

20. Principal Stratification (Part 2)

Ian Lundberg

Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

3 Nov 2022

Learning goals for today

At the end of class, you will be able to:

1. Finish the [class exercise](#) we started on Tuesday [[solutions](#)]
2. See principal stratification in action:
quantifying racial bias in policing

Administrative Records Mask Racially Biased Policing

DEAN KNOX *Princeton University*

WILL LOWE *Hertie School of Governance*

JONATHAN MUMMOLO *Princeton University*

A police officer encounters a person

A police officer encounters a person

1. Stop them? Or not?

A police officer encounters a person

1. Stop them? Or not?
2. Use force? Or not?

A police officer encounters a person

1. Stop them? Or not?
2. Use force? Or not?

Effect of race:

Would the outcome of this encounter differ if the civilian were of a different race

A police officer encounters a person

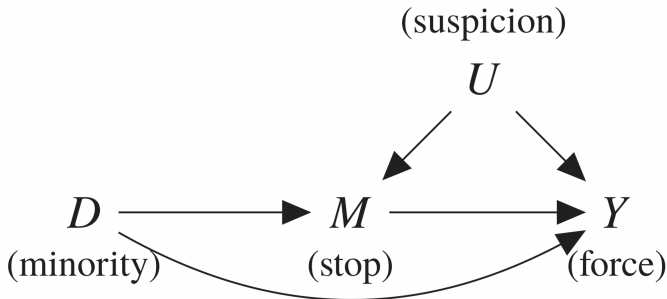
1. Stop them? Or not?
2. Use force? Or not?

Effect of race:

Would the outcome of this encounter differ
if the civilian were of a different race

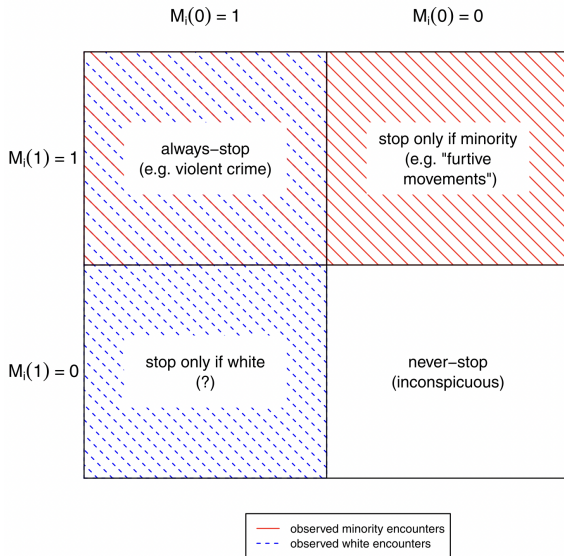
Unit of analysis is an **encounter** not a **person**

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police



Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

FIGURE 2. Principal Strata and Observed Police–Civilian Encounters



We would want the ATE

$$E(Y^{1M^1} - Y^{0M^0})$$

To estimate that, the authors say we need two things

1. Count of minority encounters¹
2. Count of white encounters

within strata of X

¹(including all four strata)

Point estimates

Note: All steps are within X . Notation dropped.

Point estimates

Note: All steps are within X . Notation dropped.

Important caveat:

The following is my reconstruction of one of the simplest of many results in Knox, Lowe, & Mummolon 2020.

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1)$$

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

$$= \underbrace{P(M = 1 \mid D = 1)}_{\text{}} \underbrace{E(Y \mid D = 1, M = 1)}_{\text{}} + \underbrace{P(M = 0 \mid D = 1)}_{\text{}} \underbrace{E(Y \mid D = 1, M = 0)}_{\text{}}$$

Law of Total
Probability

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

$$= \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)} + \underbrace{P(M = 0 \mid D = 1)} \underbrace{E(Y \mid D = 1, M = 0)}$$

Law of Total Probability

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

$$= \underbrace{P(M = 1 \mid D = 1)}_{\text{Stop rate among minority encounters}} \underbrace{E(Y \mid D = 1, M = 1)}_{\text{Use of force among stopped minority encounters}} + \underbrace{P(M = 0 \mid D = 1)}_{\text{Stop rate among minority encounters}} \underbrace{E(Y \mid D = 1, M = 0)}_{\text{Use of force among stopped minority encounters}}$$

Law of Total Probability

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

$$\begin{aligned} &= \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}} \\ &\quad + \underbrace{P(M = 0 \mid D = 1)}_{\text{Non-stop rate among minority encounters}} \underbrace{E(Y \mid D = 1, M = 0)} \end{aligned}$$

Law of Total Probability

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

$$\begin{aligned} &= \underbrace{P(M = 1 \mid D = 1)}_{\text{Stop rate among minority encounters}} \underbrace{E(Y \mid D = 1, M = 1)}_{\text{Use of force among stopped minority encounters}} \\ &\quad + \underbrace{P(M = 0 \mid D = 1)}_{\text{Non-stop rate among minority encounters}} \underbrace{E(Y \mid D = 1, M = 0)}_{\text{Use of force among non-stopped minority encounters (=0)}} \end{aligned}$$

Law of Total Probability

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1)$$

Exchangeability

$$= E(Y \mid D = 1)$$

Consistency

$$\begin{aligned} &= \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}} \\ &\quad + \overbrace{P(M = 0 \mid D = 1)}^{\text{Non-stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 0)}^{\text{Use of force among non-stopped minority encounters (=0)}} \end{aligned}$$

Law of Total Probability

$$= \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

$$E(Y^0) = \overbrace{P(M = 1 \mid D = 0)}^{\text{Stop rate among non-minority encounters}} \overbrace{E(Y \mid D = 0, M = 1)}^{\text{Use of force among stopped non-minority encounters}}$$

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

$$E(Y^0) = \overbrace{P(M = 1 \mid D = 0)}^{\text{Stop rate among non-minority encounters}} \overbrace{E(Y \mid D = 0, M = 1)}^{\text{Use of force among stopped non-minority encounters}}$$

Difference is the ATE.

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

$$E(Y^0) = \overbrace{P(M = 1 \mid D = 0)}^{\text{Stop rate among non-minority encounters}} \overbrace{E(Y \mid D = 0, M = 1)}^{\text{Use of force among stopped non-minority encounters}}$$

Difference is the ATE.

You just needed to augment the data with stop rates!

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

$$E(Y^0) = \overbrace{P(M = 1 \mid D = 0)}^{\text{Stop rate among non-minority encounters}} \overbrace{E(Y \mid D = 0, M = 1)}^{\text{Use of force among stopped non-minority encounters}}$$

Difference is the ATE.

You just needed to augment the data with stop rates!

Works because of two key factors:

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

$$E(Y^0) = \overbrace{P(M = 1 \mid D = 0)}^{\text{Stop rate among non-minority encounters}} \overbrace{E(Y \mid D = 0, M = 1)}^{\text{Use of force among stopped non-minority encounters}}$$

Difference is the ATE.

You just needed to augment the data with stop rates!

Works because of two key factors:

- Race is assumed exchangeable given X

Point estimates

Note: All steps are within X . Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = \overbrace{P(M = 1 \mid D = 1)}^{\text{Stop rate among minority encounters}} \overbrace{E(Y \mid D = 1, M = 1)}^{\text{Use of force among stopped minority encounters}}$$

vs if they involved a non-minority civilian?

$$E(Y^0) = \overbrace{P(M = 1 \mid D = 0)}^{\text{Stop rate among non-minority encounters}} \overbrace{E(Y \mid D = 0, M = 1)}^{\text{Use of force among stopped non-minority encounters}}$$

Difference is the ATE.

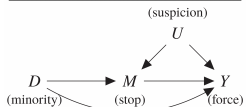
You just needed to augment the data with stop rates!

Works because of two key factors:

- ▶ Race is assumed exchangeable given X
- ▶ When $M = 0$ (no stop), then $Y = 0$ (no force)

Many possible estimands

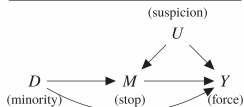
FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police



Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

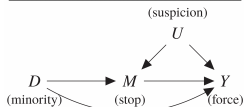


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

► ATE: $E(Y^{1M^1} - Y^{0M^0})$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

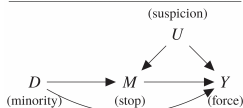


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - Racial bias, where non-stops are coded $Y = 0$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

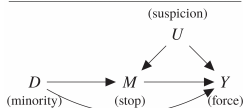


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ▶ ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - ▶ Racial bias, where non-stops are coded $Y = 0$
- ▶ CDE: $E(Y^{11} - Y^{01})$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

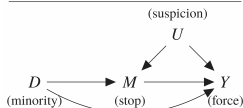


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ▶ ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - ▶ Racial bias, where non-stops are coded $Y = 0$
- ▶ CDE: $E(Y^{11} - Y^{01})$
 - ▶ Racial bias if we stopped everyone

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

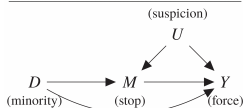


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ▶ ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - ▶ Racial bias, where non-stops are coded $Y = 0$
- ▶ CDE: $E(Y^{11} - Y^{01})$
 - ▶ Racial bias if we stopped everyone
- ▶ ATE among the stopped

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

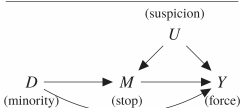


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ▶ ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - ▶ Racial bias, where non-stops are coded $Y = 0$
- ▶ CDE: $E(Y^{11} - Y^{01})$
 - ▶ Racial bias if we stopped everyone
- ▶ ATE among the stopped
 - ▶ $ATE_{M=1} = E(Y^{1M^1} | M = 1) - E(Y^{0M^0} | M = 1)$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

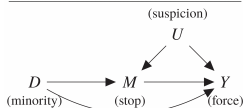


Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ▶ ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - ▶ Racial bias, where non-stops are coded $Y = 0$
- ▶ CDE: $E(Y^{11} - Y^{01})$
 - ▶ Racial bias if we stopped everyone
- ▶ ATE among the stopped
 - ▶ $ATE_{M=1} = E(Y^{1M^1} | M = 1) - E(Y^{0M^0} | M = 1)$
- ▶ Proportion of minority stops due to race

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police



Notes: Observed X is left implicit; these covariates may be causally prior to any subset of D , M , and Y .

- ▶ ATE: $E(Y^{1M^1} - Y^{0M^0})$
 - ▶ Racial bias, where non-stops are coded $Y = 0$
- ▶ CDE: $E(Y^{11} - Y^{01})$
 - ▶ Racial bias if we stopped everyone
- ▶ ATE among the stopped
 - ▶ $ATE_{M=1} = E(Y^{1M^1} | M = 1) - E(Y^{0M^0} | M = 1)$
- ▶ Proportion of minority stops due to race
 - ▶ $E(Y^{1M^1} - Y^{0M^0} | D = 1, M = 1)$

Many estimands: Necessary Assumptions

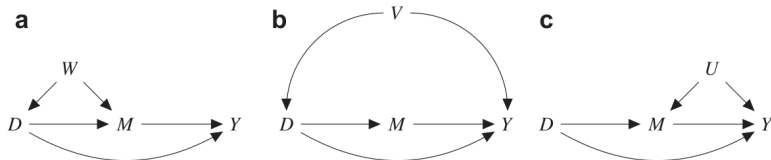
1. Mandatory reporting: $Y_i^{d0} = 0$ for all i and d
2. Mediator monotonicity: $M_i^1 \geq M_i^0$
3. Relative nonseverity of racial stops:

$$\begin{aligned} & \text{Always Stop Stratum} \\ & E(Y^{dm} \mid D = d', \overbrace{M^1 = 1, M^0 = 1}^{\text{Always Stop Stratum}}, X) \\ & \geq E(Y^{dm} \mid D = d', \underbrace{M^1 = 1, M^0 = 0}_{\text{Racial Stop Stratum}}, X) \end{aligned}$$

4. Treatment ignorability
 - ▶ $M^d \perp\!\!\!\perp D \mid X$
 - ▶ $Y^{dm} \perp\!\!\!\perp D \mid M^0, M^1, X$

Many Estimands: Necessary Assumptions

Assume absence of W and V . Ok to have U .



Many Estimands: Strong (As In Doubtful) Assumptions

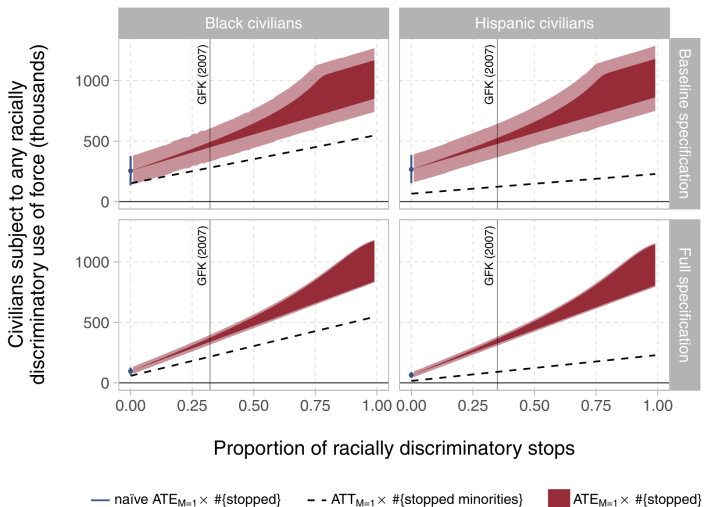
Studies about the effect of race conditional on an interaction implicitly assume these things:

1. Mediator ignorability: $Y^{dm} \perp\!\!\!\perp M^0 \mid D = d, M^1 = 1, X$
 - ▶ “violence rates in always-stop encounters must be identical to those in observationally equivalent racial stops”
2. No racial stops: $M^0 = M^1 \mid M = 1$
 - ▶ “all reported encounters were of the always-stop kind”

Knox, Lowe, & Mummolo argue that the above are implausible assumptions in the context of policing.

Without the strong assumptions, things can be learned

FIGURE 4. Bounds for Racially Discriminatory Use of Force, any Severity



Learning goals for today

At the end of class, you will be able to:

1. Finish the [class exercise](#) we started on Tuesday [[solutions](#)]
2. See principal stratification in action:
quantifying racial bias in policing

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!