In-person session 5

September 20, 2021

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

Colliders

Potential outcomes vs. do() notation

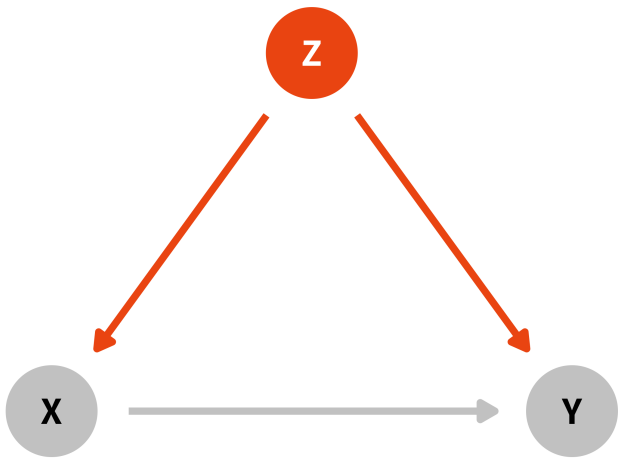
do-calculus, adjustment, and CATEs

Logic models, DAGs, and measurement

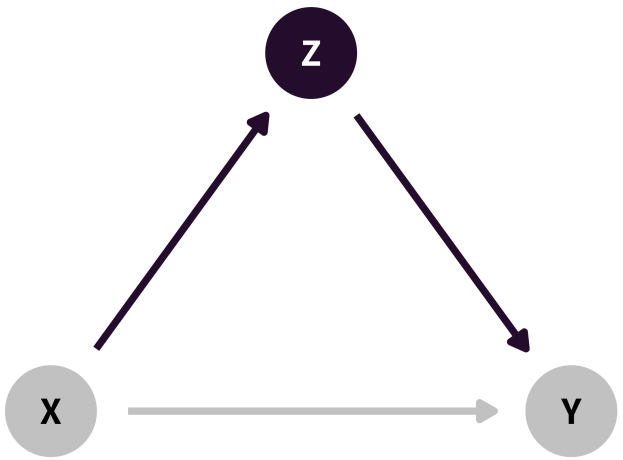
Colliders

How do I know which of these is which?

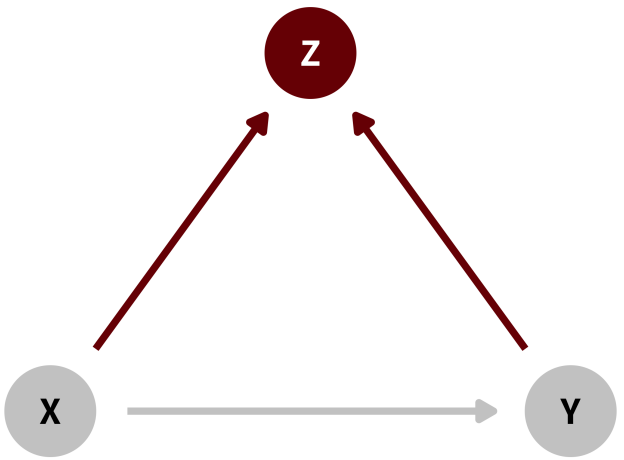
Confounder
(Fork)



Mediator
(Chain)



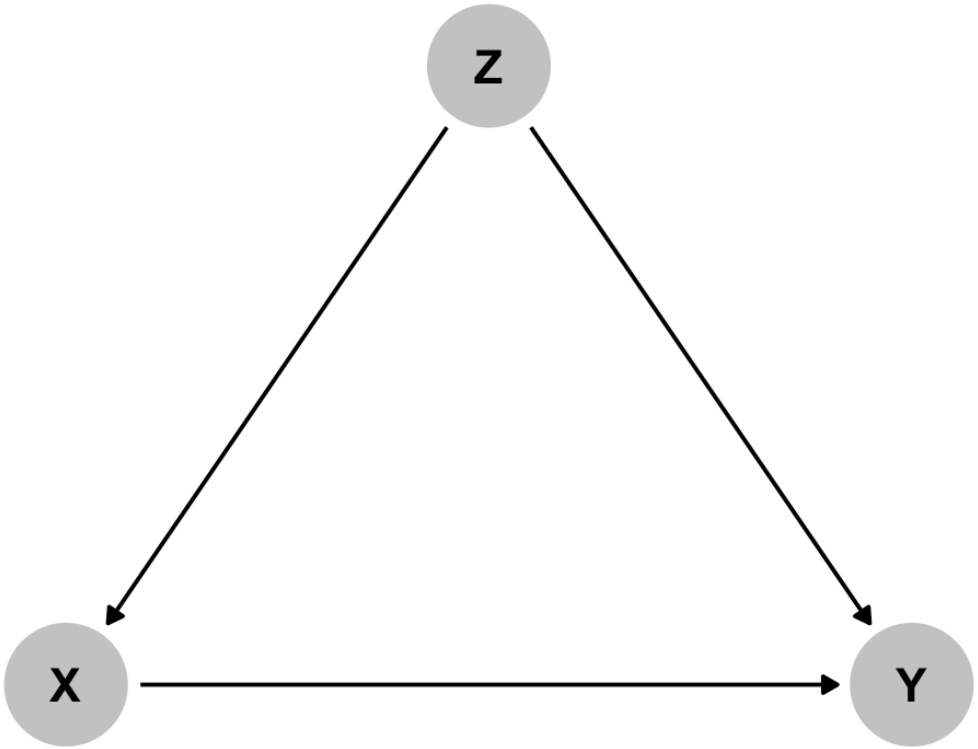
Collider
(Inverted fork)



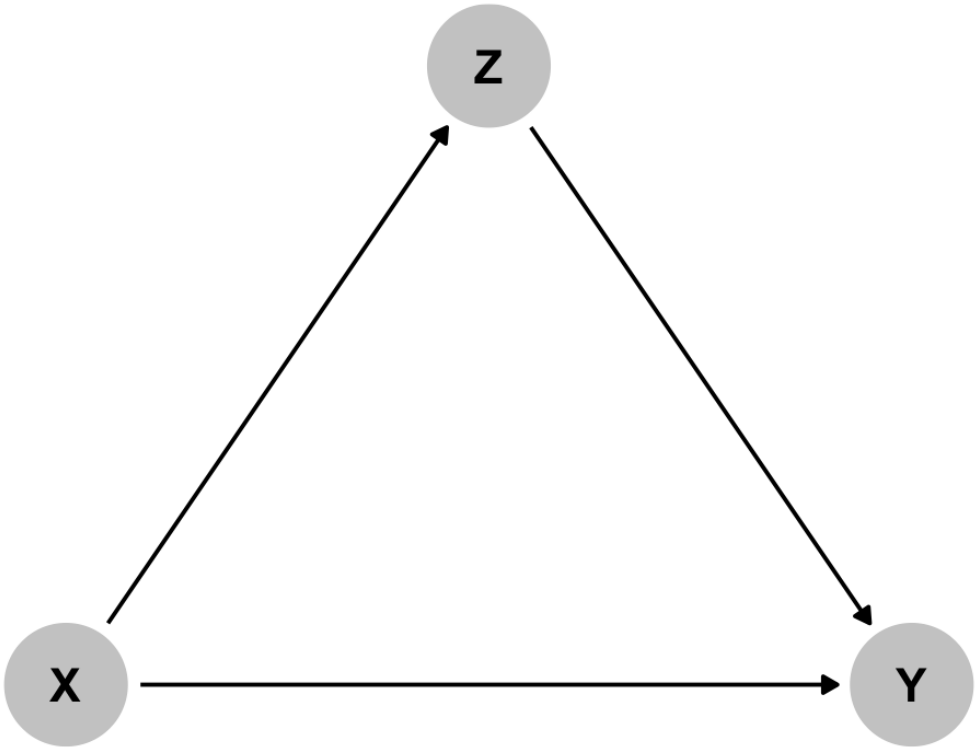
But what does that mean, "opening a backdoor path"?

How does statistical association get passed through paths?

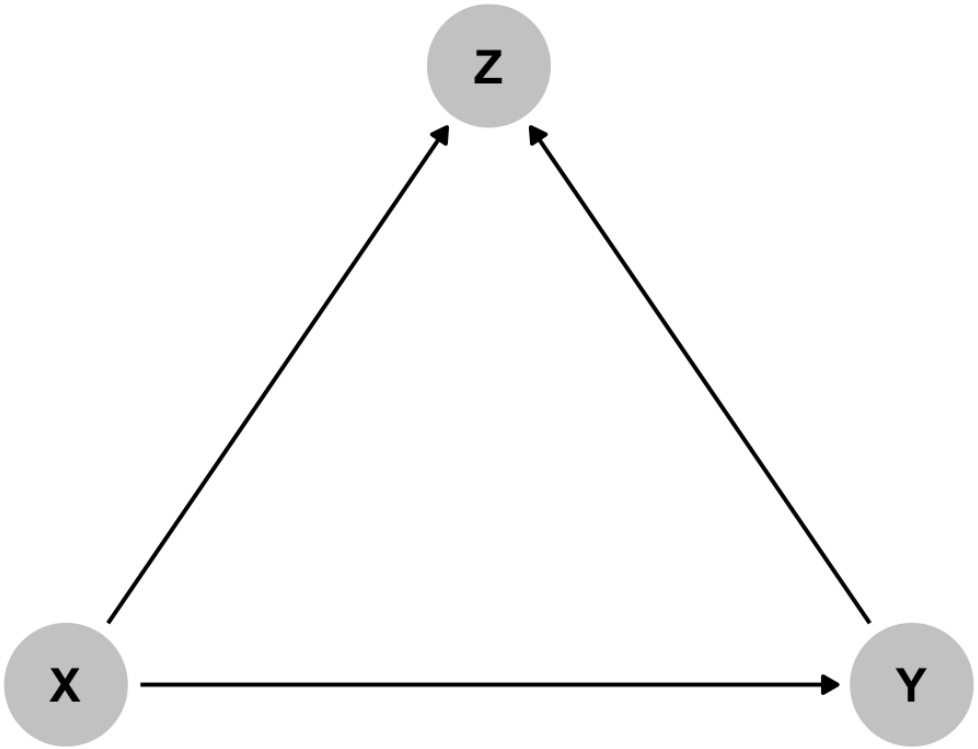

LUTRON


Z
X
Y


LUTRON


Z
X
Y


LUTRON


Z
X
Y

d-separation

How can you be sure you include everything in a DAG?

Is there a rule of thumb for the number of nodes?

How exactly do colliders mess up your results?

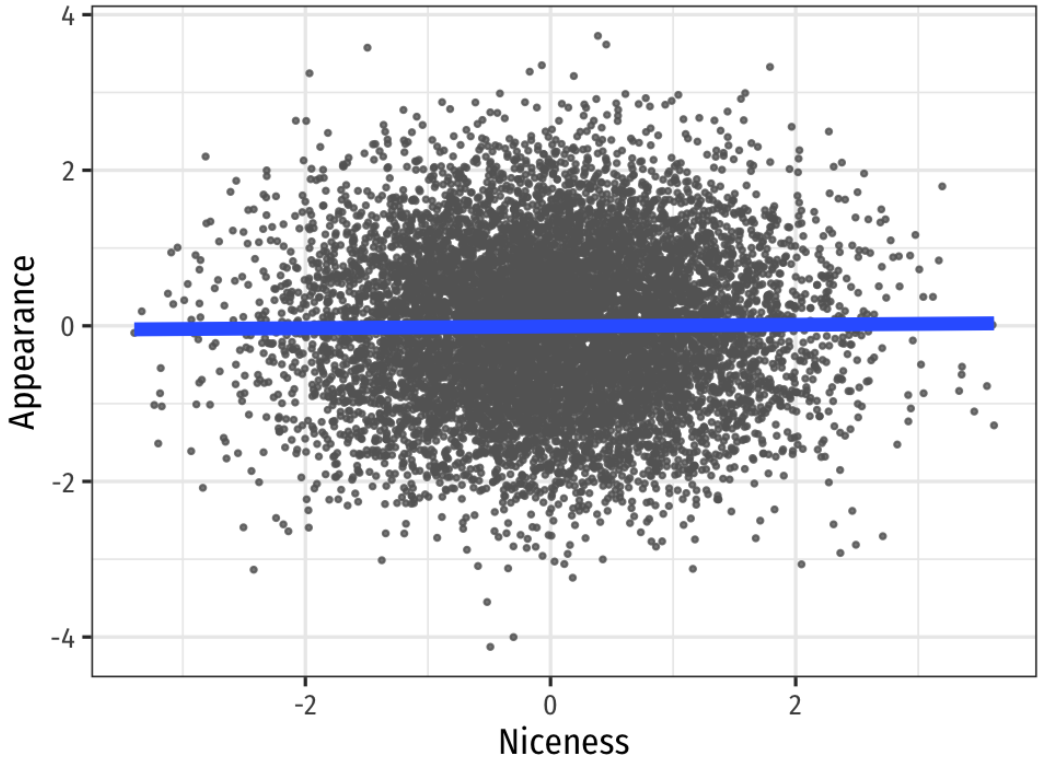
It looks like you can still get the effect of X on Y

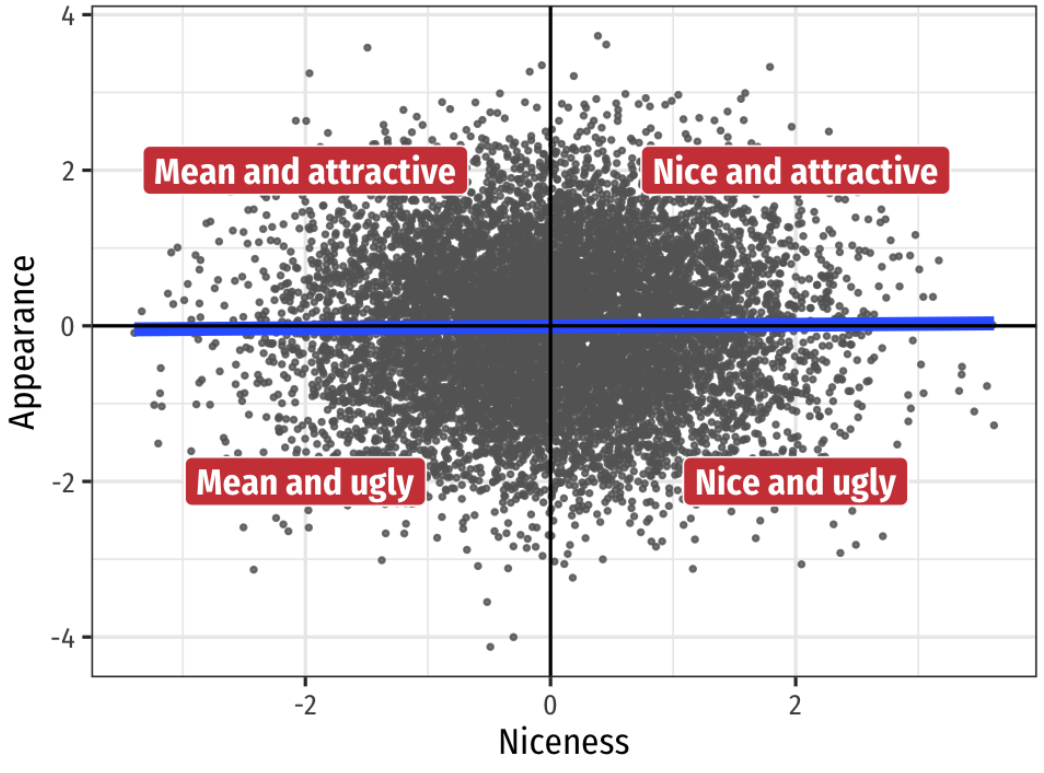
Sept. 10, 2021, 3:58 p.m. ET
By Davey Alba

Facebook sent flawed data to misinformation researchers.

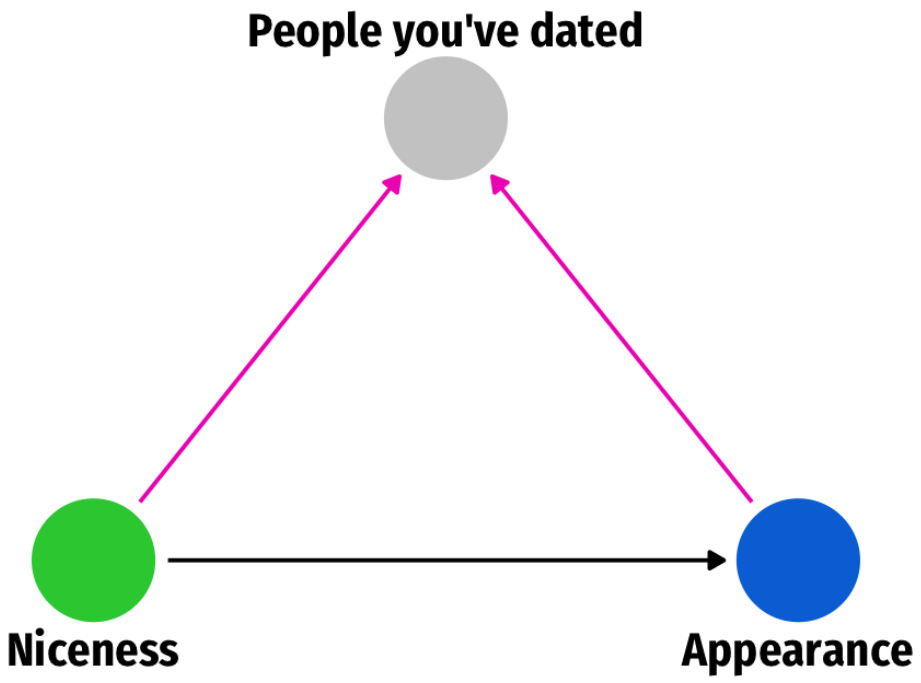
Mark Zuckerberg, chief executive of Facebook, testifying in Washington in 2018. Tom Brenner/The New York Times

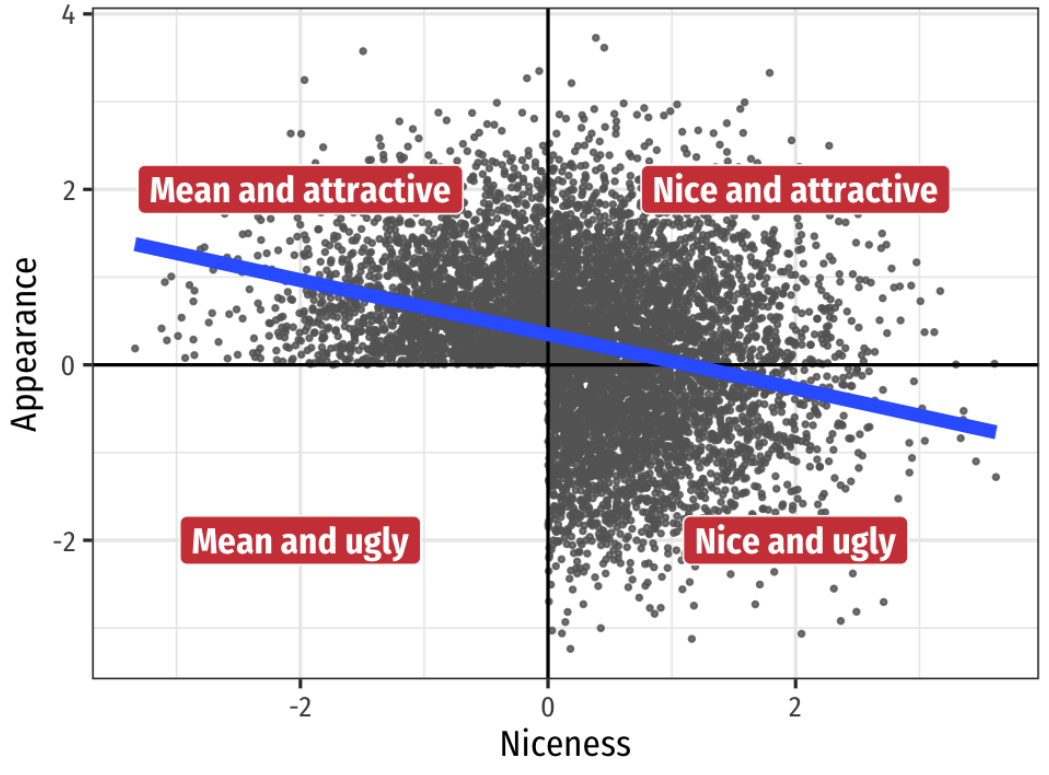
Does niceness improve appearance?





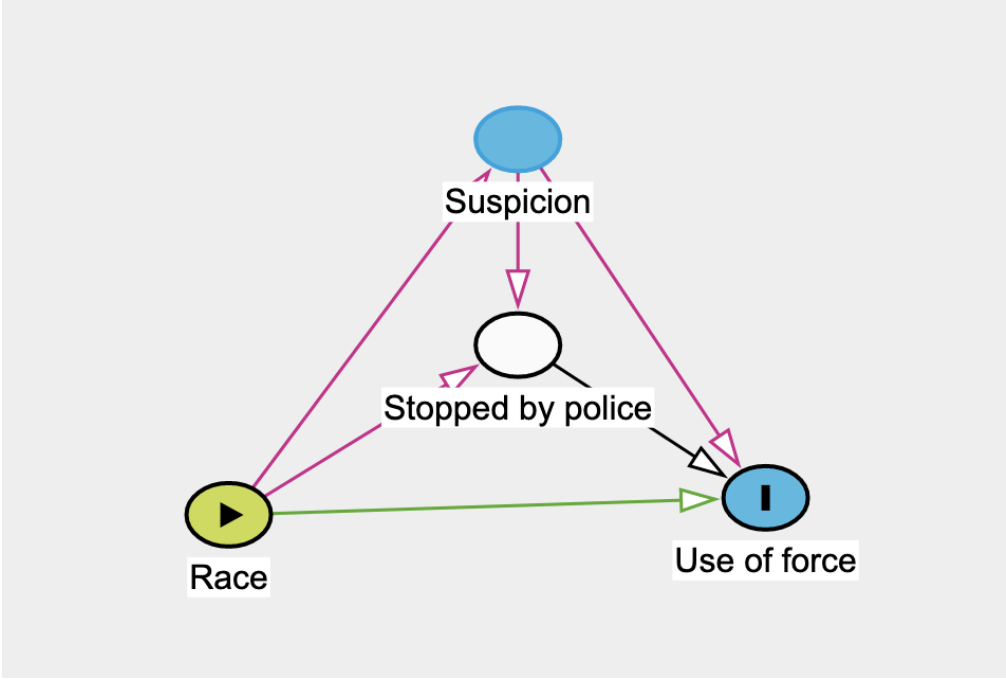
Collider distorts the true effect!





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

Administrative Records Mask Racially Biased Policing

DEAN KNOX *Princeton University*
WILL LOWE *Hertie School of Governance*
JONATHAN MUMMOLO *Princeton University*

Researchers 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-

Potential outcomes vs. do() notation

$$\mathrm{E}(\cdot), \mathbf{E}(\cdot), \mathbb{E}(\cdot) \quad \textbf{vs.} \quad \mathrm{P}(\cdot)$$

Causal effects with potential outcomes

Potential outcomes notation:

$$\delta = \frac{1}{n} \sum_{i=1}^{n} Y_i(1) - Y_i(0)$$

or alternatively with $\mathbf{E}$

$$\delta = \mathbf{E}[Y_i(1) - Y_i(0)]$$

Causal effects with do()

Pearl notation:

$$\delta = \mathbf{E}[Y_i \mid \text{do}(X = 1) - Y_i \mid \text{do}(X = 0)]$$

or more simply

$$\delta = \mathbf{E}[Y_i \mid \text{do}(X)]$$

$$\delta \quad = \quad \mathbf{E}[Y_i \mid \mathrm{do}(X)] \quad = \quad \mathbf{E}[Y_i(1) - Y_i(0)]$$

We can't see this

$$\mathbf{E}[Y_i \mid \text{do}(X)] \quad \text{or} \quad \mathbf{E}[Y_i(1) - Y_i(0)]$$

So we find the average causal effect (ACE)

$$\hat{\delta} = \mathbf{E}[Y_i \mid X = 1] - \mathbf{E}[Y_i \mid X = 0]$$

The average population-level change in y when *directly intervening* (or doing) x

$$\mathbf{E}(y \mid \mathrm{do}(x)) \neq \mathbf{E}(y \mid x)$$

The average population-level change in y when accounting for *observed* x

Causation

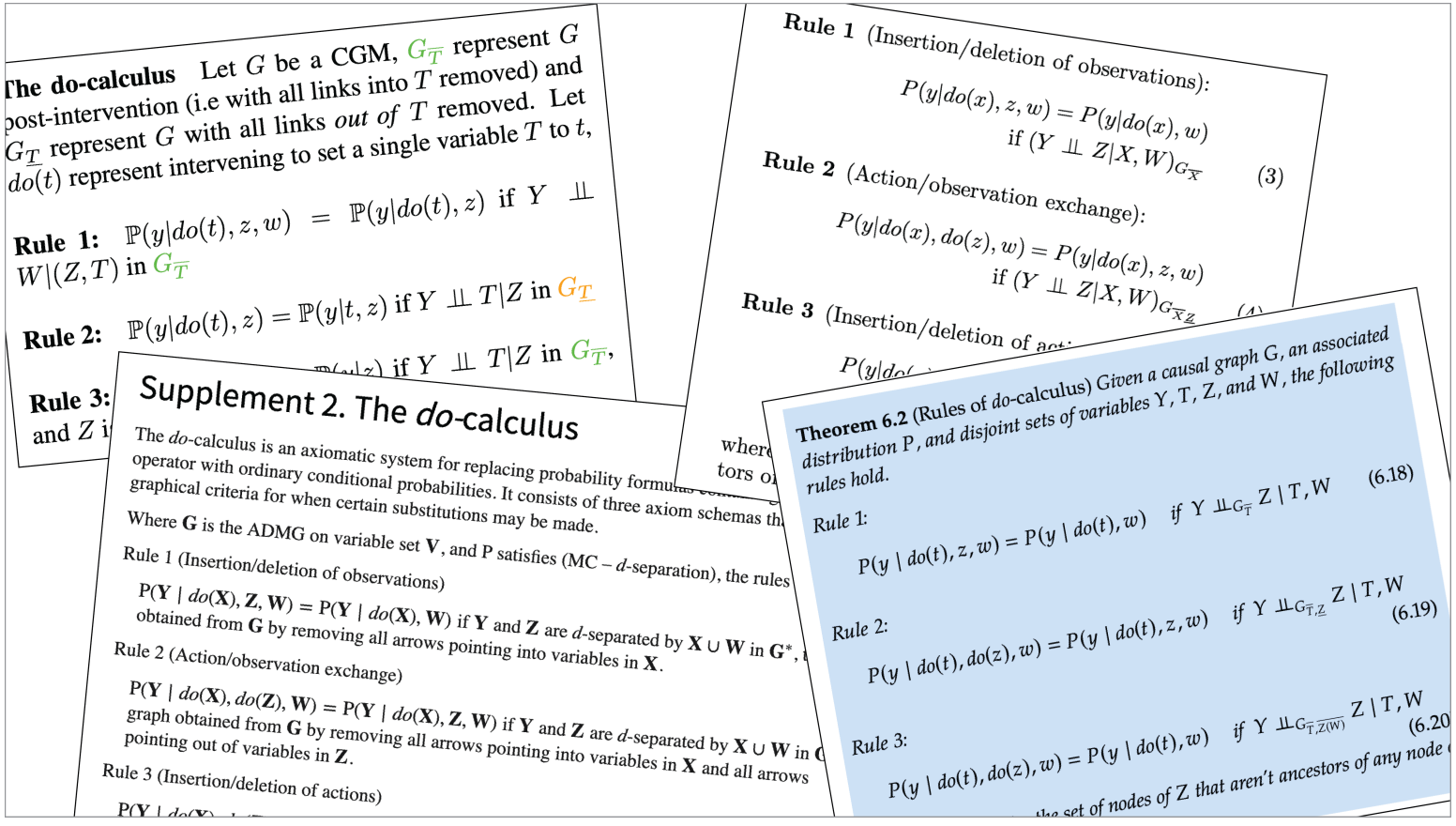
Correlation

do-calculus, adjustment, and CATEs

How can you create a DAG/transform the expression and do all the backdoor adjustments to determine causality if you can't possibly think of all variables that may impact your outcome variable?

Where can we learn more about *do*-calculus?

Here!



The do-calculus Let G be a CGM, $G_{\overline{T}}$ represent G post-intervention (i.e with all links into T removed) and $G_{\underline{T}}$ represent G with all links *out of* T removed. Let $do(t)$ represent intervening to set a single variable T to t,

Rule 1: $\mathbb{P}(y|do(t),z,w) = \mathbb{P}(y|do(t),z)$ if $Y \perp\!\!\!\perp W|(Z,T)$ in $G_{\overline{T}}$

Rule 2: $\mathbb{P}(y|do(t),z) = \mathbb{P}(y|t,z)$ if $Y \perp\!\!\!\perp T|Z$ in $G_{\underline{T}}$

Rule 3: $\mathbb{P}(y|do(t),z) = \mathbb{P}(y|do(t))$ if $Y \perp\!\!\!\perp T|Z$ in $G_{\overline{T}}$, and Z i

Rule 1 (Insertion/deletion of observations):

$$P(y|do(x),z,w) = P(y|do(x),w) \quad \text{if } (Y \perp\!\!\!\perp Z|X,W)_{G_{\overline{X}}} \tag{3}$$

Rule 2 (Action/observation exchange):

$$P(y|do(x),do(z),w) = P(y|do(x),z,w) \quad \text{if } (Y \perp\!\!\!\perp Z|X,W)_{G_{\overline{X}\underline{Z}}}$$

Rule 3 (Insertion/deletion of

where

tors o

Supplement 2. The *do*-calculus

The *do*-calculus is an axiomatic system for replacing probability formu
operator with ordinary conditional probabilities. It consists of three axiom schemas th
graphical criteria for when certain substitutions may be made.

Where $\mathbf{G}$ is the ADMG on variable set $\mathbf{V}$, and P satisfies (MC – d-separation), the rules

Rule 1 (Insertion/deletion of observations)

$P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{Z}, \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X} \cup \mathbf{W}$ in $\mathbf{G}^*$,
obtained from $\mathbf{G}$ by removing all arrows pointing into variables in $\mathbf{X}$.

Rule 2 (Action/observation exchange)

$P(\mathbf{Y} \mid do(\mathbf{X}), do(\mathbf{Z}), \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{Z}, \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X} \cup \mathbf{W}$ in
graph obtained from $\mathbf{G}$ by removing all arrows pointing into variables in $\mathbf{X}$ and all arrows
pointing out of variables in $\mathbf{Z}$.

Rule 3 (Insertion/deletion of actions)

Theorem 6.2 (Rules of do-calculus) *Given a causal graph* G*, an associated distribution* P*, and disjoint sets of variables* Y*,* T*,* Z*, and* W*, the following rules hold.*

Rule 1:

$$P(y \mid do(t), z, w) = P(y \mid do(t), w) \quad \text{if } Y \perp\!\!\!\perp_{G_{\overline{T}}} Z \mid T, W \tag{6.18}$$

Rule 2:

$$P(y \mid do(t), do(z), w) = P(y \mid do(t), z, w) \quad \text{if } Y \perp\!\!\!\perp_{G_{\overline{T},\underline{Z}}} Z \mid T, W \tag{6.19}$$

Rule 3:

$$P(y \mid do(t), do(z), w) = P(y \mid do(t), w) \quad \text{if } Y \perp\!\!\!\perp_{G_{\overline{T},\overline{Z(W)}}} Z \mid T, W \tag{6.20}$$

the set of nodes of Z that aren't ancestors of any node

Adjusting for backdoor confounding

Adjusting for frontdoor confounding

When you're making groups for CATE, how do you decide what groups to put people in?

Logic models, DAGs, and measurement

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

Berkeley Will Fully Close Its Streets to Create Giant Outdoor Dining Rooms

3

Berkeley is moving fast to expand outdoor dining

by Eve Batey | May 14, 2020, 1:02pm PDT

SHARE

How cities can add accessible green space in a post-coronavirus world

June 11, 2020 3.40pm EDT

Cities can prepare for climate change emergencies by adding green spaces to help manage stormwater, heat stress and air quality. (Shutterstock)

The COVID-19 pandemic has forced governments to weigh the benefits of keeping green spaces open against the public health concerns that come from their use. During the pandemic, playgrounds have been taped off, parks locked and access to outdoor spaces for recreation cut off.

Green spaces have positive effects on mental health, physical fitness, social cohesion and spiritual wellness. Although researchers say the coronavirus spreads more easily indoors than outdoors, they also believe the concentrated use of green spaces will increase the transmission of COVID-19.

Authors

Ryan Plummer
Professor, Environmental Sustainability Research Centre, Brock University

Darby McGrath
Adjunct professor, Environmental Sustainability Research Centre, Brock University

Sivajanani Sivarajah
Research Associate, Department of Architectural Science, Ryerson University