What is Your Estimand?

Defining the Target Quantity Connects Statistical Evidence to Theory

Ian Lundberg
UCLA
Sociology
ianlundberg.org

Rebecca Johnson
Dartmouth Quantitative
Social Science
rebeccajohnson.io

Brandon M. Stewart
Princeton
Sociology
brandonstewart.org

23 September 2021.

Replication code on Dataverse. Research reported in this publication was supported by The Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number P2CHD047879.

There is a question every quantitative study must answer:

There is a question every quantitative study must answer:

What is your estimand?

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

There is a question every quantitative study must answer:

What is your estimand?

A common answer:

The **purpose** of the statistical analysis

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

β_1 is an estimand that assumes a model

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

What if the model is wrong?

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

β_1 is an estimand that assumes a model

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

β_1 is an estimand that assumes a model

What if the model is wrong?

The model is an approximation

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

β_1 is an estimand that assumes a model

What if the model is wrong?

So β_1 is an approximation to...

The model is an approximation

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

β_1 is an estimand that assumes a model

What if the model is wrong?

The model is an approximation

So β_1 is an approximation to...

Something. I just haven't said what.

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

$\uparrow$

β_1 is an estimand that assumes a model

What if the model is wrong?

So β_1 is an approximation to...

Is it a good approximation?

The model is an approximation

Something. I just haven't said what.

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A common answer:

— We took [data source]

— We estimated β_1

$$Y = \beta_0 + X_1 \underset{\substack{\uparrow \\ \beta_1 \text{ is an estimand} \\ \text{that assumes a model}}}{\beta_1} + X_2\beta_2 + \epsilon$$

What if the model is wrong?

So β_1 is an approximation to...

Is it a good approximation?

The model is an approximation

Something. I just haven't said what.

Epistemological crisis

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

A **unit-specific quantity**

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

Y_i

A **unit-specific quantity**

There is a question every quantitative study must answer:

What is your estimand?

The **purpose** of the statistical analysis

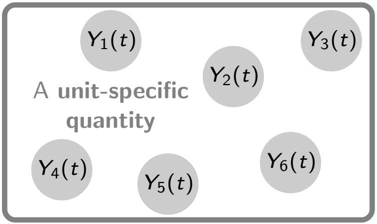
$Y_i(t)$

A **unit-specific quantity**

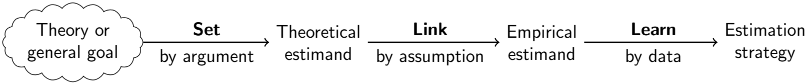
There is a question every quantitative study must answer:

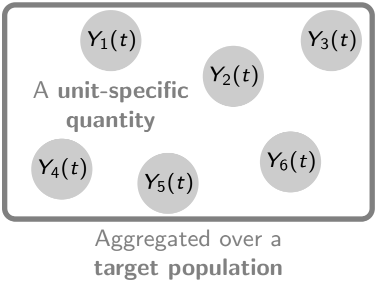
What is your estimand?

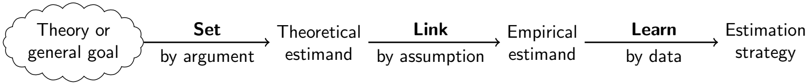
The **purpose** of the statistical analysis



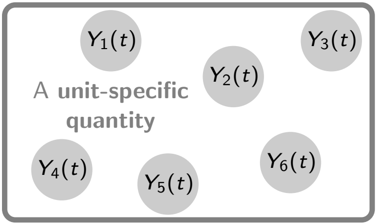
Aggregated over a **target population**


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy

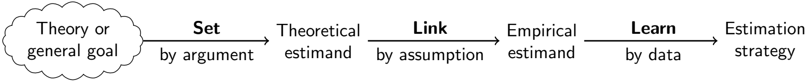

$Y_1(t)$
$Y_2(t)$
$Y_3(t)$
A **unit-specific quantity**
$Y_4(t)$
$Y_5(t)$
$Y_6(t)$
Aggregated over a **target population**



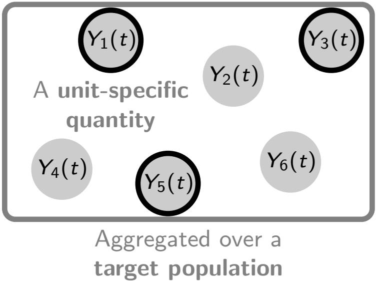
Set

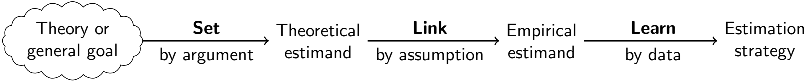


Aggregated over a **target population**


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy

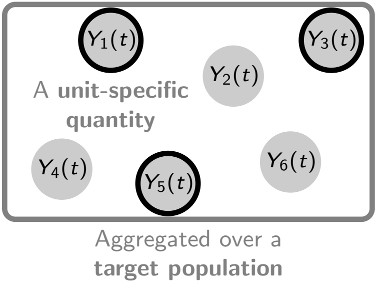
Set


$Y_1(t)$
$Y_2(t)$
$Y_3(t)$
A **unit-specific quantity**
$Y_4(t)$
$Y_5(t)$
$Y_6(t)$
Aggregated over a **target population**

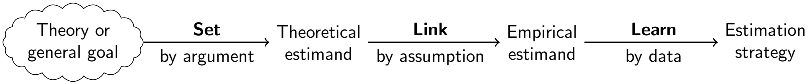


Set

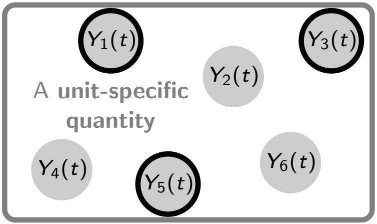
Link



$$\vec{X} \to T \to Y$$



Set



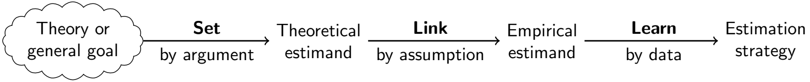
Aggregated over a **target population**

Link

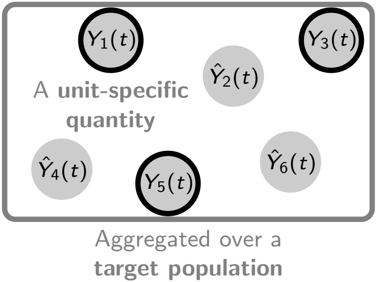
$$\begin{tikzcd} \vec{X} \arrow[r] \arrow[rr, bend left] & T \arrow[r] & Y \end{tikzcd}$$

Learn

$$\hat{Y}_i(t) = \hat{f}(\vec{X}_i, t)$$



Set



Link

$$\begin{tikzcd} \vec{X} \arrow[r] \arrow[rr, bend left] & T \arrow[r] & Y \end{tikzcd}$$

Learn

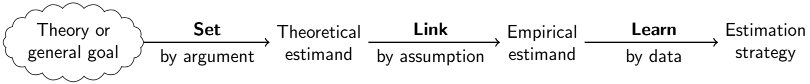
$$\hat{Y}_i(t) = \hat{f}(\vec{X}_i, t)$$

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

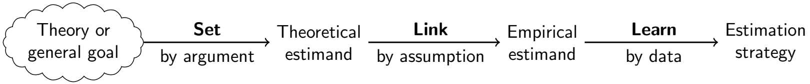
Roadmap of the talk:

1) Examples of how things can go wrong

2) An example that follows our framework from top to bottom



Example 1: An influential study with a narrow theoretical estimand



Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Effect of motherhood
on employment

Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Effect of motherhood
on employment

First two births
are the same sex

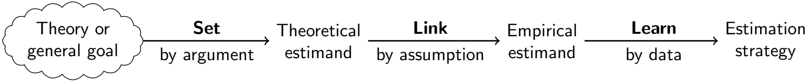
Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

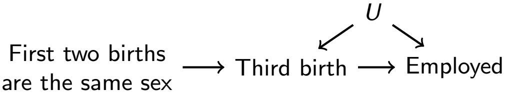
Effect of motherhood on employment

First two births are the same sex $\longrightarrow$ Third birth

Angrist and Evans 1998



Effect of motherhood
on employment



Angrist and Evans 1998

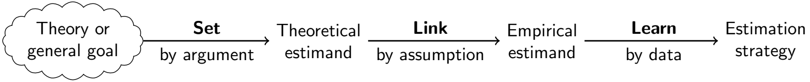
Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Vague estimand

Effect of motherhood
on employment

$$
\begin{tikzcd}
 & & U \arrow[dl] \arrow[dr] & \\
\text{First two births are the same sex} \arrow[r] & \text{Third birth} \arrow[rr] & & \text{Employed}
\end{tikzcd}
$$

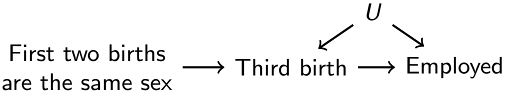
Angrist and Evans 1998



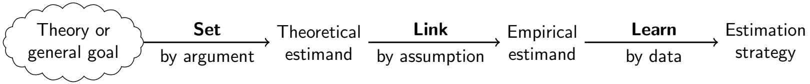
Vague estimand

Effect of motherhood on employment

Precise estimand



Angrist and Evans 1998



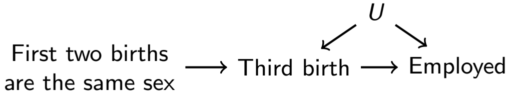
unit-specific quantity

Vague estimand

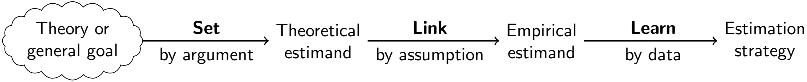
Effect of motherhood on employment

Precise estimand

Effect of having **3 vs. 2 children**



Angrist and Evans 1998



Vague estimand

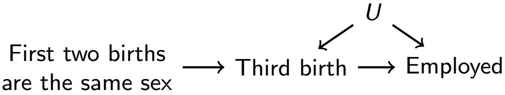
Effect of motherhood on employment

Precise estimand

Effect of having 3 vs. 2 children

[among those with at least two children who would have a third birth if and only if the first two were of the same sex]

target population



Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Precise estimand

Effect of having 3 vs. 2 children

among those with at least two children who would have a third birth if and only if the first two were of the same sex

$\approx$ 4% of all mothers

Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Precise estimand

Effect of having 3 vs. 2 children

among those with at least two children who would have a third birth if and only if the first two were of the same sex

$\approx$ 4% of all mothers

You have to argue either:

1)

2)

Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Precise estimand

Effect of having 3 vs. 2 children

among those with at least two children who would have a third birth if and only if the first two were of the same sex

$\approx$ 4% of all mothers

You have to argue either:

1) That estimand matters for theory, or

2)

Angrist and Evans 1998

Precise estimand

Effect of having 3 vs. 2 children

among those with at least two children who would have a third birth if and only if the first two were of the same sex

$\approx$ 4% of all mothers

You have to argue either:

1) That estimand matters for theory, or

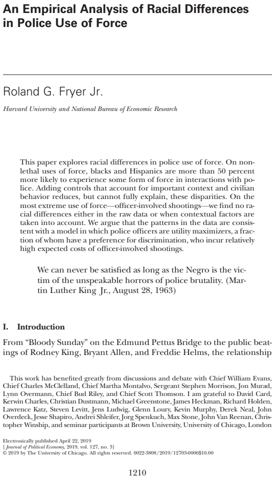
2) It speaks to some broader estimand

Angrist and Evans 1998

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Example 2: An influential study with a misleading link to evidence

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy



An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

This paper explores racial differences in police use of force. On nonlethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of whom have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

> We can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality. (Martin Luther King Jr., August 28, 1963)

I. Introduction

From "Bloody Sunday" on the Edmund Pettus Bridge to the public beatings of Rodney King, Bryant Allen, and Freddie Helms, the relationship

This work has benefited greatly from discussions and debate with Chief William Evans, Chief Charles McClelland, Chief Martha Montalvo, Sergeant Stephen Morrison, Jon Murad, Lynn Overmann, Chief Bud Riley, and Chief Scott Thomson. I am grateful to David Card, Kerwin Charles, Christian Dustmann, Michael Greenstone, James Heckman, Richard Holden, Lawrence Katz, Steven Levitt, Jens Ludwig, Glenn Loury, Kevin Murphy, Derek Neal, John Overdeck, Jesse Shapiro, Andrei Shleifer, Jorg Spenkuch, Max Stone, John Van Reenan, Christopher Winship, and seminar participants at Brown University, University of Chicago, London

Electronically published April 22, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 3]

Theory or general goal —**Set** by argument→ Theoretical estimand —**Link** by assumption→ Empirical estimand —**Learn** by data→ Estimation strategy

It is the most surprising result of my career.
— Roland Fryer

An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

This paper explores racial differences in police use of force. On nonlethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of whom have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

> We can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality. (Martin Luther King Jr., August 28, 1963)

I. Introduction

From "Bloody Sunday" on the Edmund Pettus Bridge to the public beatings of Rodney King, Bryant Allen, and Freddie Helms, the relationship

This work has benefited greatly from discussions and debate with Chief William Evans, Chief Charles McClelland, Chief Martha Montalvo, Sergeant Stephen Morrison, Jon Murad, Lynn Overmann, Chief Bud Riley, and Chief Scott Thomson. I am grateful to David Card, Kerwin Charles, Christian Dustmann, Michael Greenstone, James Heckman, Richard Holden, Lawrence Katz, Steven Levitt, Jens Ludwig, Glenn Loury, Kevin Murphy, Derek Neal, John Overdeck, Jesse Shapiro, Andrei Shleifer, Jorg Spenkuch, Max Stone, John Van Reenan, Christopher Winship, and seminar participants at Brown University, University of Chicago, London

Electronically published April 22, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 3]

1210

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

> It is the most surprising result of my career.
> — Roland Fryer

An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

This paper explores racial differences in police use of force. On nonlethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of whom have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

> We can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality. (Martin Luther King Jr., August 28, 1963)

I. Introduction

From "Bloody Sunday" on the Edmund Pettus Bridge to the public beatings of Rodney King, Bryant Allen, and Freddie Helms, the relationship

This work has benefited greatly from discussions and debate with Chief William Evans, Chief Charles McClelland, Chief Martha Montalvo, Sergeant Stephen Morrison, Jon Murad, Lynn Overmann, Chief Bud Riley, and Chief Scott Thomson. I am grateful to David Card, Kerwin Charles, Christian Dustmann, Michael Greenstone, James Heckman, Richard Holden, Lawrence Katz, Steven Levitt, Jens Ludwig, Glenn Loury, Kevin Murphy, Derek Neal, John Overdeck, Jesse Shapiro, Andrei Shleifer, Jorg Spenkuch, Max Stone, John Van Reenan, Christopher Winship, and seminar participants at Brown University, University of Chicago, London

Electronically published April 22, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 3]

TheUpshot

DATA DIVE

Surprising New Evidence Shows Bias in Police Use of Force but Not in Shootings

Theory or general goal → **Set** by argument → Theoretical estimand → **Link** by assumption → Empirical estimand → **Learn** by data → Estimation strategy

> It is the most surprising result of my career.
> — Roland Fryer

An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

This paper explores racial differences in police use of force. On nonlethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of whom have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

> We can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality. (Martin Luther King Jr., August 28, 1963)

I. Introduction

From "Bloody Sunday" on the Edmund Pettus Bridge to the public beatings of Rodney King, Bryant Allen, and Freddie Helms, the relationship

This work has benefited greatly from discussions and debate with Chief William Evans, Chief Charles McClelland, Chief Martha Montalvo, Sergeant Stephen Morrison, Jon Murad, Lynn Overmann, Chief Bud Riley, and Chief Scott Thomson. I am grateful to David Card, Kerwin Charles, Christian Dustmann, Michael Greenstone, James Heckman, Richard Holden, Lawrence Katz, Steven Levitt, Jens Ludwig, Glenn Loury, Kevin Murphy, Derek Neal, John Overdeck, Jesse Shapiro, Andrei Shleifer, Jorg Spenkuch, Max Stone, John Van Reenan, Christopher Winship, and seminar participants at Brown University, University of Chicago, London

Electronically published April 22, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 3]

1210

TheUpshot

DATA DIVE

Surprising New Evidence Shows Bias in Police Use of Force but Not in Shootings

OPINION | COMMENTARY

The Myth of Systemic Police Racism

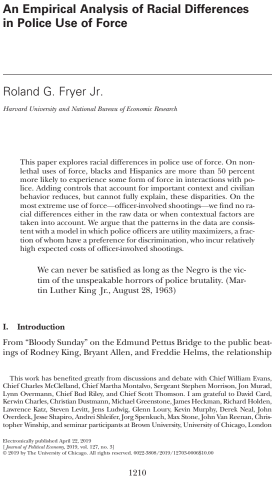
WSJ | OPINION

Hold officers accountable who use excessive force. But there's no evidence of widespread racial bias.

By Heather Mac Donald
June 2, 2020 1:44 pm ET

Theory or general goal —**Set** by argument→ Theoretical estimand —**Link** by assumption→ Empirical estimand —**Learn** by data→ Estimation strategy

> It is the most surprising result of my career.
> — Roland Fryer



An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

This paper explores racial differences in police use of force. On nonlethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police. Adding controls that account for important context and civilian behavior reduces, but cannot fully explain, these disparities. On the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account. We argue that the patterns in the data are consistent with a model in which police officers are utility maximizers, a fraction of whom have a preference for discrimination, who incur relatively high expected costs of officer-involved shootings.

> We can never be satisfied as long as the Negro is the victim of the unspeakable horrors of police brutality. (Martin Luther King Jr., August 28, 1963)

I. Introduction

From "Bloody Sunday" on the Edmund Pettus Bridge to the public beatings of Rodney King, Bryant Allen, and Freddie Helms, the relationship

This work has benefited greatly from discussions and debate with Chief William Evans, Chief Charles McClelland, Chief Martha Montalvo, Sergeant Stephen Morrison, Jon Murad, Lynn Overmann, Chief Bud Riley, and Chief Scott Thomson. I am grateful to David Card, Kerwin Charles, Christian Dustmann, Michael Greenstone, James Heckman, Richard Holden, Lawrence Katz, Steven Levitt, Jens Ludwig, Glenn Loury, Kevin Murphy, Derek Neal, John Overdeck, Jesse Shapiro, Andrei Shleifer, Jorg Spenkuch, Max Stone, John Van Reenan, Christopher Winship, and seminar participants at Brown University, University of Chicago, London

Electronically published April 22, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 3]
© 2019 by The University of Chicago. All rights reserved. 0022-3808/2019/12703-0006$10.00

1210

$$\text{Theory or general goal} \xrightarrow[\text{by argument}]{\textbf{Set}} \text{Theoretical estimand} \xrightarrow[\text{by assumption}]{\textbf{Link}} \text{Empirical estimand} \xrightarrow[\text{by data}]{\textbf{Learn}} \text{Estimation strategy}$$

Evidence:

Claim:

Why wrong:

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

Claim:

Why wrong:

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

Claim: Police are unbiased

Why wrong:

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

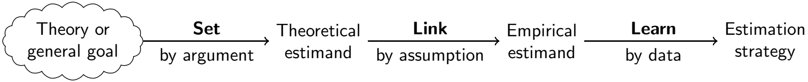
Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

Claim: Police are unbiased

Why wrong:

Perceived as black $\longrightarrow$ Lethal force

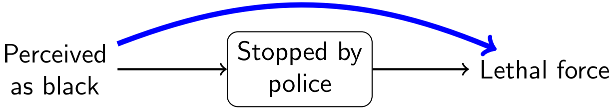
Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.



Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

Claim: Police are unbiased

Why wrong:



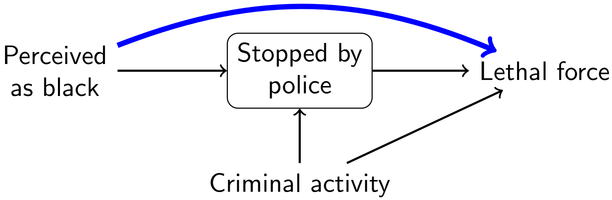
Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

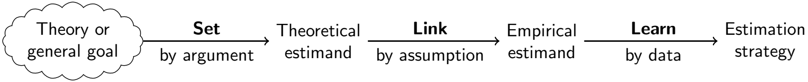
Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

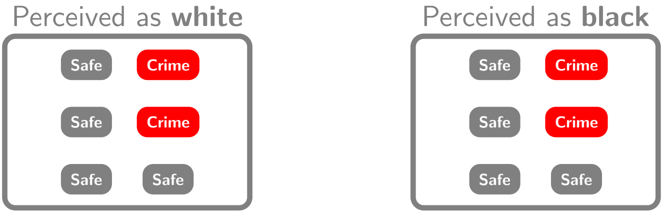
Claim: Police are unbiased

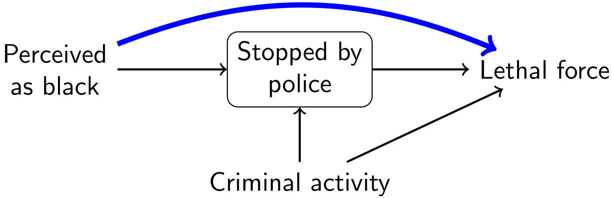
Why wrong:



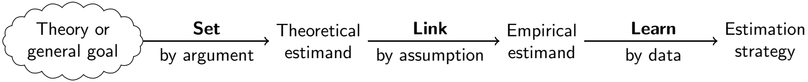
Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

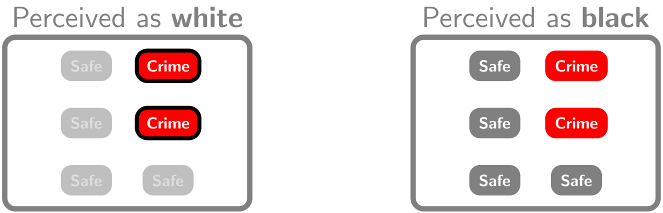


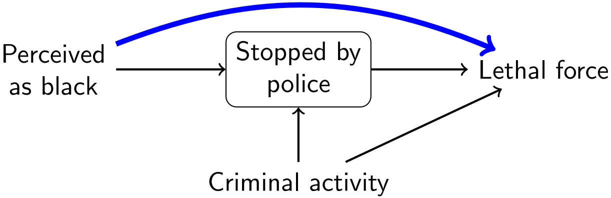




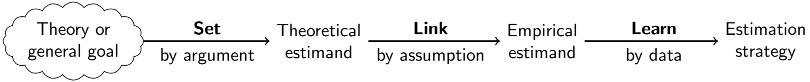
Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

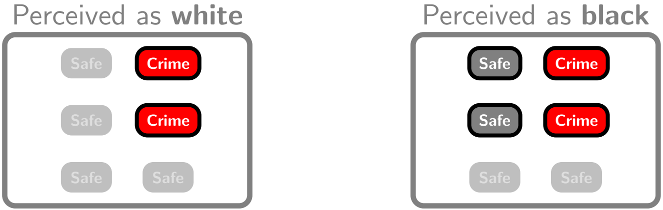


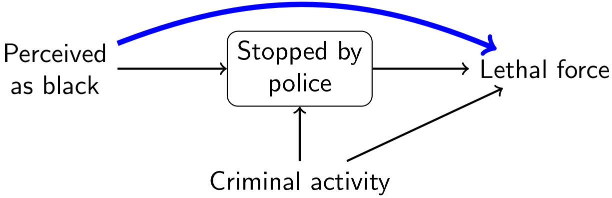




Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.







Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

Claim: Police are unbiased

Why wrong: Causal sample selection. Among those stopped, the white civilians may be more dangerous.

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

Claim: Police are unbiased

Why wrong: Causal sample selection. Among those stopped, the white civilians may be more dangerous.

Fryer responds:

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

~~**Claim:** Police are unbiased~~

Why wrong: Causal sample selection. Among those stopped, the white civilians may be more dangerous.

Fryer responds:

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

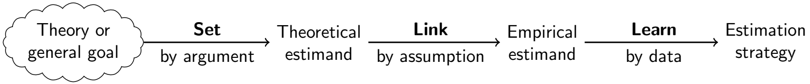
Evidence: Police use lethal force at the same rate against black and white civilians who are stopped.

~~**Claim:** Police are unbiased~~

Why wrong: Causal sample selection. Among those stopped, the white civilians may be more dangerous.

Fryer responds: "We use the term 'racial differences' 114 times in lieu of the more prescriptive wording—'racial discrimination.' We use the phrase 'conditional on an interaction' 20 times...I am not sure how many more ways we would have needed to caveat our results to satisfy [the critics]."

Fryer 2019. Fuller critique by Knox et al. 2020 and Durlauf and Heckman 2020.



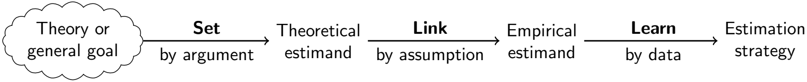
Example 3: An influential study where estimation led to confusion

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

> The emergence of a female advantage in education is attributable to a reversal in the gender-specific effects of father status.
>
> — Buchmann and DiPrete 2006

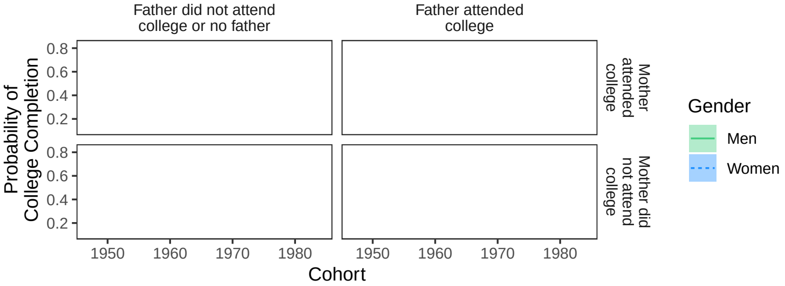
	β	(SE)
Birth Cohort 1966+ (vs. 1938–1965)	.318	(.285)
Female	–.136	(.133)
Later Cohorts × Female	–.107	(.272)
Mother Some College	.737**	(.134)
Later Cohorts × Mother Some College	.079	(.218)
No Father Present	–.031	(.129)
Father Some College	1.285**	(.113)
Later Cohorts × No Father	–.107	(.226)
Later Cohorts × Father Some College	–.390	(.211)
Mother Some College × Female	.120	(.147)
No Father Present × Female		
Father Some College × Female		
Mother Some College × No Father	.108	(.208)
Mother Some College × Father Some College	.150	(.138)
No Father or Father ≤HS × Male	.303*	(.143)
No Father or Father ≤HS × Male × Later Cohorts	–.801**	(.293)
Mother Some College × Female × Later Cohorts	.221	(.295)
No Father × Female × Later Cohorts		
Father Some College × Male × Later Cohorts		
Age Main Effects	(omitted)	
2- and 3-way Interactions between Age and (Gender, Cohort)	(omitted)	
Constant	1.695**	(.140)
N	7,024	
df	15	

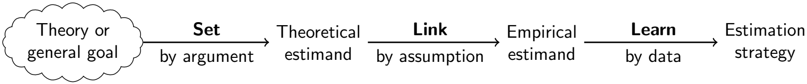
	β	(SE)
Birth Cohort 1966+ (vs. 1938–1965)	.318	(.285)
Female	–.136	(.133)
Later Cohorts × Female	–.107	(.272)
Mother Some College	.737**	(.134)
Later Cohorts × Mother Some College	.079	(.218)
No Father Present	–.031	(.129)
Father Some College	1.285**	(.113)
Later Cohorts × No Father	–.107	(.226)
Later Cohorts × Father Some College	–.390	(.211)
Mother Some College × Female	.120	(.147)
No Father Present × Female		
Father Some College × Female		
Mother Some College × No Father	.108	(.208)
Mother Some College × Father Some College	.150	(.138)
No Father or Father ≤HS × Male	.303*	(.143)
No Father or Father ≤HS × Male × Later Cohorts	–.801**	(.293)
Mother Some College × Female × Later Cohorts	.221	(.295)
No Father × Female × Later Cohorts		
Father Some College × Male × Later Cohorts		
Age Main Effects	(omitted)	
2- and 3-way Interactions between Age and (Gender, Cohort)	(omitted)	
Constant	1.695**	(.140)
N	7,024	
df	15	



Coefficient:
Gender × Cohort
× Father status

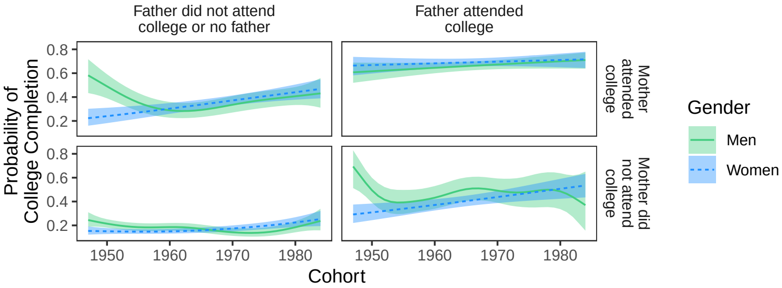
> The emergence of a female advantage in education is attributable to a reversal in the gender-specific effects of father status.
> — Buchmann and DiPrete 2006

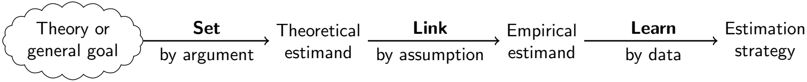




Coefficient:
Gender × Cohort × Father status

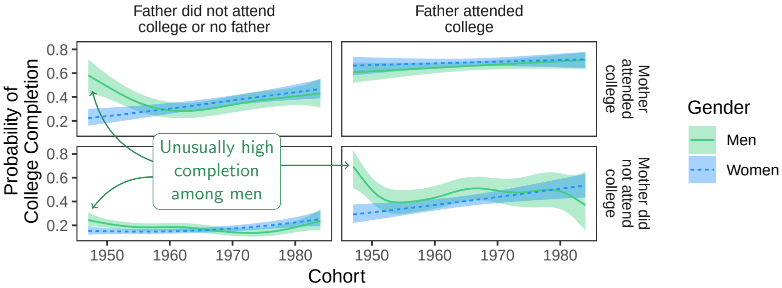
> The emergence of a female advantage in education is attributable to a reversal in the gender-specific effects of father status.
> — Buchmann and DiPrete 2006





Coefficient:
Gender × Cohort × Father status

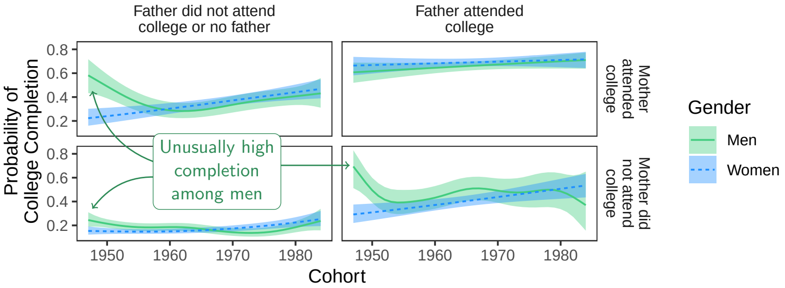
> The emergence of a female advantage in education is attributable to a reversal in the gender-specific effects of father status.
>
> — Buchmann and DiPrete 2006



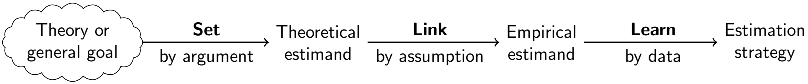
Theory or general goal → **Set** by argument → Theoretical estimand → **Link** by assumption → Empirical estimand → **Learn** by data → Estimation strategy

Coefficient:
Gender × Cohort
× Father status

> The emergence of a female advantage in education is attributable to a reversal in the gender-specific effects of father status.
> — Buchmann and DiPrete 2006



Alternate theory: The Vietnam War



Meta-example: Vague estimands are widespread

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Review
All 32 articles in ASR 2018 using quantitative data

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Review
All 32 articles in ASR 2018 using quantitative data

Clarity about **unit-specific quantity**

3 types of articles with distinct issues

- 18 articles | 56% of *ASR* → Causal target: Observational data
- 7 articles | 22% of *ASR* → Causal target: Experiments
- 7 articles | 22% of *ASR* → Descriptive target

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Review
All 32 articles in ASR 2018 using quantitative data

Clarity about **unit-specific quantity**

3 types of articles with distinct issues

18 articles | 56% of *ASR*

Causal target: Observational data

0 out of 18
articles was clear

7 articles | 22% of *ASR*

Causal target: Experiments

4 out of 7
articles were clear

7 articles | 22% of *ASR*

Descriptive target

1 out of 7
articles was clear

Theory or general goal → **Set** by argument → Theoretical estimand → **Link** by assumption → Empirical estimand → **Learn** by data → Estimation strategy

Review
All 32 articles in ASR 2018 using quantitative data

Clarity about **unit-specific quantity**

3 types of articles with distinct issues

18 articles · 56% of *ASR*	7 articles · 22% of *ASR*	7 articles · 22% of *ASR*
Causal target: Observational data	Causal target: Experiments	Descriptive target
0 out of 18 articles was clear	**4 out of 7** articles were clear	**1 out of 7** articles was clear

Clarity about the **target population**

3 types of articles with distinct issues

16 articles · 50% of *ASR*	6 articles · 19% of *ASR*	10 articles · 31% of *ASR*
Probability samples	Administrative records	Other samples

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Review
All 32 articles in ASR 2018 using quantitative data

Clarity about **unit-specific quantity**

3 types of articles with distinct issues

18 articles | 56% of *ASR*	7 articles | 22% of *ASR*	7 articles | 22% of *ASR*
Causal target: Observational data	Causal target: Experiments	Descriptive target
0 out of 18 articles was clear	**4 out of 7** articles were clear	**1 out of 7** articles was clear

Clarity about the **target population**

3 types of articles with distinct issues

16 articles | 50% of *ASR*	6 articles | 19% of *ASR*	10 articles | 31% of *ASR*
Probability samples	Administrative records	Other samples
6 out of 16 articles were clear	**1 out of 6** articles was clear	**2 out of 10** articles were clear

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Roadmap of the talk:

1) Examples of how things can go wrong
2) An example that follows our framework from top to bottom

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Roadmap of the talk:

1) Examples of how things can go wrong

2) An example that follows our framework from top to bottom

General Method

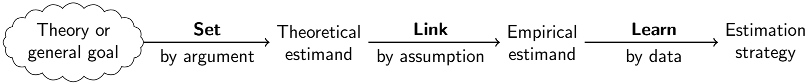
The Gap-Closing Estimand:
The Disparity Across Social Categories
That Would Persist if We Equalized a Treatment

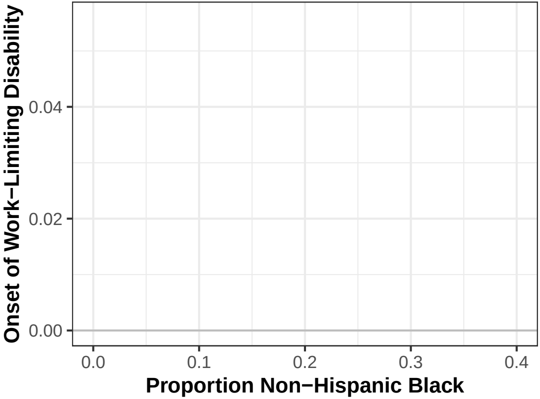
Specific Example

Quantifying the Contribution of Occupational
Segregation to Racial Disparities in Health:
A Gap-Closing Perspective

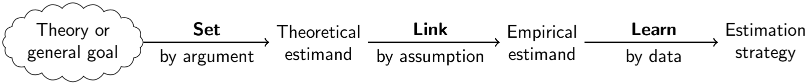
Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

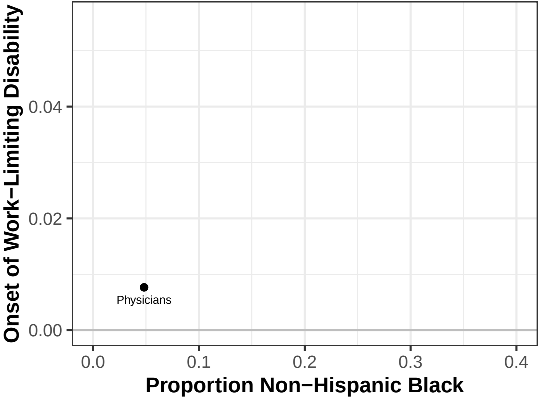
Current
Population Survey
Annual Social and
Economic Supplement
Ages 25–60
2005–2020



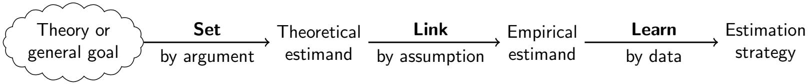


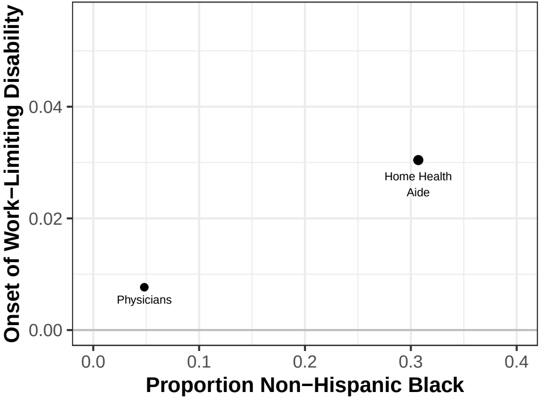
Current Population Survey Annual Social and Economic Supplement
Ages 25–60
2005–2020



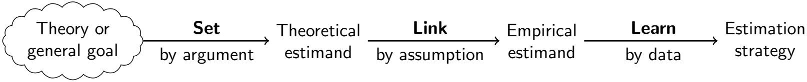


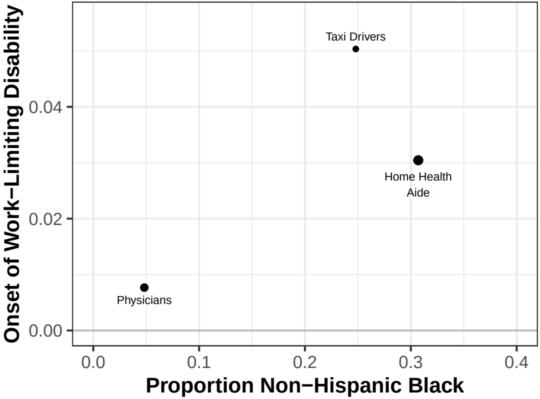
Current
Population Survey
Annual Social and
Economic Supplement
Ages 25–60
2005–2020





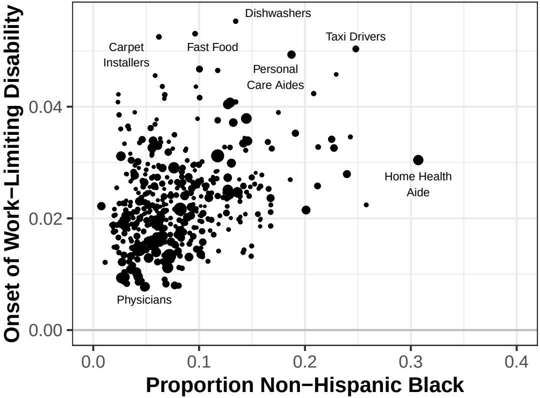
Current Population Survey Annual Social and Economic Supplement
Ages 25–60
2005–2020





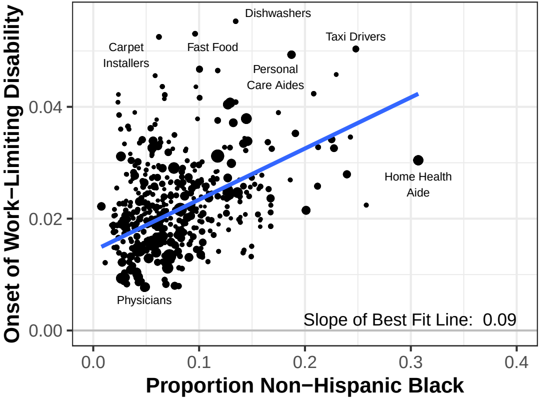
Current Population Survey Annual Social and Economic Supplement
Ages 25–60
2005–2020

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy



Current Population Survey Annual Social and Economic Supplement
Ages 25–60
2005–2020

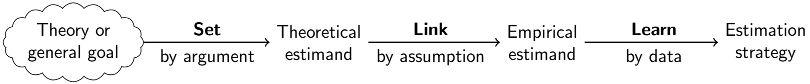
Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy



Current Population Survey Annual Social and Economic Supplement
Ages 25–60
2005–2020

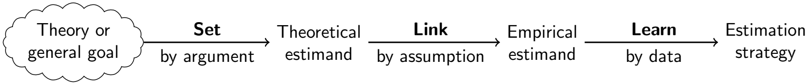
$$\text{Theory or general goal} \xrightarrow[\text{by argument}]{\textbf{Set}} \text{Theoretical estimand} \xrightarrow[\text{by assumption}]{\textbf{Link}} \text{Empirical estimand} \xrightarrow[\text{by data}]{\textbf{Learn}} \text{Estimation strategy}$$

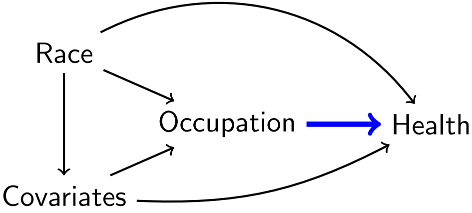
Theoretical estimand



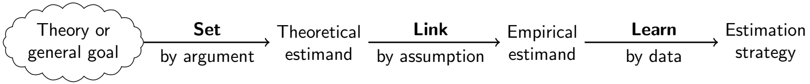
Theoretical estimand

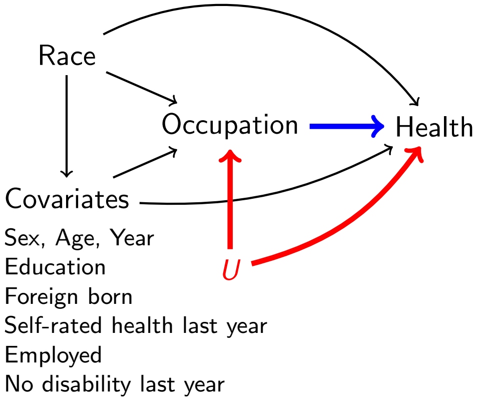
- **Unit-specific quantity:**
  Onset of work-limiting disability
  if occupations were randomly shuffled
  within educational strata
- Target population
  - U.S. Ages 25–60
  - Employed at baseline survey
  - No work-limiting disability at baseline survey
  - Never previously quit a job for health reasons

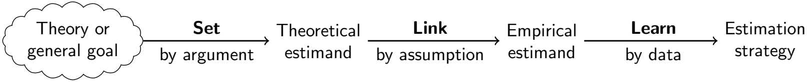




Sex, Age, Year
Education
Foreign born
Self-rated health last year
Employed
No disability last year


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy

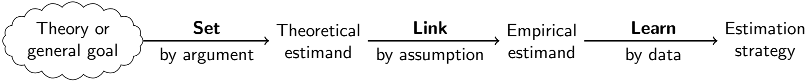

Race
Occupation
Health
Covariates
Sex, Age, Year
Education
Foreign born
Self-rated health last year
Employed
No disability last year
U



		Administrative Assistant	Home Health Aid	Sales Supervisor
White	Person 1	y		
	Person 2			y
	Person 3		y	
	Person 4		y	
Black	Person 5			y
	Person 6	y		
	Person 7	y		
	Person 8			y

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

		Administrative Assistant	Home Health Aid	Sales Supervisor
White	Person 1	y	•	•
White	Person 2	•	•	y
White	Person 3	•	y	•
White	Person 4	•	y	•
Black	Person 5	•	•	y
Black	Person 6	y	•	•
Black	Person 7	y	•	•
Black	Person 8	•	•	y



		Administrative Assistant	Home Health Aid	Sales Supervisor	Average outcome if occupations were shuffled
White	Person 1	y	•	•	•
	Person 2	•	•	y	
	Person 3	•	y	•	
	Person 4	•	y	•	
Black	Person 5	•	•	y	
	Person 6	y	•	•	
	Person 7	y	•	•	
	Person 8	•	•	y	

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

$$\hat{y}_i(t) = \hat{f}(\vec{X}_i, t)$$

		Administrative Assistant	Home Health Aid	Sales Supervisor	Average outcome if occupations were shuffled ↓
White	Person 1	y	•	•	•
	Person 2	•	•	y	
	Person 3	•	y	•	
	Person 4	•	y	•	
Black	Person 5	•	•	y	
	Person 6	y	•	•	
	Person 7	y	•	•	
	Person 8	•	•	y	

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

$$\hat{y}_i(t) = \hat{f}(\vec{X}_i, t)$$

		Administrative Assistant	Home Health Aid	Sales Supervisor	Average outcome if occupations were shuffled $\downarrow$
White	Person 1	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 2	•	•	y	
	Person 3	•	y	•	
	Person 4	•	y	•	
Black	Person 5	•	•	y	
	Person 6	y	•	•	
	Person 7	y	•	•	
	Person 8	•	•	y	

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

$$\hat{y}_i(t) = \hat{f}(\vec{X}_i, t)$$

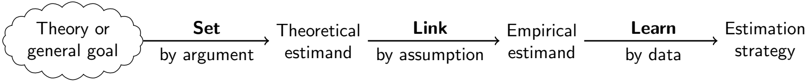
		Administrative Assistant	Home Health Aid	Sales Supervisor	Average outcome if occupations were shuffled ↓
White	Person 1	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
White	Person 2	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
White	Person 3	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
White	Person 4	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
Black	Person 5	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
Black	Person 6	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
Black	Person 7	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
Black	Person 8	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$

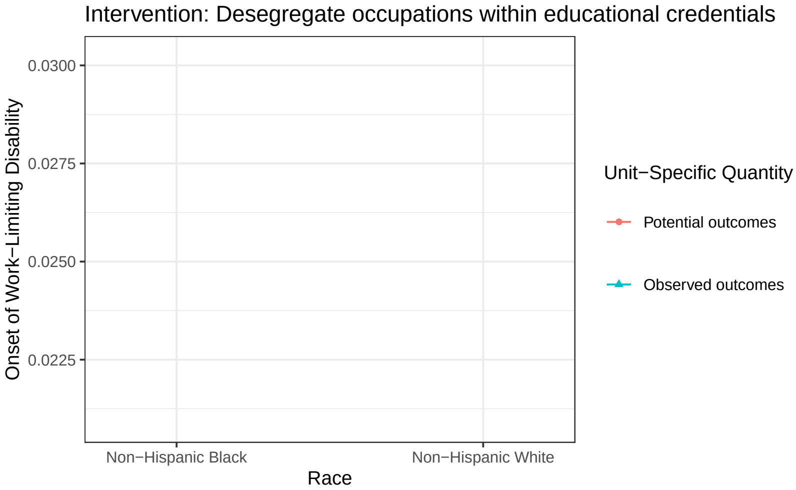
Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

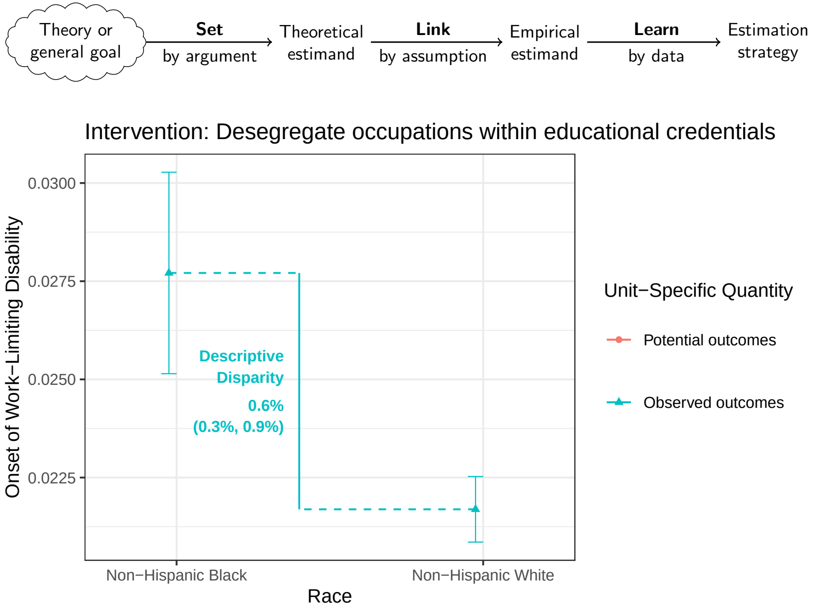
$$\hat{y}_i(t) = \hat{f}(\vec{X}_i, t)$$

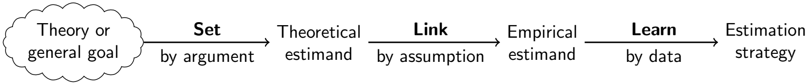
		Administrative Assistant	Home Health Aid	Sales Supervisor	Average outcome if occupations were shuffled $\downarrow$
White	Person 1	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 2	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 3	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 4	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
Black	Person 5	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 6	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 7	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$
	Person 8	$\hat{y}$	$\hat{y}$	$\hat{y}$	$\hat{\bar{y}}$

Gap-Closing Estimand

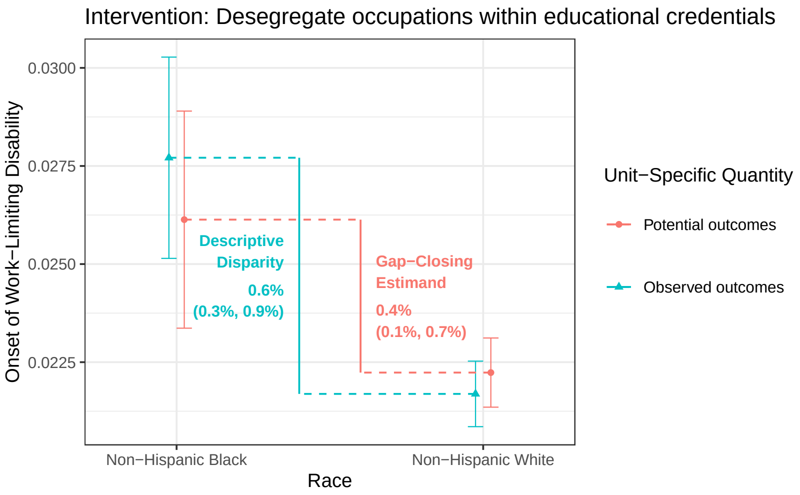

Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy


Intervention: Desegregate occupations within educational credentials
Onset of Work−Limiting Disability
0.0300
0.0275
0.0250
0.0225
Non−Hispanic Black
Non−Hispanic White
Race
Unit−Specific Quantity
Potential outcomes
Observed outcomes


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy
Intervention: Desegregate occupations within educational credentials
Onset of Work−Limiting Disability
0.0300
0.0275
0.0250
0.0225
Descriptive Disparity
0.6%
(0.3%, 0.9%)
Non−Hispanic Black
Non−Hispanic White
Race
Unit−Specific Quantity
Potential outcomes
Observed outcomes

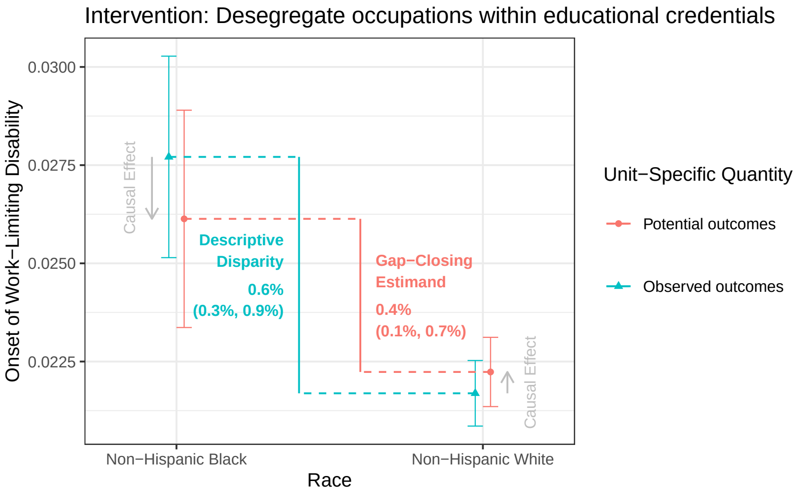


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy



Intervention: Desegregate occupations within educational credentials
Onset of Work−Limiting Disability
0.0300
0.0275
0.0250
0.0225
Descriptive Disparity
0.6%
(0.3%, 0.9%)
Gap−Closing Estimand
0.4%
(0.1%, 0.7%)
Non−Hispanic Black
Non−Hispanic White
Race
Unit−Specific Quantity
Potential outcomes
Observed outcomes

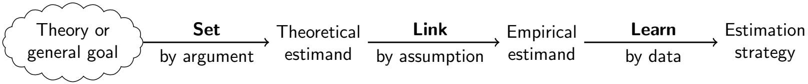
Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

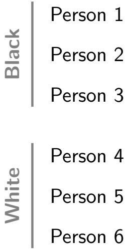


Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

Treatment Condition

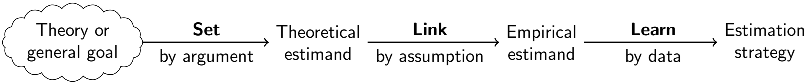
		Black	White
Population	Person 1	$Y_1(\text{Black})$	$Y_1(\text{White})$
	Person 2	$Y_2(\text{Black})$	$Y_2(\text{White})$
	Person 3	$Y_3(\text{Black})$	$Y_3(\text{White})$
	Person 4	$Y_4(\text{Black})$	$Y_4(\text{White})$
	Person 5	$Y_5(\text{Black})$	$Y_5(\text{White})$
	Person 6	$Y_6(\text{Black})$	$Y_6(\text{White})$


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy

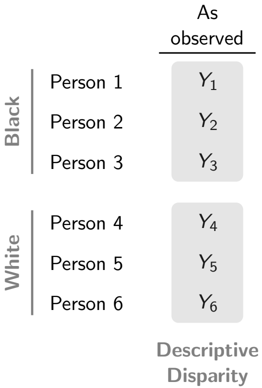

Black
Person 1
Person 2
Person 3
White
Person 4
Person 5
Person 6

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

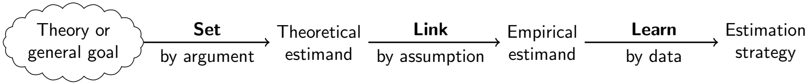
		As observed
Black	Person 1	Y_1
	Person 2	Y_2
	Person 3	Y_3
White	Person 4	Y_4
	Person 5	Y_5
	Person 6	Y_6

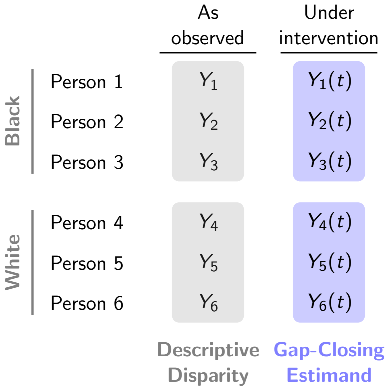


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy



As observed
Black
Person 1
Person 2
Person 3
Y_1
Y_2
Y_3
White
Person 4
Person 5
Person 6
Y_4
Y_5
Y_6
Descriptive Disparity


Theory or general goal
Set
by argument
Theoretical estimand
Link
by assumption
Empirical estimand
Learn
by data
Estimation strategy


As observed
Under intervention
Black
Person 1
Person 2
Person 3
White
Person 4
Person 5
Person 6
Y_1
Y_2
Y_3
Y_4
Y_5
Y_6
$Y_1(t)$
$Y_2(t)$
$Y_3(t)$
$Y_4(t)$
$Y_5(t)$
$Y_6(t)$
Descriptive Disparity
Gap-Closing Estimand

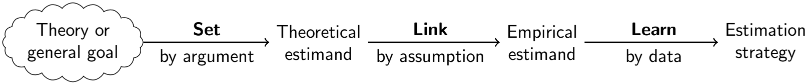
Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

		As observed	Under intervention
Black	Person 1	Y_1	$Y_1(t)$
	Person 2	Y_2	$Y_2(t)$
	Person 3	Y_3	$Y_3(t)$
White	Person 4	Y_4	$Y_4(t)$
	Person 5	Y_5	$Y_5(t)$
	Person 6	Y_6	$Y_6(t)$
		Descriptive Disparity	**Gap-Closing Estimand**

Vanderweele & Robinson 2014

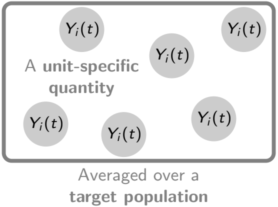
Jackson & Vanderweele 2018

Lundberg 2021



What is your estimand?

⟵ Every quantitative study should answer this question

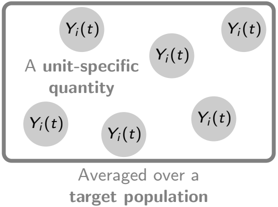


Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

What is your estimand?

⟵ Every quantitative study should answer this question

When you **write** a quantitative paper, the estimand allows you to



Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

What is your estimand?

⟵ Every quantitative study should answer this question

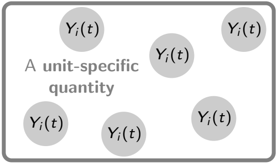
When you **write** a quantitative paper, the estimand allows you to

— Motivate the question

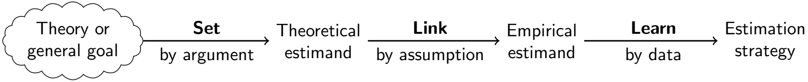
— Address selection

— Use machine learning

— Speak to a broad audience



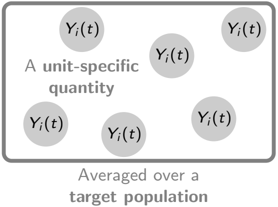
Averaged over a **target population**



What is your estimand?

⟵ Every quantitative study should answer this question

When you **read** a quantitative paper, the estimand allows you to



Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

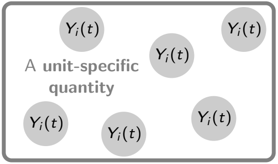
What is your estimand?

⟵ Every quantitative study should answer this question

When you **read** a quantitative paper, the estimand allows you to

— Understand the author's aim

— Pinpoint your concerns



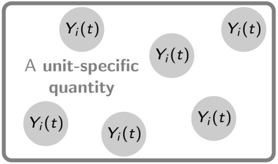
Averaged over a **target population**

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

What is your estimand?

$\longleftarrow$ Every quantitative study should answer this question

In the future, estimands will only become more important



Averaged over a **target population**

Theory or general goal → **Set** by argument → Theoretical estimand → **Link** by assumption → Empirical estimand → **Learn** by data → Estimation strategy

What is your estimand?

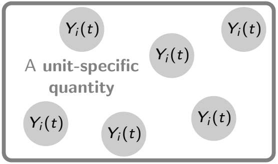
⟵ Every quantitative study should answer this question

In the future, estimands will only become more important

New data have missing values

— Non-probability samples

— Administrative records



Averaged over a **target population**

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

What is your estimand?

⟵ Every quantitative study should answer this question

In the future, estimands will only become more important

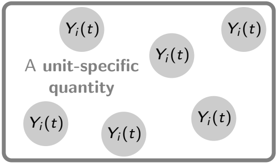
New data have missing values

— Non-probability samples

— Administrative records

New methods flourish with a clear goal

— Machine learning



Averaged over a **target population**

Theory or general goal $\xrightarrow[\text{by argument}]{\textbf{Set}}$ Theoretical estimand $\xrightarrow[\text{by assumption}]{\textbf{Link}}$ Empirical estimand $\xrightarrow[\text{by data}]{\textbf{Learn}}$ Estimation strategy

What is your estimand?

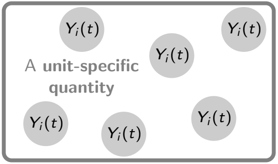
⟵ Every quantitative study should answer this question

In the future, estimands will only become more important

New data have missing values

— Non-probability samples

— Administrative records



Averaged over a **target population**

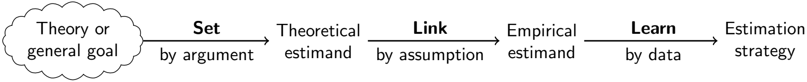
New methods flourish with a clear goal

— Machine learning

New questions can be found beyond coefficients

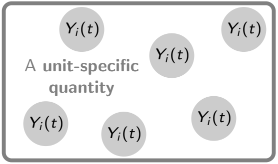
— Describe counterfactual disparities

— Predict the effects of targeted interventions



What is your estimand?

⟵ Every quantitative study should answer this question



Averaged over a **target population**

Ian Lundberg

ilundberg@princeton.edu
ianlundberg.org

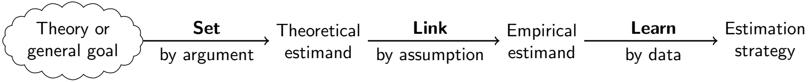
Rebecca Johnson

rebecca.ann.johnson@
dartmouth.edu
rebeccajohnson.io

Brandon Stewart

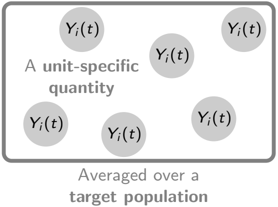
bms4@princeton.edu
brandonstewart.org

American Sociological Review, 2021.
Open access on SocArxiv
Code on Dataverse



What is your estimand?

⟵ Every quantitative study should answer this question



Ian Lundberg

ilundberg@princeton.edu
ianlundberg.org

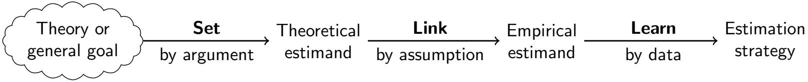
Rebecca Johnson

rebecca.ann.johnson@
dartmouth.edu
rebeccajohnson.io

Brandon Stewart

bms4@princeton.edu
brandonstewart.org

American Sociological Review, 2021.
Open access on SocArxiv
Code on Dataverse



What is your estimand?

⟵ Every quantitative study should answer this question

Defining the Target Quantity Connects Statistical Evidence to Theory

Ian Lundberg

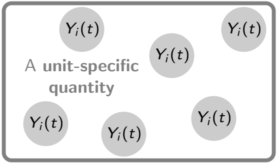
ilundberg@princeton.edu
ianlundberg.org

Rebecca Johnson

rebecca.ann.johnson@
dartmouth.edu
rebeccajohnson.io

Brandon Stewart

bms4@princeton.edu
brandonstewart.org



Averaged over a **target population**

American Sociological Review, 2021.
Open access on SocArxiv
Code on Dataverse