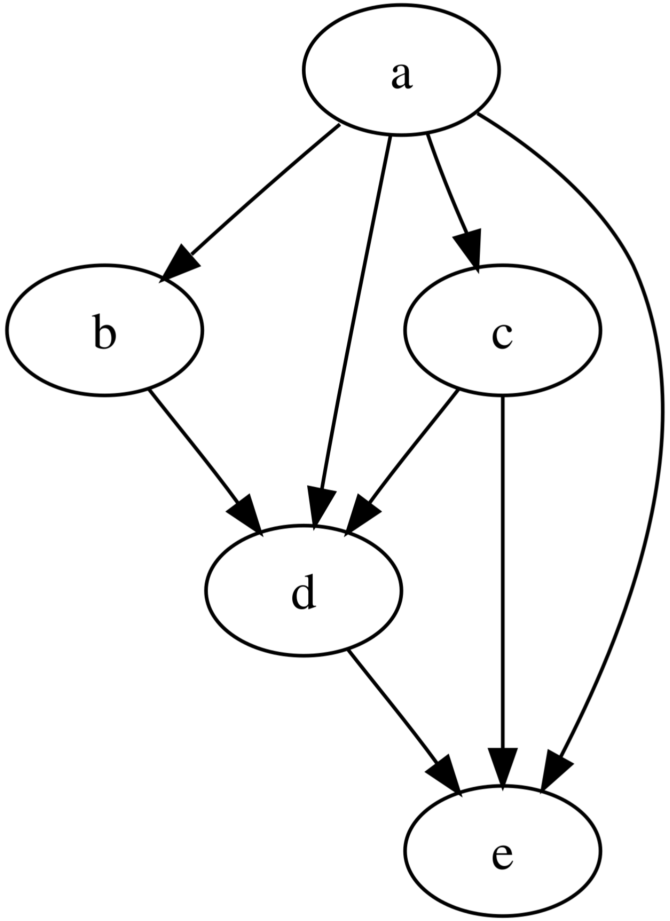
Causal graphs as a simple yet effective reasoning tool

Julia M. Rohrer
Leipzig University
www.juliarohrer.com
@dingdingpeng.the100.ci
www.the100.ci

Slides: Resources

https://slate.com/technology/2017/04/heres-why-people-saw-the-dress-differently.html

Causal Graphs



» As a means to visually represent and explicate our model of how the data were generated

» As a tool to reason more clearly about the implications of the data, given our model

Directed Acyclic Graphs

- » Developed from Bayesian networks
- » Comes with the axiomatic system of do-calculus
- » Equivalent with the Potential Outcomes notation (?) [(Galles & Pearl, 1998)]

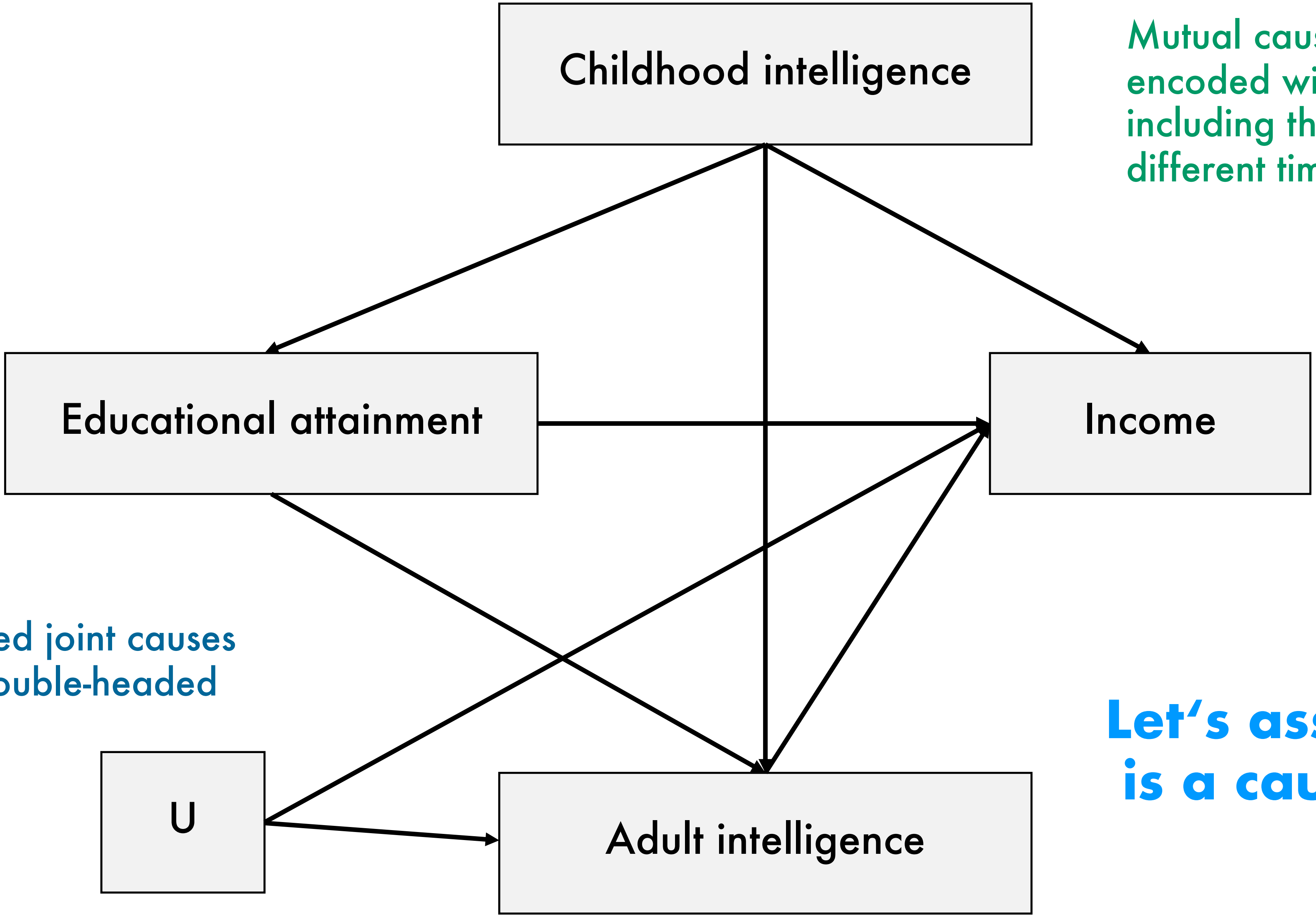
Judea Pearl

Causal inference issues

Identification	**Estimation**
Given your assumptions, if your sample was infinitely large—would you be able to estimate the causal effect of interest?	Actually estimating the effect with the data you got + your assumptions

Elwert (2013)


Directed Acyclic Graphs



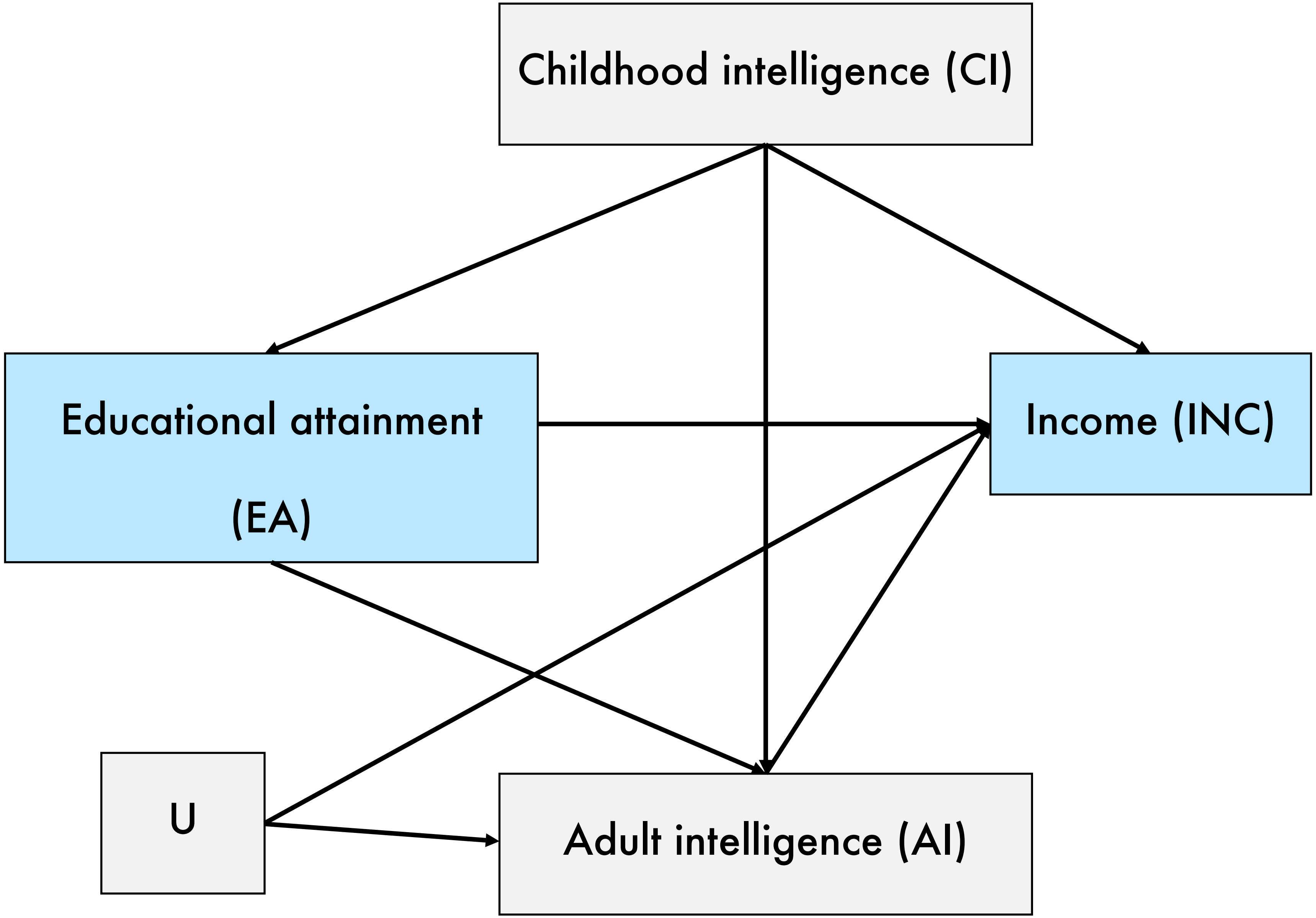
Mutual causation can be encoded without cycles by including the same variable at different time points

Unobserved joint causes replace double-headed arrows

Let's assume this is a causal DAG

Rohrer (2018)

Backdoor Criterion



EA -> INC
EA <- CI -> INC

EA <- CI -> AI -> INC
EA <- CI -> AI <- U -> INC

EA -> AI -> CI -> INC
EA -> AI -> INC
EA -> AI <- U -> INC

Which of these paths lead to associations between EA and INC?

Are these associations causal or non-causal?

→ keep the causal ones, block the non-causal ones

3 Fundamental structures

- Chains: X → M → Y
  - transmits a causal association from X to Y
  - conditioning on (control for) M blocks transmission

Elwert (2013), Rohrer (2018)

What does it mean to condition on a variable?

- Any means of introducing information about the variable into the analysis (using it as a control variable, a covariate…)
- For example
  - Statistical adjustment in an ANCOVA, regression, SEM…
  - Using the variable for matching, weighting
  - Stratification by the variable
  - Only including participants with a certain level of a variable

3 Fundamental structures

- Chains: X → M → Y
  - transmits a causal association from X to Y
  - conditioning on (control for) M blocks transmission
    - if you're interested in the (total) effect of X on Y, you usually don't want to do that, because the chain is part of the causal effect of interest → control for mediator leads to overcontrol bias
  - aka mediation

Elwert (2013), Rohrer (2018)

3 Fundamental structures

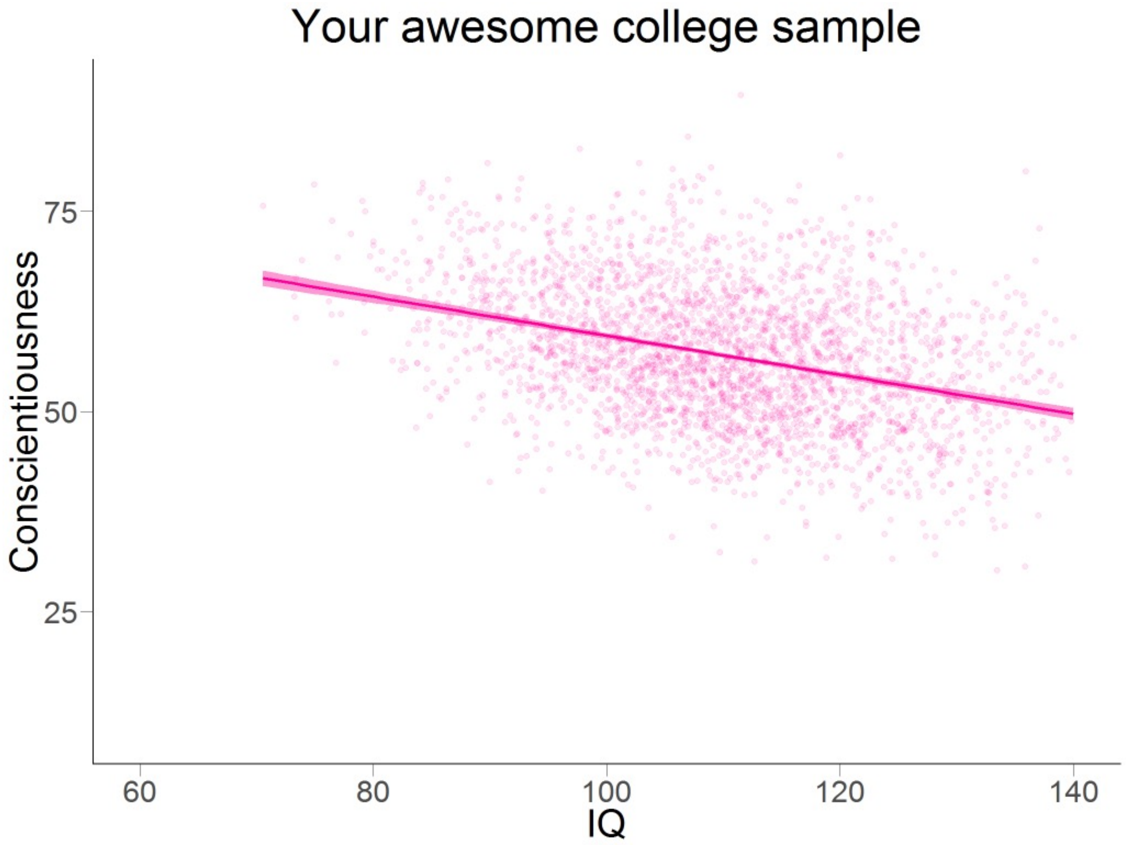
- Forks: X ← C → Y
  - induces a non-causal association between X and Y
  - conditioning on C removes the non-causal association
    - you usually want to get rid of these, so conditioning on confounders is usually the way to go
  - aka confounding

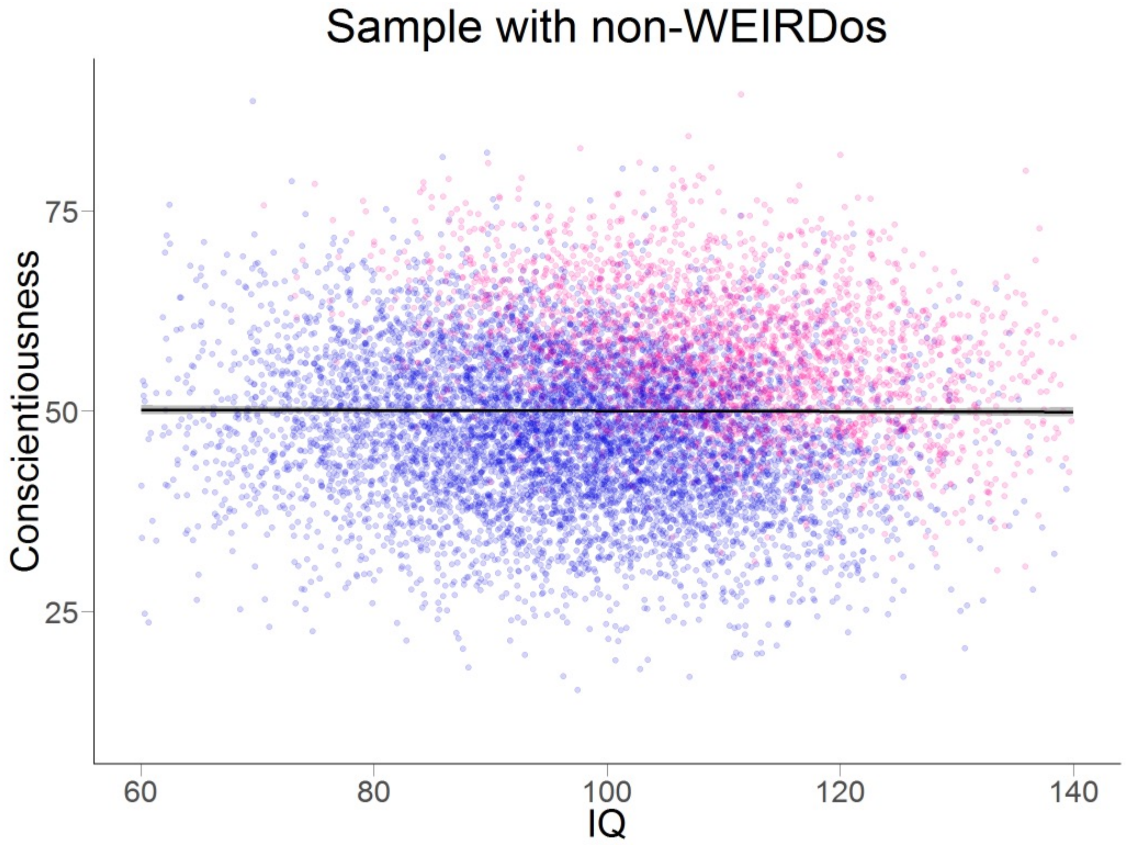
Elwert (2013), Rohrer (2018)

Inverted forks X → Z ← Y

- does not transmit any association
  - e.g., travel time → attendance of this symposium ← interest in the presented topics
- conditioning on Z *opens* transmission of a non-causal association
  - e.g., conditioning on attendance
  - among the people who attend this symposium, those who had to travel particularly far are probably especially interested in the presented topics
- in other words, third-variable control *introduces* a new spurious association
- aka collider bias

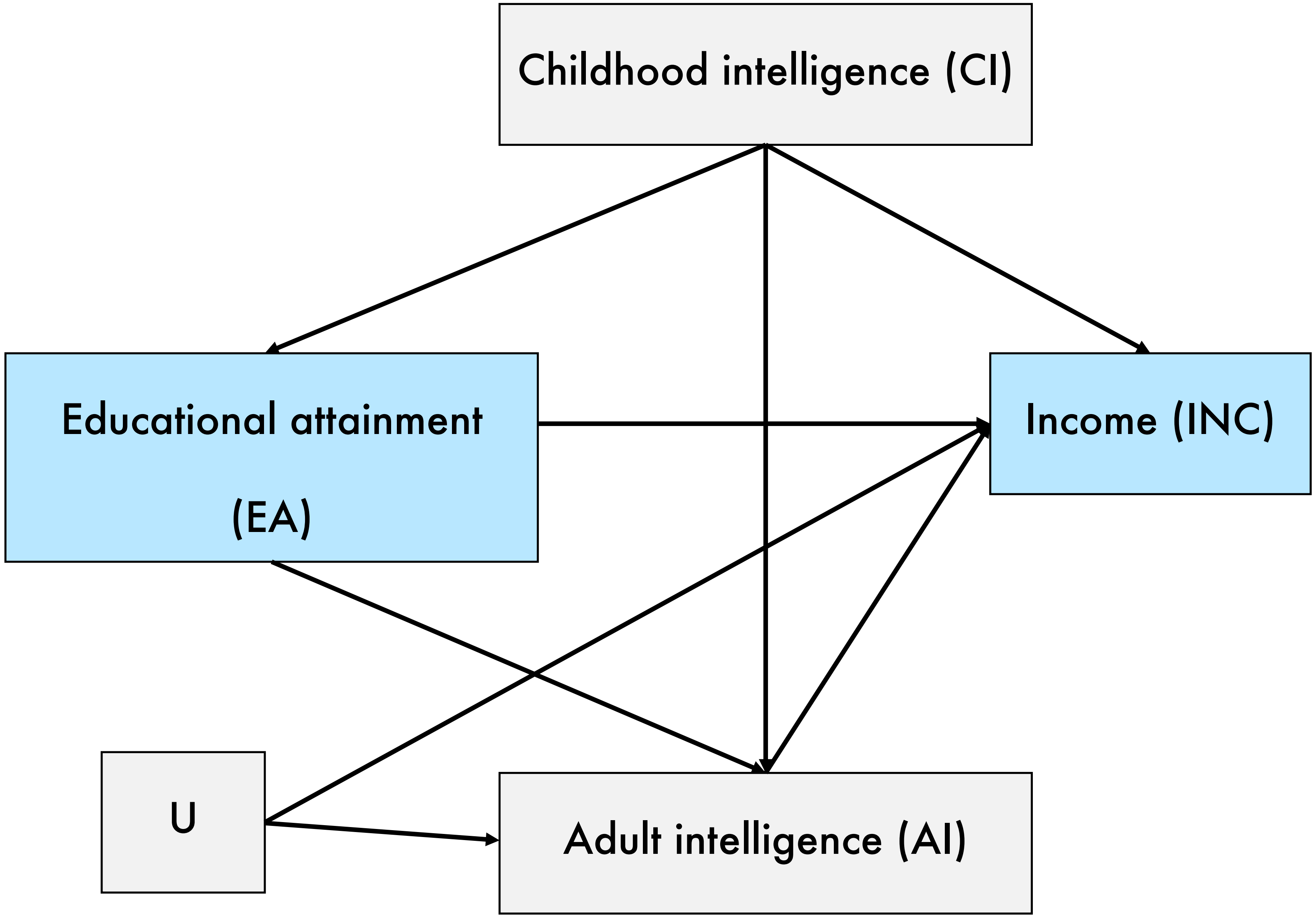
Conscientiousness → College ← IQ





[Rohrer (2017)]

	Association between X and Y when not conditioning on anything	Association between X and Y when conditioning on the variable in the middle
Chain: X → Mediator → Y	**Causal association**	No association (overcontrol bias)
Fork: X ← Confounder → Y	Non-causal association	**No association**
Inverted fork: X → Collider ← Y	**No association**	Non-causal association (collider bias)



EA -> INC
EA <- CI -> INC

EA <- CI -> AI -> INC
EA <- CI -> AI <- U -> INC

EA -> AI -> INC
EA -> AI <- U -> INC

Paths

» Paths that transmit non-causal associations and need to be blocked (backdoor paths)

» EA <- CI -> INC

» EA <- CI -> AI -> INC

Mediator passes on the non-causal association

Conditioning on the mediator AI would close this one backdoor path (but there are other reasons why we wouldn't want to do that)

Confounder that introduces the non-causal association

Conditioning on the confounder closes both backdoor paths

Paths

- Paths that transmit causal associations and should not be blocked
  - EA -> INC
  - EA -> AI -> INC

Arrows all flow from cause to outcome – all fine, don't control away the good stuff

Paths

- Paths that are blocked thanks to a collider…
- …but would lead to non-causal associations *if* the collider was conditioned on
  - EA -> AI <- CI -> INC
  - EA <- CI -> AI <- U -> INC
  - EA -> AI <- U -> INC

Two arrows pointing into node → collider → this path doesn't do anything, you're all good
UNLESS you condition on the collider

Paths

» What happens if you conditioning on the collider?

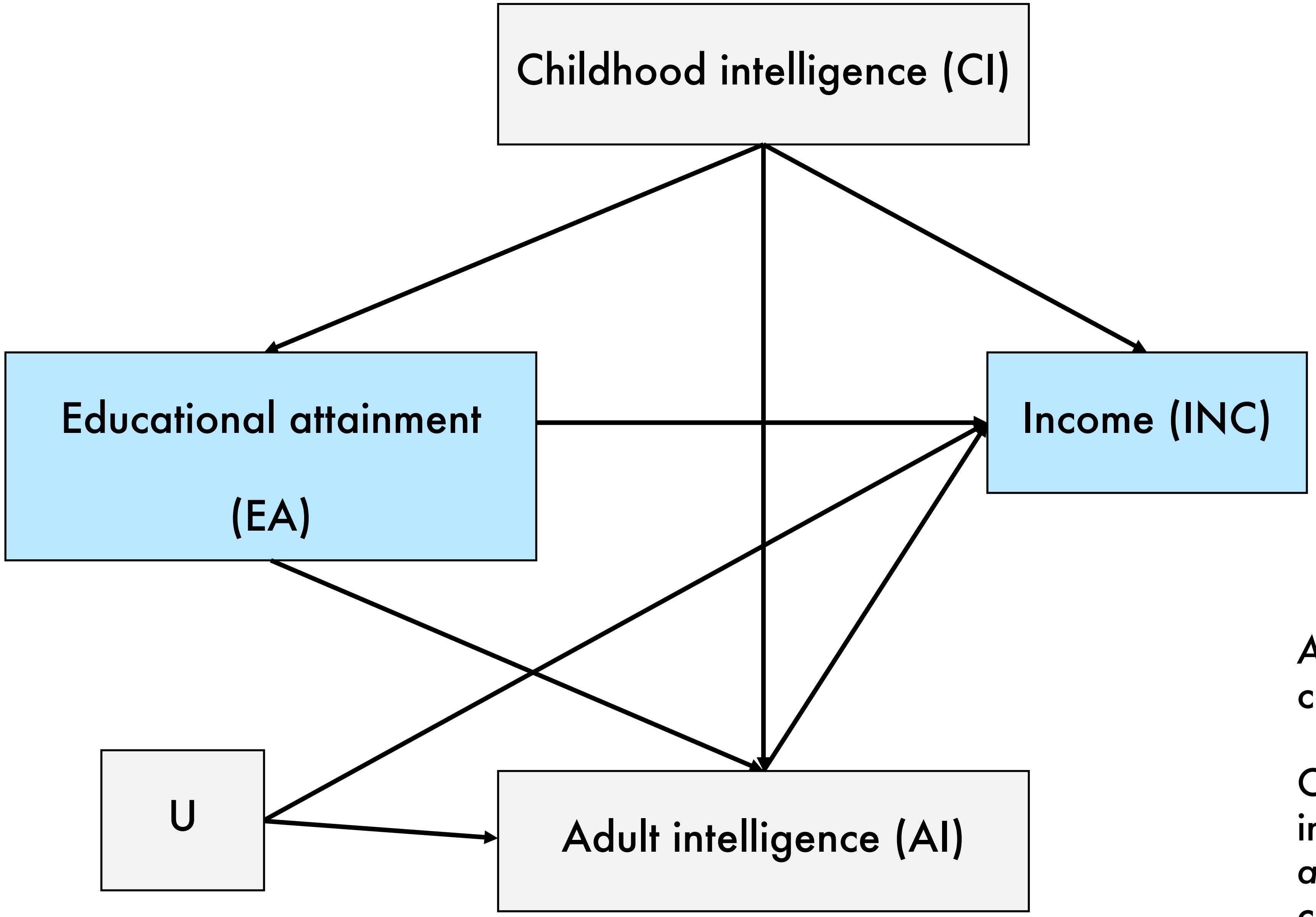
EA -> AI <- CI -> INC

EA <- CI -> AI <- U -> INC

New backdoor paths that could be fixed by conditioning on CI

EA -> AI <- U -> INC

New backdoor paths that can only be fixed if U is measured and conditioning on



Assuming that this is the correct causal structure:

Conditioning on childhood intelligence is both sufficient *and* necessary to identify the causal effect of interest

Causal estimation

- different ways to condition on a variable, for example
  - stratification, sub-group analysis
  - regression adjustment
  - weighting & matching approaches
- if your causal identification strategy was wrong, no amount of estimation can rescue you
- if your causal identification strategy was right, things can still go wrong
  - Insufficient control due to misspecified models, measurement error in covariates...

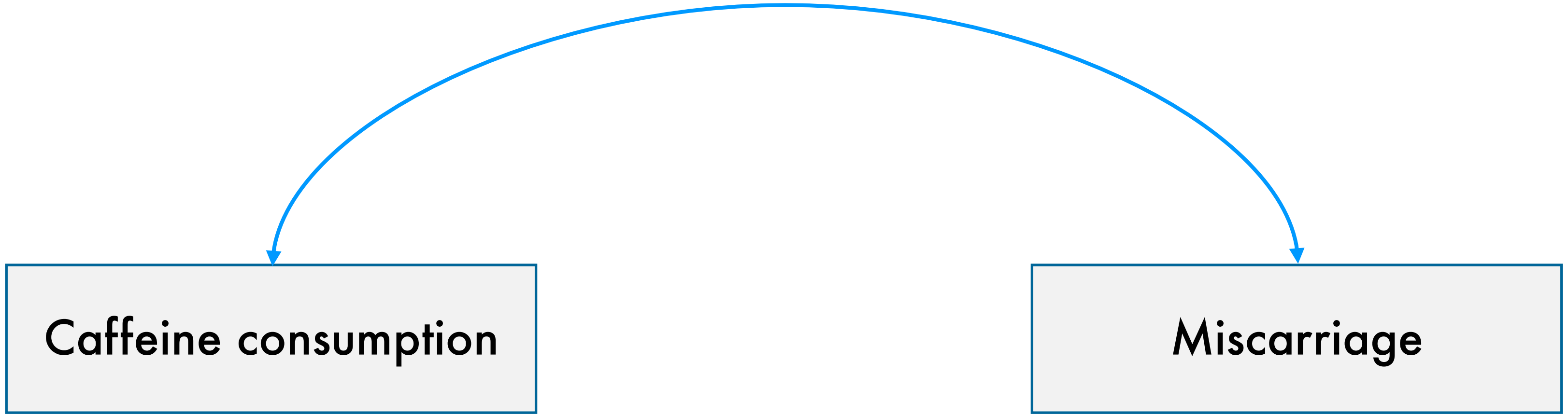
Three examples


COFFEE...OK DURING PREGNANCY?
Too much coffee and caffeine while pregnant or trying to conceive can affect birthweight and time to conceive. Limit to <100 mg caffeine daily.
WATCH YOUR PORTION SIZE
Using a smaller cup can make a huge difference to caffeine content.
ADD MILK TO CUT THE CAFFEINE
MILK
It also pumps up your calcium, protein and vitamin D.
TRY TEA INSTEAD
Less than half the caffeine of coffee. Plus it has plenty of healthy antioxidants.
DON'T SUB SODA
Regular & diet soda may cause problems with fertility & pregnancy.
TRY DECAF
Gradually switch to decaf or half the caffeine coffee.
EATRIGHTMAMA.COM/COFFEE_PREGNANT


Instant coffee
100mg in a mug
Filter coffee
140mg in a mug
Tea
75mg in a mug
Cola
COLA
40mg in a mug
Energy drink
80mg in a 250ml can
Herbal tea
No more than 4 cups a day
Plain chocolate
25mg in a 50g bar
Mother&Baby

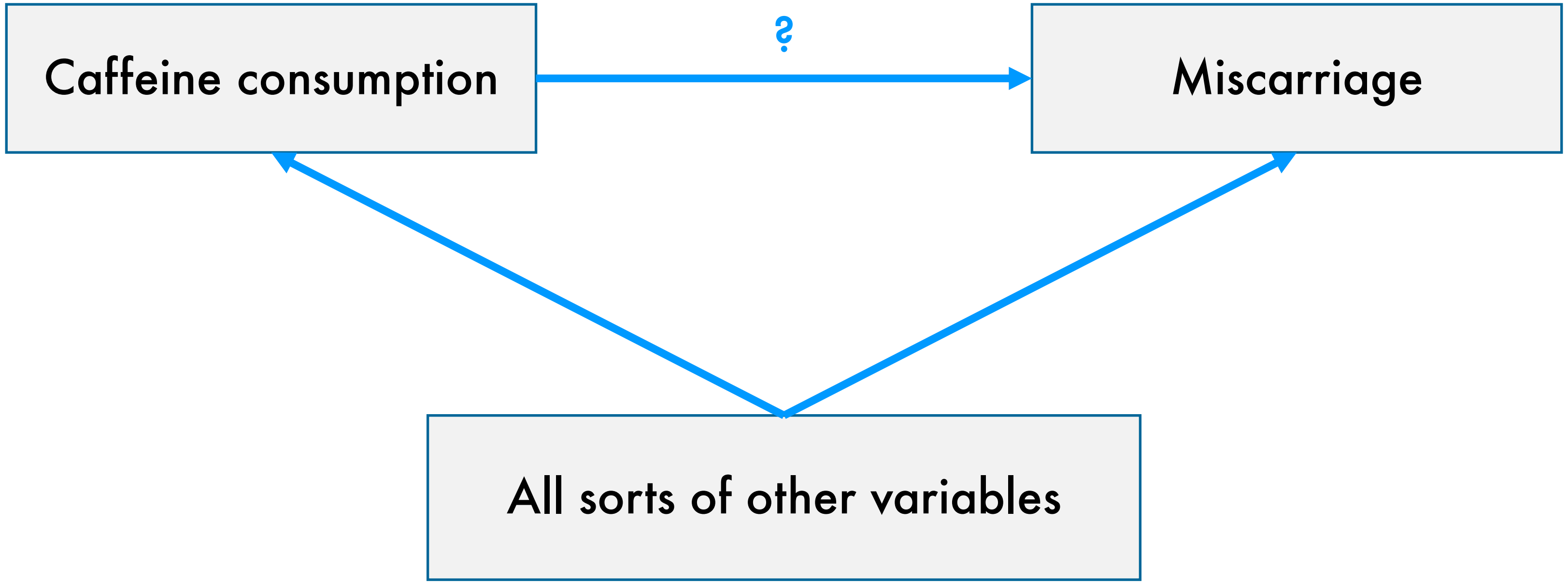
Does caffeine cause miscarriages?

(from Emily Oster‘s „Expecting Better“)



Does caffeine cause miscarriages?

(from Emily Oster‘s „Expecting Better“)



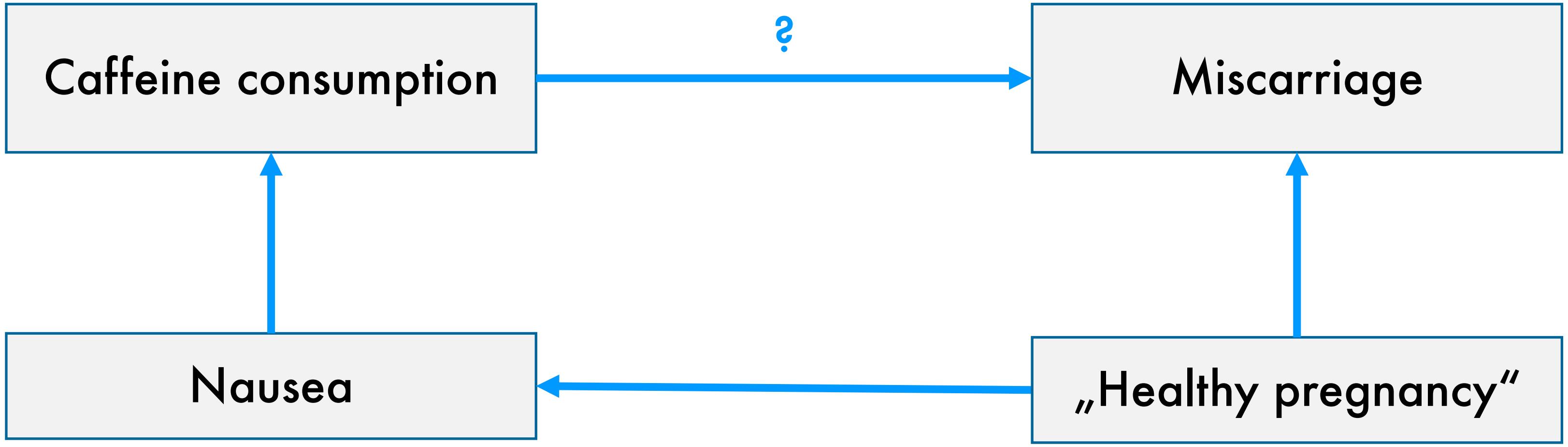
Does caffeine cause miscarriages?

(from Emily Oster‘s „Expecting Better“)

- Most miscarriages happen in the first trimester
- Morning sickness is also quite common in the first trimester
- Morning sickness makes coffee a lot less enjoyable
- Morning sickness does seem to be an indicator of a *healthy* pregnancy
  - One idea: evolved protective mechanism triggered by hormones

Does caffeine cause miscarriages?

(from Emily Oster's „Expecting Better")



Miscarriage ← Healthy pregnancy → Nausea → Caffeine consumption

Path that induces a non-causal association between caffeine consumption and miscarriages

To identify the causal effect of interest, we have to statistically adjust for either „healthy pregnancy" or nausea

Confusing stats

Today in the UK and the US infant mortality rates are well below 1%, so to speak of an average (or mean) age of death makes perfect sense. Those numbers really do tell us how long, on average, the people in that community live. For hunter-gatherer communities, we need a different mathematical model because, without healthcare, infant mortality rates could be in the region of 20-40%.

If nearly half a group dies in infancy and the other half live well into their early fifties that yields a mean life expectancy of 25: misleading and not in the least helpful for telling us about how long people live for.

In hunter-gatherer groups, life was, and is, undeniably hard, but their lifespan was not as short as the numbers press us to think. If you were a hunter-gatherer and you made it to adolescence, there was a strong likelihood that you would live a long and healthy life – not so different from modern humans.

Search analysis, research, academics…

Hunter-gatherers live nearly as long as we do but with limited access to healthcare

Published: October 31, 2018 1.57pm CET

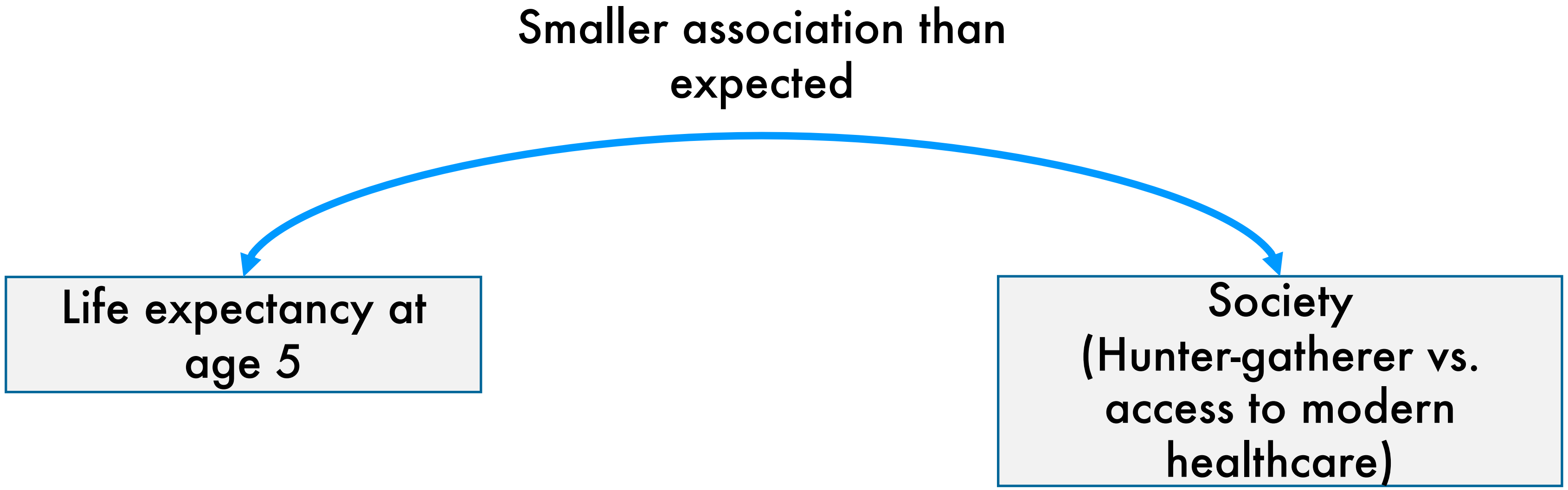
erichon/Shutterstock

Modern life has many benefits. Transport, comfy furniture, smartphones, TV, the internet, dentistry and advanced medicine would be at the top of most people's lists. Our bodies also show signs of responding positively to modern life. In almost every

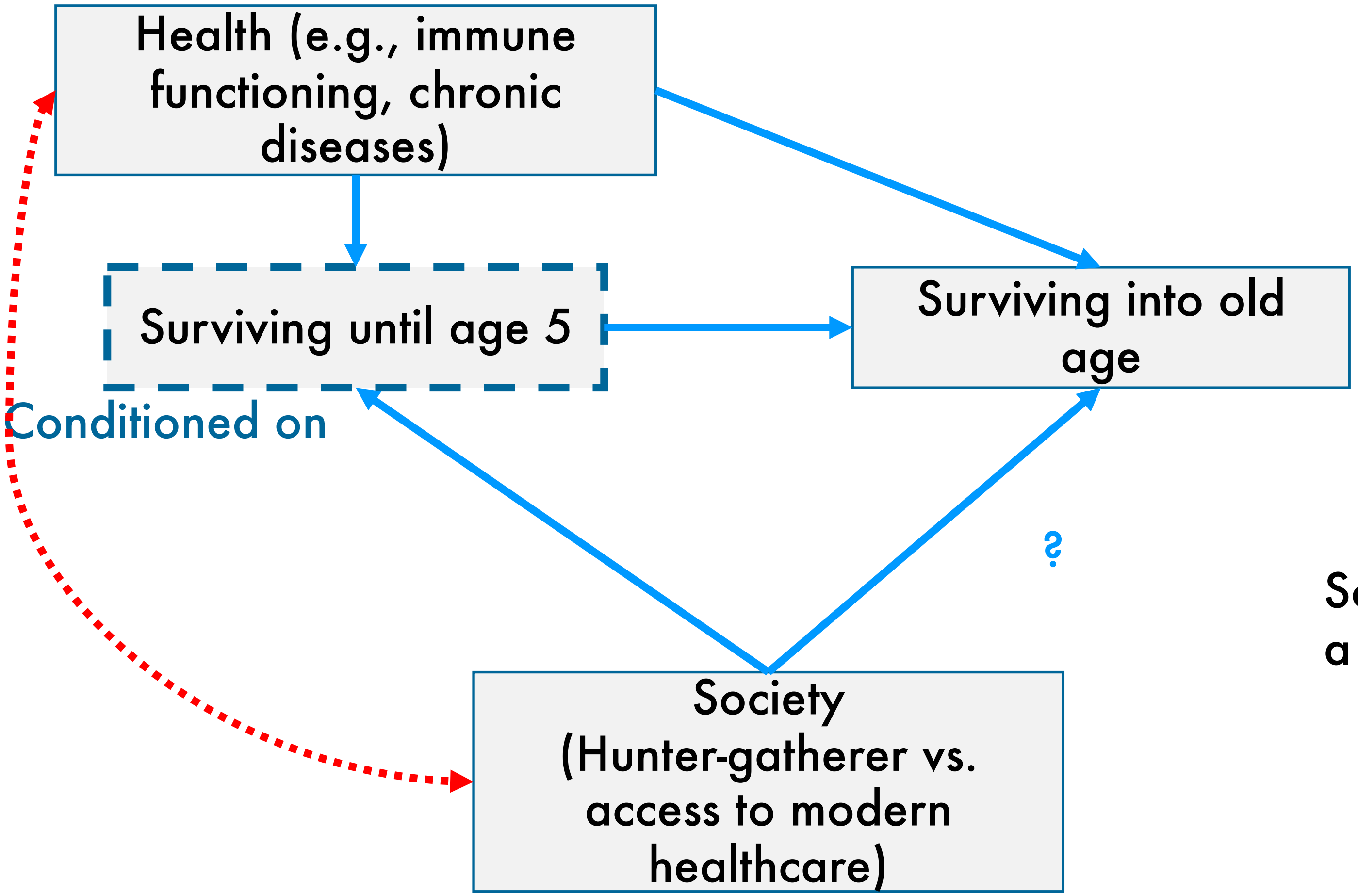
Author

Vybarr Cregan-Reid
Reader in Environmental Humanities, University of Kent

Life expectancy in hunter-gatherers



Life expectancy in hunter-gatherers

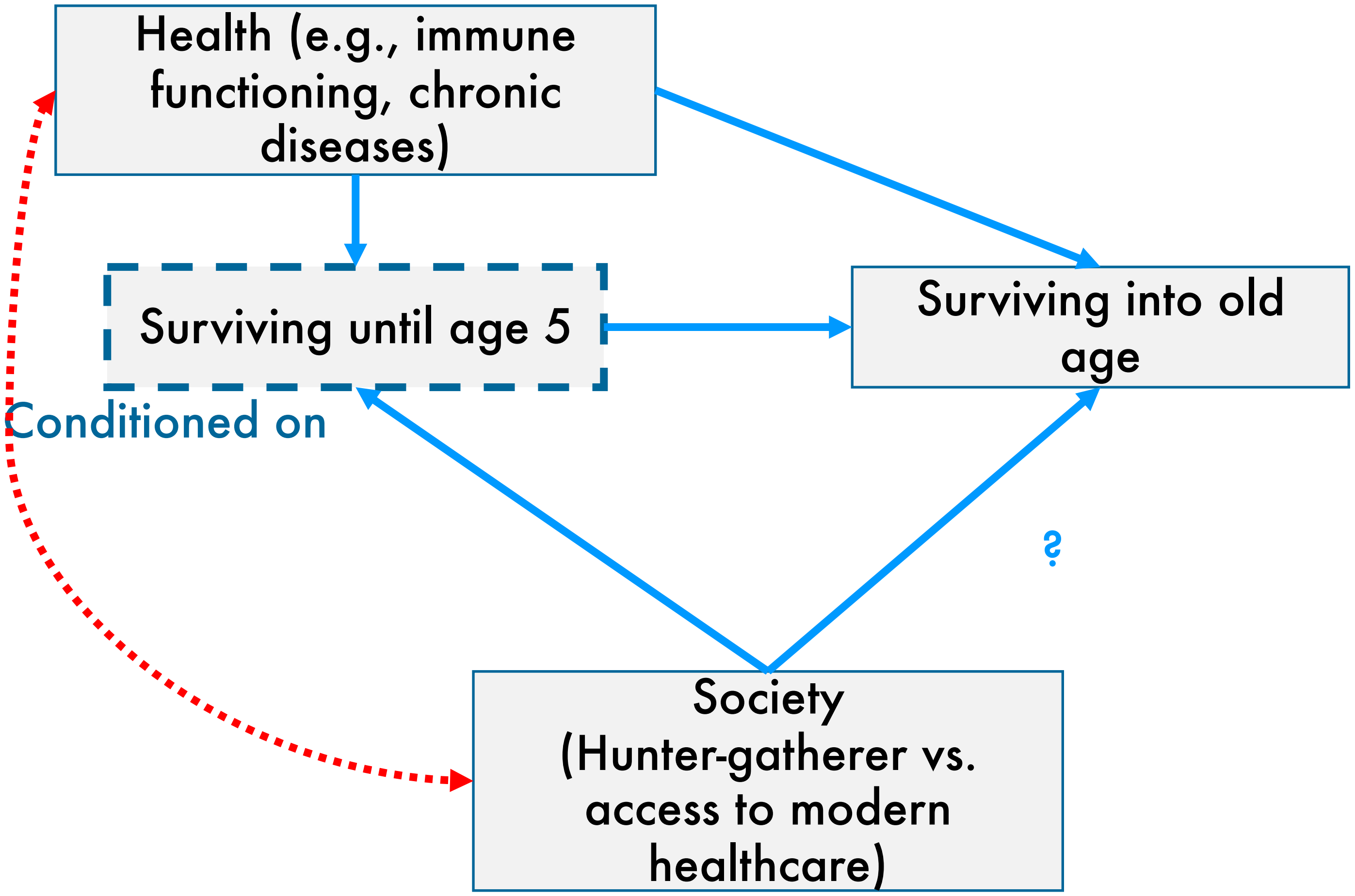


Health → <u>Surviving until age 5</u> ← Society

We have conditioned on a collider which induces non-causal associations between health and society

Society ←→ Health → Surviving into old age is a non-causal backdoor path

Life expectancy in hunter-gatherers



One way to think about this in more narrative terms:

- high childhood mortality constitutes a survival filter
- the population that made it to adolescence in a hunter-gatherer society has already been filtered more heavily
  - thus, it isn't that much surprising if those people prove to be quite sturdy
  - comparison at that point: apples vs oranges
- relevant counterfactual: if you took the people who survived high childhood mortality and then transplanted them to a modern society…

Virologie

Corona-Studie: Kinder genauso infektiös wie Erwachsene

Stand: 1.5.2020, 12:02 Uhr
Von Anja Braun

Teilen:

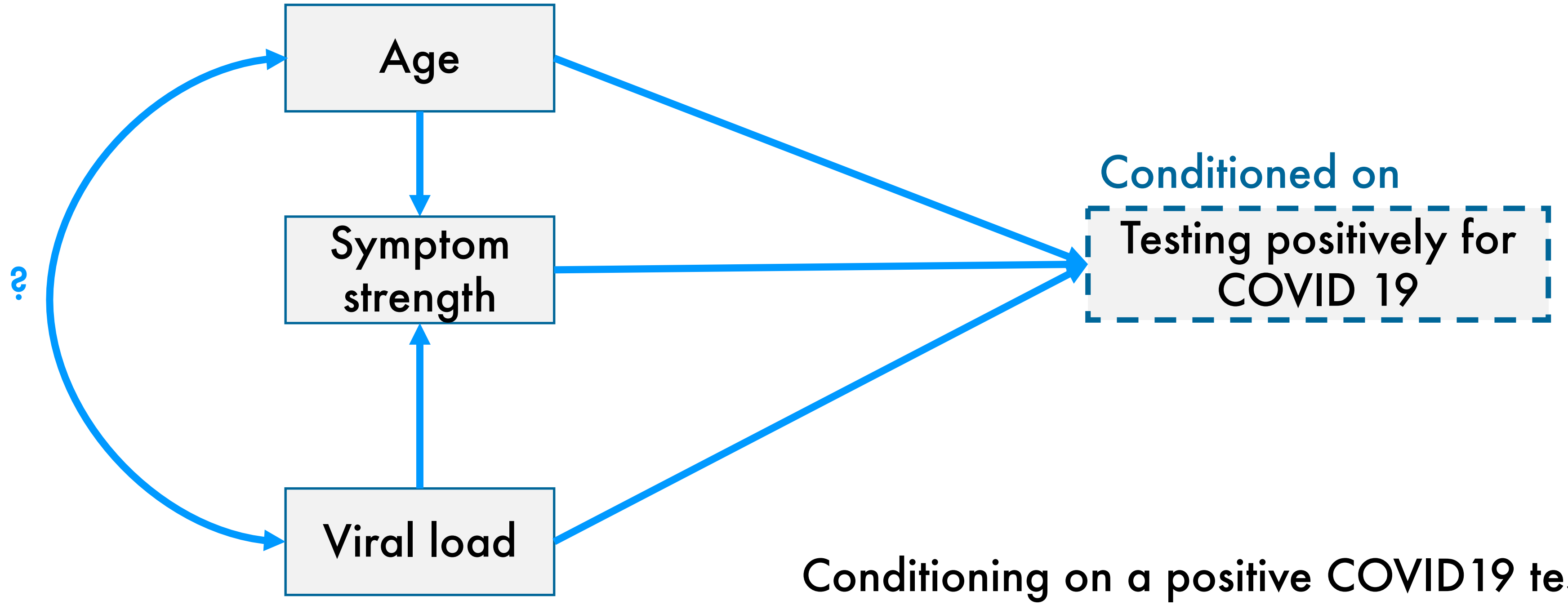
Wissenschaftler der Berliner Charité haben untersucht, welche Rolle Kinder bei der Übertragung des Coronavirus spielen. Erste Untersuchungen von Rachenabstrichen ergaben, dass Kinder offenbar genauso infektiös sind wie Erwachsene.

Kinder haben zwar viel seltener Symptome von Covid-19, aber sie haben offenbar die gleiche Viruslast im Rachen wie Erwachsene. Das ist das Ergebnis einer Laborauswertung der Charité. Virologen untersuchten 3712 Abstriche von positiv auf Corona getesteten Patienten auf ihre Altersverteilung. Darunter waren 37 Kindergartenkinder, 16 Grundschüler und 74 Jugendliche.

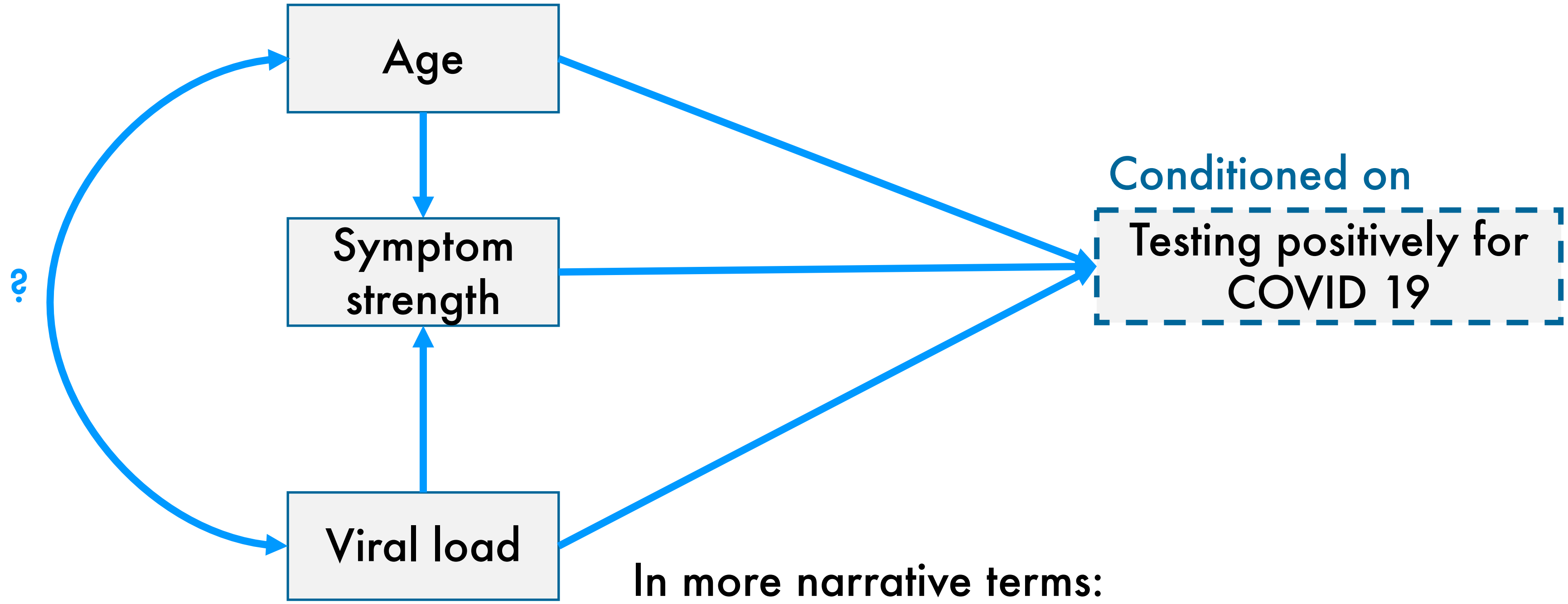
Corona study: Kids are just as infectious as adults

Scientists at Charité (Berlin) have investigated the role that children play in the transmission of corona. Initial investigations of throat swabs show, that children are just as infectious as adults.

Children are much less likely to have symptoms of COVID19, but it appears they have the same viral load in their throat as adults. This has been shown in a laborary study of Charité. Virologists investigated 3712 throat swabs of patients **who have tested positively for COVID**



Conditioning on a positive COVID19 test will induce (non-causal) associations between age, symptom strength, and viral load that do not hold in the general (unconditioned) population



In more narrative terms:

- only kids with an exceptionally high viral load may develop symptoms and thus get tested in the first place
- adults may already develop symptoms and get tested at very low viral loads

→ not a „fair“ comparison of whether kids are as contagious as adults

nature communications

Article | Open access | Published: 12 November 2020

Collider bias undermines our understanding of COVID-19 disease risk and severity

Gareth J. Griffith, Tim T. Morris, Matthew J. Tudball, Annie Herbert, Giulia Mancano, Lindsey Pike, Gemma C. Sharp, Jonathan Sterne, Tom M. Palmer, George Davey Smith, Kate Tilling, Luisa Zuccolo, Neil M. Davies & Gibran Hemani

Nature Communications **11**, Article number: 5749 (2020) | Cite this article

Abstract

Numerous observational studies have attempted to identify risk factors for infection with SARS-CoV-2 and COVID-19 disease outcomes. Studies have used datasets sampled from patients admitted to hospital, people tested for active infection, or people who volunteered to participate. Here, we highlight the challenge of interpreting observational evidence from such non-representative samples. Collider bias can induce associations between two or more variables which affect the likelihood of an individual being sampled, distorting associations between these variables in the sample. Analysing UK Biobank data, compared to the wider cohort the participants tested for COVID-19 were highly selected for a range of genetic, behavioural, cardiovascular, demographic, and anthropometric traits. We discuss the

DAGitty — draw and analyze causal diagrams

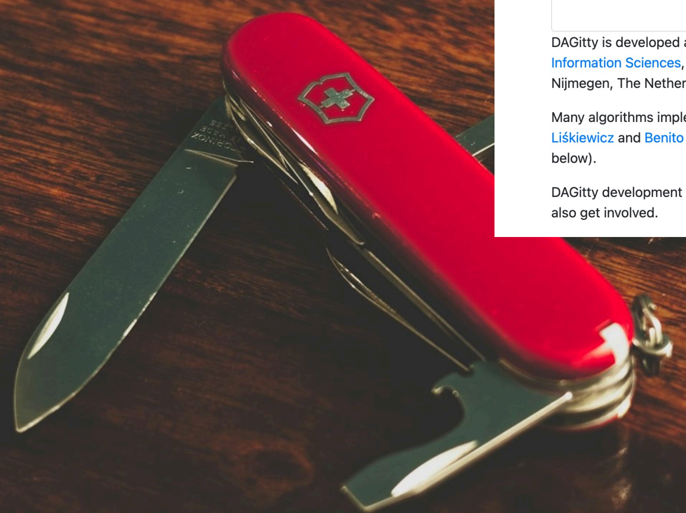
DAGitty is a browser-based environment for creating, editing, and analyzing causal diagrams (also known as directed acyclic graphs or causal Bayesian networks). The focus is on the use of causal diagrams for minimizing bias in empirical studies in epidemiology and other disciplines. For background information, see the "learn" page.

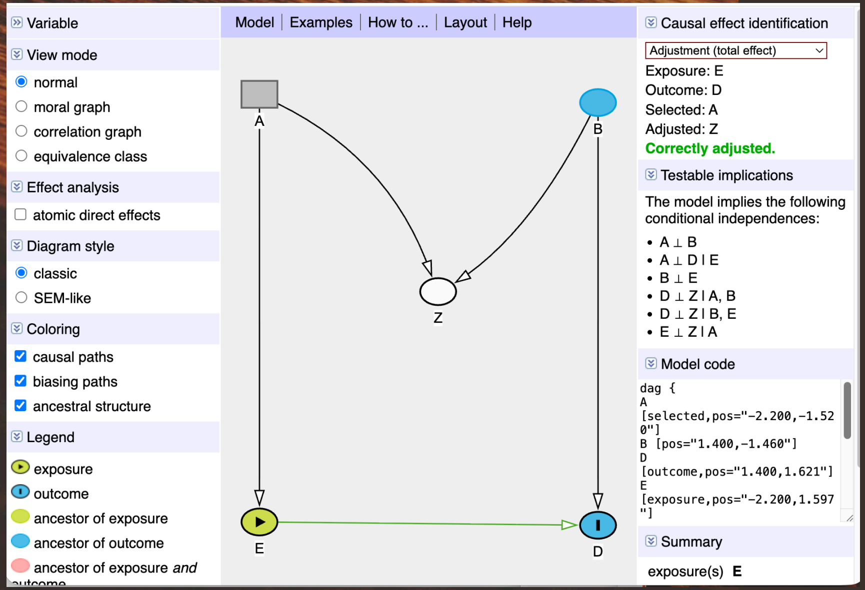
Launch	Download	Learn	Code
Launch DAGitty online in your browser.	Download DAGitty's source for offline use.	Learn more about DAGs and DAGitty.	The R package "dagitty" is available on CRAN or github.

DAGitty is developed and maintained by Johannes Textor (Institute for Computing and Information Sciences, Radboud University, and Medical BioSciences department, Radboudumc, Nijmegen, The Netherlands).

Many algorithms implemented in DAGitty were developed in close collaboration with Maciej Liśkiewicz and Benito van der Zander, University of Lübeck, Germany (see literature references below).

DAGitty development happens on github. You can download all source code from there and also get involved.





Thank you for your attention!

Julia M. Rohrer
Leipzig University
www.juliarohrer.com
@dingdingpeng.the100.ci
www.the100.ci

Slides: Resources