made. Copyright © 2006 John Wiley & Sons, Ltd.

Van Breukelen 2006¹³



Journal of Clinical Epidemiology 59 (2006) 920–925

Journal of Clinical Epidemiology

ORIGINAL ARTICLES

ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies

Gerard J.P. Van Breukelen*

Department of Methodology & Statistics, Research Institute Caphri, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
Accepted 13 July 2005

Abstract

Background and Objective: For inferring a treatment effect from the difference between a treated and untreated group on a quantitative outcome measured before and after treatment, current methods are analysis of covariance (ANCOVA) of the outcome with the baseline as covariate, and analysis of variance (ANOVA) of change from baseline. This article compares both methods on power and bias, for randomized and nonrandomized studies.

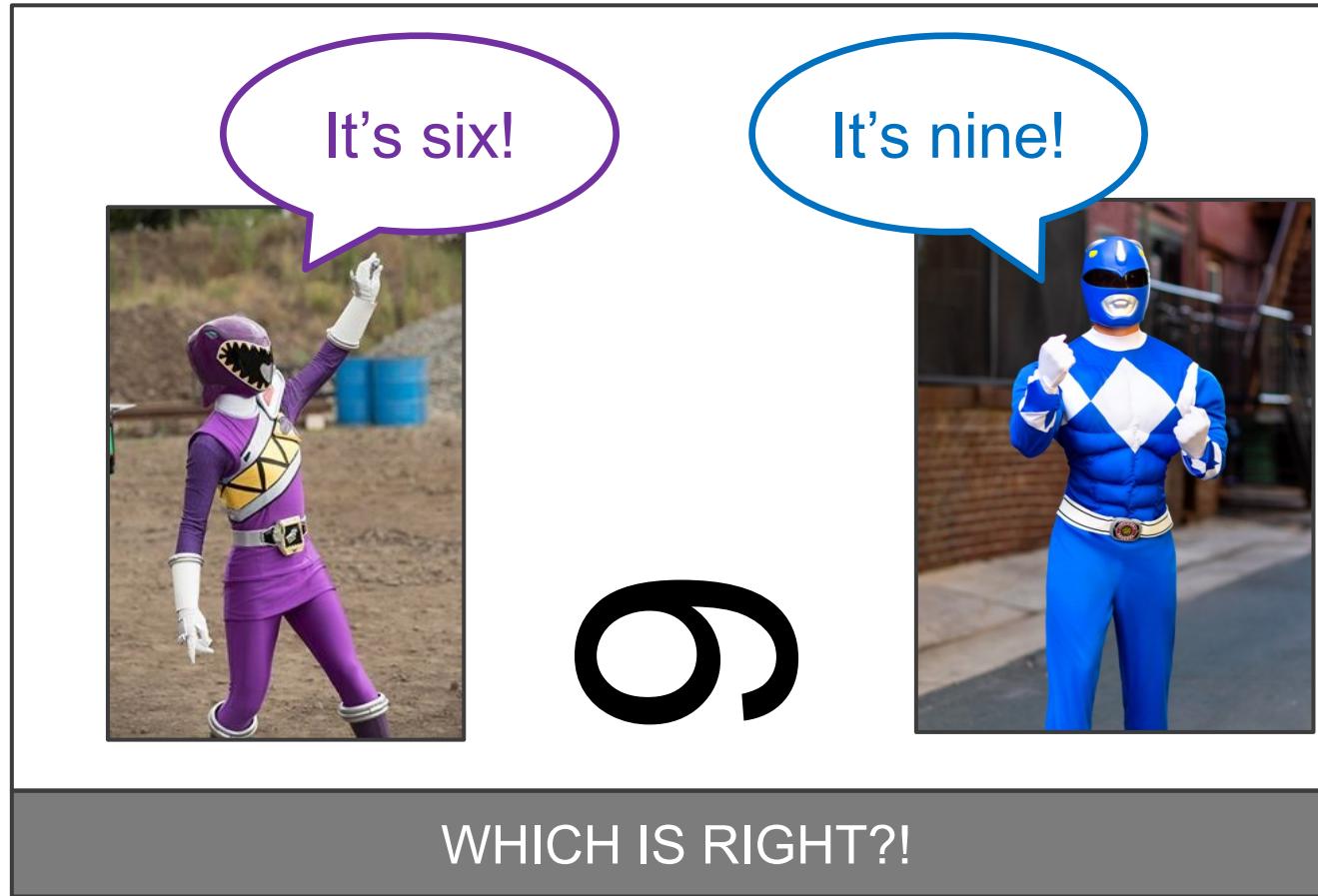
Methods: The methods are compared by writing both as a regression model and as a repeated measures model, and are applied to a nonrandomized study of preventing depression.

Results: In randomized studies both methods are unbiased, but ANCOVA has more power. If treatment assignment is based on the baseline, only ANCOVA is unbiased. In nonrandomized studies with preexisting groups differing at baseline, the two methods cannot both be unbiased, and may contradict each other. In the study of depression, ANCOVA suggests absence, but ANOVA of change suggests presence, of a treatment effect. The methods differ because ANCOVA assumes absence of a baseline difference.

Conclusion: In randomized studies and studies with treatment assignment depending on the baseline, ANCOVA must be used. In non-randomized studies of preexisting groups, ANOVA of change seems less biased than ANCOVA, but two control groups and two baseline measurements are recommended. © 2006 Elsevier Inc. All rights reserved.

NON-EXPERIMENTAL DATA

- In non-experimental data: **Change-score analyses** return different values (sometimes sign-opposed) to **follow-up adjusted for baseline analyses**



NON-EXPERIMENTAL DATA

- Advice on the right approach differs...

Van Breukelen 2006¹³



Journal of Clinical Epidemiology 59 (2006) 920–925

**Journal of
Clinical
Epidemiology**

ORIGINAL ARTICLES

ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies

Gerard J.P. Van Breukelen*

Department of Methodology & Statistics, Research Institute Caphri, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Accepted 13 July 2005

“In non-randomised trials **analyses of the change score** seems less biased than **analyses of follow-up**”

Glymour et al 2005¹⁴



American Journal of Epidemiology
Copyright © 2005 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved

Vol. 162, No. 3
Printed in U.S.A.
DOI: 10.1093/aje/kwi187

When Is Baseline Adjustment Useful in Analyses of Change? An Example with Education and Cognitive Change

M. Maria Glymour¹, Jennifer Weuve², Lisa F. Berkman¹, Ichiro Kawachi¹, and James M. Robins³

¹ Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA.

² Department of Environmental Health, Harvard School of Public Health, Boston, MA.

³ Department of Epidemiology, Harvard School of Public Health, Boston, MA.

“In some cases, **change-score analyses without baseline adjustment** provide unbiased... estimates when **baseline-adjusted estimates** are biased”

NON-EXPERIMENTAL DATA

Senn 2006¹²

STATISTICS IN MEDICINE
Statist. Med. 2006; 25:4334–4344
Published online 21 August 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2682



Change from baseline and analysis of covariance revisited

Stephen Senn*,†

“Although many situations can be envisaged where **analysis of follow-up** is biased it is very difficult to imagine circumstances under which... (**change-score analyses**) would then be unbiased”

Shahar & Shahar 2010¹⁵

Causal diagrams and change variables

Eyal Shahar MD MPH¹ and Doron J. Shahar²

¹Professor, Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

²Undergraduate student, Departments of Physics and Mathematics, College of Science, University of Arizona, Tucson, AZ, USA

Keywords

causal diagrams, change score, change variables, directed acyclic graphs

Correspondence

Dr Eyal Shahar
Division of Epidemiology and Biostatistics
Mel and Enid Zuckerman College of Public Health
The University of Arizona
1295 N. Martin Avenue
Tucson, AZ
USA
E-mail: shahar@email.arizona.edu

Abstract

Background The true change in the value of a variable between two time points is often assumed to be a cause or an effect of interest. To our knowledge, this assumption is based on intuition, rather than on any formal theoretical justification.

Methods We used causal directed acyclic graphs to explore the causal properties of a change variable, and critically examined competing structures.

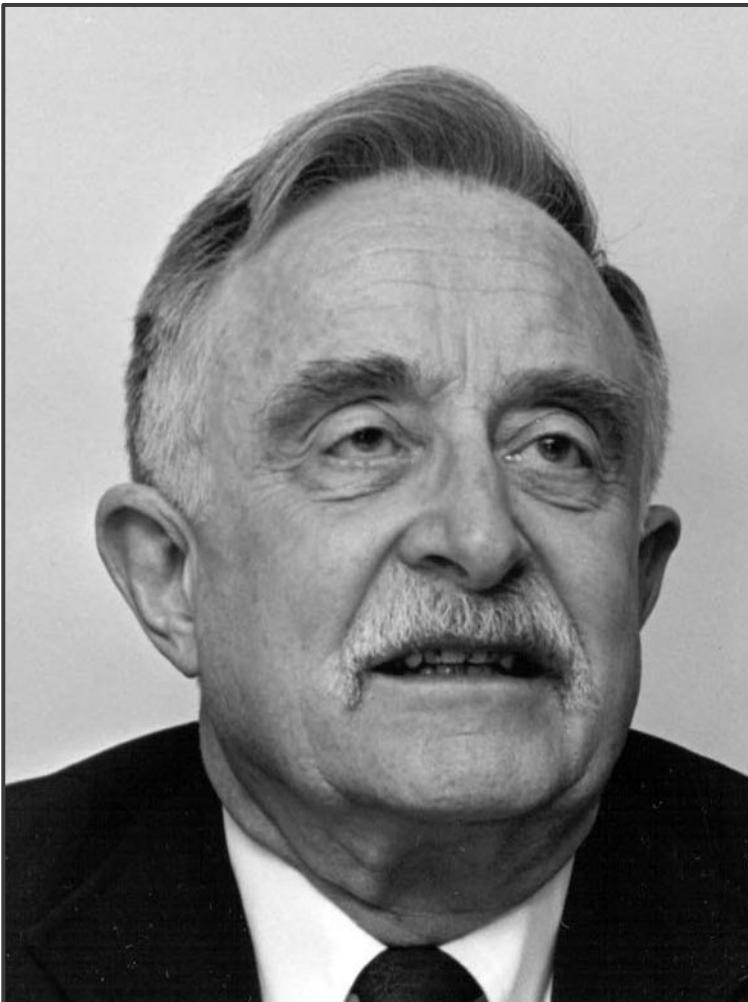
Results Based on the proposed causal structure, a change variable (true change) is no more than a derived variable. It does not cause anything and is not of causal interest.

Conclusions A true change is not a variable in the physical world. Therefore, modelling the change between two time points is justified only in a few situations.

“Modelling the **change between two time points** is justified only in a few situations”

LORD'S PARADOX

FREDERIC M LORD



LORD 1967¹

Psychological Bulletin
1967, Vol. 68, No. 5, 304-305

A PARADOX IN THE INTERPRETATION OF GROUP COMPARISONS

FREDERIC M. LORD

Educational Testing Service

Attention is called to a basic source of confusion in the interpretation of certain types of group comparison data.

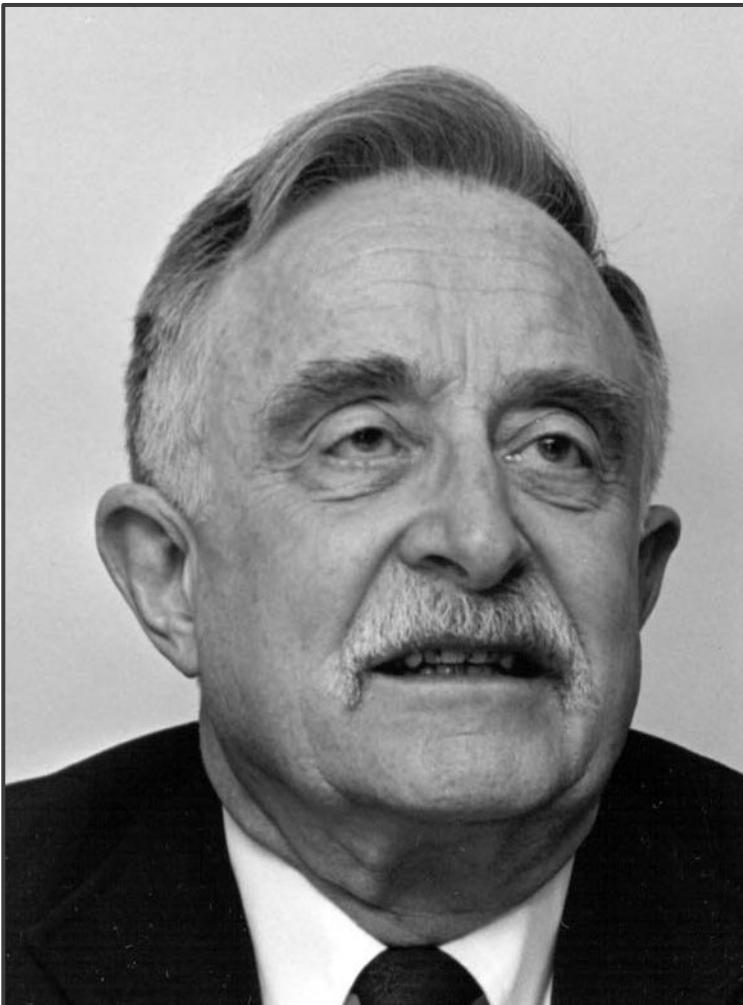
It is common practice in behavioral research, and in other areas, to apply the analysis of covariance in the investigation of preexisting natural groups. The research worker is usually interested in some criterion variable (y) and would like to make allowances for the fact that his groups are not matched on some important independent variable or "control" variable (x). The situation is such that observed differences in the

parison of the educational achievements of different racial groups. The research worker usually uses analysis of covariance regardless, or he may try to resort to a simple and direct interpretation of group means.

The present note points out a type of problem that arises in interpreting data on preexisting groups. The difficulty can most easily be pointed out with the help of a hypothetical illustrative example.

LORD'S PARADOX

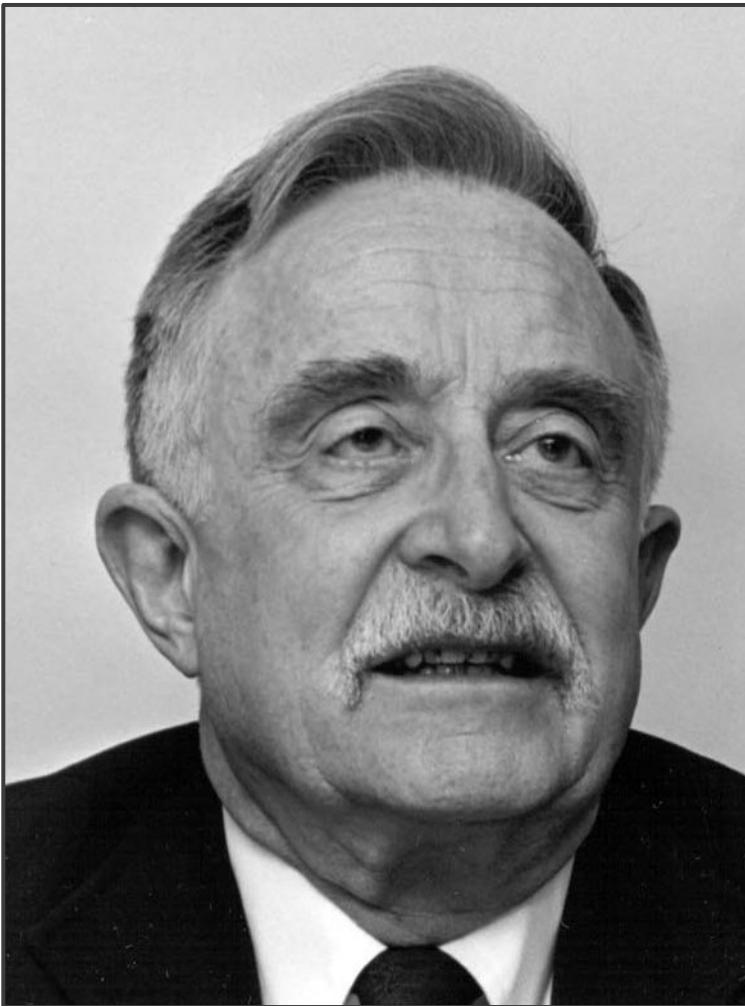
FREDERIC M LORD



- “A large university is interested in investigating the effects... of the **diet** provided in the university dining halls and any **sex difference** in these effects.”
- “The **weight** of each student at the **time of his [her/their] arrival in September** and... the **following June** are recorded”

LORD'S PARADOX

FREDERIC M LORD



- “At the end of the school year, the data are independently examined by two statisticians”

STATISTICIAN 1



STATISTICIAN 2



LORD'S PARADOX

STATISTICIAN 1

- Analysis of **change-scores**
- Difference-in-difference

$$[Y_1 - Y_0] = \alpha_0 + \alpha_1 X$$



STATISTICIAN 2

- Analysis of **follow-up adjusted for baseline**
- ANCOVA

$$Y_1 = \beta_0 + \beta_1 X + \beta_2 Y_0$$



LORD'S PARADOX



STATISTICIAN 1

“There is no evidence of... change during the year for either boys or girls, and... no evidence of a differential change between the sexes”



STATISTICIAN 2

“Wherever boys and girls start with the same...[baseline]...weight, it is... obvious... that the subgroup of boys gains more than the subgroup of girls.”

“The college dietitian is having some difficulty reconciling the conclusions of the two statisticians”

EXPLORATIONS & EXPLANATIONS

O'CONNOR 1973²

UNRAVELING LORD'S PARADOX: THE APPROPRIATE USE OF MULTIPLE REGRESSION ANALYSIS IN QUASI-EXPERIMENTAL RESEARCH

Edward F. O'Connor, Jr.

LOCASCIO & CORDRAY 1983³

A RE-ANALYSIS OF "LORD'S PARADOX"¹

JOSEPH J. LOCASCIO AND DAVID S. CORDRAY

Northwestern University

HOLLAND AND RUBIN 1986⁴

RESEARCH DESIGNS AND CAUSAL INFERENCES: ON LORD'S PARADOX*

PAUL W. HOLLAND AND DONALD B. RUBIN

WAINER 1991⁵

Adjusting for Differential Base Rates: Lord's Paradox Again

Howard Wainer
Educational Testing Service, Princeton, New Jersey
and Princeton University

LUND 1999⁶

Lord's Paradox Re-examined

THORLEIF LUND
*Department of Special Education, University of Oslo, PO Box 1140 Blindern,
N-0317 Oslo 3, Norway*

PEARL 2016⁷

Judea Pearl*
Lord's Paradox Revisited – (Oh Lord! Kumbaya!)

DOI 10.1515/jci-2016-0021

XIAO 2019⁸

An Empirical Unraveling of Lord's Paradox

ZhiMin Xiao ^a, Steve Higgins^a, and Adetayo Kasim^b

^aSchool of Education, Durham University, Durham, UK; ^bDurham University, Stockton-on-Tees, UK

WRIGHT 2019⁹

Allocation to groups: Examples of Lord's paradox

Daniel B. Wright*

University of Nevada, Las Vegas, Nevada, USA

WIJAYATUNGA 2020¹⁰

Resolving the Lord's Paradox

Priyantha Wijayatunga¹

¹ Department of Statistics, Umeå University, Umeå SE-90187, Sweden

EXPLORATIONS & EXPLANATIONS

- **Locascio and Cordray 1983³:** “[METHOD 1] provided a correct assessment of treatment effects”
- **London and Wright 2001¹¹:** “Usually [METHOD 2] will be preferred”
- **Holland and Rubin 1986⁴:** “If both statisticians made only descriptive statements, they would both be correct... (for) causal statements, neither would be correct or incorrect because (of) untestable assumptions”
- **Pearl 2016⁷:** “Both statisticians were...correct...each estimated a different effect”
- **Xiao 2019⁸:** “Multilevel modeling can ameliorate the divergence in sign”

ESTIMAND CONFUSION

- **Change-scores analyses** give us the **wrong estimand!**
 - They don't capture **total causal effect** of X_0 on Y_1



Total causal effect
of X_0 on Y_1

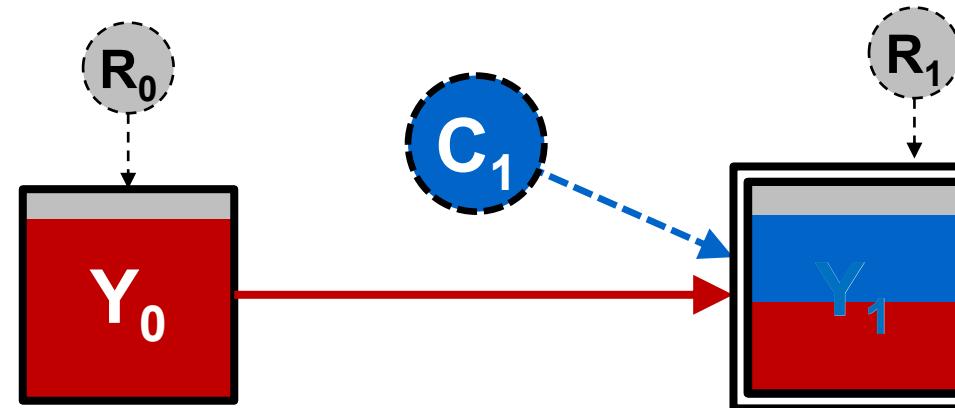


Change-score
Estimand?

- Instead they provide a *conflated* mix of effects on their parent components (Y_0 and Y_1)
- Perhaps they estimate another – more useful estimand?
 - e.g. the causal effect of the exposure on ‘change’ in the outcome?

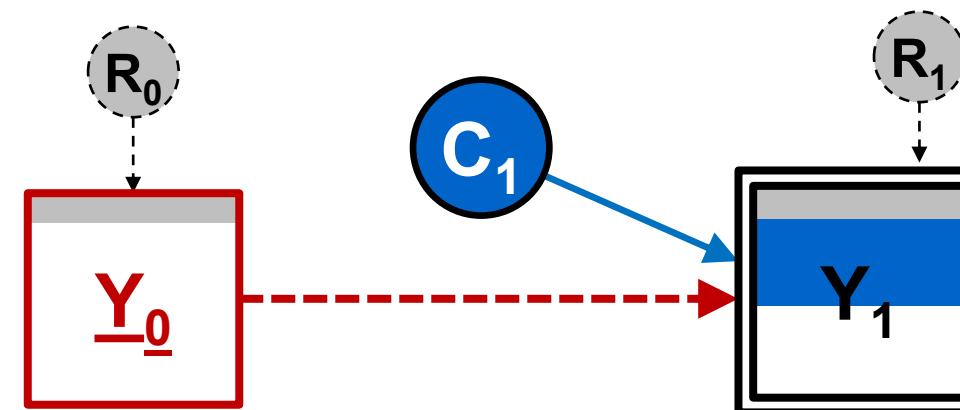
EXOGENOUS CHANGE

- The future values of a variable (Y_1) are determined by:
 - **Past values** of that variable (Y_0), i.e. '**determined change**' ($f\{Y_0\}$)
 - **Random variation** (R_1), i.e. '**random change**' ($R_1 - R_0$)
 - All ***non-random determinants of change***, i.e. '**exogenous change**' (C_1)
- C_1 is the element of Y_1 **NOT** determined by Y_0



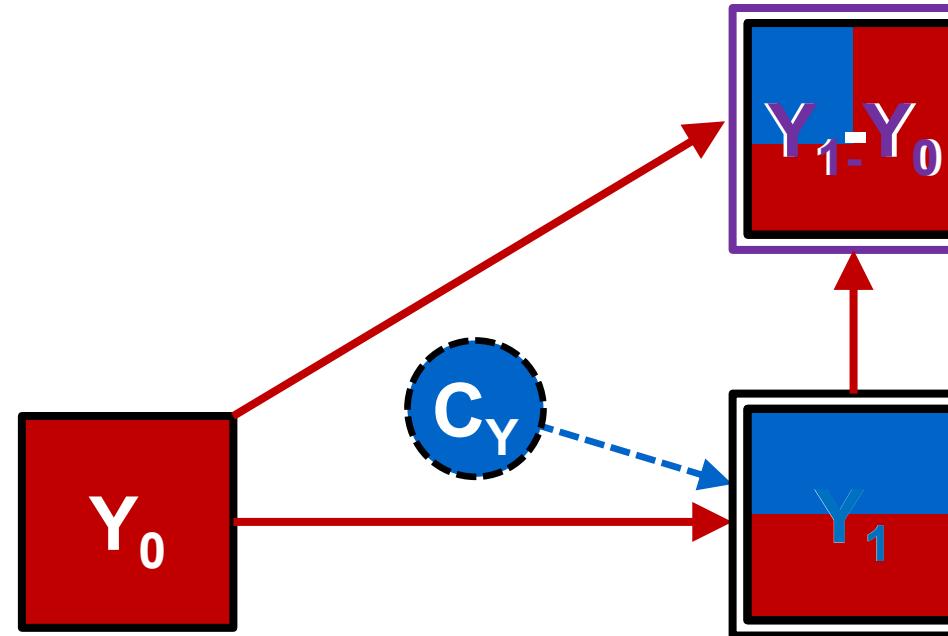
EXOGENOUS CHANGE

- The future values of a variable (Y_1) are determined by:
 - **Past values** of that variable (Y_0), i.e. '**determined change**' ($f\{Y_0\}$)
 - **Random variation** (R_1), i.e. '**random change**' ($R_1 - R_0$)
 - All ***non-random determinants of change***, i.e. '**exogenous change**' (C_1)
- C_1 is the element of Y_1 **NOT** determined by Y_0
- An estimate of C_1 can be obtained from Y_1 conditioning on Y_0



CHANGE-SCORES ARE JUNK!

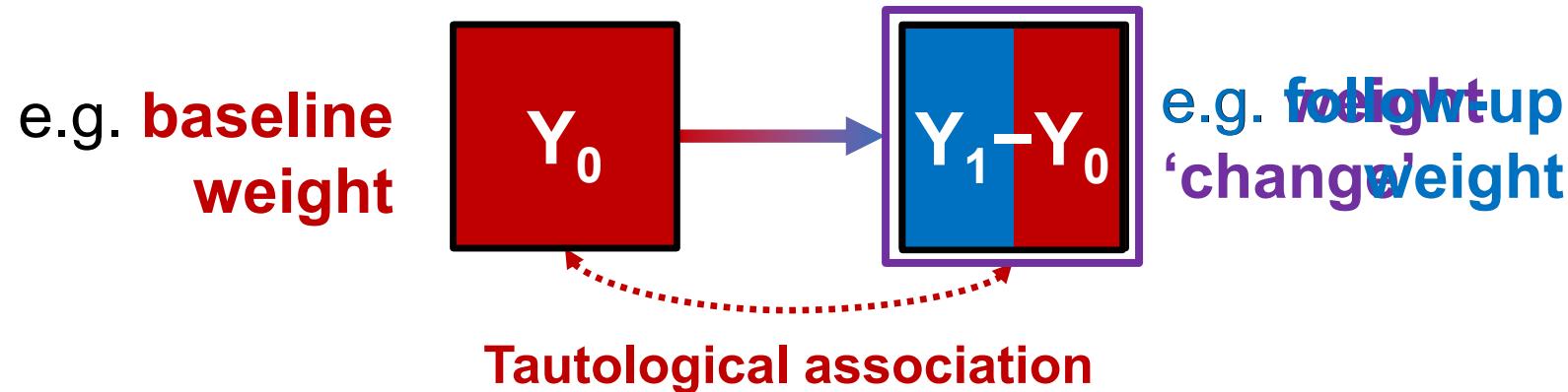
- **Change-scores** don't capture **exogenous change**; they are a misleading conflation of information from **BOTH** determining parents



- They can be **very** misleading because the **baseline** parent variable has been transformed with a **NEGATIVE** sign

TAUTOLOGICAL ASSOCIATIONS

- This is most obvious when the **change-score** is analysed in direct relation to either parent (e.g. ‘*how does **baseline** determine **change**?*’)

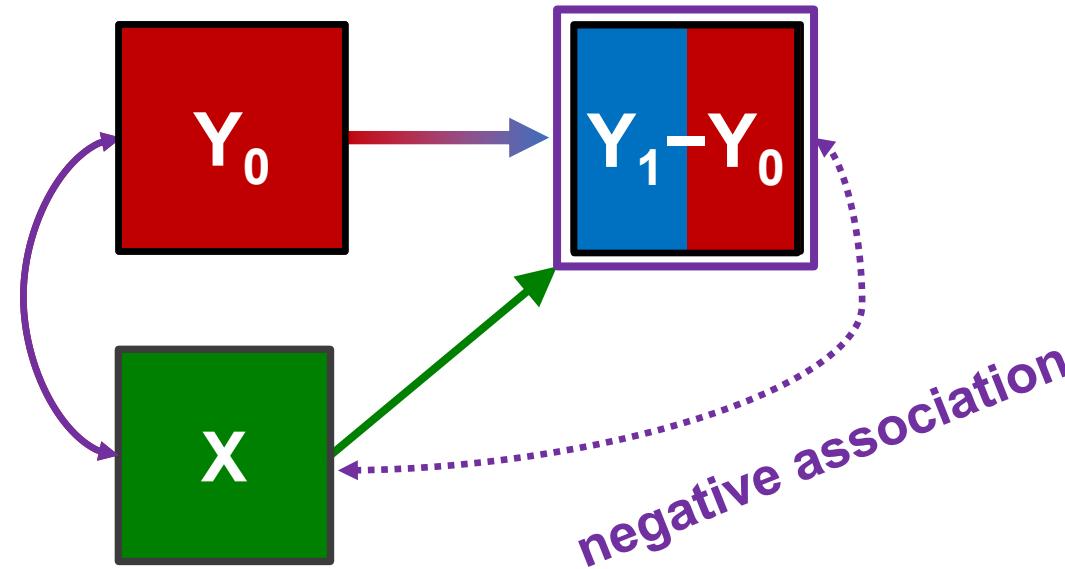


$$[Y_1 - Y_0] = \beta_0 + \beta_1 Y_0 + \dots + \varepsilon$$

- Such analyses produce strong ‘tautological associations’
- ‘**Law of Initial Value**’ = Largely a consequence of the tautological association between Y_0 (baseline) and $-Y_0$ component of change-score

COMPOSITE VARIABLE BIAS

- Problems are less visible when the **exposure** is not part of the **change-score**



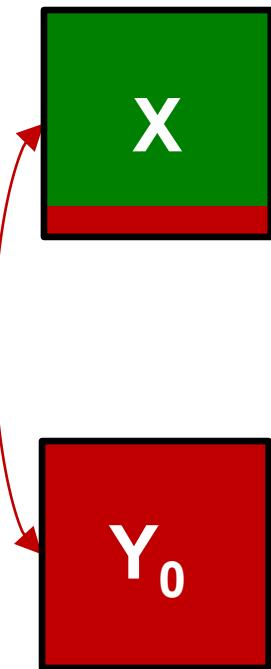
$$[Y_1 - Y_0] = \beta_0 + \beta_1 X + \dots + \varepsilon$$

- Any correlation between **exposure** (X) and **baseline outcome** (Y_0)...
- ...will produce a (strong) **negative association** with **change-score**

COMPOSITE VARIABLE BIAS

- Disagreement between **change-score analysis** and **analyses of follow-up adjusted for baseline** depends on relationship between X and Y_0

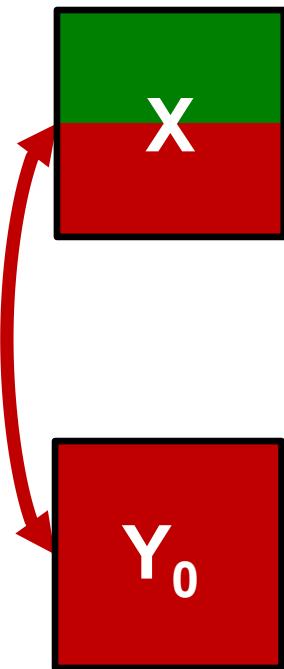
- When association between X and Y_0 small:
 - Association between X and $Y_1 - Y_0$ converges on association between X and C_1



COMPOSITE VARIABLE BIAS

- Disagreement between **change-score analysis** and **analyses of follow-up adjusted for baseline** depends on relationship between \mathbf{X} and \mathbf{Y}_0

- When association between \mathbf{X} and \mathbf{Y}_0 small:
 - Association between \mathbf{X} and $[\mathbf{Y}_1 - \mathbf{Y}_0]$ converges on association between \mathbf{X} and \mathbf{C}_Y
- As association between \mathbf{X} and \mathbf{Y}_0 increases:
 - Association between \mathbf{X} and $[\mathbf{Y}_1 - \mathbf{Y}_0]$ increasingly dominated by $-\mathbf{Y}_0$



A PARADOX UNCOVERED

STATISTICIAN 2

- **ESTIMATE:** Direct causal effect
 - Diluted by random variation



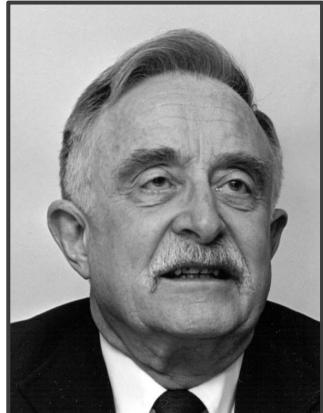
STATISTICIAN 1

- **ESTIMATE:** Obscure
 - May be extremely misleading



A DEFEATED LORD

LORD



“There simply is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled preexisting differences between groups...”

PARTLY TRUE: You cannot account for baseline differences between groups if you only have summary information

BUT: With individual-level data, you can obtain (diluted) effect estimates if you have good ***overlap*** between exposure levels and can explain the ***reasons*** for the differences

HOMER



RECOMMENDED READING

Composite variable bias

- Shahar E. The association of body mass index with health outcomes: causal, inconsistent, or confounded?. *American journal of epidemiology*. 2009 Oct 15;170(8):957-8.
- Shahar, E., 2010. Metabolic syndrome? A critical look from the viewpoints of causal diagrams and statistics. *Journal of Cardiovascular Medicine*, 11(10), pp.772-779.
- Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Current epidemiology reports*. 2016 Mar;3(1):63-71.

Change scores

- Tennant, P.W.G, Arnold, K.F., Ellison, G. and Gilthorpe, M.S., 2021. Analyses of 'change scores' do not estimate causal effects in observational data. *International journal of epidemiology*.
- Shahar, E. and Shahar, D.J., 2012. Causal diagrams and change variables. *Journal of evaluation in clinical practice*, 18(1), pp.143-148.

SUMMARY

- Composite variables are poorly suited to causal effect estimation
- By collapsing information from the multiple parent variables into a single node, their causal effects and variances get conflated
- This violates consistency; as it is no longer possible to attribute causal effects to the specific components therein
- If the parents occur at different times, it may also prevent exchangeability, as the different components may require different adjustment sets

Solution: Think!

- Unless you are deliberately summarising several contemporaneous variables and *understand* what will happen to their causal effects and variances, you will be better focussing on analysing individual parent variables