

# In-person session 4

**February 5, 2026**

PMAP 8521: Program evaluation  
Andrew Young School of Policy Studies

# Plan for today

**More regression things**

**Logic models and evaluation**

**Measuring outcomes**

**DAGs**

# Exam 1 details

# Final project details

# More regression things

# Interactive null worlds

# Statistical vs. substantive significance

# Logic models and evaluation

# The paradox of evaluation

**Evaluation is good, but expensive**

"Evaluation thinking"

**Too much evaluation is bad**

Taming programs

# Logic model

**Let's make another logic model!**



# Dolly Parton's Imagination Library

[Main site](#)

[Video intro](#)

# Measuring outcomes

# Outcomes and programs

**Outcome variable**

Thing you're measuring

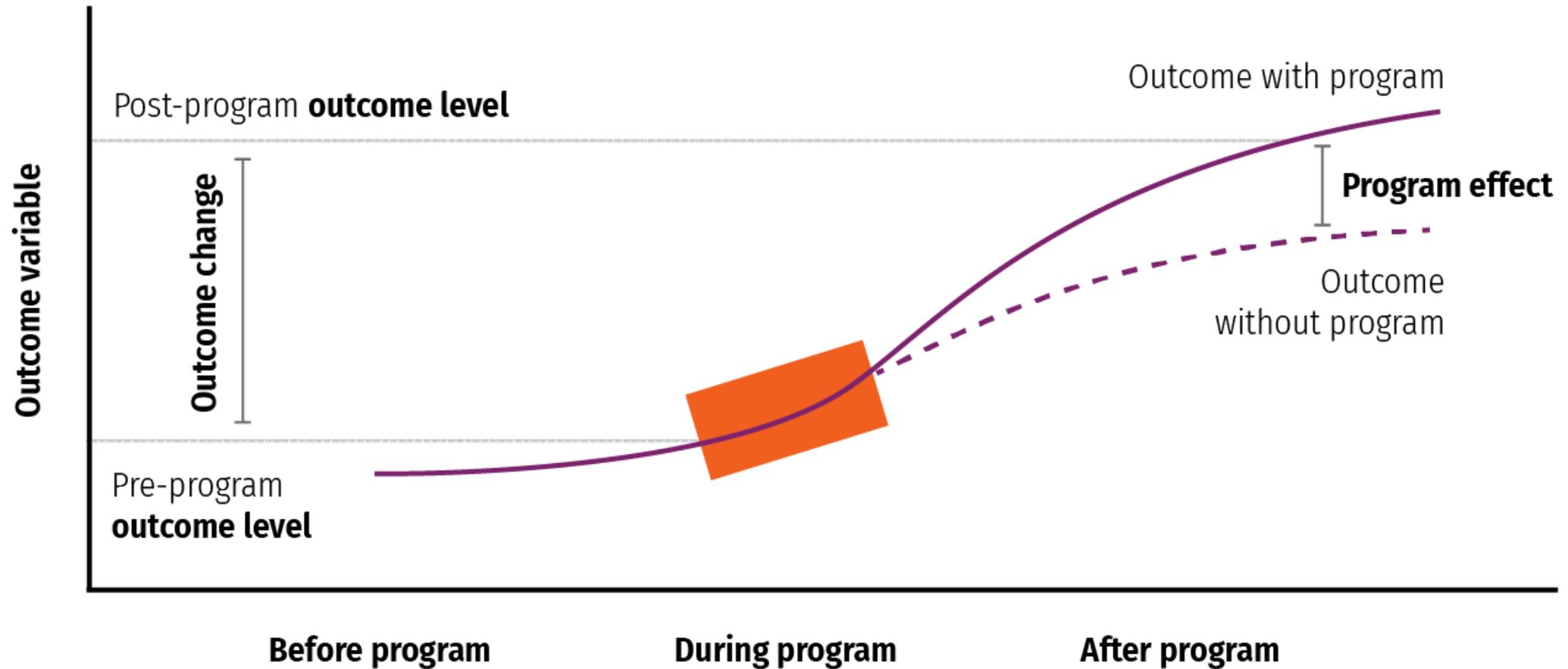
**Outcome change**

$\Delta$  in thing you're measuring over time

**Program effect**

$\Delta$  in thing you're measuring over time *because of the program*

# Outcomes and programs



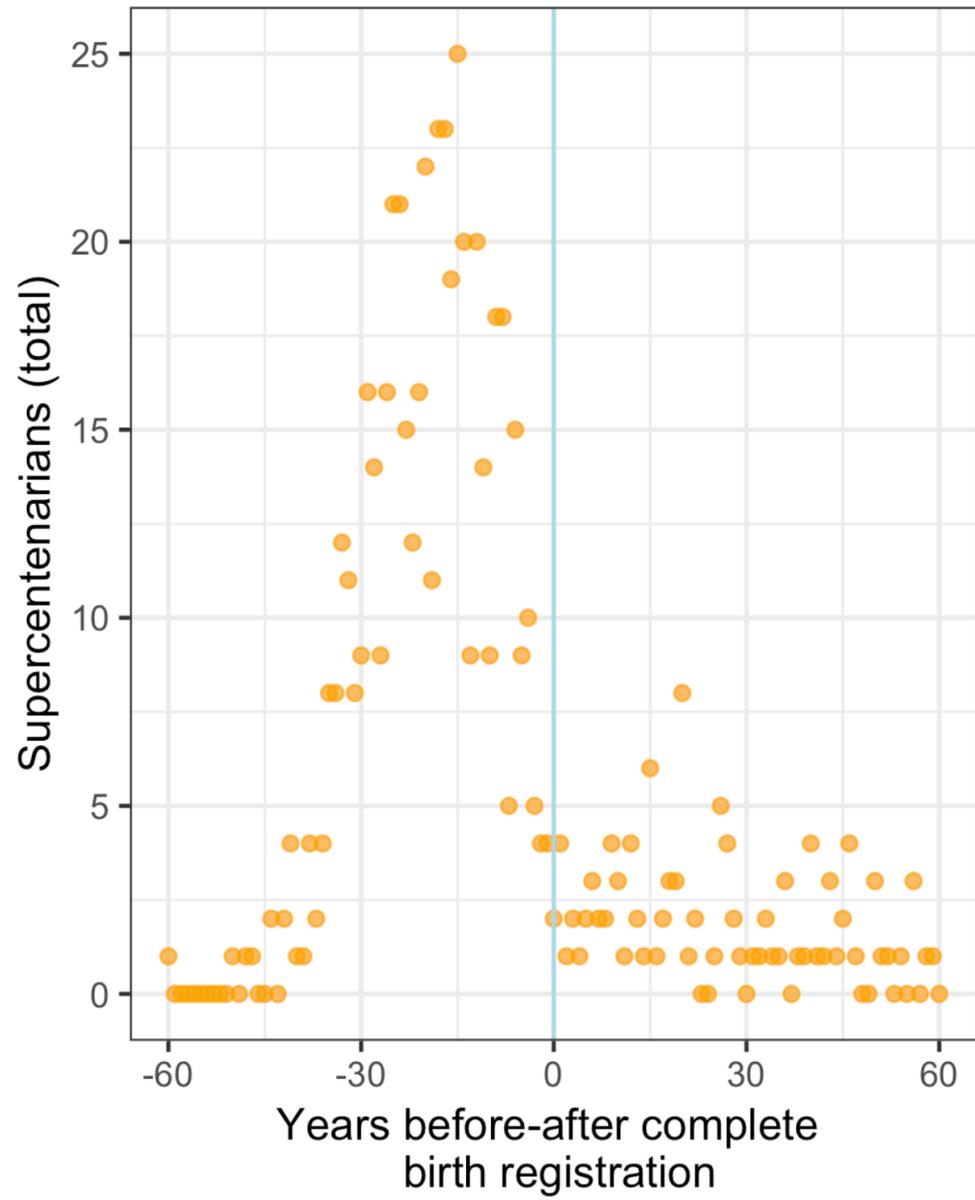
# Abstraction

Let's abstract an outcome!

Ladder of abstraction for witches

# DAGs

**Causal thinking is necessary—  
even for descriptive work!**



**"Every time I get a haircut, I become more mature!"**

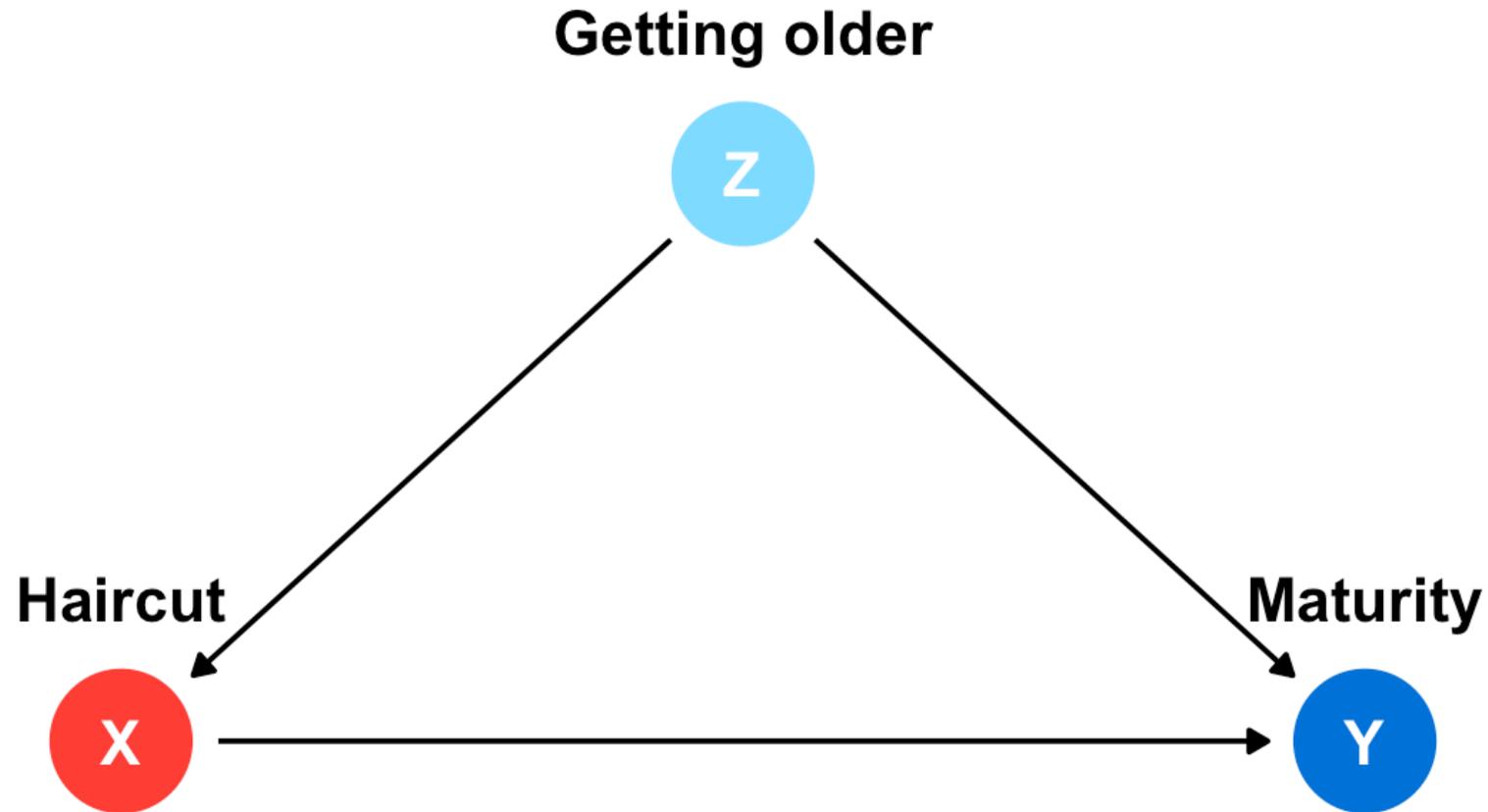


"Every time I get a haircut, I become more mature!"

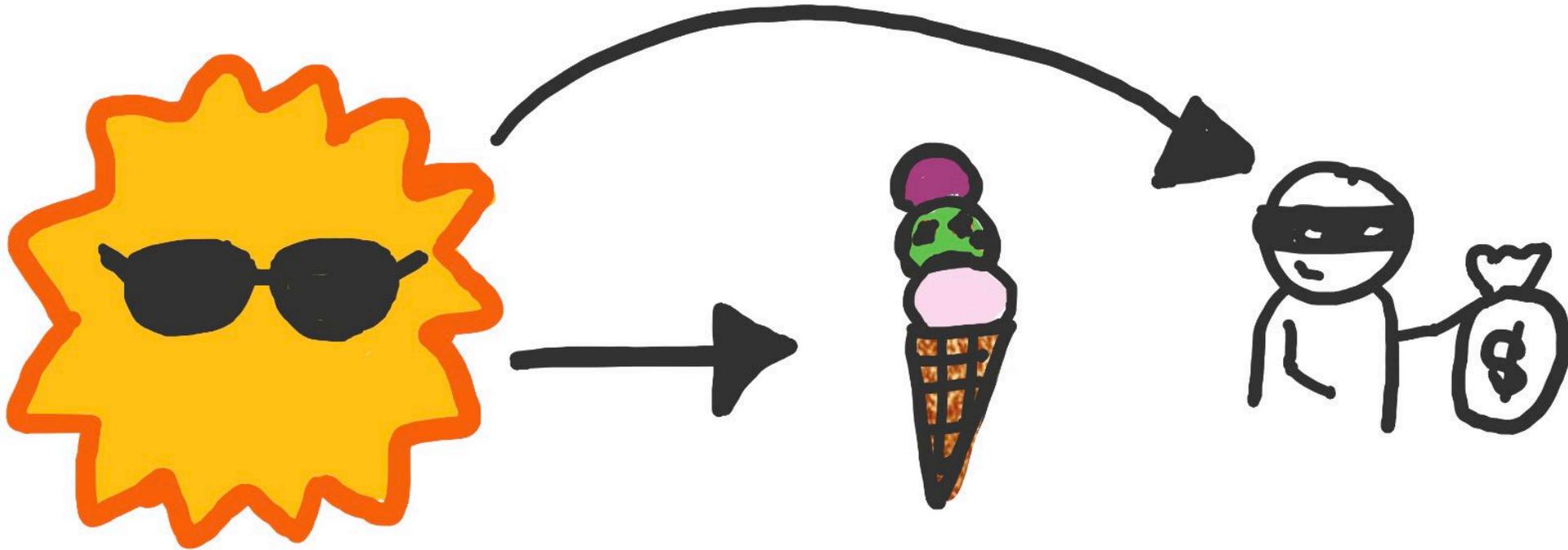


$$E[\text{Maturity} \mid \text{do}(\text{Get haircut})]$$

# Getting older opens a backdoor path



# Ice cream causes crime



Summer weather opens a backdoor path

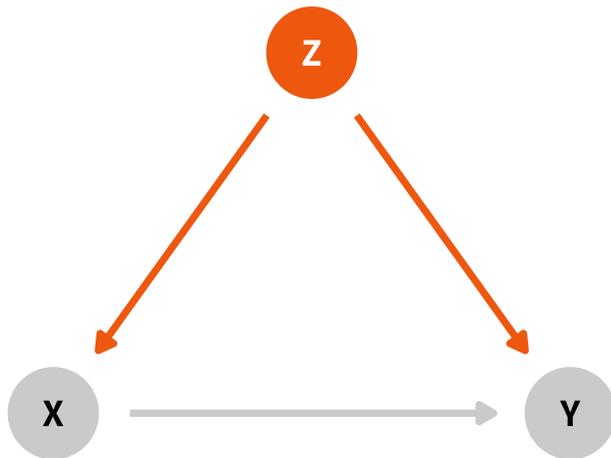
**But what does that mean,  
"opening a backdoor path"?**

**How does statistical association  
get passed through paths?**

# How do I know which of these is which?

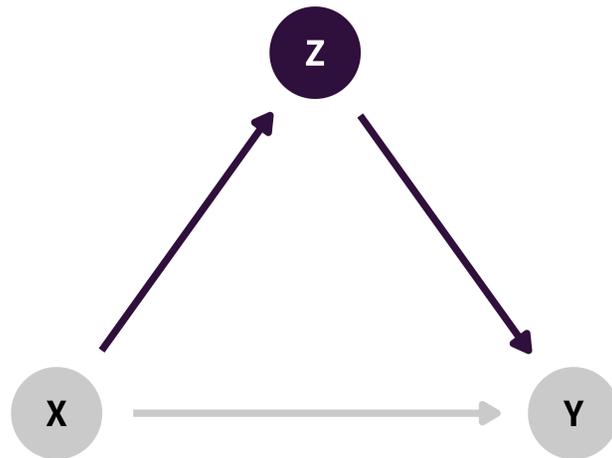
## Confounder

(Fork)



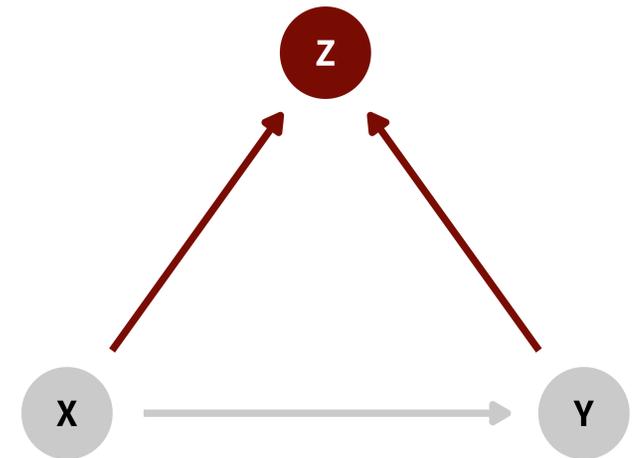
## Mediator

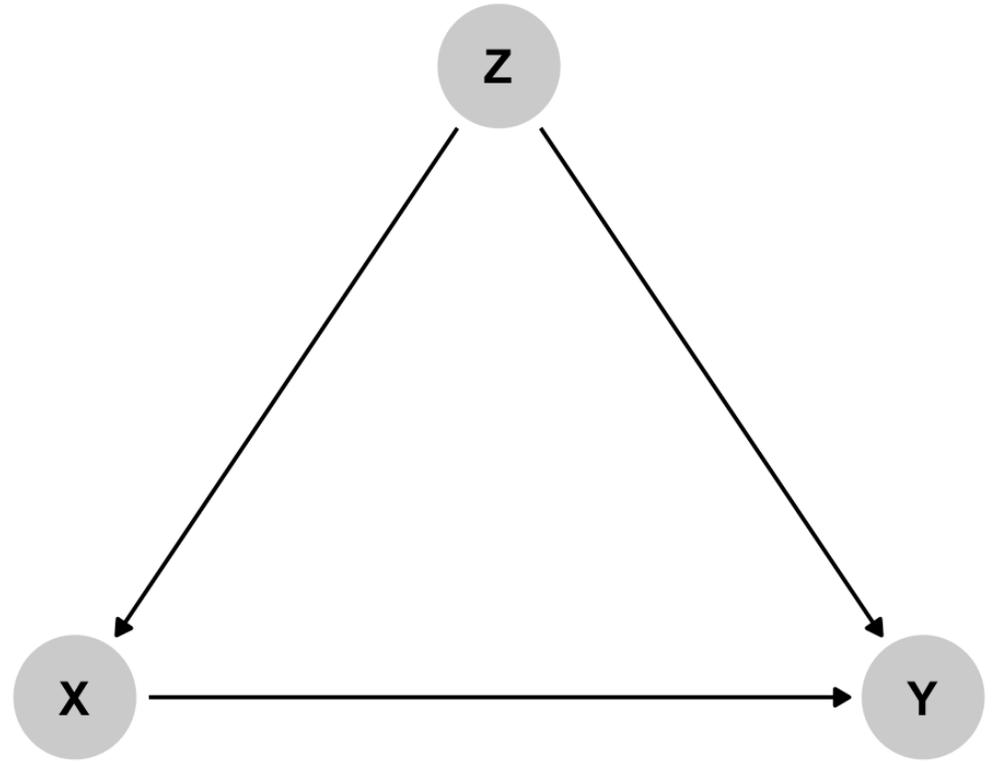
(Chain)

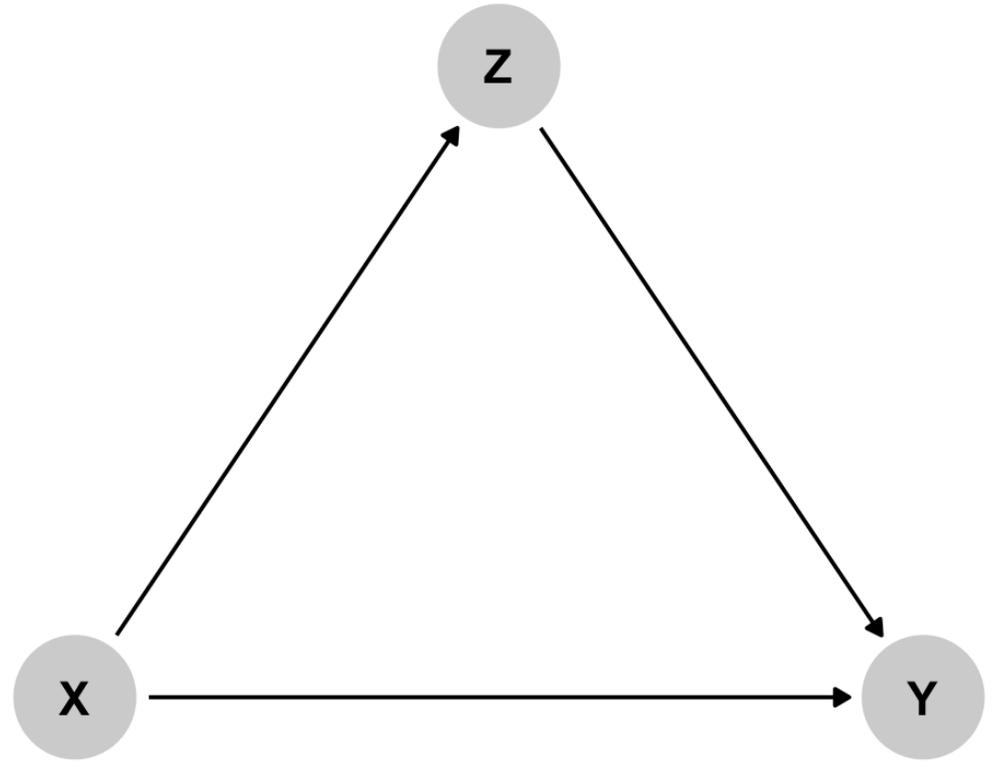


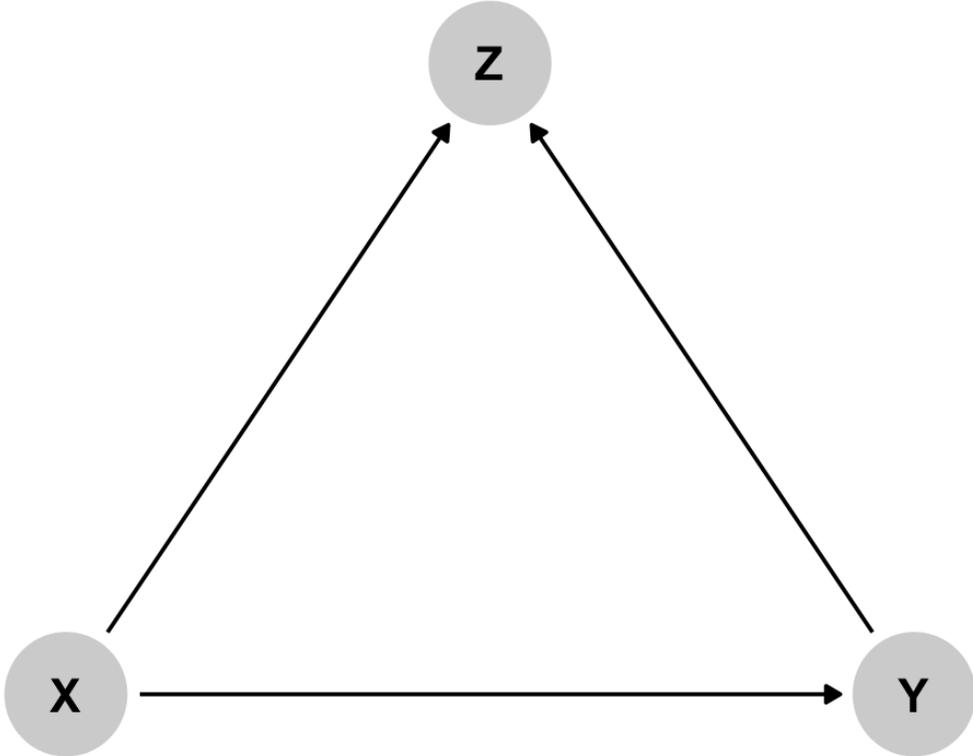
## Collider

(Inverted fork)

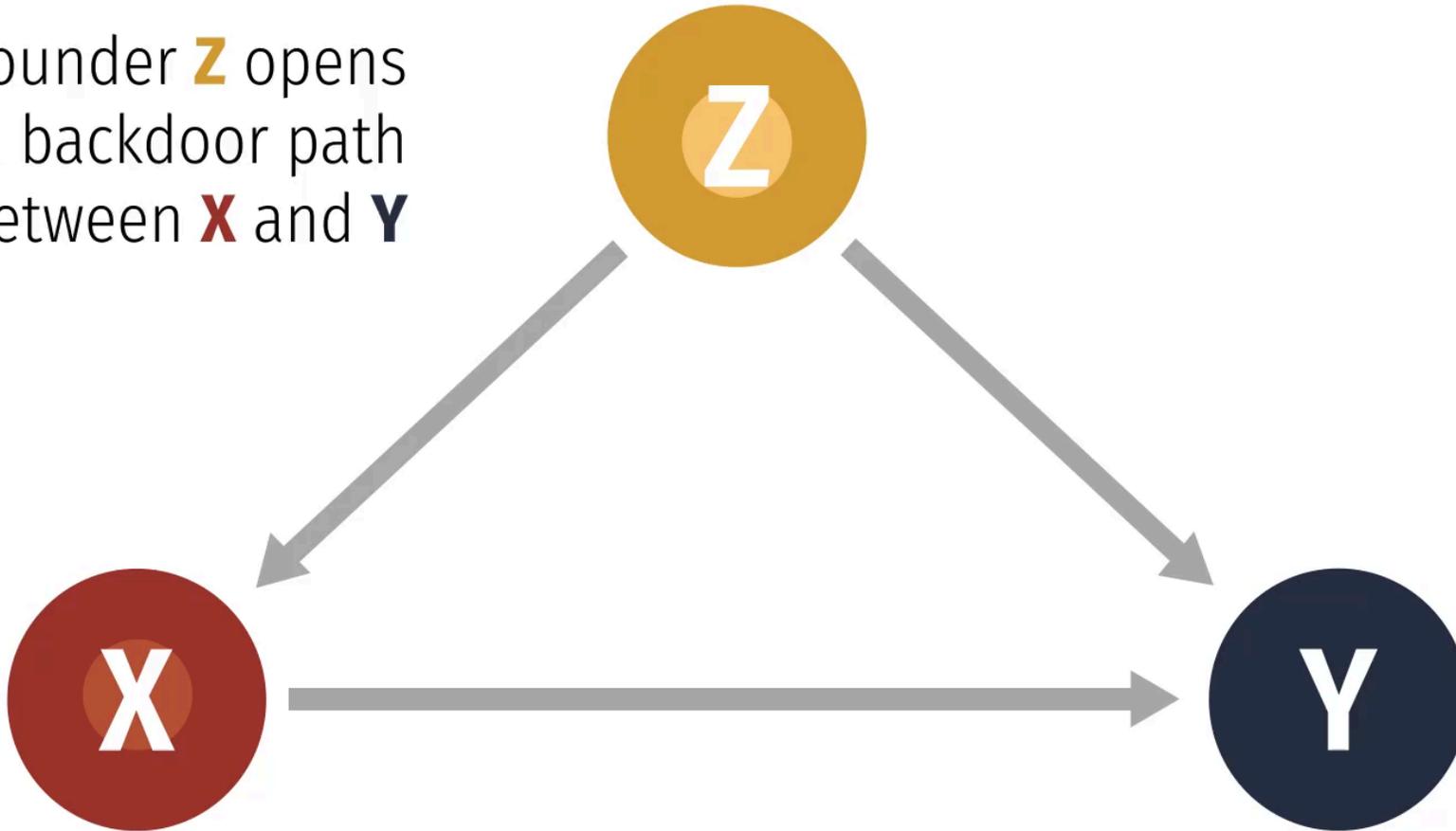








Confounder **Z** opens  
a backdoor path  
between **X** and **Y**

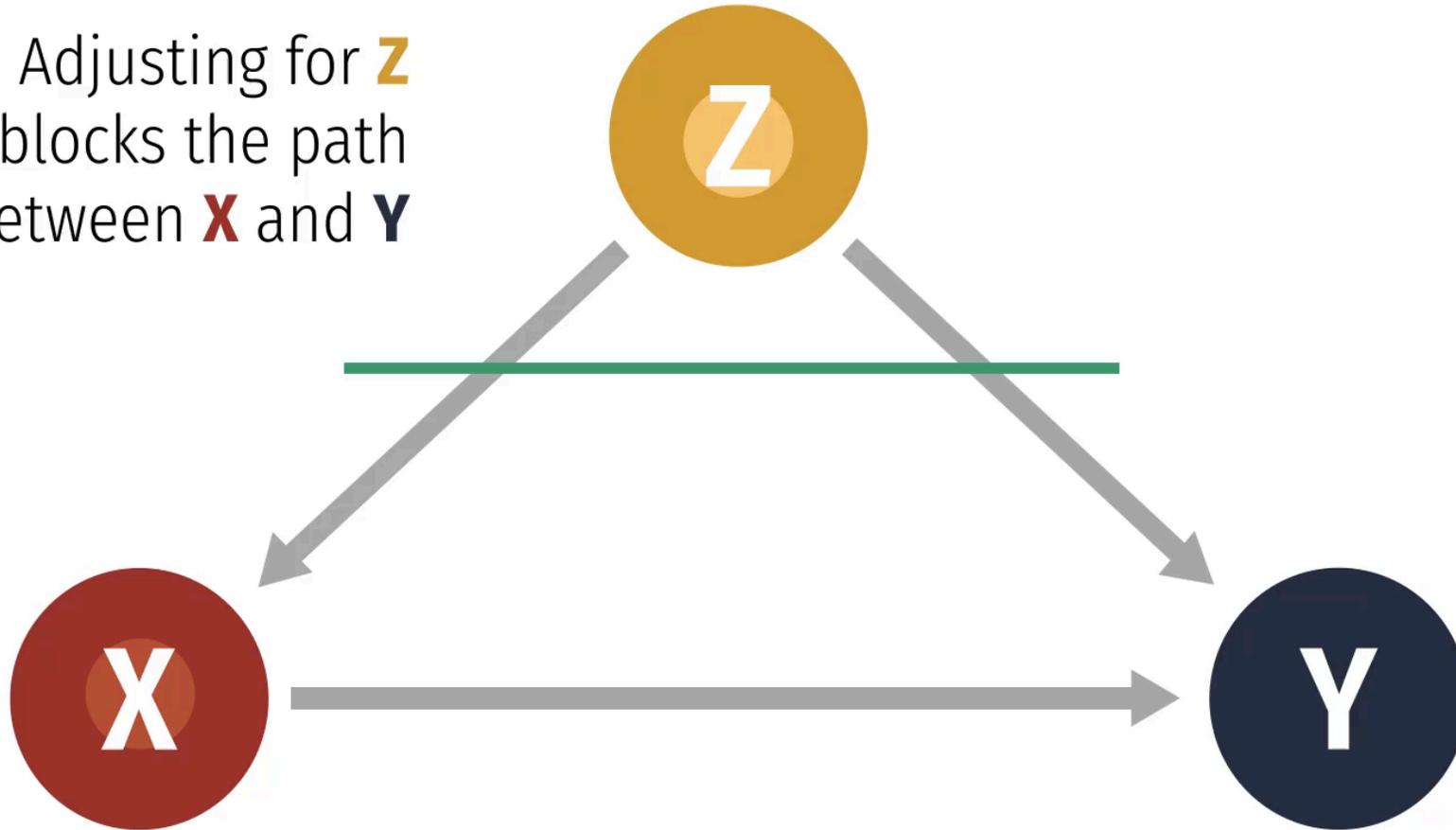


**Y** ~

▶ 0:00 / 0:06



Adjusting for **Z**  
blocks the path  
between **X** and **Y**

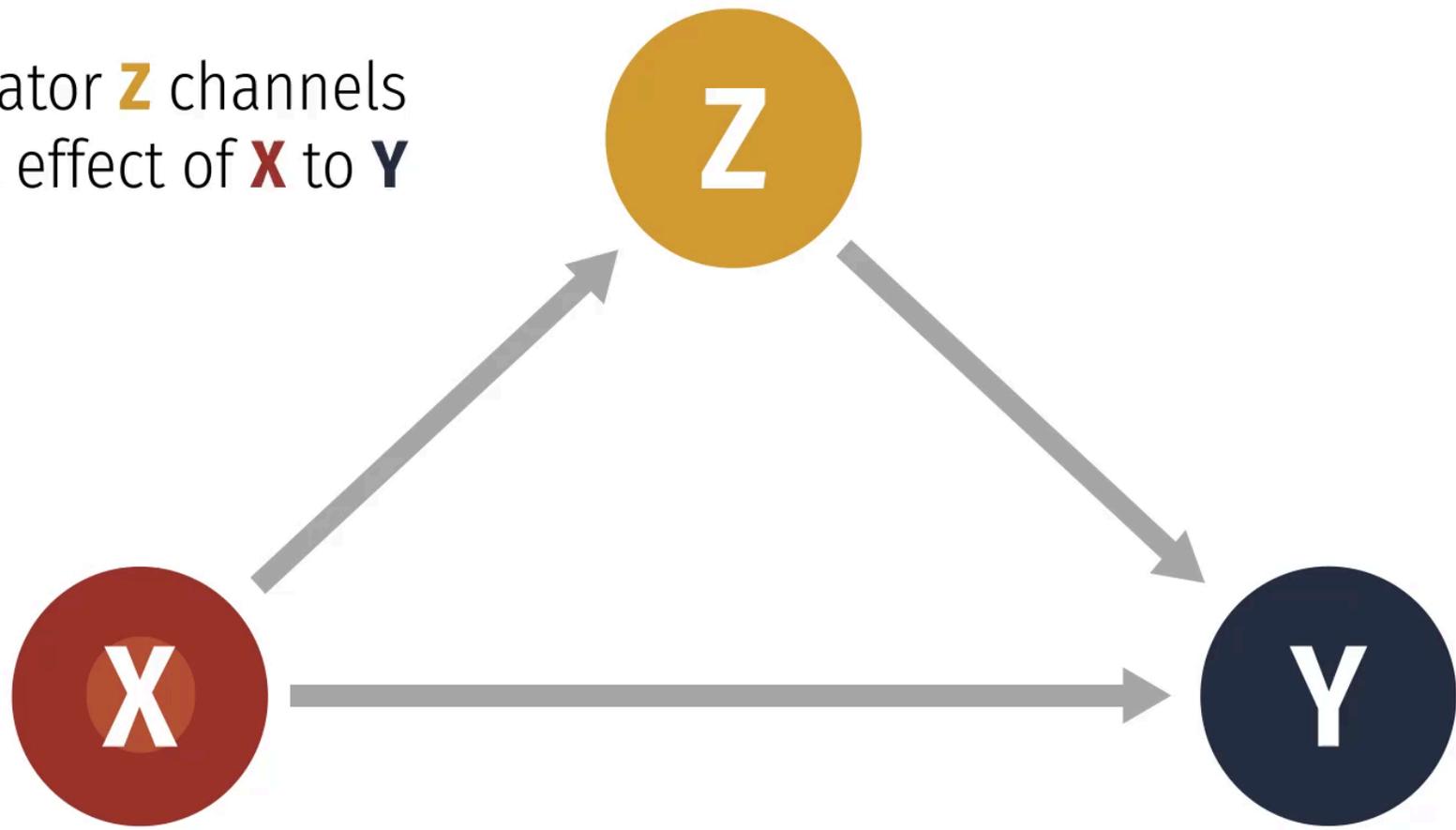


**Y** ~

▶ 0:00 / 0:03



Mediator **Z** channels indirect effect of **X** to **Y**



**Y** ~



# d-separation

Except for the one arrow between X and Y,  
no statistical association can flow between X and Y

**This is identification—**  
all alternative stories are ruled out  
and the relationship is isolated

**How do we know if we have the right DAG?**

**Right nodes? Enough nodes?  
Too many nodes?**

**Right arrows?**

**lol**

Nick Huntington-Klein et al., “The Influence of Hidden Researcher Decisions in Applied Microeconomics,”  
*Economic Inquiry* 59, no. 3 (2021): 944–60, <https://doi.org/10.1111/ecin.12992>.

Received: 5 June 2020 | Accepted: 14 February 2021

DOI: 10.1111/ecin.12992

**ORIGINAL ARTICLE**

Economic Inquiry

## The influence of hidden researcher decisions in applied microeconomics

Nick Huntington-Klein, Assistant Professor<sup>1</sup>  | Andreu Arenas, Assistant Professor<sup>2</sup> |  
Emily Beam, Assistant Professor<sup>3</sup> | Marco Bertoni, Associate Professor<sup>4</sup> |  
Jeffrey R. Bloem, Research Economist<sup>5</sup> | Pralhad Burli, Economist<sup>6</sup> |  
Naibin Chen, Graduate Student<sup>7</sup> | Paul Grieco, Associate Professor<sup>8</sup> |  
Godwin Ekpe, Graduate Student<sup>9</sup> | Todd Pugatch, Associate Professor<sup>10</sup> |  
Martin Saavedra, Associate Professor<sup>11</sup> | Yaniv Stopnitzky, Assistant Professor<sup>12</sup>

<sup>1</sup>Seattle University, Seattle,  
Washington, USA

<sup>2</sup>University of Barcelona & IEB,  
Barcelona, Spain

<sup>3</sup>University of Vermont, Burlington,  
Vermont, USA

<sup>4</sup>Department of Economics and  
Management “M. Fanno”, Padova  
University, Padova, Italy

<sup>5</sup>USDA Economic Research Service,  
Kansas City, Missouri, USA

<sup>6</sup>Idaho National Laboratory, Idaho Falls,  
Idaho, USA

### Abstract

Researchers make hundreds of decisions about data collection, preparation, and analysis in their research. We use a many-analysts approach to measure the extent and impact of these decisions. Two published causal empirical results are replicated by seven replicators each. We find large differences in data preparation and analysis decisions, many of which would not likely be reported in a publication. No two replicators reported the same sample size. Statistical significance varied across replications, and for one of the studies the effect’s sign varied as well. The standard deviation of estimates across replications was 3–4 times the mean reported standard error.

# What's the difference between logic models and DAGs?

Can't I just remake my logic model in Dagitty and be done?

# DAGs vs. Logic models

**DAGs are a *statistical* tool**

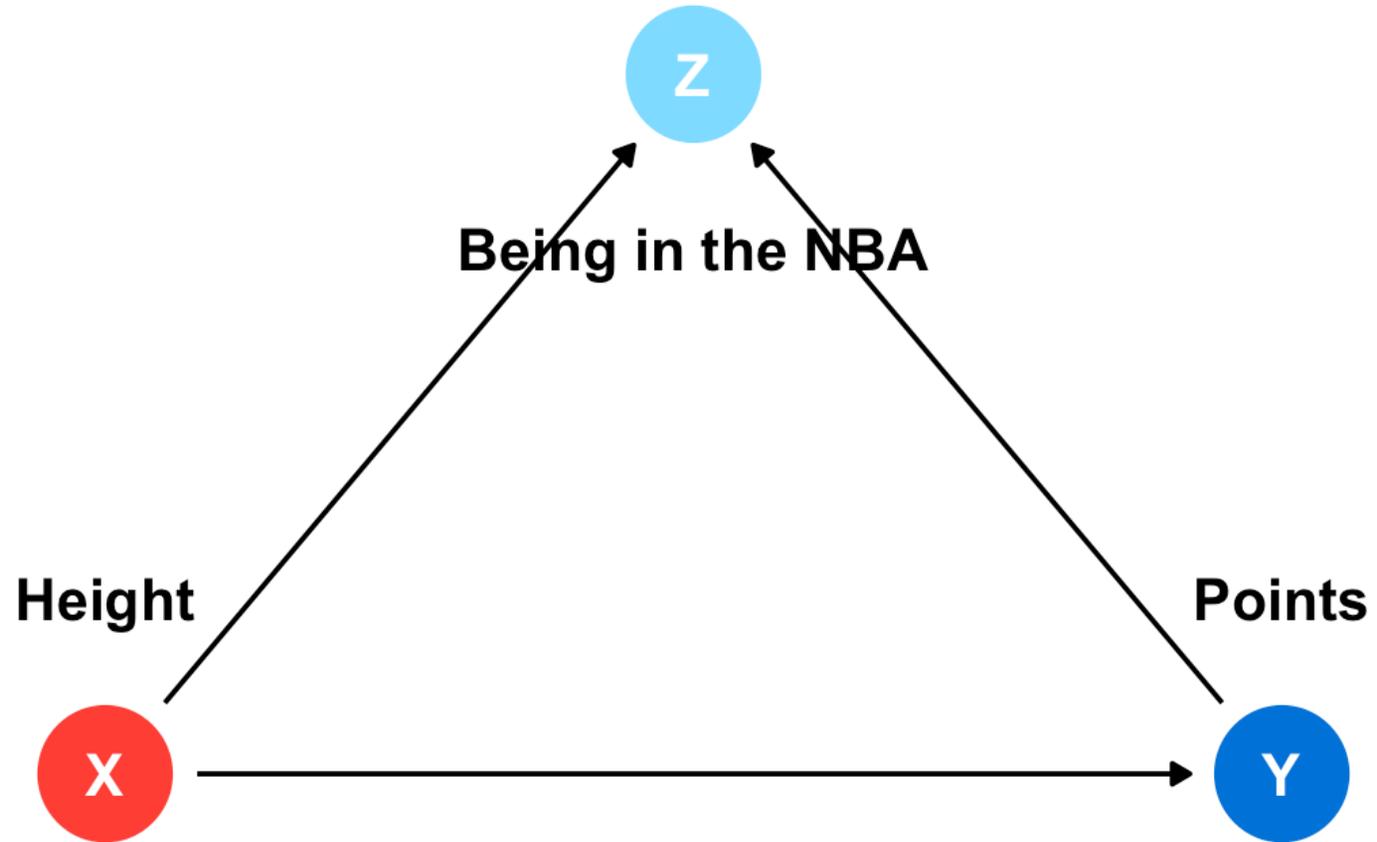
Describe a data-generating process  
and isolate/identify relationships

**Logic models are a *managerial* tool**

Oversee the inner workings of a program and its theory

**How exactly do colliders  
mess up your results?**

**It looks like you can  
still get the effect of X on Y**





Sept. 10, 2021, 3:58 p.m. ET

By [Davey Alba](#)

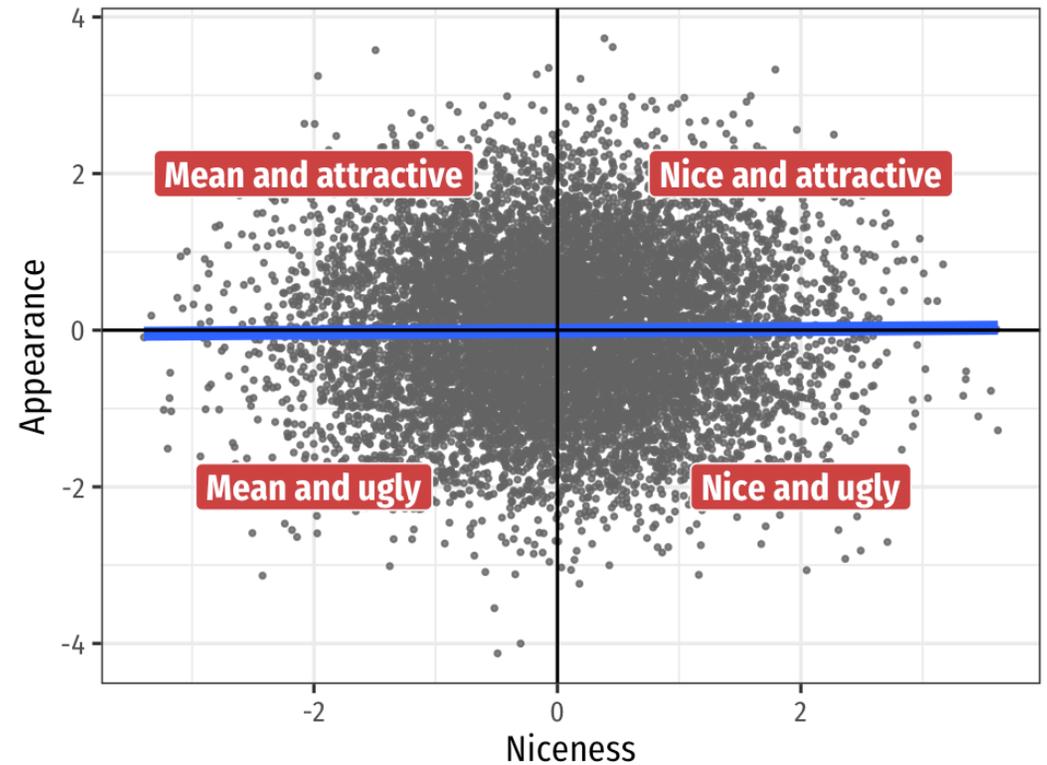
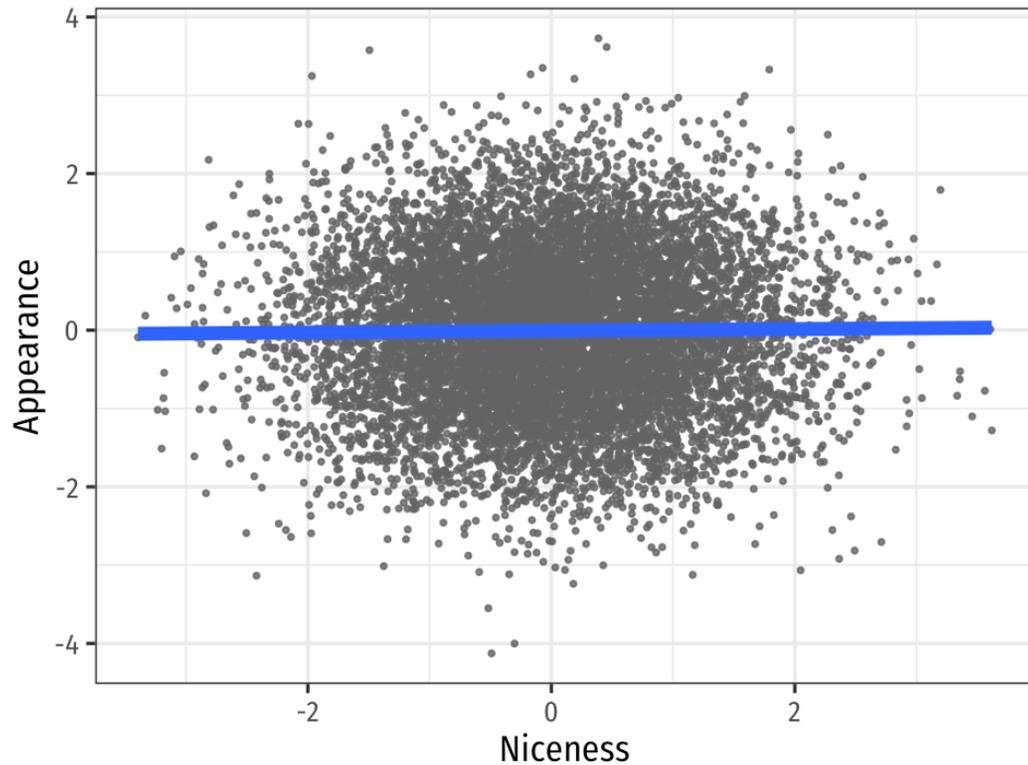


## Facebook sent flawed data to misinformation researchers.

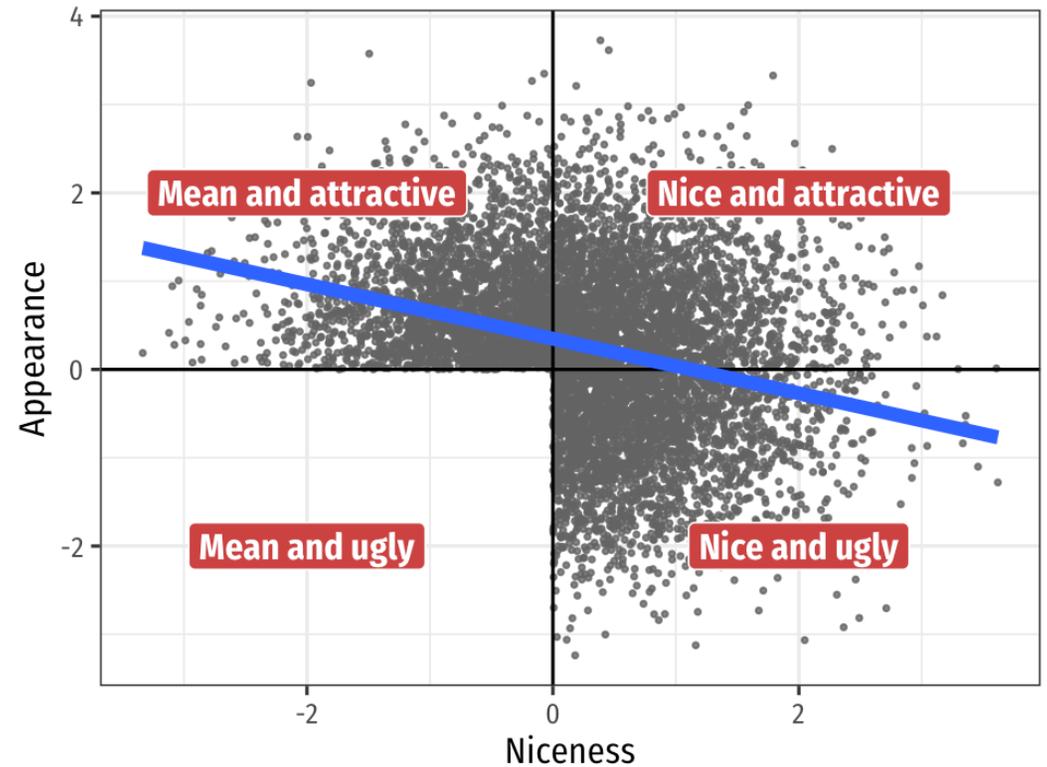
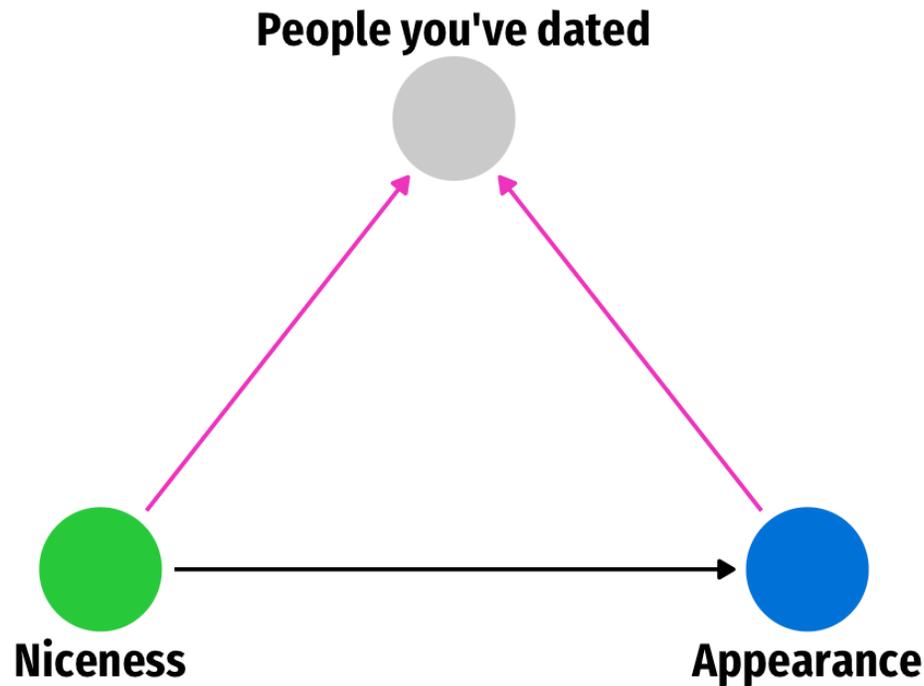


Mark Zuckerberg, chief executive of Facebook, testifying in Washington in 2018. Tom Brenner/The New York Times

# Does niceness improve appearance?

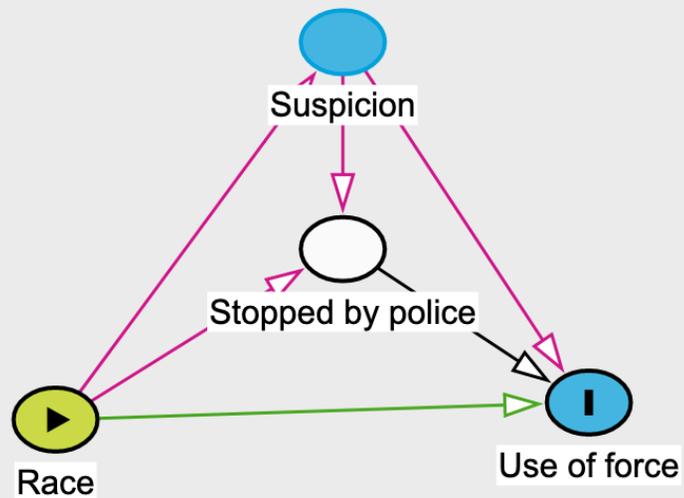


# Collider distorts the true effect!



# Effect of race on police use of force using administrative data

# Effect of race on police use of force using administrative data



American Political Science Review, Page 1 of 19  
doi:10.1017/S0003055420000039

© American Political Science Association 2020

## Administrative Records Mask Racially Biased Policing

DEAN KNOX *Princeton University*

WILL LOWE *Heriot School of Governance*

JONATHAN MUMMOLO *Princeton University*

**R**esearchers often lack the necessary data to credibly estimate racial discrimination in policing. In particular, police administrative records lack information on civilians police observe but do not investigate. In this article, we show that if police racially discriminate when choosing whom to investigate, analyses using administrative records to estimate racial discrimination in police behavior are statistically biased, and many quantities of interest are unidentified—even among investigated individuals—absent strong and untestable assumptions. Using principal stratification in a causal mediation framework, we derive the exact form of the statistical bias that results from traditional estimation. We develop a bias-correction procedure and nonparametric sharp bounds for race effects, replicate published findings, and show the traditional estimator can severely underestimate levels of racially biased policing or mask discrimination entirely. We conclude by outlining a general and feasible design for future studies that is robust to this inferential snare.

Concern over racial bias in policing, and the public availability of large administrative data sets documenting police-civilian interactions, have prompted a raft of studies attempting to quantify the effect of civilian race on law enforcement behavior. These studies consider a range of outcomes including ticketing, stop duration, searches, and the use of force (e.g., Antonovics and Knight 2009; Fryer 2019; Ridgeway 2006; Nix et al. 2017). Most research in this area attempts to adjust for omitted variables that may correlate with suspect race and the outcome of interest. In contrast, this study addresses a more fundamental problem that remains even if the vexing issue of omitted variable bias is solved: the inevitable statistical bias that results from studying racial discrimination using records that are themselves the product of racial discrimination (Angrist and Pischke 2008; Elwert and Winship 2014; Rosenbaum 1984). We show that when there is any

biased absent additional data and/or strong and untestable assumptions.

This study makes several contributions. We clarify the causal estimands of interest in the study of racially discriminatory policing—quantities that many studies appear to be targeting, but are rarely made explicit—and show that the conventional approach fails to recover any known causal quantity in reasonable settings. Next, we highlight implicit and highly implausible assumptions in prior work and derive the statistical bias when they are violated. We proceed to develop informative nonparametric sharp bounds for the range of possible race effects, apply these in a reanalysis and extension of a prominent article on police use of force (Fryer 2019), and present bias-corrected results that suggest this and similar studies drastically underestimate the level of racial bias in police-civilian interactions. Finally, we outline strategies for future data collection and re-

**Let's make a DAG!**