Studying Social Inequality with Data Science

INFO 3370 / 5371
Spring 2023

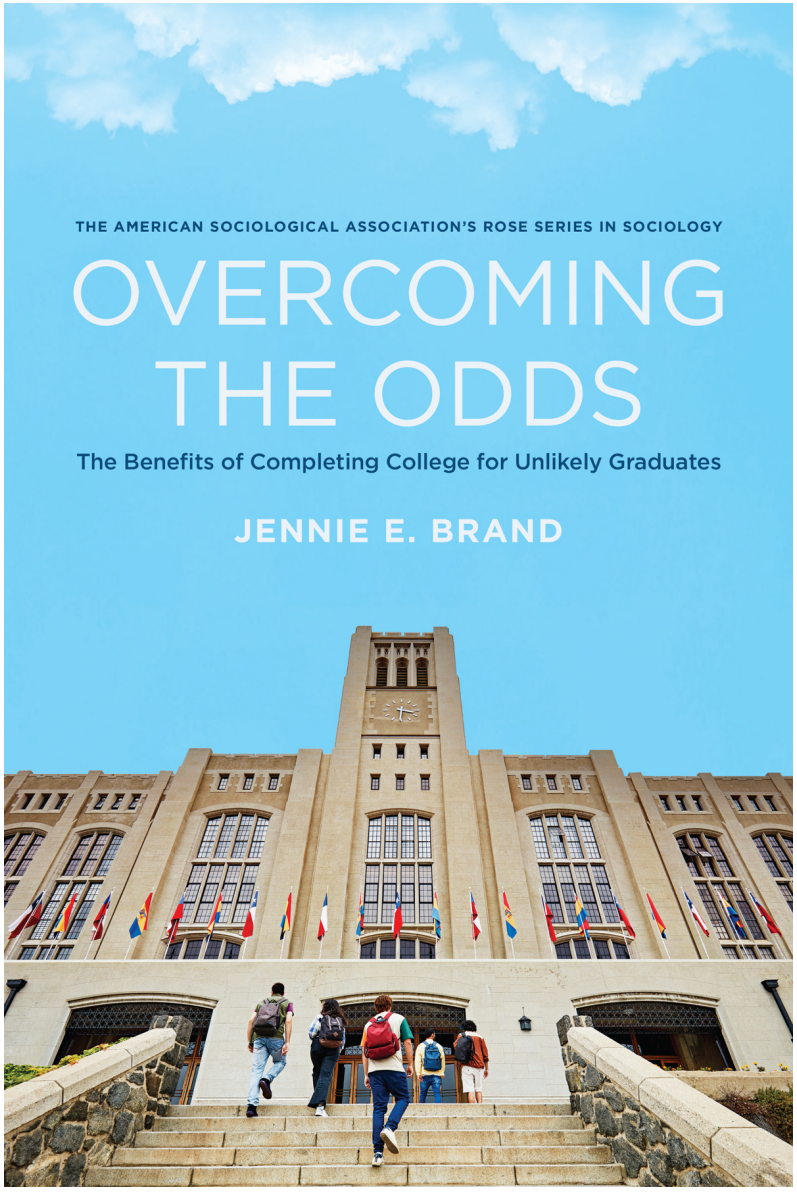
Interventions to Promote Equality: Educational Expansion

Learning goals for today

By the end of class, you will be able to

- study education as an intervention that might change inequality

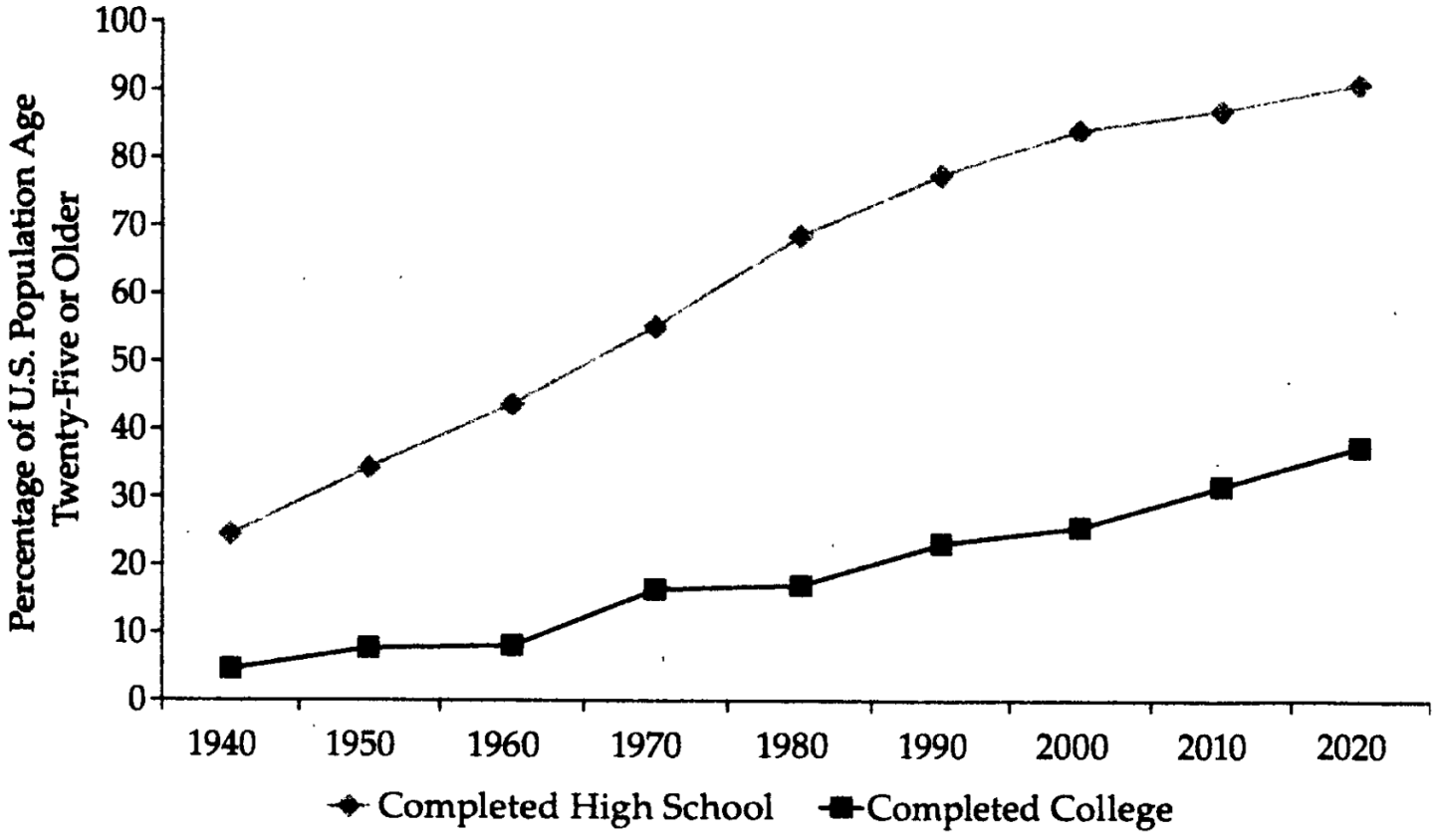
A motivating research study



Americans' education in 1900 (Brand 2023 p. 6)

- 6% graduated from high school
- 3% graduated from college

Figure 1.1 High School and Four-Year College Completion Rates, 1940–2020



Source: U.S. Census Bureau, March Current Population Survey and Annual Social and Economic Supplement to the Current Population Survey.

(Brand 2023)

Why did education expand?

Why did education expand?

- Public investment in college
  - Morrill Act (1862) sold land to establish colleges
  - G.I. Bill (1944) funded veterans' college

Why did education expand?

- Public investment in college
  - Morrill Act (1862) sold land to establish colleges
  - G.I. Bill (1944) funded veterans' college
- Rising labor market demand for skills

Investment changes: From institutions to students

Investment changes: From institutions to students

- Higher Education Act (1965) created Pell Grants
- Federal student loans

Investment changes: From institutions to students

- Higher Education Act (1965) created Pell Grants
- Federal student loans

Consequences

Investment changes: From institutions to students

- Higher Education Act (1965) created Pell Grants
- Federal student loans

Consequences

1. Growing perception that individuals (not states) ought to bear the cost of college

Investment changes: From institutions to students

- Higher Education Act (1965) created Pell Grants
- Federal student loans

Consequences

1. Growing perception that individuals (not states) ought to bear the cost of college
2. Privatization of college

Investment changes: From institutions to students

- Higher Education Act (1965) created Pell Grants
- Federal student loans

Consequences

1. Growing perception that individuals (not states) ought to bear the cost of college
2. Privatization of college

Open question: Should we invest more? How would we decide?

We would like to know whether **college pays off**:
does it have positive effects on desired outcomes?

Does college pay off? A hypothetical experiment

Does college pay off? A hypothetical experiment

- randomly sample high school students

Does college pay off? A hypothetical experiment

- randomly sample high school students
- randomize them to either
  - stop education after a high school degree, or
  - complete a 4-year degree

Does college pay off? A hypothetical experiment

- randomly sample high school students
- randomize them to either
  - stop education after a high school degree, or
  - complete a 4-year degree
- compare adult wages

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	
Top quartile of family income	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	
Top quartile of family income	
College prep courses	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	
Top quartile of family income	
College prep courses	
Outcome: Low wage job in 0% of adulthood	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	Neither parent finished high school
Top quartile of family income	
College prep courses	
Outcome: Low wage job in 0% of adulthood	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	Neither parent finished high school
Top quartile of family income	Bottom quartile of family income
College prep courses	
Outcome: Low wage job in 0% of adulthood	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	Neither parent finished high school
Top quartile of family income	Bottom quartile of family income
College prep courses	No college prep courses
Outcome: Low wage job in 0% of adulthood	

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Nick	Javier
4-year degree	**No degree**
Both parents finished college	Neither parent finished high school
Top quartile of family income	Bottom quartile of family income
College prep courses	No college prep courses
Outcome: Low wage job in 0% of adulthood	**Outcome:** Low wage job in 80% of adulthood

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Rich	Nick	Javier
No degree	**4-year degree**	**No degree**
	Both parents finished college	Neither parent finished high school
	Top quartile of family income	Bottom quartile of family income
	College prep courses	No college prep courses
	Outcome: Low wage job in 0% of adulthood	**Outcome:** Low wage job in 80% of adulthood

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Rich	Nick	Javier
No degree	**4-year degree**	**No degree**
Both parents finished college	Both parents finished college	Neither parent finished high school
Top quartile of family income	Top quartile of family income	Bottom quartile of family income
College prep courses	College prep courses	No college prep courses
	Outcome: Low wage job in 0% of adulthood	**Outcome:** Low wage job in 80% of adulthood

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Rich	Nick	Javier
No degree	**4-year degree**	**No degree**
Both parents finished college	Both parents finished college	Neither parent finished high school
Top quartile of family income	Top quartile of family income	Bottom quartile of family income
College prep courses	College prep courses	No college prep courses
Outcome: Low wage job in 73% of adulthood	**Outcome:** Low wage job in 0% of adulthood	**Outcome:** Low wage job in 80% of adulthood

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Rich	Nick	Javier
No degree	**4-year degree**	**No degree**
Both parents finished college	Both parents finished college	Neither parent finished high school
Top quartile of family income	Top quartile of family income	Bottom quartile of family income
College prep courses	College prep courses	No college prep courses
Outcome: Low wage job in 73% of adulthood	**Outcome:** Low wage job in 0% of adulthood	**Outcome:** Low wage job in 80% of adulthood

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Rich	Nick	Javier	Diego
No degree	**4-year degree**	**No degree**	**4-year degree**
Both parents finished college	Both parents finished college	Neither parent finished high school	Neither parent finished high school
Top quartile of family income	Top quartile of family income	Bottom quartile of family income	Bottom quartile of family income
College prep courses	College prep courses	No college prep courses	No college prep courses
Outcome: Low wage job in 73% of adulthood	**Outcome:** Low wage job in 0% of adulthood	**Outcome:** Low wage job in 80% of adulthood	**Outcome:** Low wage job in 7% of adulthood

Does college pay off? An observational study

Brand 2023 p. 44, 69, 116–117

Rich	Nick
No degree	**4-year degree**
Both parents finished college	Both parents finished college
Top quartile of family income	Top quartile of family income
College prep courses	College prep courses
Outcome: Low wage job in 73% of adulthood	**Outcome:** Low wage job in 0% of adulthood

Javier	Diego
No degree	**4-year degree**
Neither parent finished high school	Neither parent finished high school
Bottom quartile of family income	Bottom quartile of family income
No college prep courses	No college prep courses
Outcome: Low wage job in 80% of adulthood	**Outcome:** Low wage job in 7% of adulthood

Mathematical notation for two types of claims

Mathematical notation for two types of claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Mathematical notation for two types of claims

People with
college degrees
earn more

A college degree
causes
higher earnings

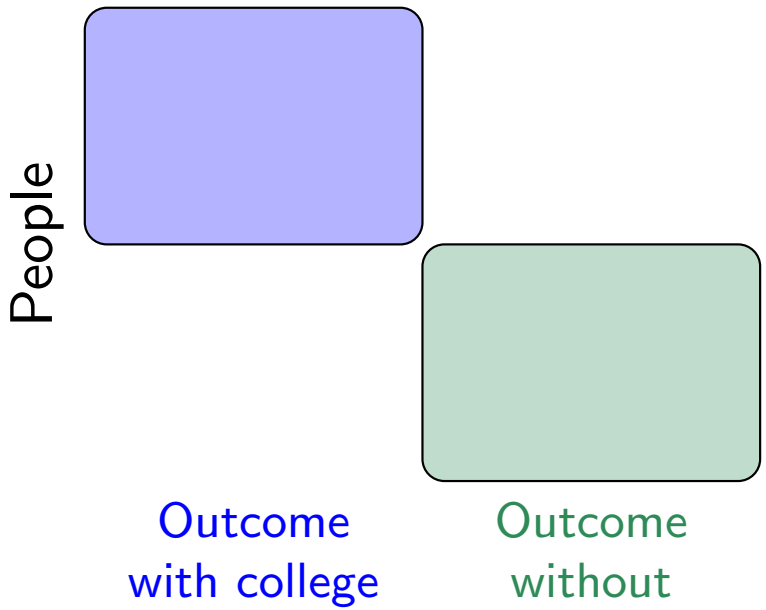
Two sets of people
Two treatments

Mathematical notation for two types of claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Two sets of people
Two treatments



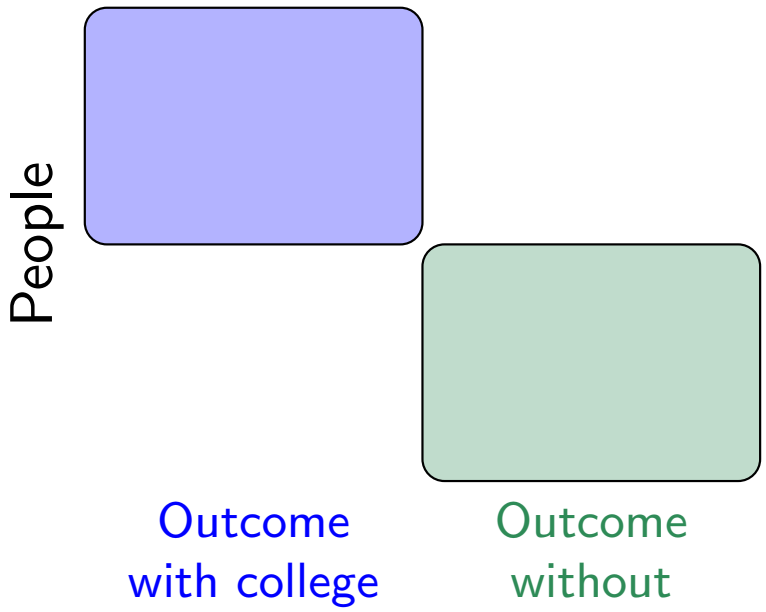
Mathematical notation for two types of claims

People with
college degrees
earn more

A college degree
causes
higher earnings

Two sets of people
Two treatments

Same people
Two treatments



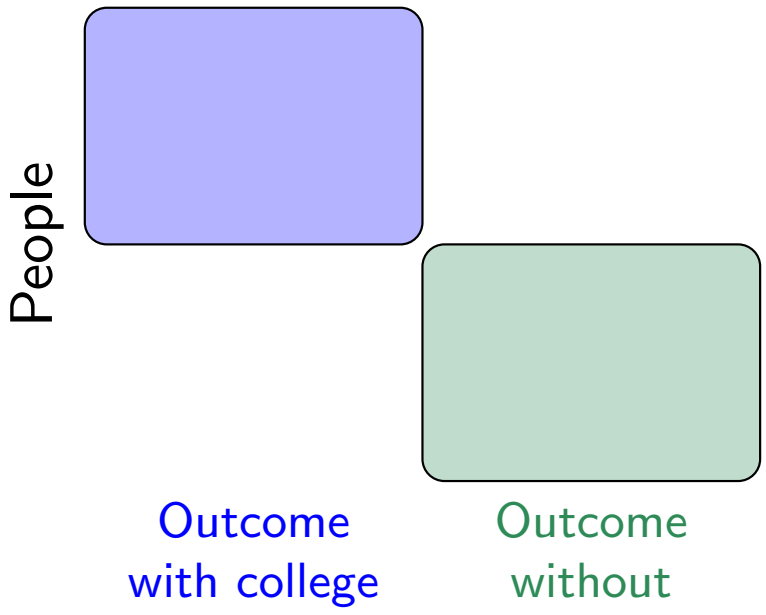
Mathematical notation for two types of claims

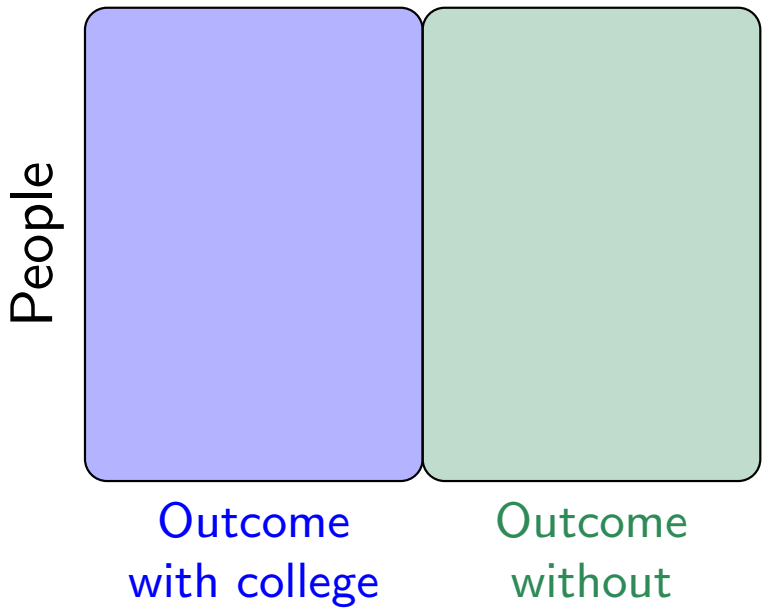
People with
college degrees
earn more

A college degree
causes
higher earnings

Two sets of people
Two treatments

Same people
Two treatments

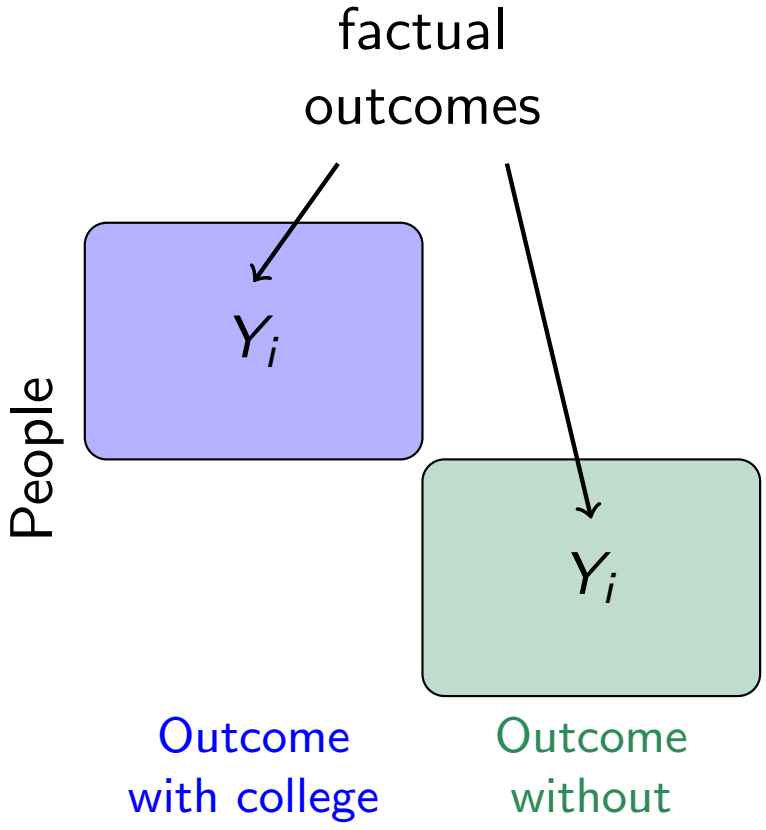




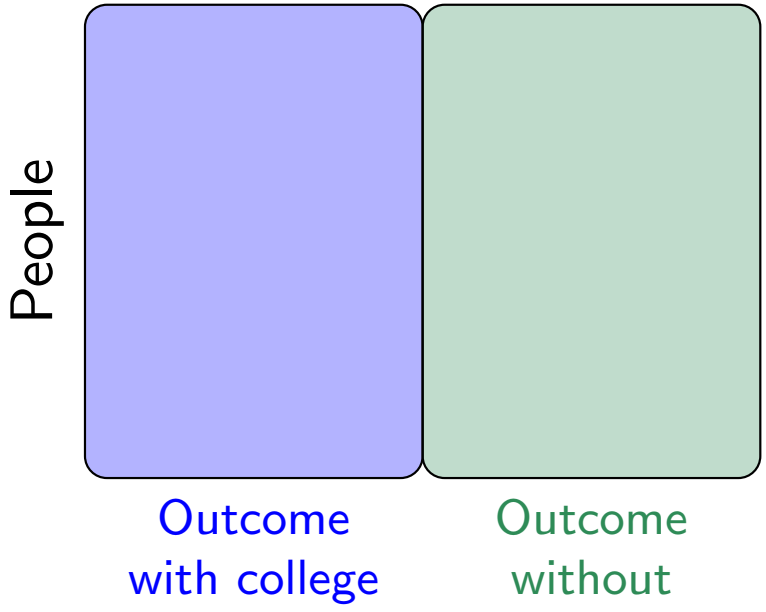
Mathematical notation for two types of claims

People with
college degrees
earn more

A college degree
causes
higher earnings



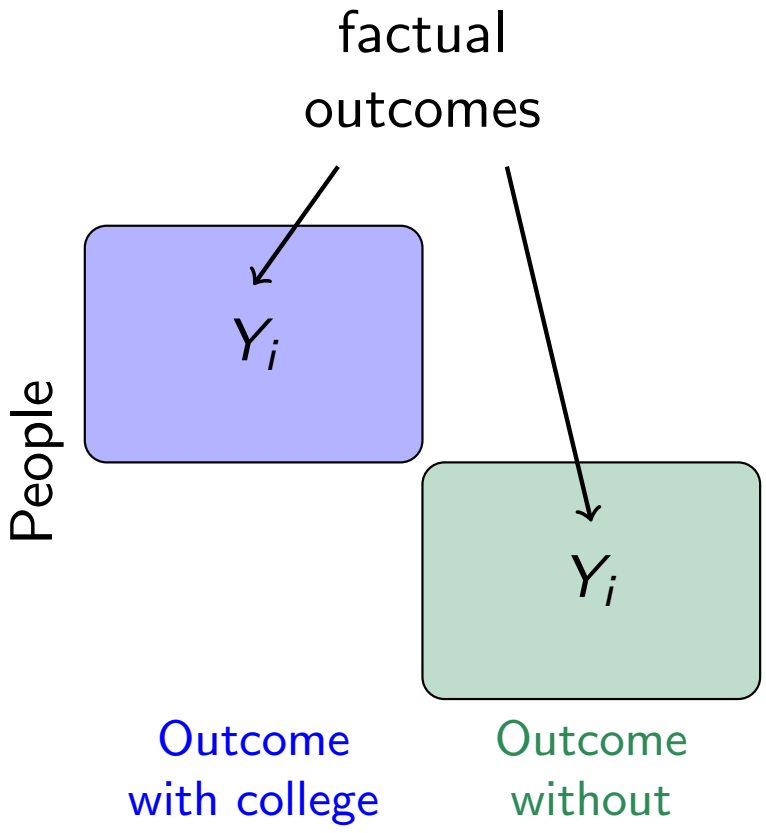
Same people
Two treatments

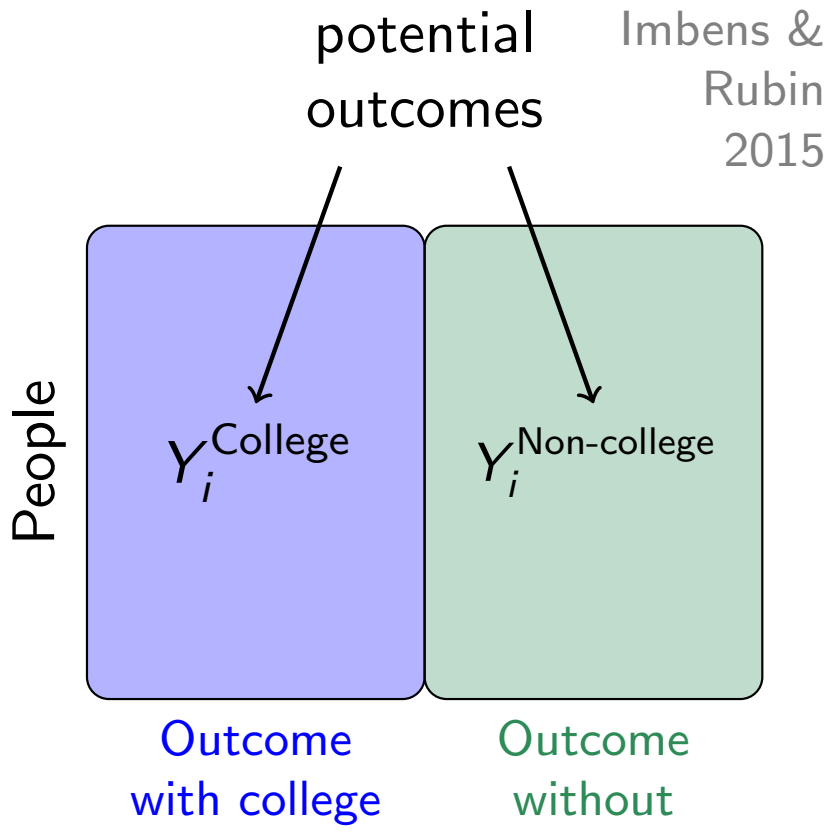


Mathematical notation for two types of claims

People with
college degrees
earn more

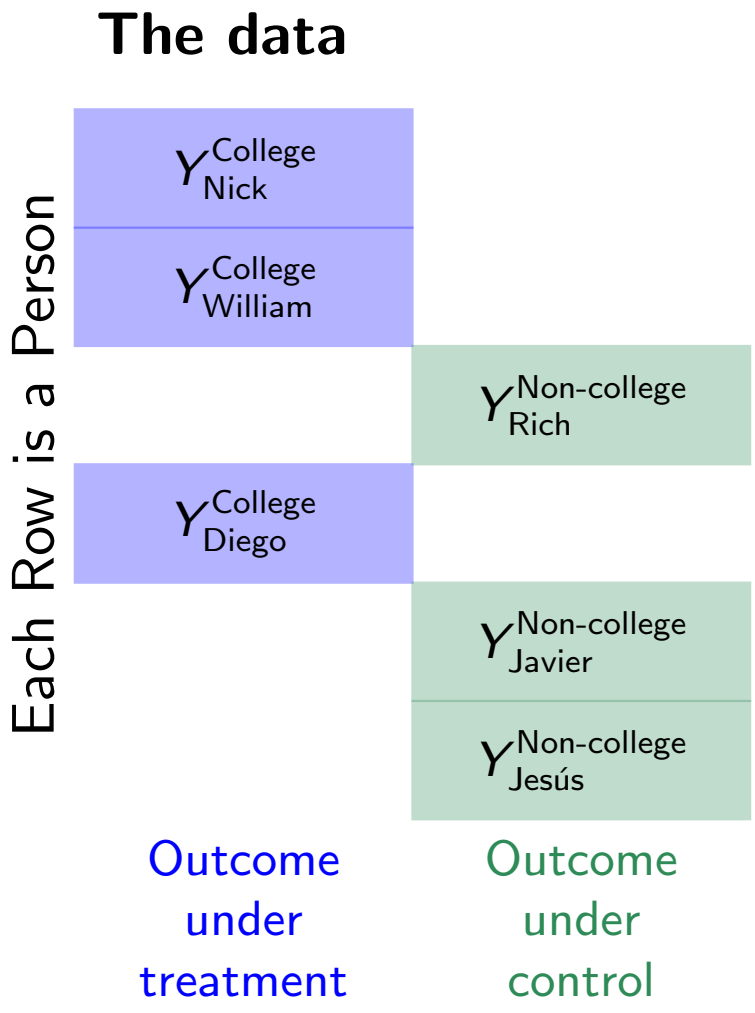
A college degree
causes
higher earnings





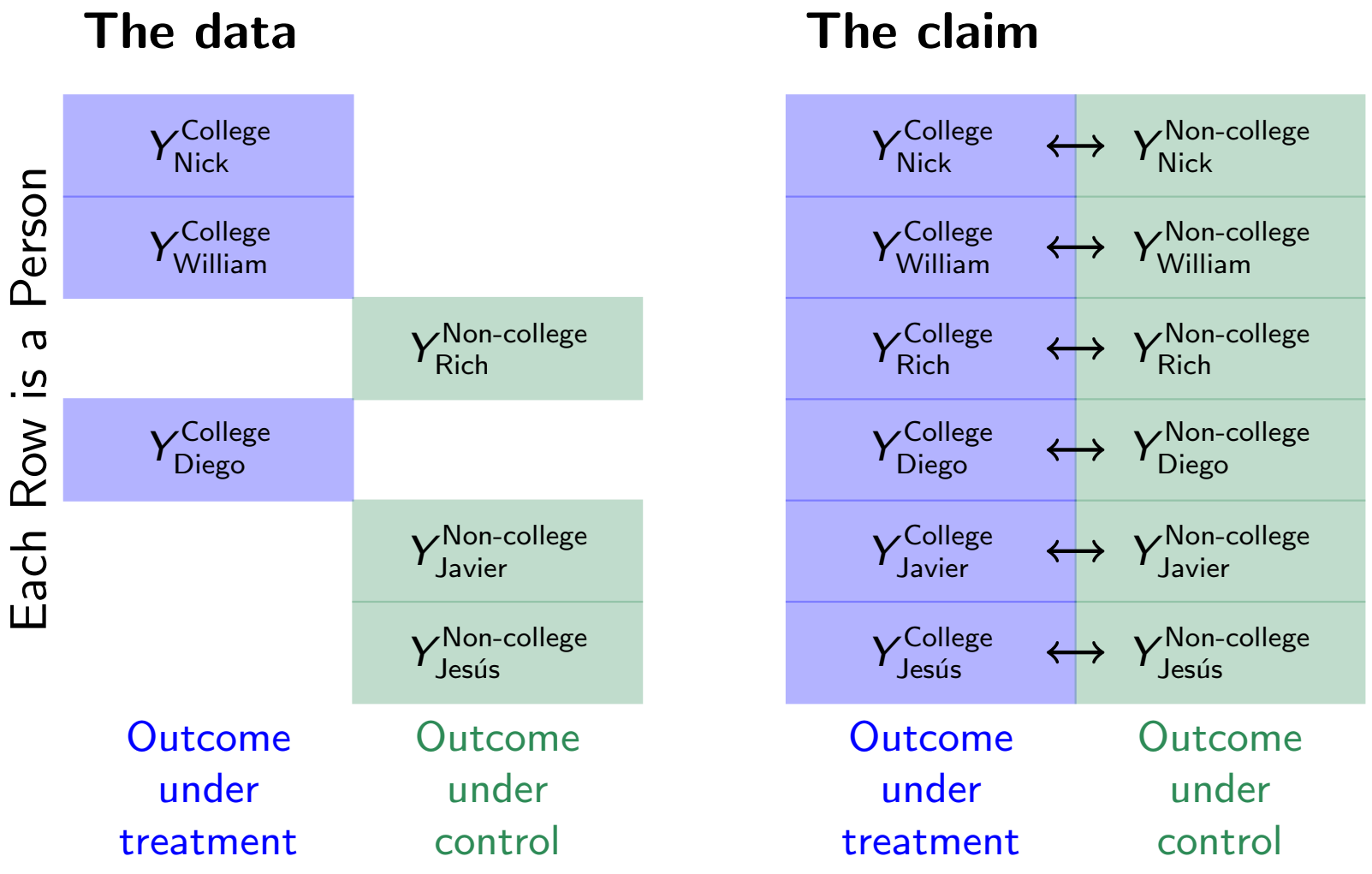
Imbens &
Rubin
2015

The fundamental problem of causal inference



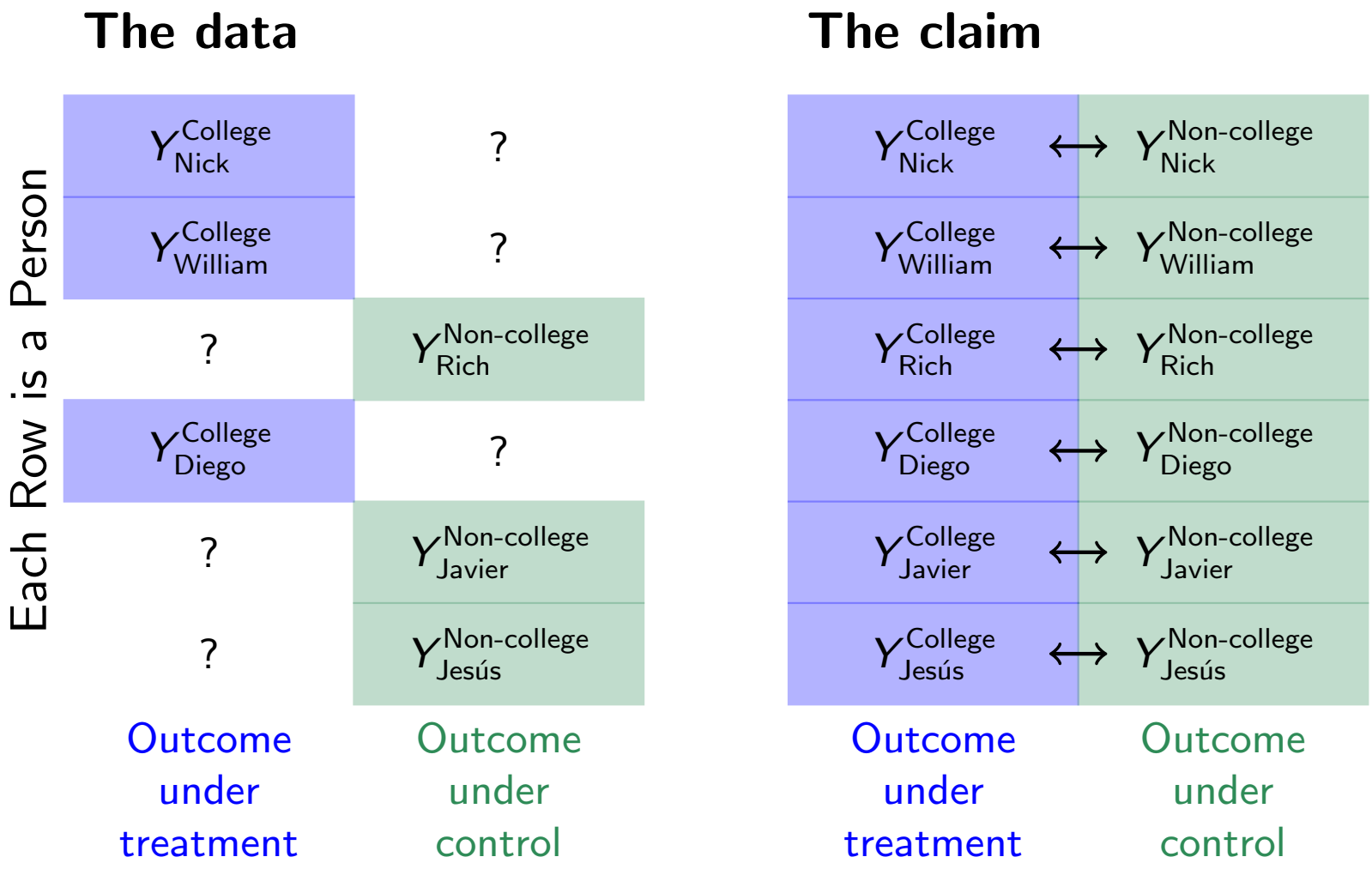
Holland 1986

The fundamental problem of causal inference



Holland 1986

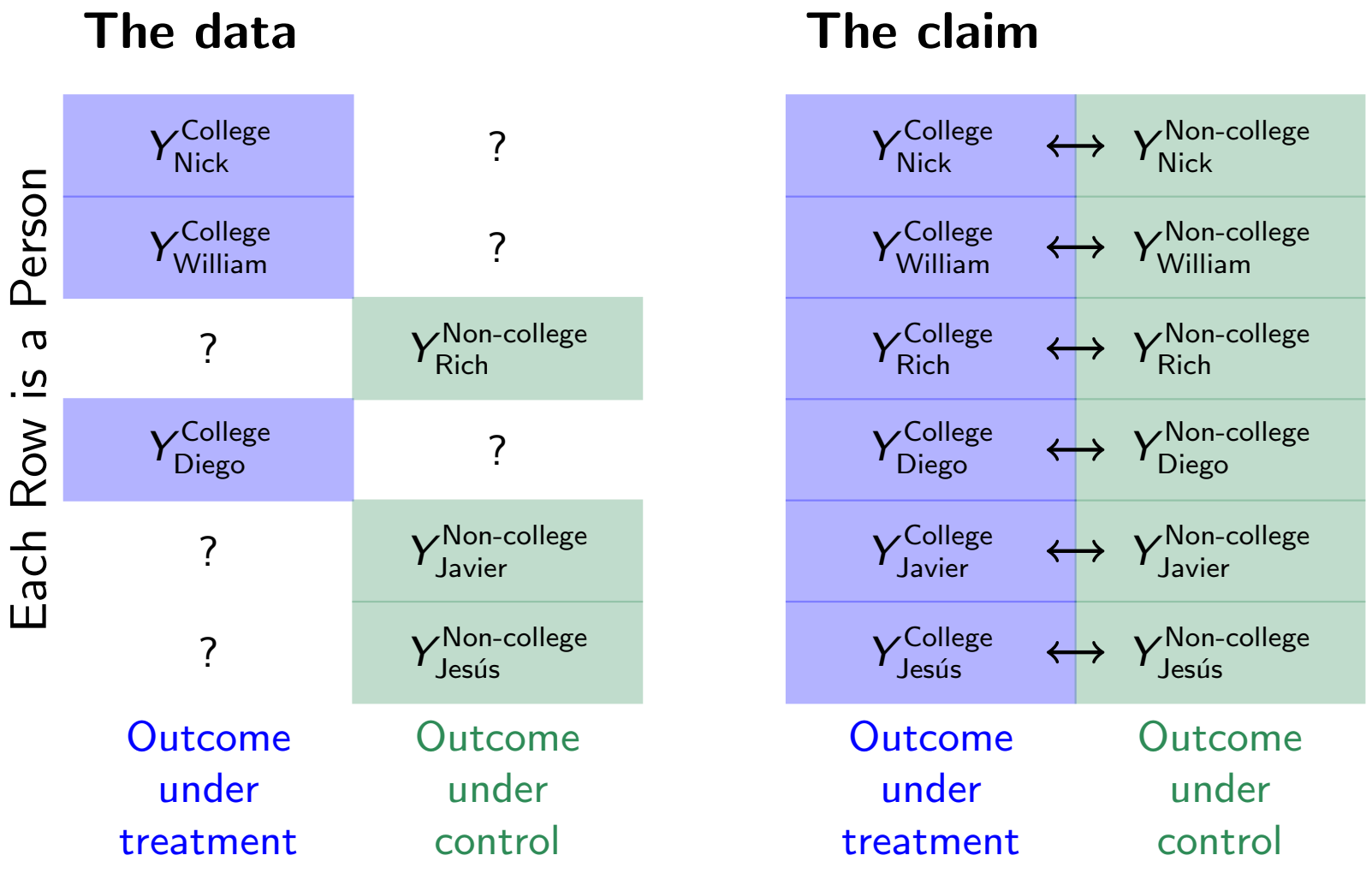
The fundamental problem of causal inference



Counterfactuals are **not observed**

Holland 1986

Solving the problem: Assumptions (Exchangeability)



Solving the problem: Assumptions (Exchangeability)

The data

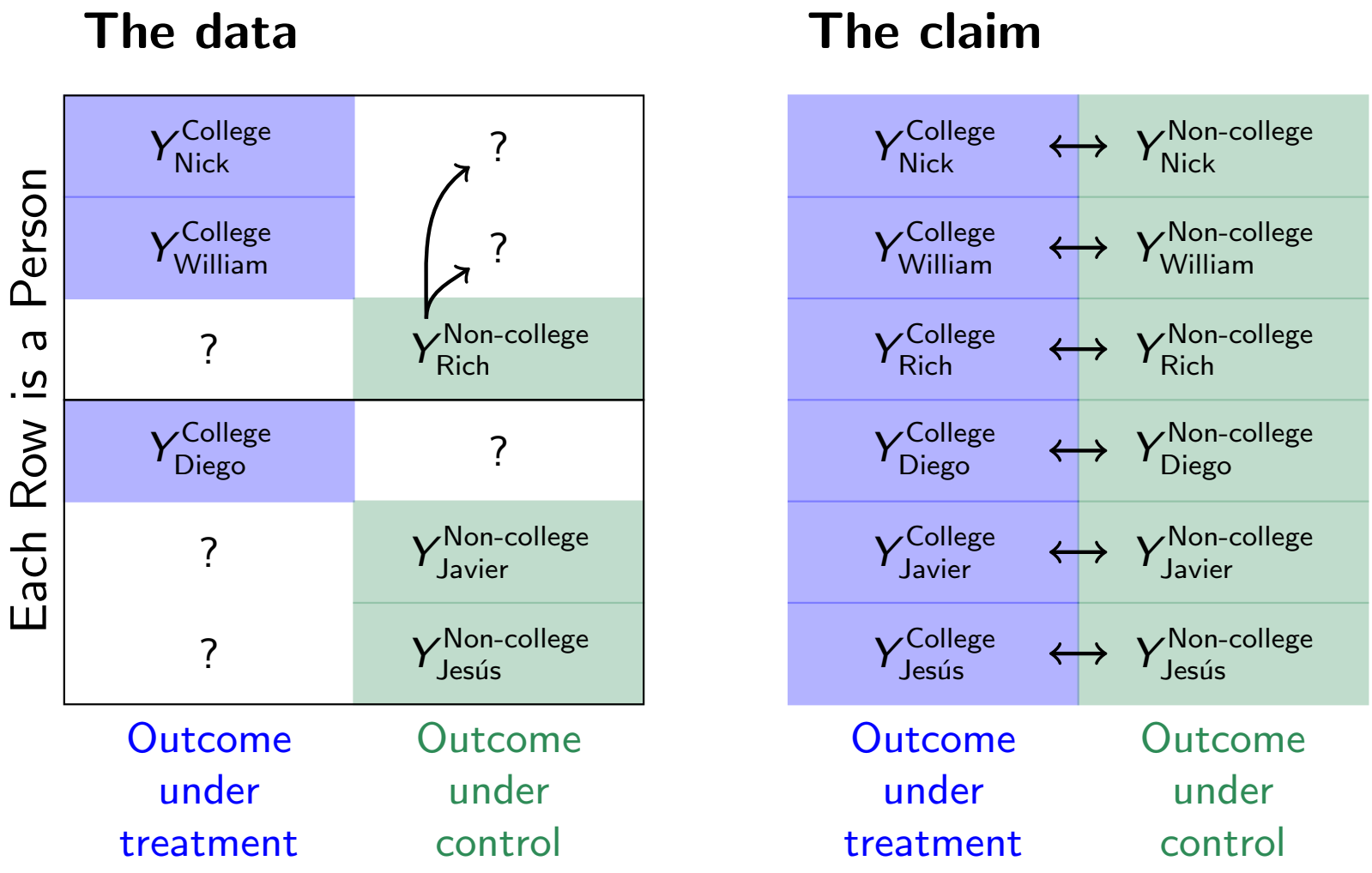
Each Row is a Person

Outcome under treatment	Outcome under control
$Y^{\text{College}}_{\text{Nick}}$	?
$Y^{\text{College}}_{\text{William}}$	?
?	$Y^{\text{Non-college}}_{\text{Rich}}$
$Y^{\text{College}}_{\text{Diego}}$	?
?	$Y^{\text{Non-college}}_{\text{Javier}}$
?	$Y^{\text{Non-college}}_{\text{Jesús}}$

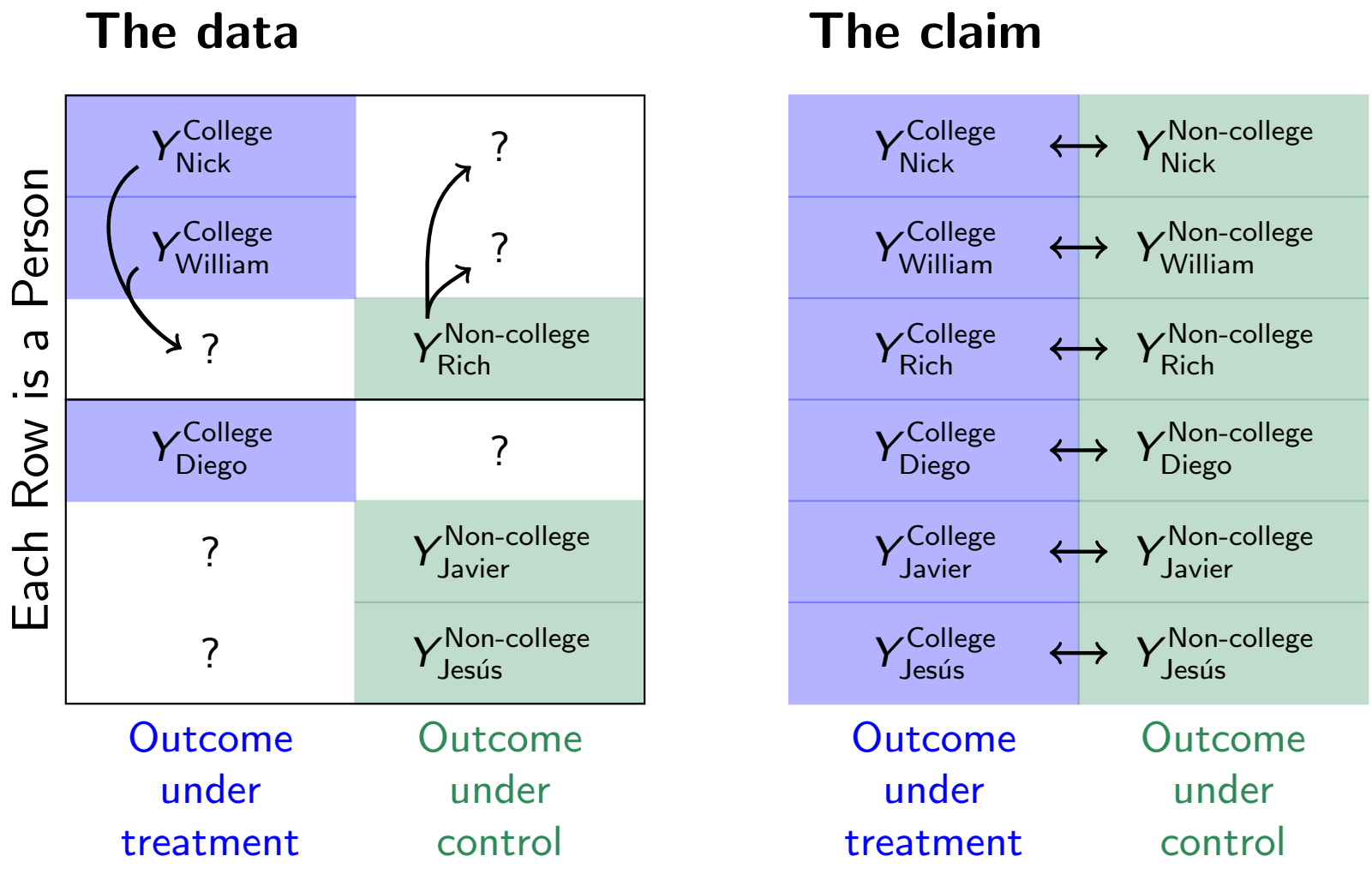
The claim

Outcome under treatment		Outcome under control
$Y^{\text{College}}_{\text{Nick}}$	$\leftrightarrow$	$Y^{\text{Non-college}}_{\text{Nick}}$
$Y^{\text{College}}_{\text{William}}$	$\leftrightarrow$	$Y^{\text{Non-college}}_{\text{William}}$
$Y^{\text{College}}_{\text{Rich}}$	$\leftrightarrow$	$Y^{\text{Non-college}}_{\text{Rich}}$
$Y^{\text{College}}_{\text{Diego}}$	$\leftrightarrow$	$Y^{\text{Non-college}}_{\text{Diego}}$
$Y^{\text{College}}_{\text{Javier}}$	$\leftrightarrow$	$Y^{\text{Non-college}}_{\text{Javier}}$
$Y^{\text{College}}_{\text{Jesús}}$	$\leftrightarrow$	$Y^{\text{Non-college}}_{\text{Jesús}}$

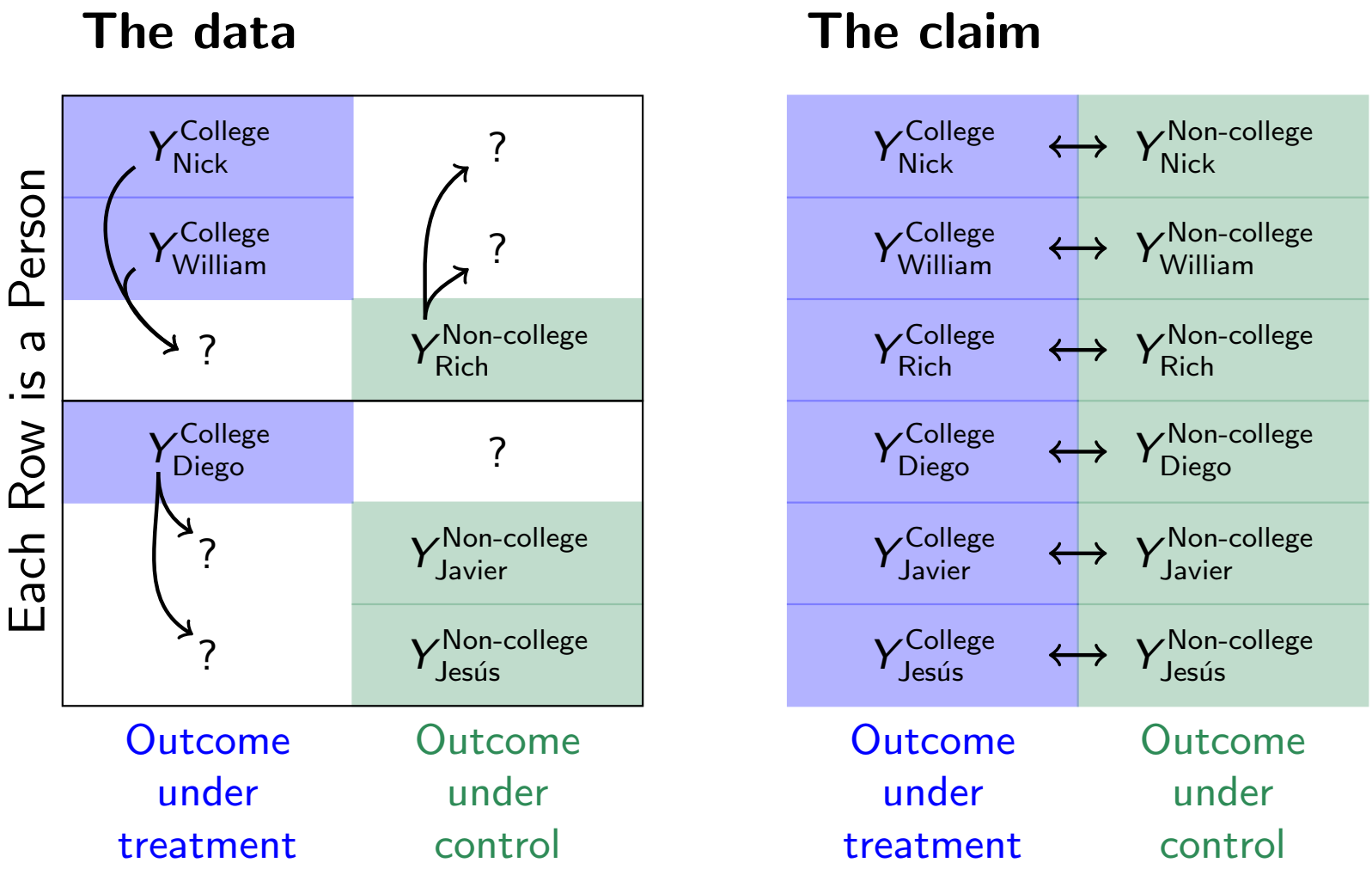
Solving the problem: Assumptions (Exchangeability)



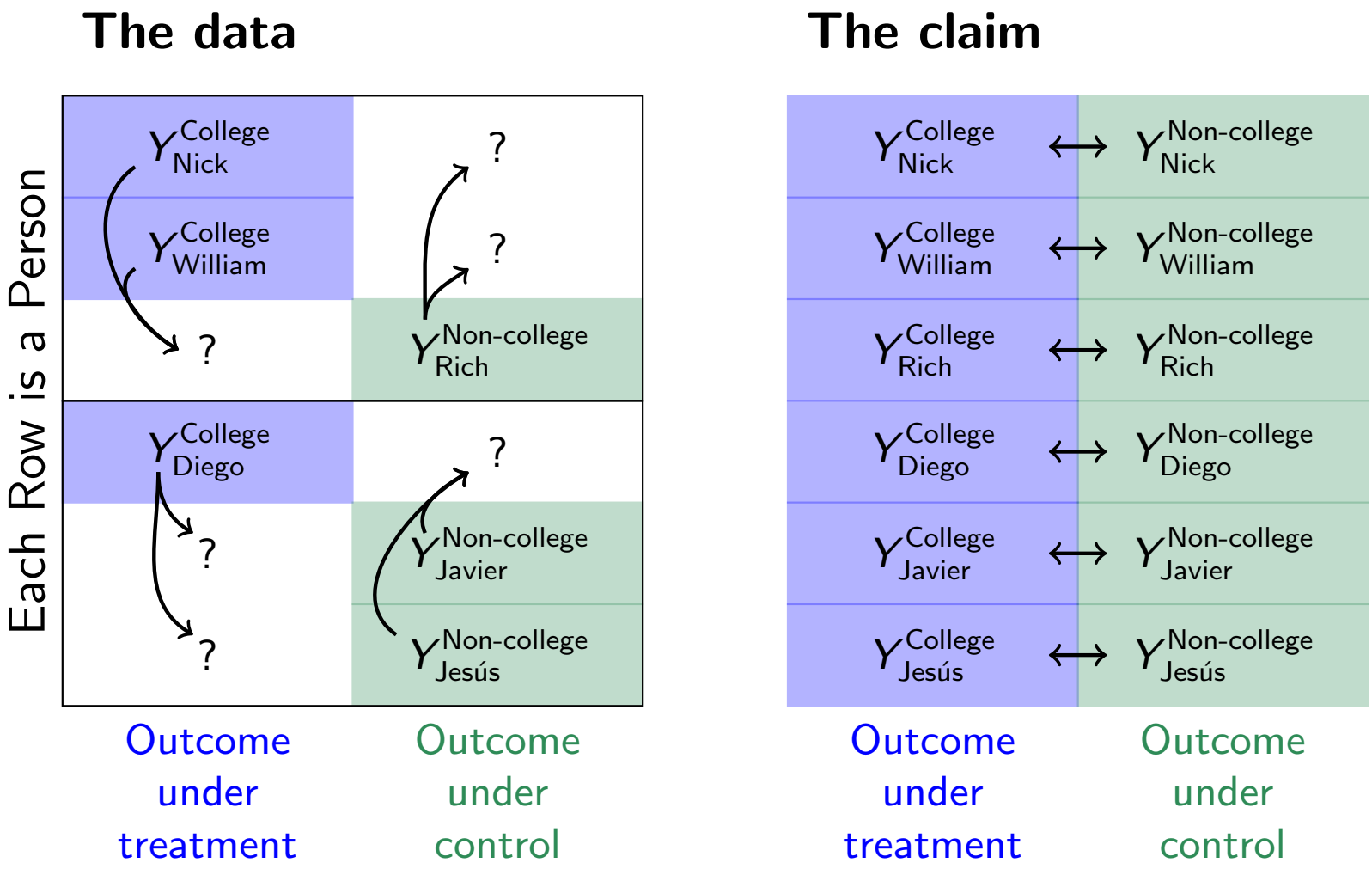
Solving the problem: Assumptions (Exchangeability)



Solving the problem: Assumptions (Exchangeability)



Solving the problem: Assumptions (Exchangeability)



Quick review

Quick review

1. causal effects involve missing data
   - Nick finished college college
   - outcome without college is unobserved

Quick review

1. causal effects involve missing data
   - Nick finished college college
   - outcome without college is unobserved
2. randomization solves the missing data problem by design
   - treated and control groups are exchangeable

Quick review

1. causal effects involve missing data
   - Nick finished college college
   - outcome without college is unobserved
2. randomization solves the missing data problem by design
   - treated and control groups are exchangeable
3. observational studies solve the missing data problem by assumptions
   - find population subgroups who look similar before treatment
   - assume it is like an experiment within the subgroups

Concrete example: Research questions and empirical evidence

Studying the effect of college

Source: Brand 2023

Data

- probability samples of children 14–22 in 1979 and 12–17 in 1997
- measurement of pre-college characteristics
- measurement of four-year completion
- many outcome variables in adulthood
  - earnings, unemployment, low-wage work, etc
  - household income, family poverty, marriage, etc
  - social assistance receipt
  - volunteering, voting, etc

Who benefits the most from college?

Source: Brand 2023

Who benefits the most from college?

Source: Brand 2023

A summary of advantage: The propensity score

$$P(\text{College} \mid \text{Measured Pre-College Variables})$$

(technically: the measured variables must be sufficient for conditional exchangeability to hold)

Who benefits the most from college?

Source: Brand 2023

Who benefits most from college:

- those likely to finish college (advantaged)
- those unlikely to finish college (disadvantaged)

Who benefits the most from college?

Source: Brand 2023

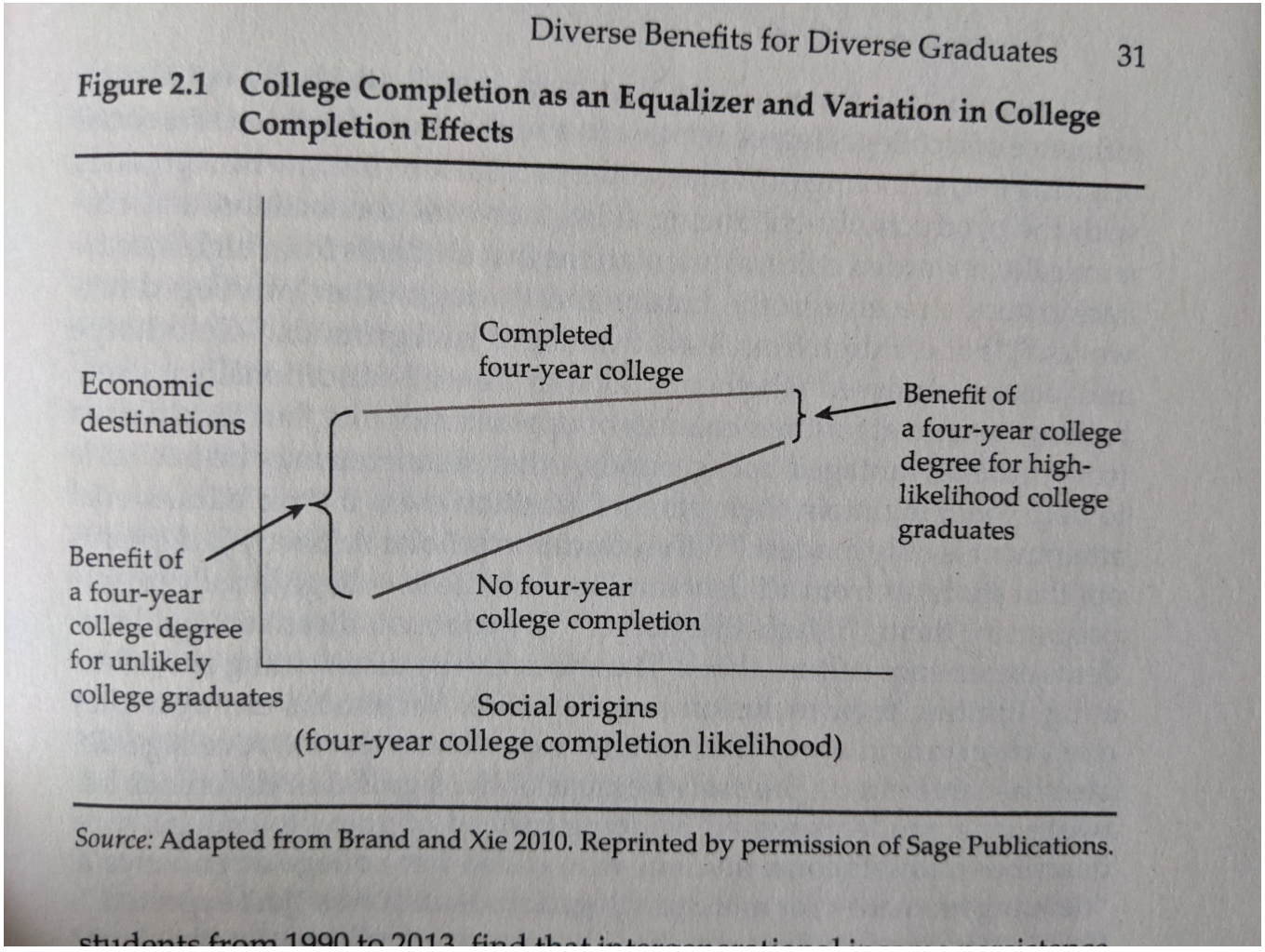
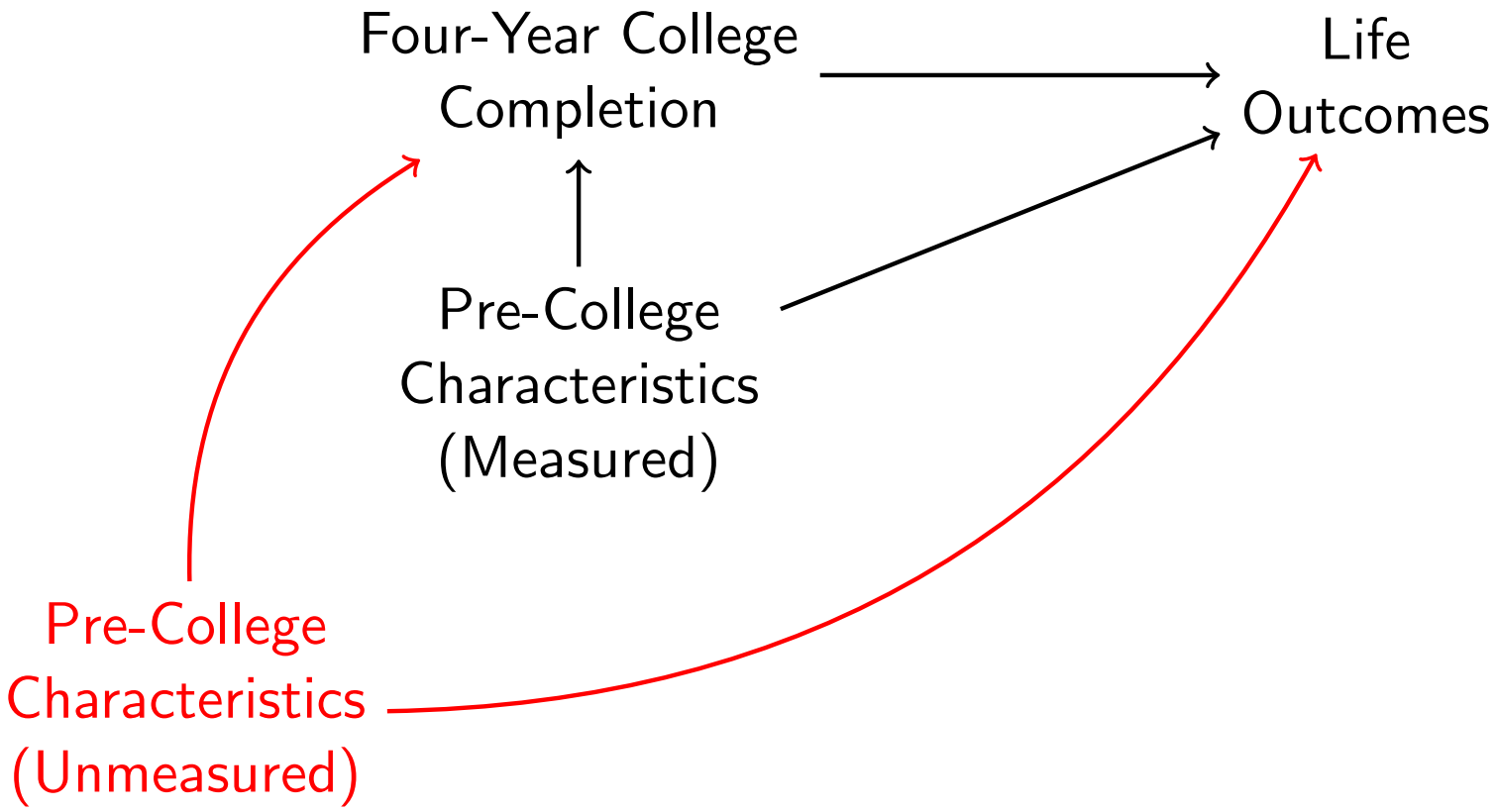


Figure 2.1 College Completion as an Equalizer and Variation in College Completion Effects

Source: Adapted from Brand and Xie 2010. Reprinted by permission of Sage Publications.

Causal identification

Brand 2023 p. 74 for measured variable list



Estimation

Brand 2023 p. 84

Estimation

Brand 2023 p. 84

- logistic regression to predict college completion

Estimation

Brand 2023 p. 84

- logistic regression to predict college completion
- restrict to region of common support
  - $\geq$ 0.3% chance of college completion
  - $\leq$ 92.3% chance of college completion

Estimation

Brand 2023 p. 84

- logistic regression to predict college completion
- restrict to region of common support
  - $\geq$ 0.3% chance of college completion
  - $\leq$ 92.3% chance of college completion
- match each college graduate to a non-graduate with similar probability of college completion

Estimation

Brand 2023 p. 84

- logistic regression to predict college completion
- restrict to region of common support
  - $\geq$ 0.3% chance of college completion
  - $\leq$ 92.3% chance of college completion
- match each college graduate to a non-graduate with similar probability of college completion
- report estimates for three groups
  - low propensity of college completion 0.3–20%
  - middle propensity of college completion 20–50%
  - high propensity of college completion 50–92.3%

An intervention to **reduce inequality**

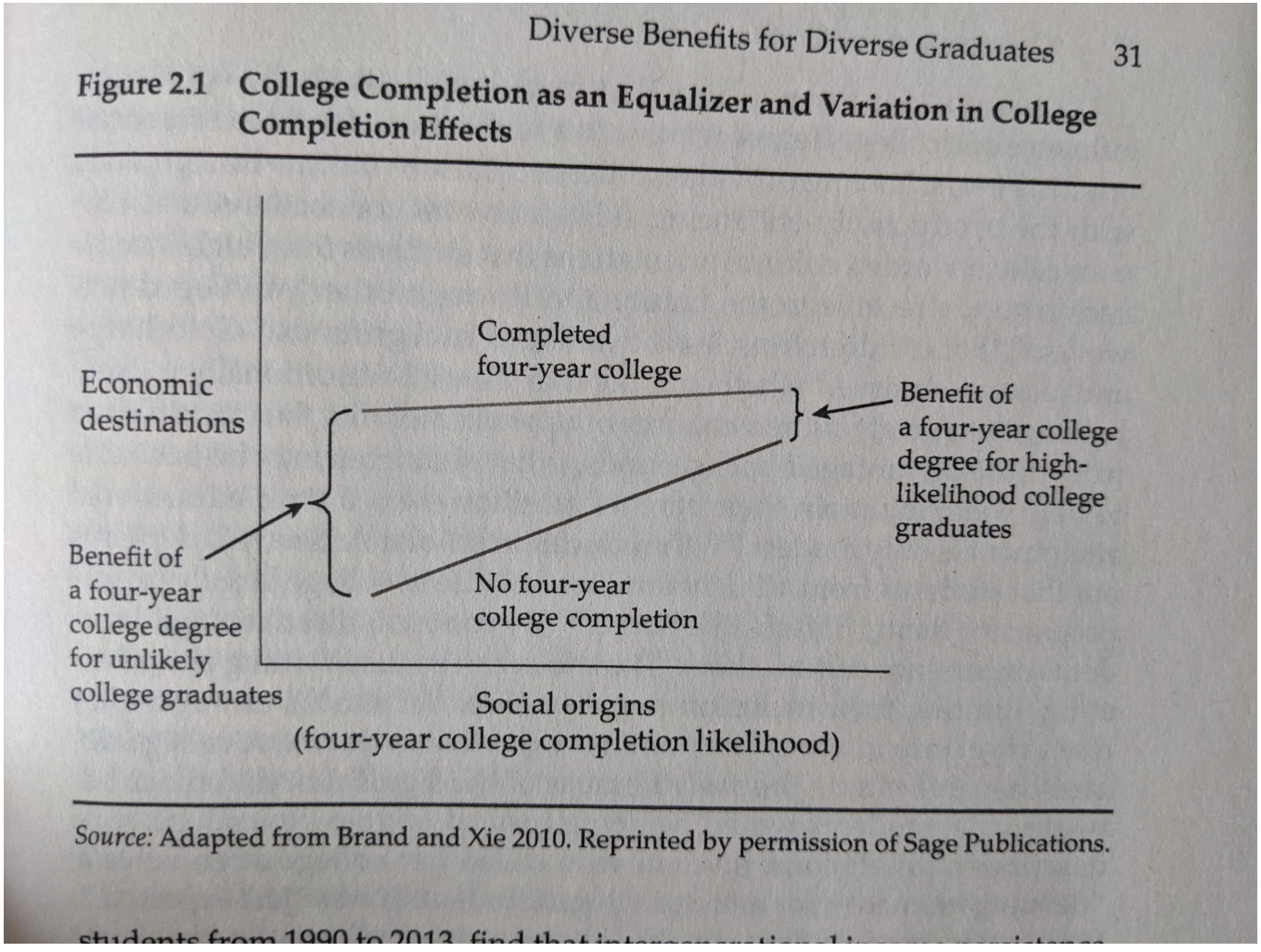


Figure 2.1 College Completion as an Equalizer and Variation in College Completion Effects

Source: Adapted from Brand and Xie 2010. Reprinted by permission of Sage Publications.

students from 1990 to 2013, find that

An intervention to **reduce inequality**

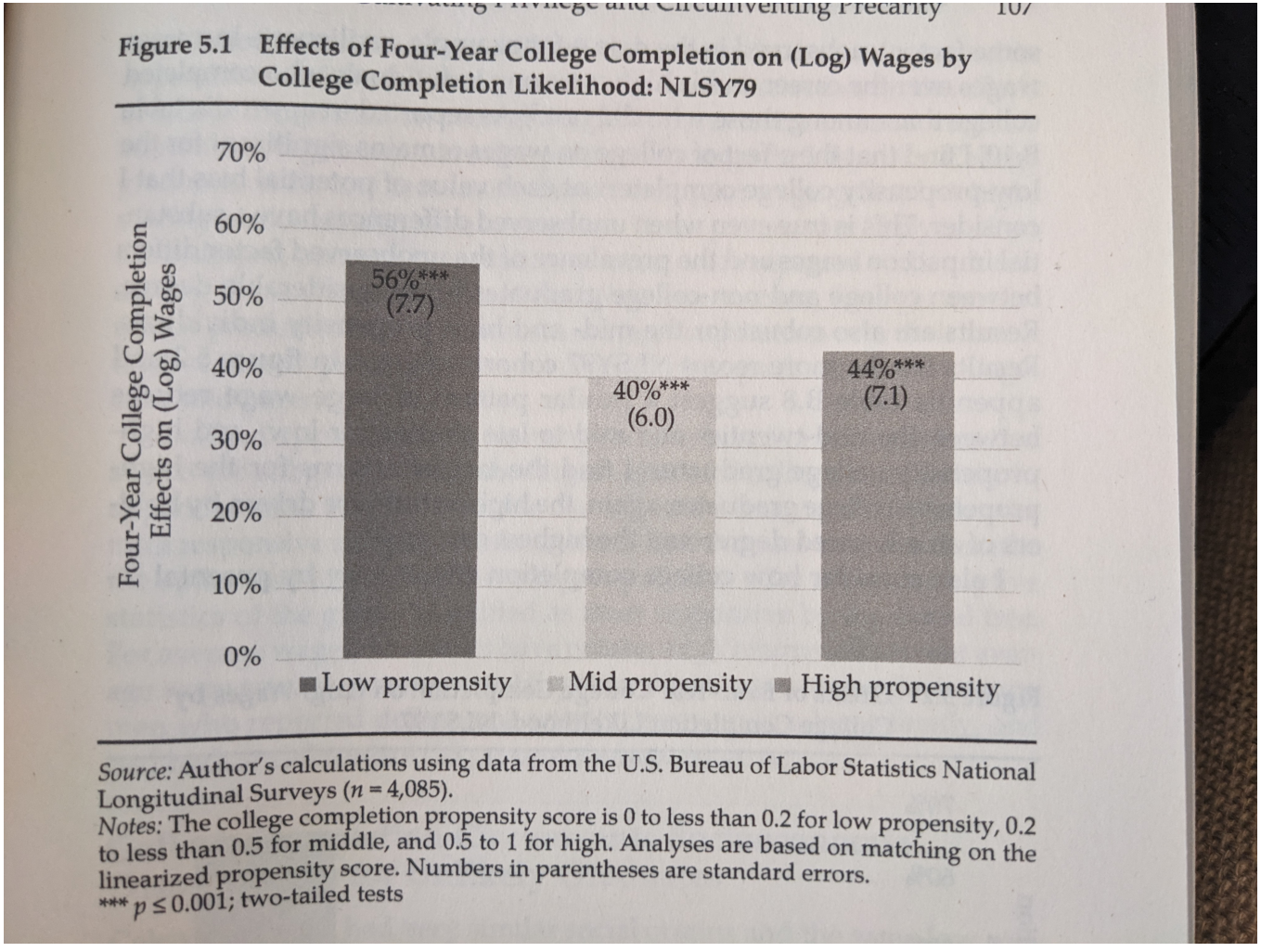


Figure 5.1 Effects of Four-Year College Completion on (Log) Wages by College Completion Likelihood: NLSY79

Source: Author's calculations using data from the U.S. Bureau of Labor Statistics National Longitudinal Surveys (n = 4,085).
Notes: The college completion propensity score is 0 to less than 0.2 for low propensity, 0.2 to less than 0.5 for middle, and 0.5 to 1 for high. Analyses are based on matching on the linearized propensity score. Numbers in parentheses are standard errors.
*** $p \leq 0.001$; two-tailed tests

An intervention to **reduce inequality**

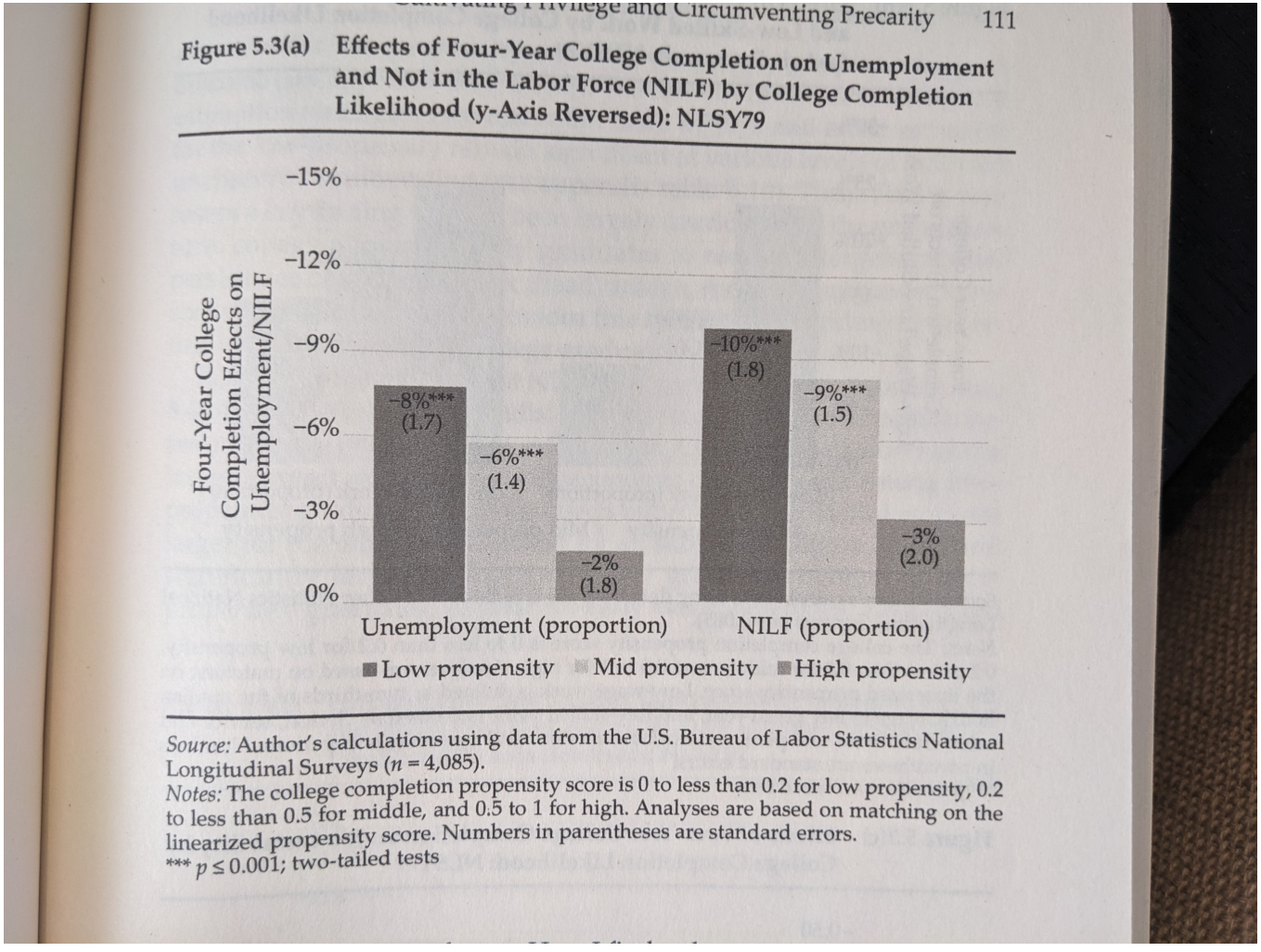


Figure 5.3(a) Effects of Four-Year College Completion on Unemployment and Not in the Labor Force (NILF) by College Completion Likelihood (y-Axis Reversed): NLSY79

Source: Author's calculations using data from the U.S. Bureau of Labor Statistics National Longitudinal Surveys ($n = 4{,}085$).
Notes: The college completion propensity score is 0 to less than 0.2 for low propensity, 0.2 to less than 0.5 for middle, and 0.5 to 1 for high. Analyses are based on matching on the linearized propensity score. Numbers in parentheses are standard errors.
*** $p \leq 0.001$; two-tailed tests

An intervention to **reduce inequality**

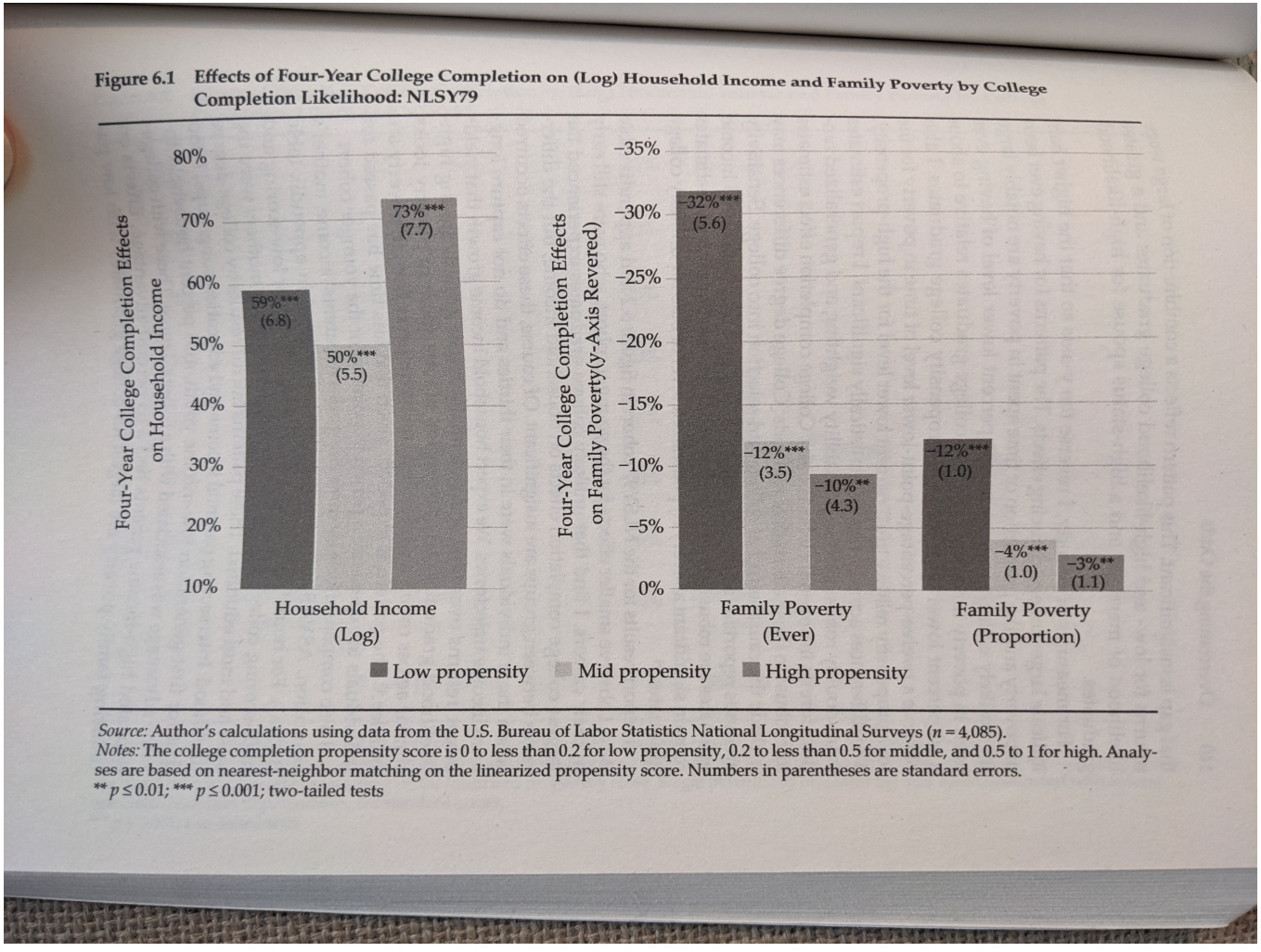


Figure 6.1 **Effects of Four-Year College Completion on (Log) Household Income and Family Poverty by College Completion Likelihood: NLSY79**

Source: Author's calculations using data from the U.S. Bureau of Labor Statistics National Longitudinal Surveys ($n = 4{,}085$).
Notes: The college completion propensity score is 0 to less than 0.2 for low propensity, 0.2 to less than 0.5 for middle, and 0.5 to 1 for high. Analyses are based on nearest-neighbor matching on the linearized propensity score. Numbers in parentheses are standard errors.
** $p \leq 0.01$; *** $p \leq 0.001$; two-tailed tests

Next week: On what should we match?

Causal origins of statistical associations

Causal origins of statistical associations

There are two reasons A and Y can be associated

Causal origins of statistical associations

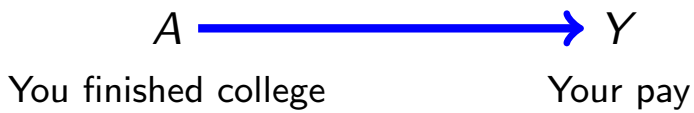
There are two reasons A and Y can be associated

- A **causal path**: $A \rightarrow Y$

Causal origins of statistical associations

There are two reasons A and Y can be associated

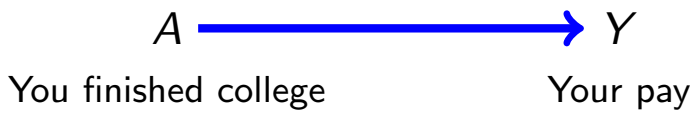
- A **causal path**: $A \rightarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

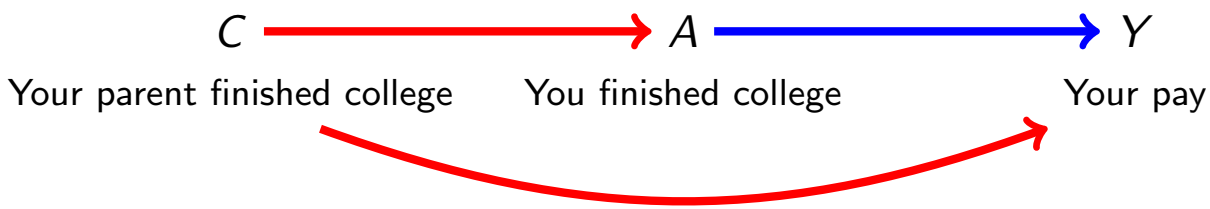
- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

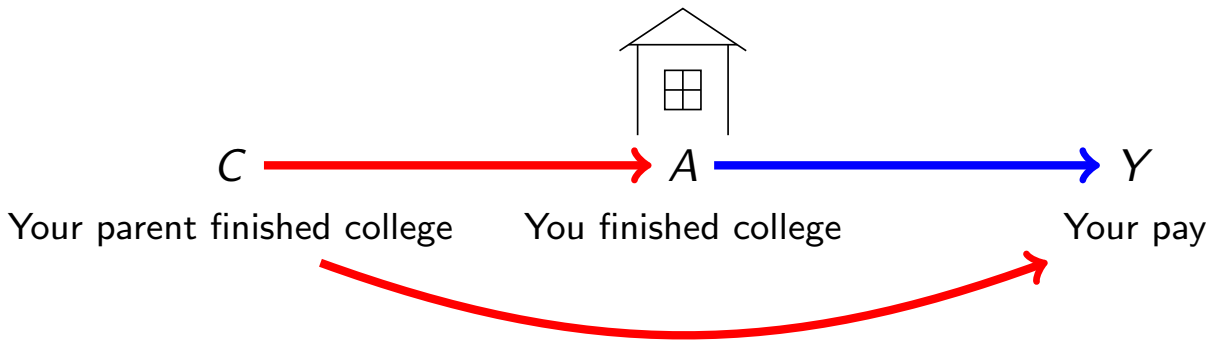
- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

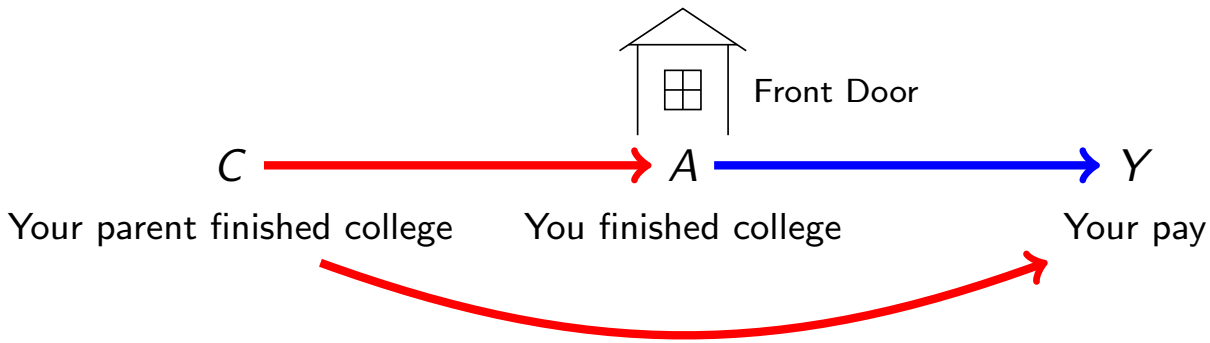
- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

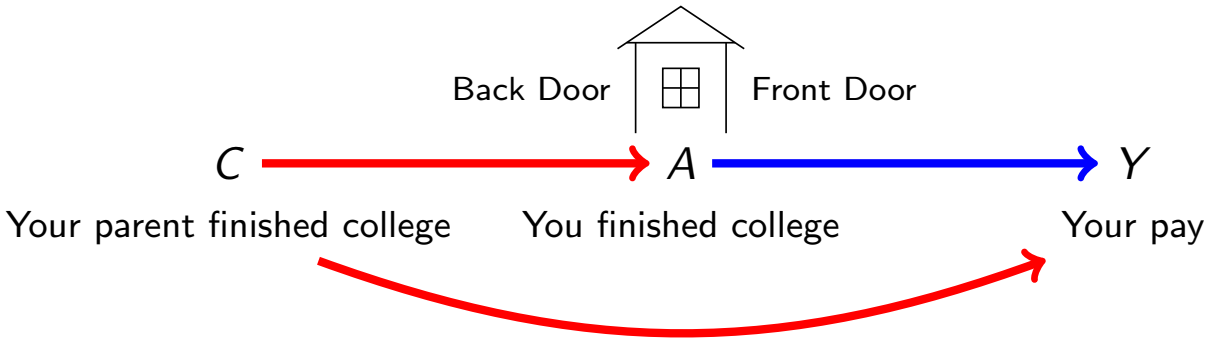
- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

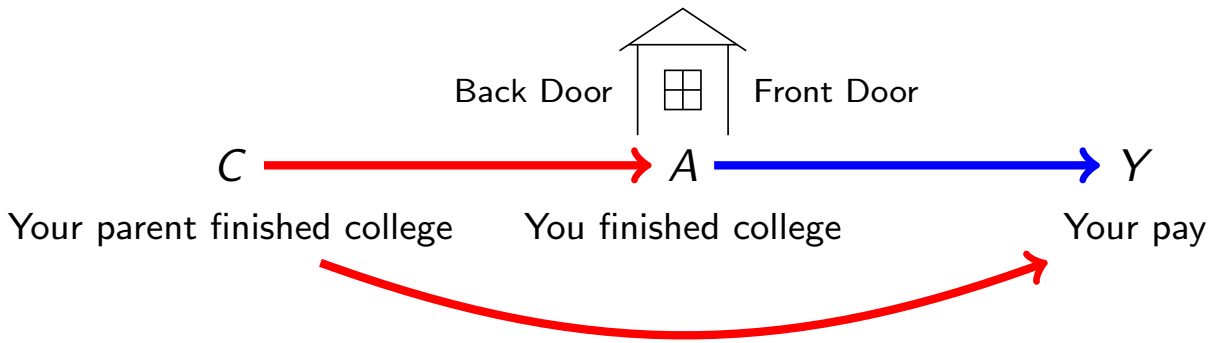
- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$



Causal origins of statistical associations

There are two reasons A and Y can be associated

- A **causal path**: $A \to Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \to Y$
  - or blocked colliders $A \to \boxed{C} \leftarrow Y$

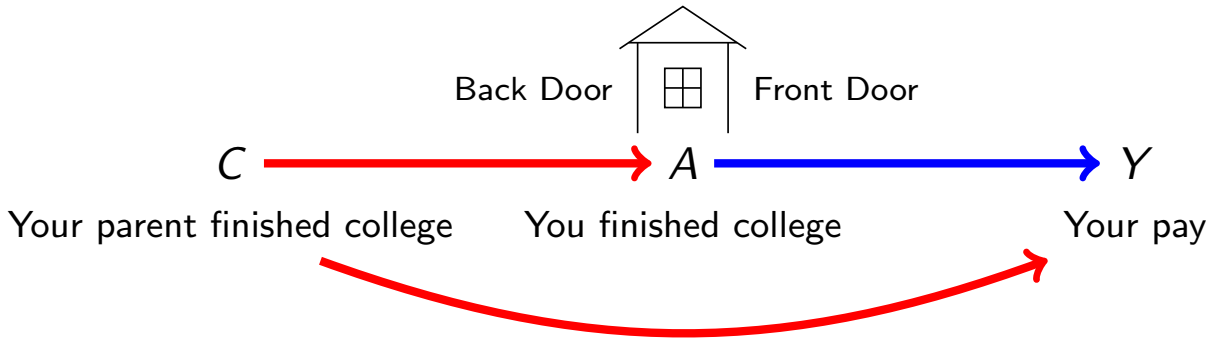


To block this backdoor path, condition on C (a **confounder**)

Causal origins of statistical associations

There are two reasons A and Y can be associated

- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



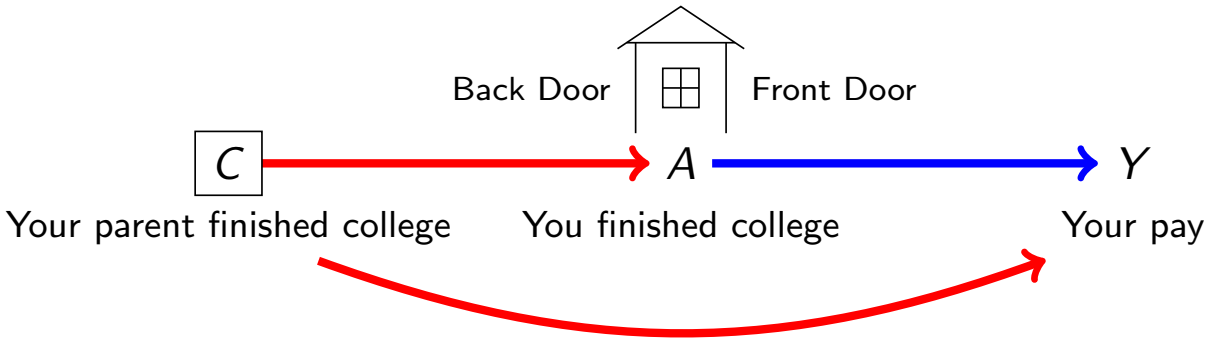
To block this backdoor path, condition on C (a **confounder**)

- Analyze within subgroups defined by C

Causal origins of statistical associations

There are two reasons A and Y can be associated

- A **causal path**: $A \rightarrow Y$
- A **backdoor path** involving
  - unblocked forks $A \leftarrow C \rightarrow Y$
  - or blocked colliders $A \rightarrow \boxed{C} \leftarrow Y$



To block this backdoor path, condition on C (a **confounder**)

- Analyze within subgroups defined by C

Colliders[1]

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

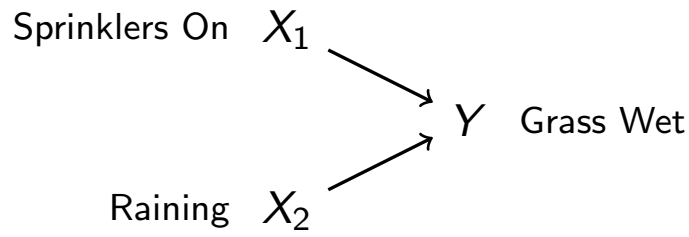
Colliders[1]

Suppose I have sprinklers on a timer.

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

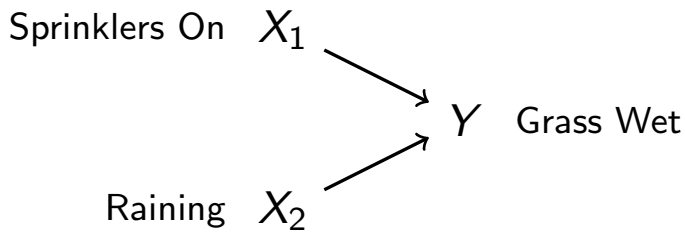
Suppose I have sprinklers on a timer.



[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.

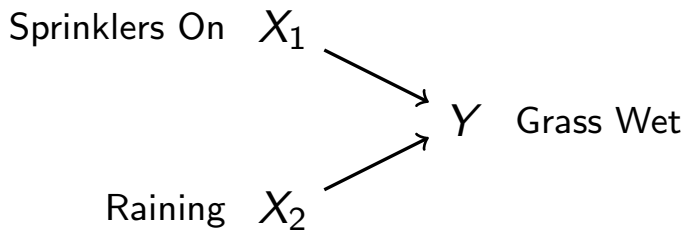


We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.



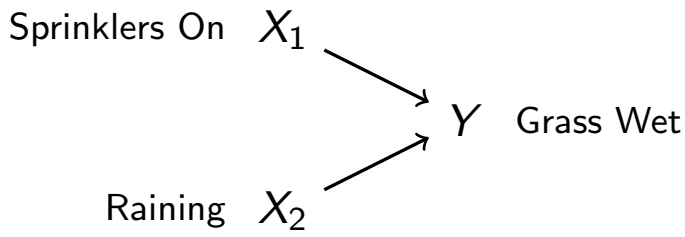
We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

- The collider blocks the path

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.



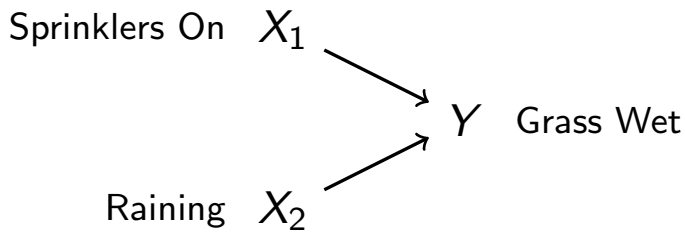
We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

- The collider blocks the path
- X_1 is independent of X_2
  - (Sprinklers On) is uninformative about (Raining)

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Colliders[1]

Suppose I have sprinklers on a timer.



We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

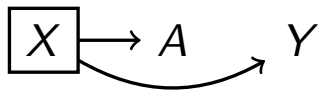
- The collider blocks the path
- X_1 is independent of X_2
  - (Sprinklers On) is uninformative about (Raining)
- Conditioning on Y opens the path
  - If the grass is wet (conditional on $Y = 1$), then either (Sprinklers On) or (Raining)

---

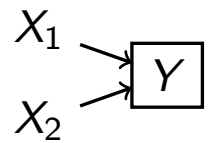
[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

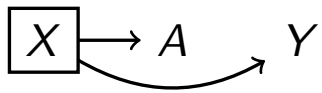


Conditioning on an collider **opens** a closed path



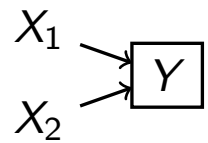
Ancestors vs. Colliders

Conditioning on an ancestor
closes an open path



Example
— X is your parent's education
— A is your education
— Y is your pay

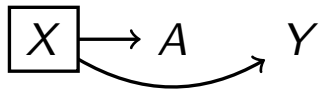
Conditioning on an collider
opens a closed path



Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

Ancestors vs. Colliders

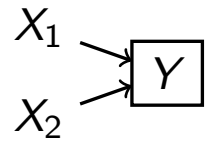
Conditioning on an ancestor **closes** an open path



Example
— X is your parent's education
— A is your education
— Y is your pay

In the population,
A and Y are **related**

Conditioning on an collider **opens** a closed path

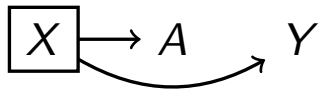


Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

In the population,
X_1 and X_2 are **independent**

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

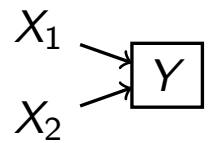


Example
— X is your parent's education
— A is your education
— Y is your pay

In the population,
A and Y are **related**

Within strata of X,
A and Y are **independent**

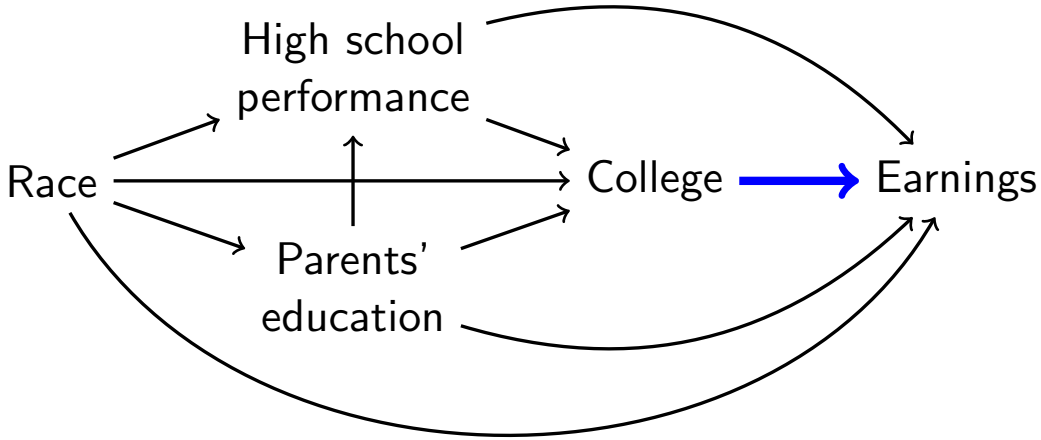
Conditioning on an collider **opens** a closed path



Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

In the population,
X_1 and X_2 are **independent**

Within strata of Y,
X_1 and X_2 are **related**


High school
performance
Race
College
Earnings
Parents'
education

How to find adjustment variables to identify causal effects

Goal:

Block all backdoor paths so treatment A and outcome Y are associated only by the causal path

How to find adjustment variables to identify causal effects

Goal:
Block all backdoor paths so treatment A and outcome Y are associated only by the causal path

Backdoor path: Any sequence of edges $A \leftarrow \texttt{nodes} \rightarrow Y$

Blocked if it contains an adjusted variable along a fork

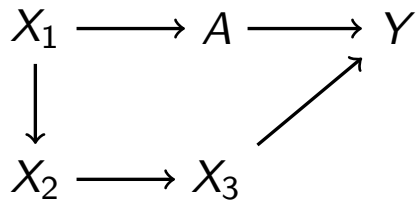
$$A \leftarrow \boxed{C} \rightarrow Y$$
$$A \leftarrow \boxed{C} \leftarrow \cdots \rightarrow Y$$
$$A \leftarrow \cdots \rightarrow \boxed{C} \rightarrow Y$$

Blocked if it contains an unadjusted collider

$$A \rightarrow C \leftarrow Y$$

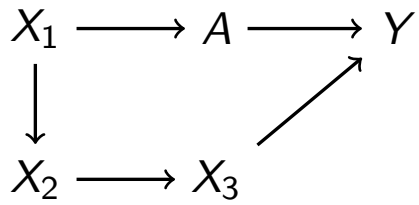
Exercise 1

Find adjustment sets that identify the effect of A on Y



Exercise 1

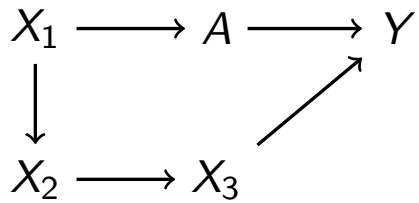
Find adjustment sets that identify the effect of A on Y



We can block the backdoor path in several ways:

Exercise 1

Find adjustment sets that identify the effect of A on Y

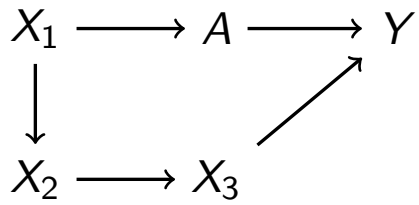


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$

Exercise 1

Find adjustment sets that identify the effect of A on Y

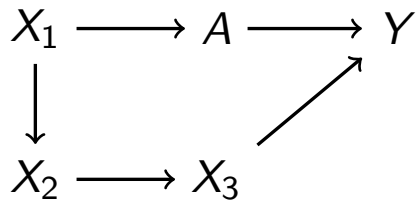


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$

Exercise 1

Find adjustment sets that identify the effect of A on Y

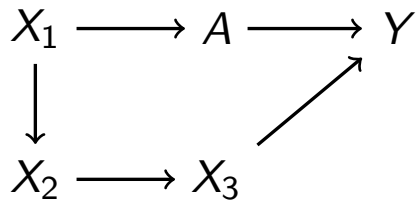


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$
- Condition on X_3: $A \leftarrow X_1 \rightarrow X_2 \rightarrow \boxed{X_3} \rightarrow Y$

Exercise 1

Find adjustment sets that identify the effect of A on Y

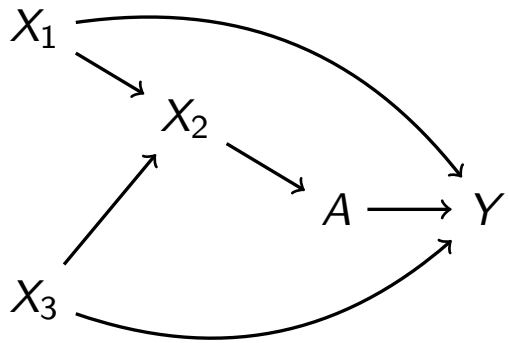


We can block the backdoor path in several ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$
- Condition on X_3: $A \leftarrow X_1 \rightarrow X_2 \rightarrow \boxed{X_3} \rightarrow Y$
- Any combination of the above

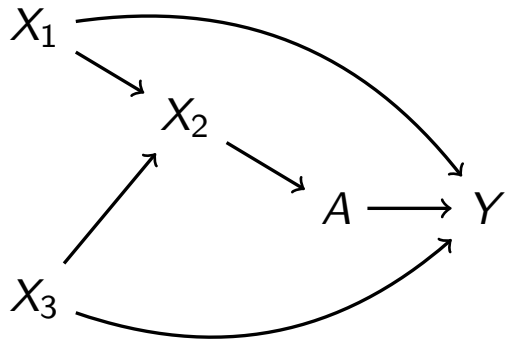
Exercise 2

Find 3 sufficient adjustment sets to identify $A \rightarrow Y$



Exercise 2

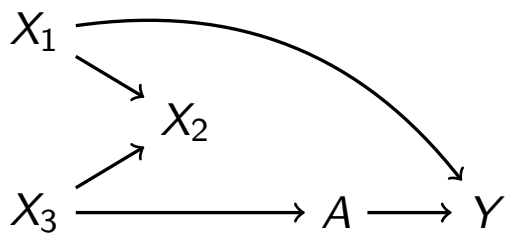
Find 3 sufficient adjustment sets to identify $A \to Y$



Answer: $\{X_2\}, \{X_1, X_3\}, \{X_1, X_2, X_3\}$

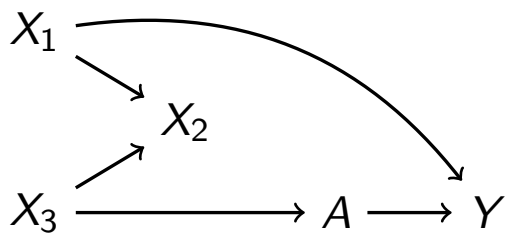
Exercise 3

What is the smallest adjustment set that identifies $A \to Y$?



Exercise 3

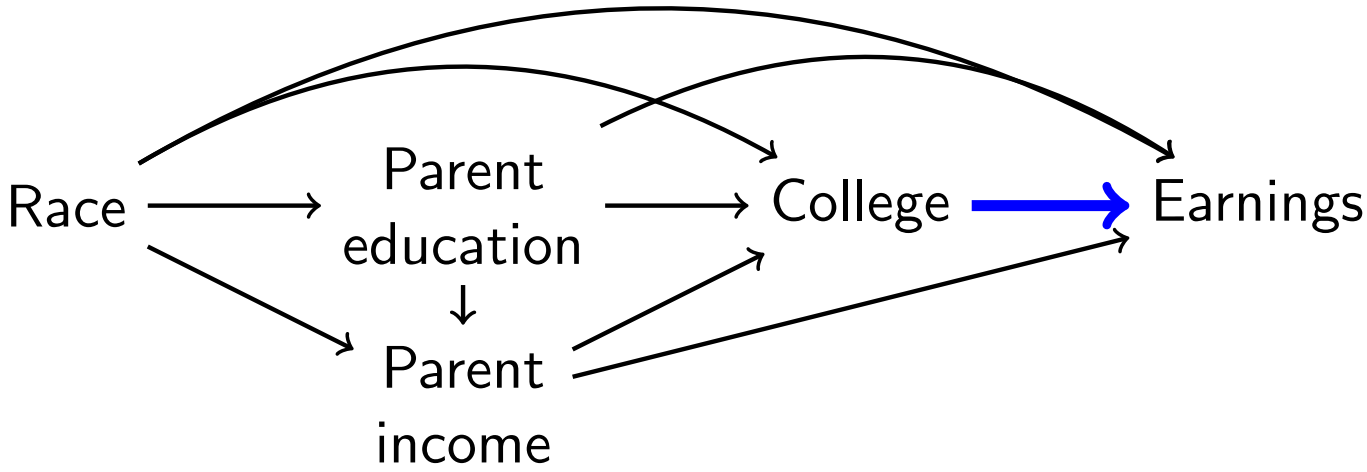
What is the smallest adjustment set that identifies $A \to Y$?



Answer: The empty set! Don't condition on anything.
The collider X_2 already blocks the path.

Recap: Causal effect of college on earnings

1. Fundamental problem: Counterfactuals not observed
2. Assume a DAG: enumerate causal processes creating association between education and earnings



3. Block the pathways that are not of interest
   - e.g. estimate within subgroups of {race, parent education, parent income}
   - model-based
     - regress earnings on everything
     - change the value of `College`
     - predict the counterfactuals

Causal inference is hard. What do we gain?