What is Your Estimand?

A workshop on specifying quantitative research goals

Ian Lundberg
UCLA
Sociology
ianlundberg.org

Rebecca Johnson
Dartmouth Quantitative
Social Science
rebeccajohnson.io

Brandon M. Stewart
Princeton
Sociology
brandonstewart.org

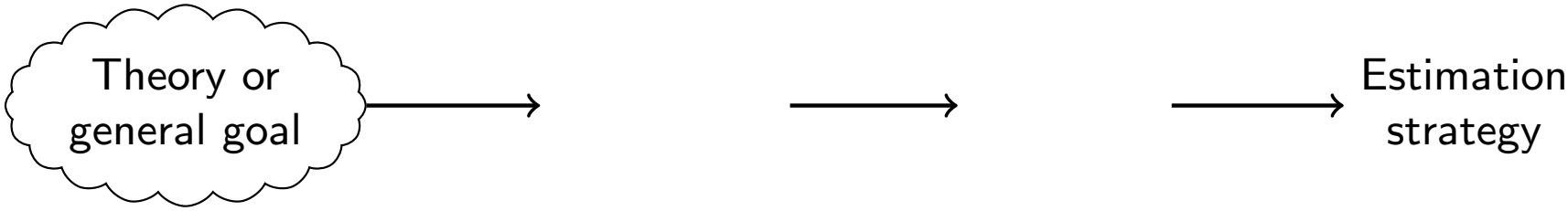
NOTE: These slides are for a pedagogical workshop.
The main slides for this paper are here.

15 October 2021. Cornell Population Center Training Workshop.
Paper in *American Sociological Review*. Preprint on SocArxiv. 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.

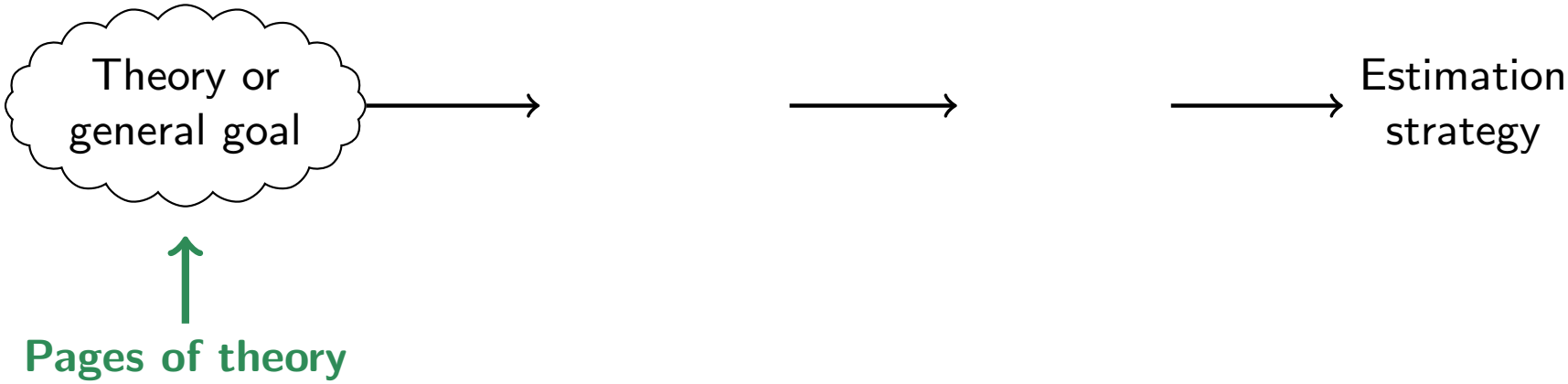
Plan for today

- Brief recap of paper: "What is Your Estimand?"
- In groups, we will discuss
  - Defining the intervention in a causal estimand
  - Defining the target population in a descriptive estimand
  - Policy implications of descriptive and causal estimands
- Professional development: Open for questions

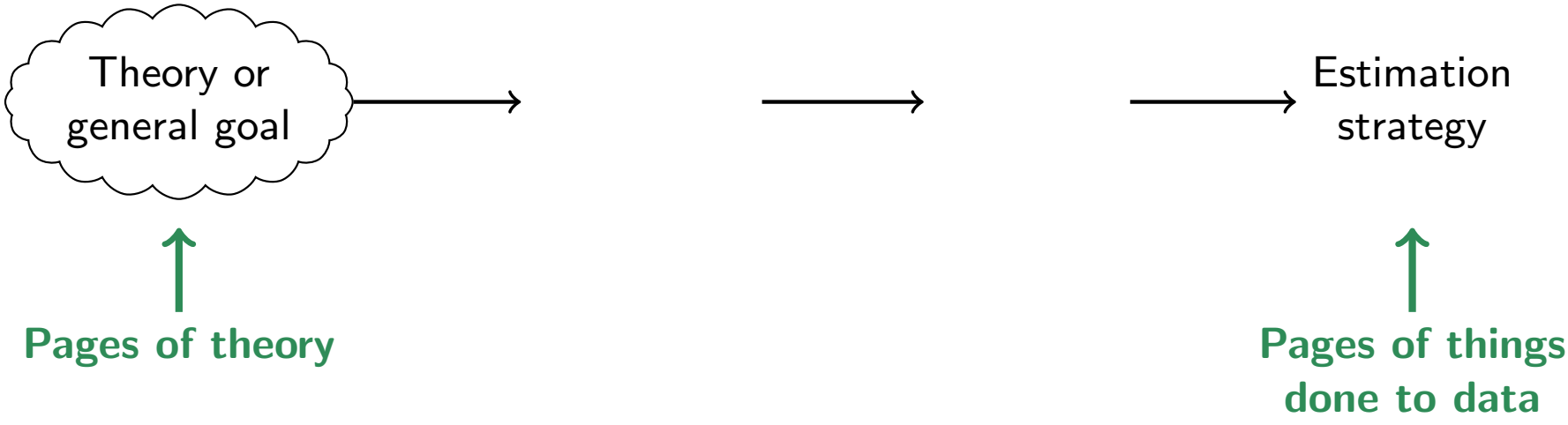
Research framework: Estimands connect theory to evidence



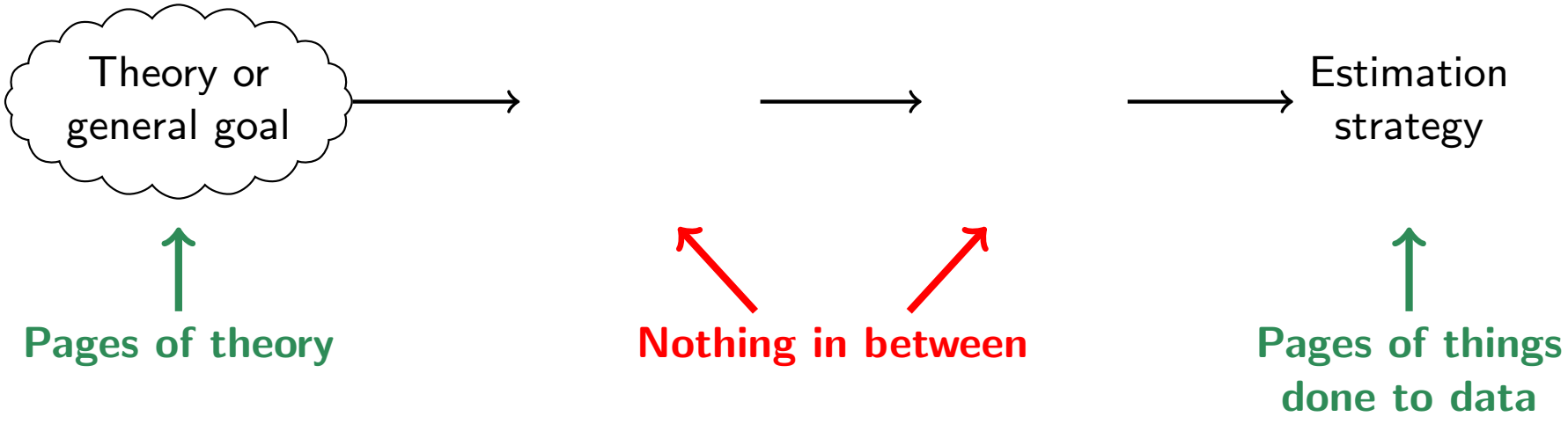
Research framework: Estimands connect theory to evidence



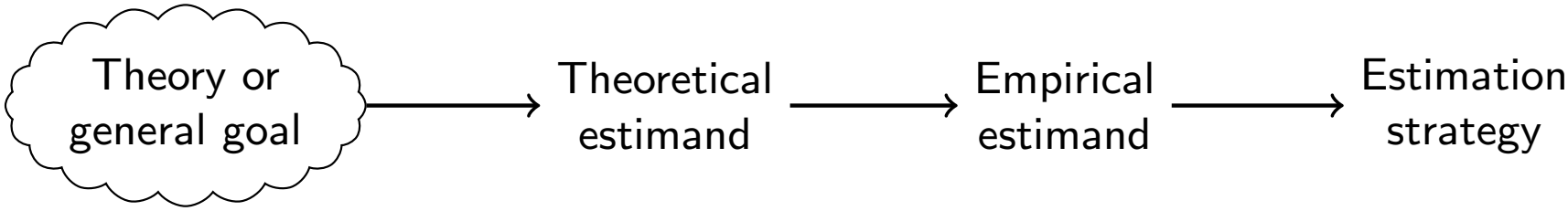
Research framework: Estimands connect theory to evidence



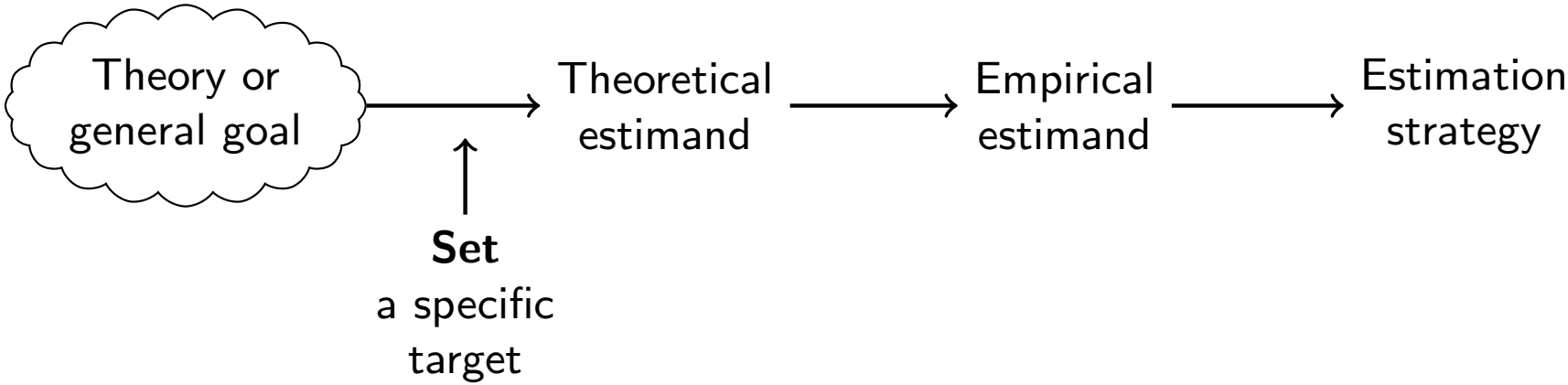
Research framework: Estimands connect theory to evidence



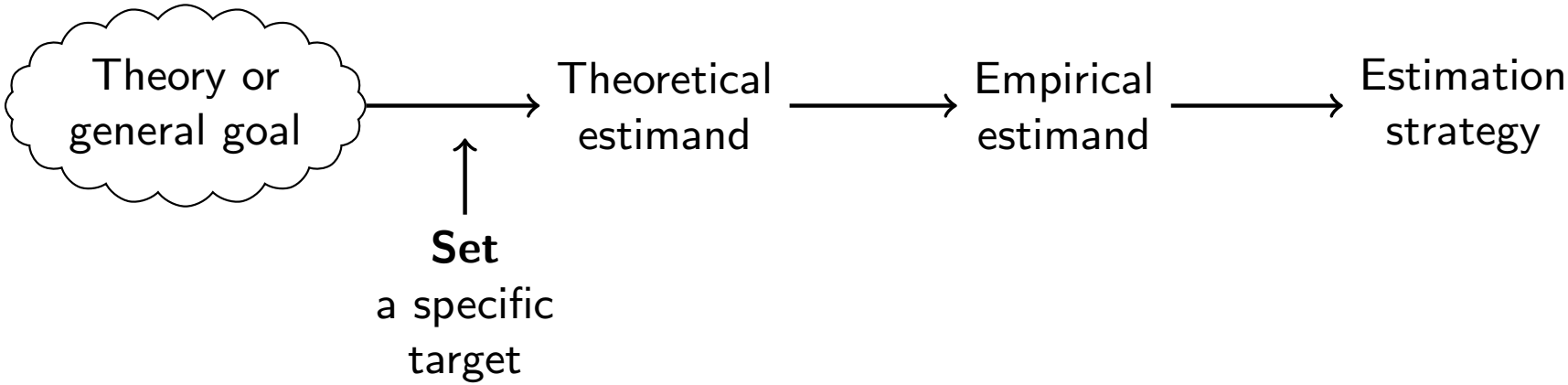
Research framework: Estimands connect theory to evidence



Research framework: Estimands connect theory to evidence



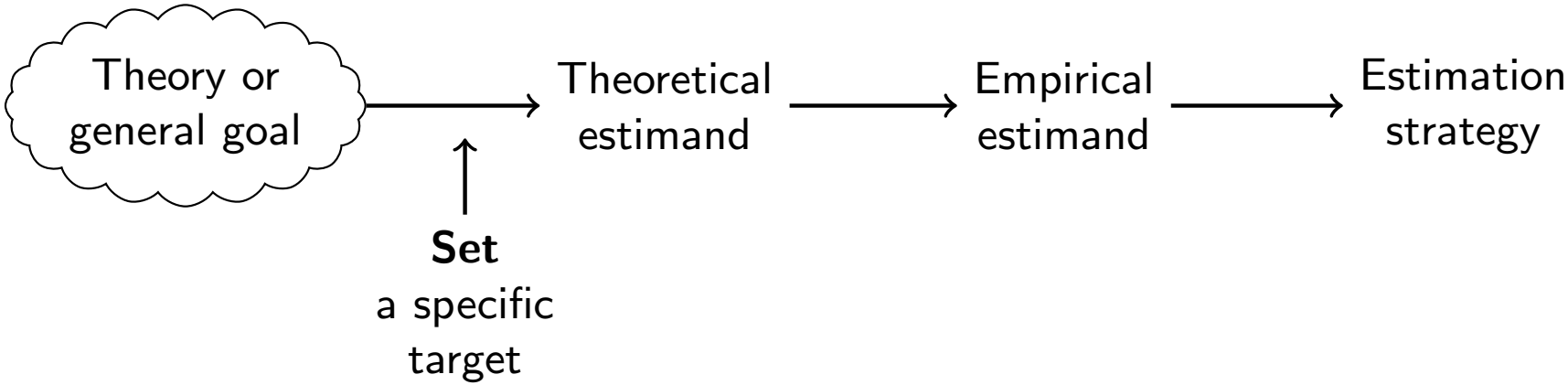
Research framework: Estimands connect theory to evidence



Definition

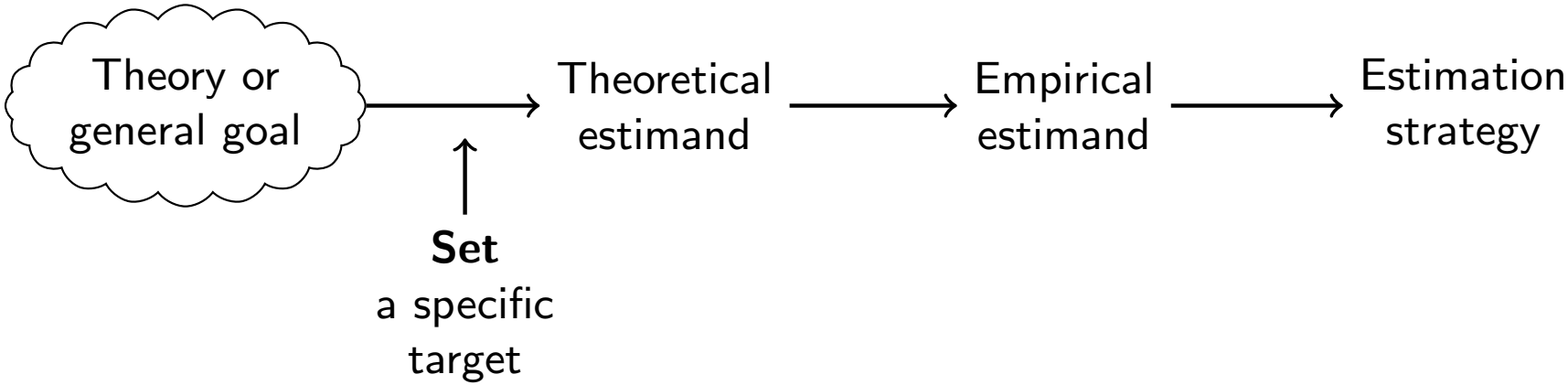
A **unit-specific quantity** aggregated over a **target population**

Research framework: Estimands connect theory to evidence



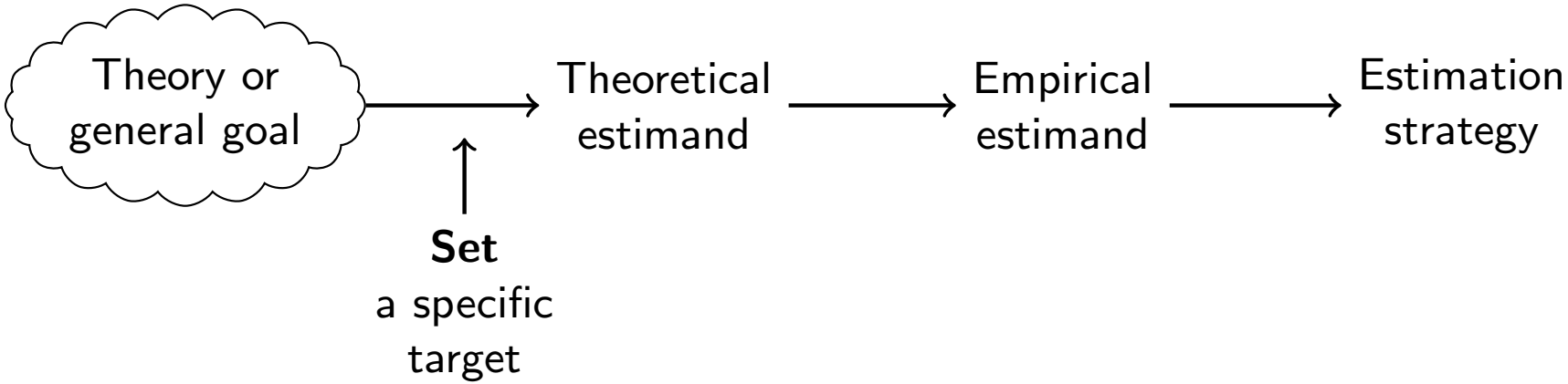
Definition	Example
A **unit-specific quantity** aggregated over a **target population**	$\frac{1}{\text{Size of U.S. adult population}} \sum_{i \text{ in U.S. adult population}} \left(\text{Employed}_i \right)$

Research framework: Estimands connect theory to evidence



Definition	Example
A **unit-specific quantity** aggregated over a **target population**	$\frac{1}{\text{Size of U.S. adult population}} \sum_{i \text{ in U.S. adult population}} \Big(\underbrace{\text{Employed}_i(\text{Job training})}_{\text{Employment if received job training}} - \underbrace{\text{Employed}_i(\text{No job training})}_{\text{Employment if did not receive job training}} \Big)$

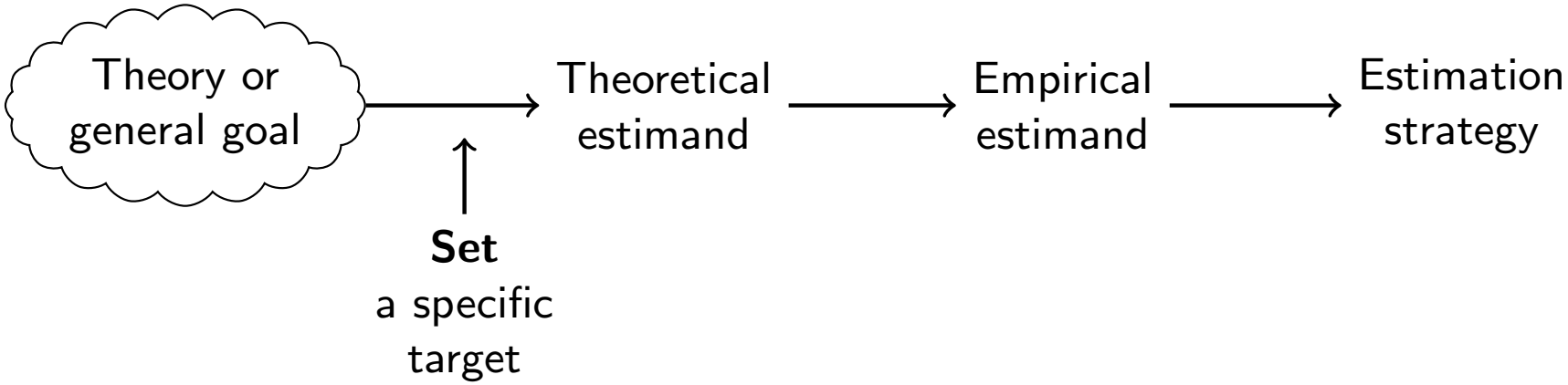
Research framework: Estimands connect theory to evidence



Definition	Example
A **unit-specific quantity** aggregated over a **target population**	$\frac{1}{\text{Size of U.S. adult population}} \sum_{i \text{ in U.S. adult population}} \Big(\underbrace{\text{Employed}_i(\text{Job training})}_{\text{Employment if received job training}} - \underbrace{\text{Employed}_i(\text{No job training})}_{\text{Employment if did not receive job training}} \Big)$

Lieberson 1987, Abbott 1988, Freedman 1991, Xie 2013, Hernán 2018

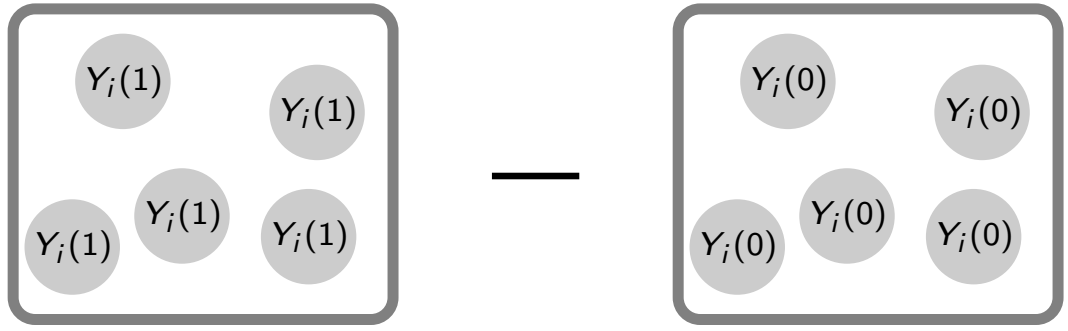
Research framework: Estimands connect theory to evidence



Definition

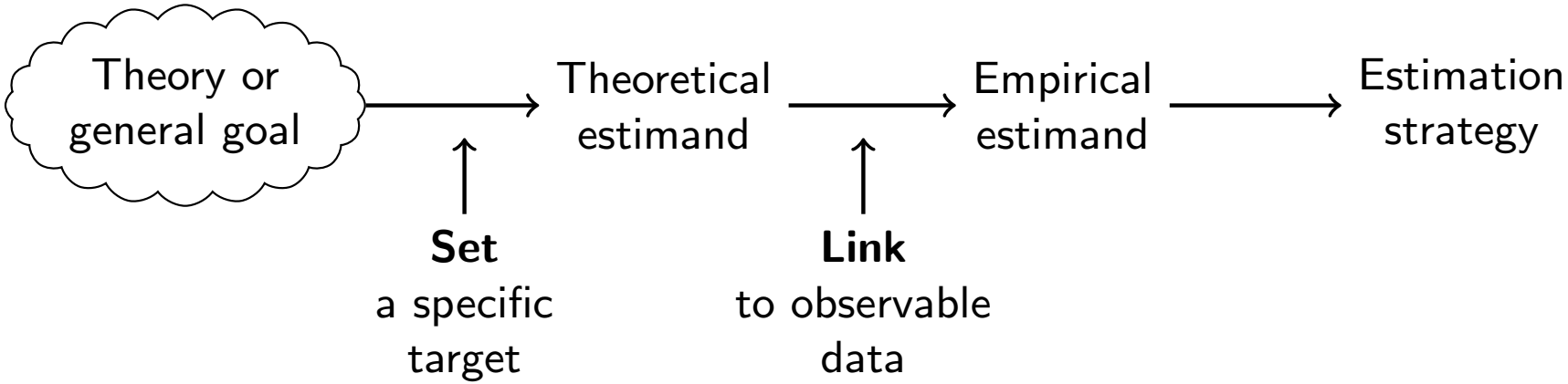
A **unit-specific quantity** aggregated over a **target population**

Example



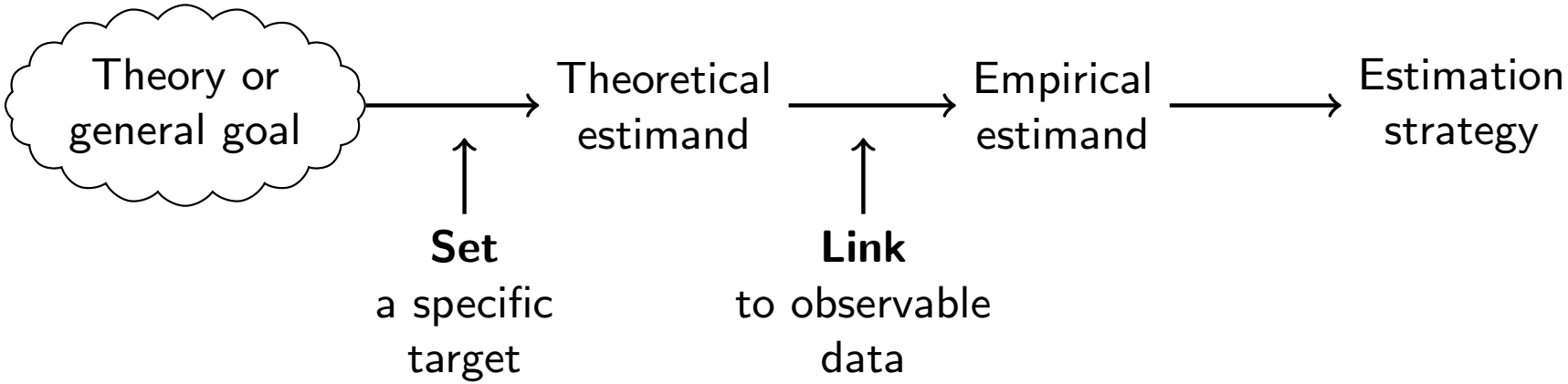
Lieberson 1987, Abbott 1988, Freedman 1991, Xie 2013, Hernán 2018

Research framework: Estimands connect theory to evidence



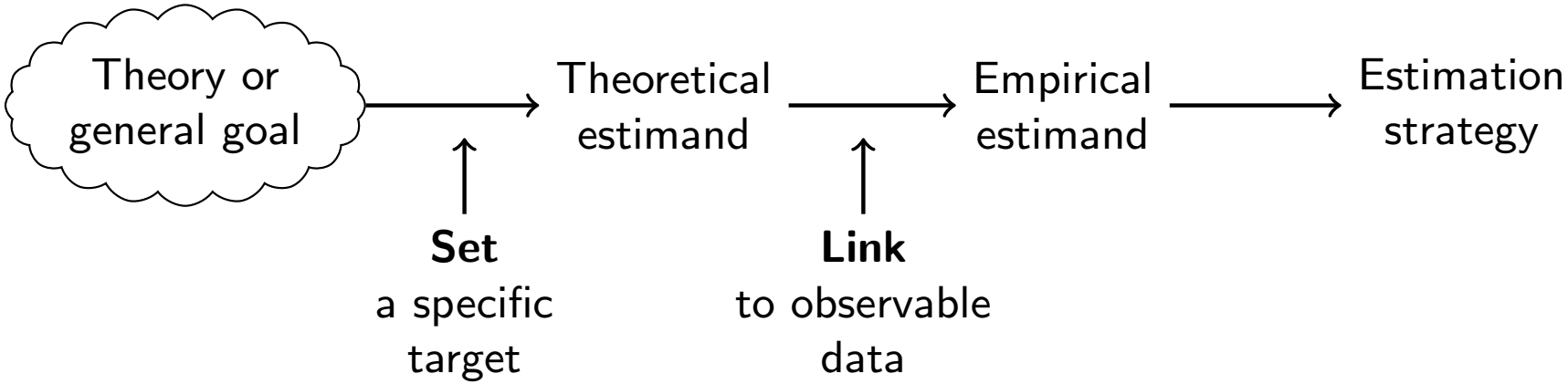
Definition	Example
A quantity involving **observable data**	$Y_i(1)$ $Y_i(1)$ $Y_i(1)$ $Y_i(1)$ $Y_i(1)$ − $Y_i(0)$ $Y_i(0)$ $Y_i(0)$ $Y_i(0)$ $Y_i(0)$

Research framework: Estimands connect theory to evidence



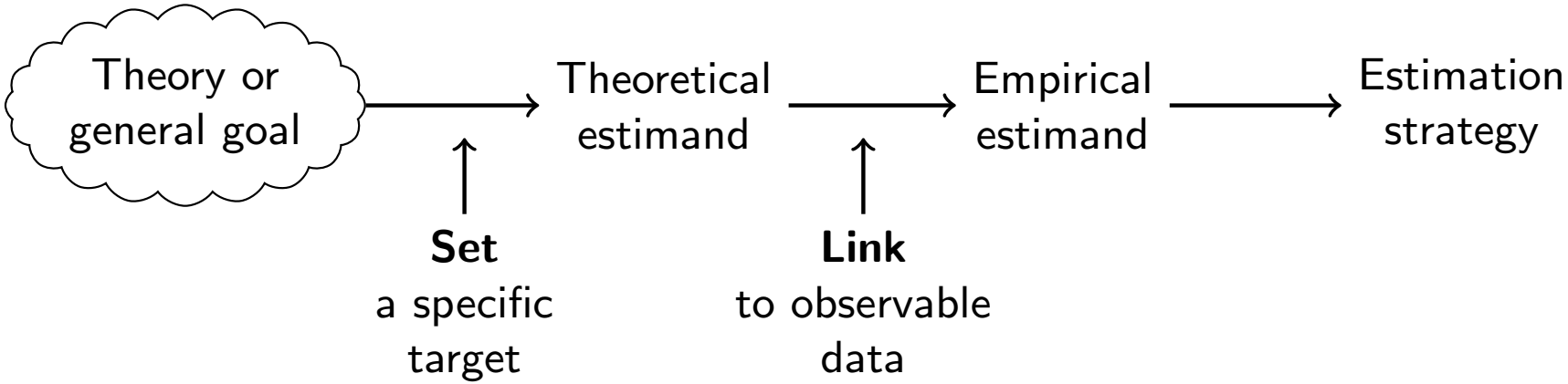
Definition	Example
A quantity involving **observable data**	$Y_i(1)$ $Y_i(1)$ $Y_i(1)$ $Y_i(1)$ $Y_i(1)$ — $Y_i(0)$ $Y_i(0)$ $Y_i(0)$ $Y_i(0)$ $Y_i(0)$

Research framework: Estimands connect theory to evidence



Definition	Example
A quantity involving **observable data**	$Y_i(1)$ $Y_i(1)$ $Y_i(1)$ $Y_i(1)$ $Y_i(1)$ — $Y_i(0)$ $Y_i(0)$ $Y_i(0)$ $Y_i(0)$ $Y_i(0)$

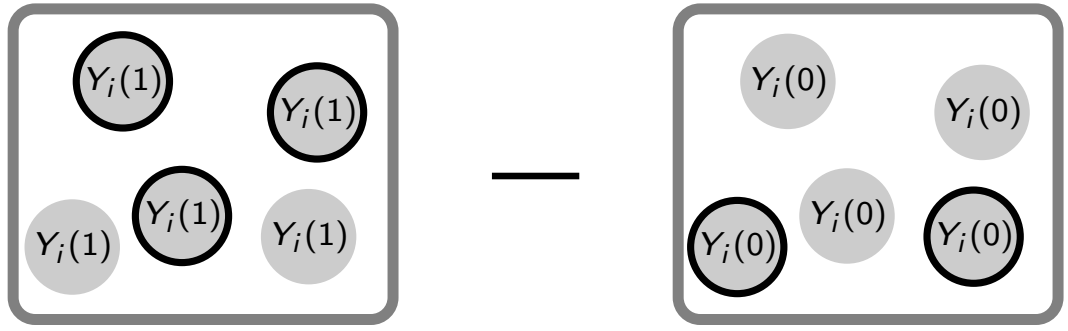
Research framework: Estimands connect theory to evidence

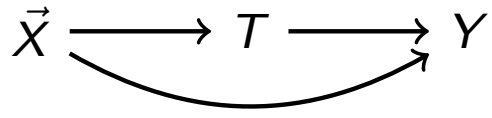


Definition

A quantity involving **observable data**

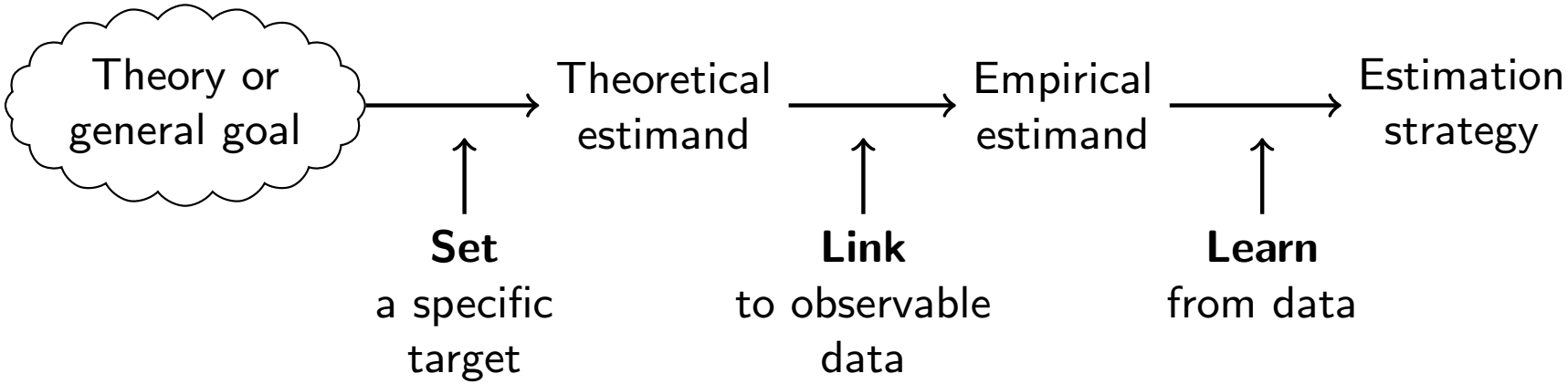
Example

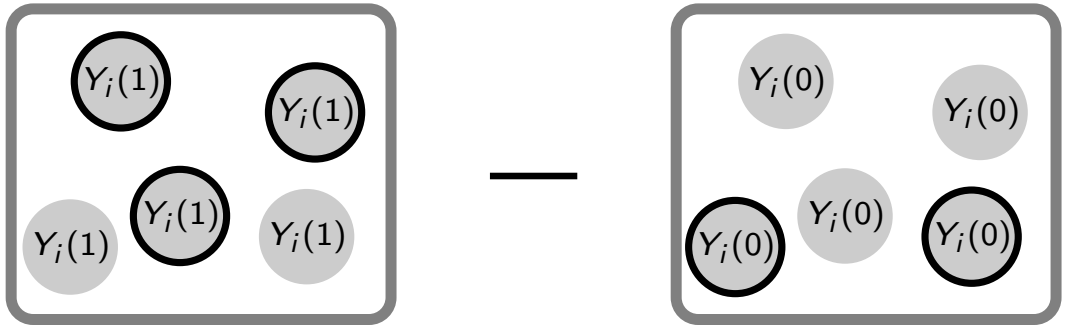




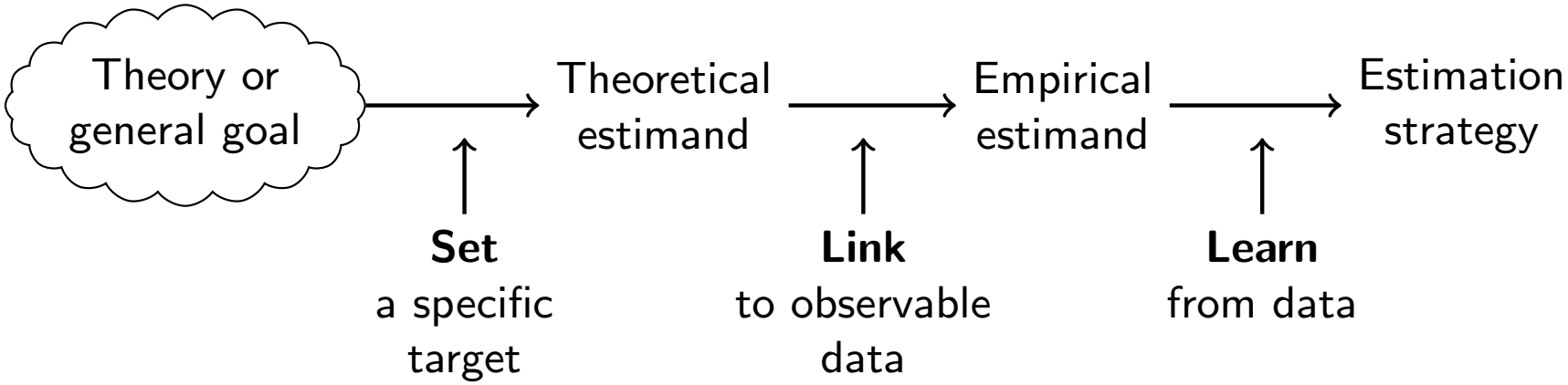
Pearl 2009, Imbens and Rubin 2015,
Morgan and Winship 2015, Elwert and Winship 2014

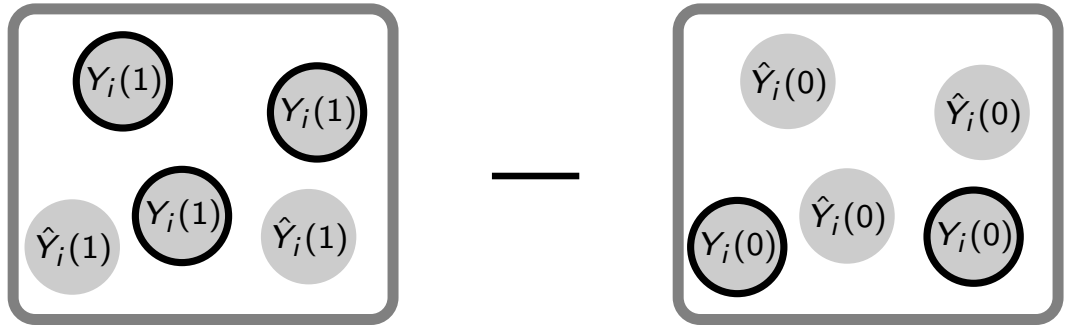
Research framework: Estimands connect theory to evidence



Definition	Example
An algorithm applied to data	

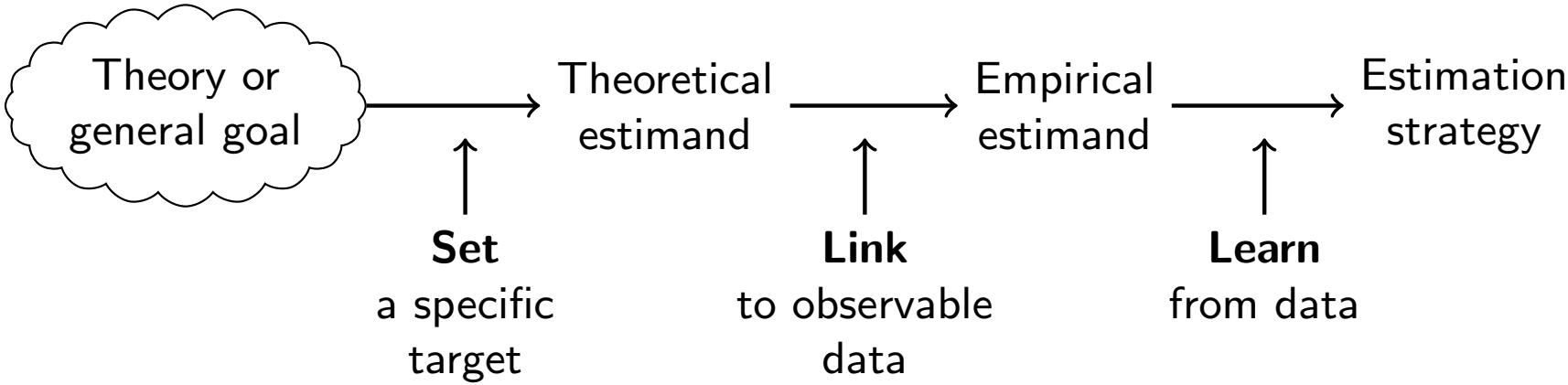
Research framework: Estimands connect theory to evidence



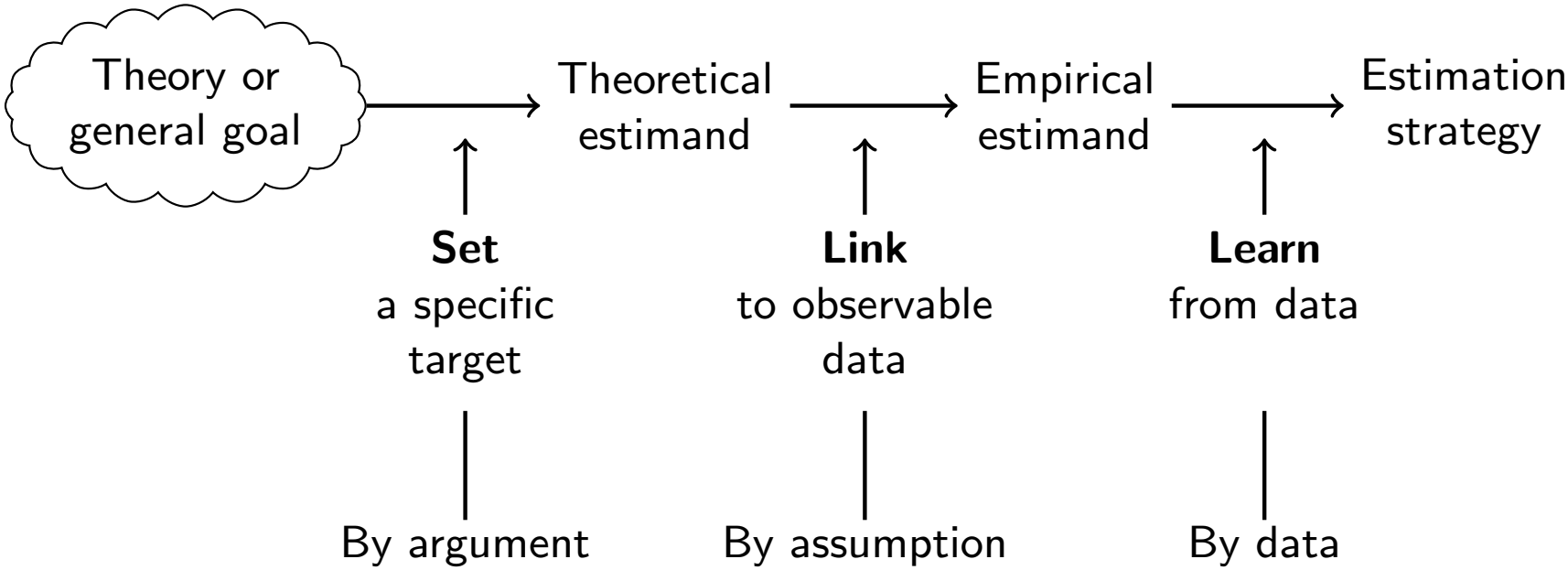
Definition	Example
An algorithm applied to data	

Young 2009, Watts 2014, Berk et al. 2019, Molina and Garip 2019

Research framework: Estimands connect theory to evidence



Research framework: Estimands connect theory to evidence



1. Set the target quantity.

Describe a population

What is the proportion employed
among U.S. resident women ages 21–35?

Describe a population

What is the proportion employed
among U.S. resident women ages 21–35?

Woman 1
Woman 2
Woman 3
Woman 4

Describe a population

What is the proportion employed
among U.S. resident women ages 21–35?

	Employed?
Woman 1	1
Woman 2	0
Woman 3	1
Woman 4	1

Describe population subgroups

What is the proportion employed
among U.S. resident women ages 21–35,
comparing mothers to non-mothers?

Describe population subgroups

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?

	Employed?
Mother 1	0
Mother 2	0
Mother 3	0
Mother 4	1

	Employed?
Non-Mother 1	1
Non-Mother 2	0
Non-Mother 3	1
Non-Mother 4	1

Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

Woman 1
Woman 2
Woman 3
Woman 4

Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$
Woman 1	0
Woman 2	0
Woman 3	0
Woman 4	1

Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$
Woman 1	0	1
Woman 2	0	0
Woman 3	0	1
Woman 4	1	1

Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0

Describe population subgroups

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?

	Employed?
Mother 1	0
Mother 2	0
Mother 3	0
Mother 4	1

	Employed?
Non-Mother 1	1
Non-Mother 2	0
Non-Mother 3	1
Non-Mother 4	1

Very **different** research goals

Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0

If the estimand is causal, you have to tell the reader:

what is the intervention?

(Hernán 2016)

If the estimand is causal, you have to tell the reader:

what is the intervention?

Example: Effect of motherhood on employment

(Hernán 2016)

If the estimand is causal, you have to tell the reader:

what is the intervention?

Example: Effect of motherhood on employment

— How many kids?

(Hernán 2016)

If the estimand is causal, you have to tell the reader:

what is the intervention?

Example: Effect of motherhood on employment

— How many kids?

— Adopted? Biological?

(Hernán 2016)

If the estimand is causal, you have to tell the reader:

what is the intervention?

Example: Effect of motherhood on employment

— How many kids?

— Adopted? Biological?

— Sex mix of the kids?

(Hernán 2016)

If the estimand is causal, you have to tell the reader:

what is the intervention?

Example: Effect of motherhood on employment

— How many kids?

— Adopted? Biological?

— Sex mix of the kids?

— Born this year? 5 years ago?

(Hernán 2016)

If the estimand is causal, you have to tell the reader:

what is the intervention?

Example: Effect of motherhood on employment

— How many kids?

— Adopted? Biological?

— Sex mix of the kids?

— Born this year? 5 years ago?

One resolution: Effect **among mothers** with their factual configuration of kids, compared with a counterfactual of having no kids

(Hernán 2016)

Group Exercise

If the estimand is causal, you have to tell the reader:

what is the intervention?

Group Exercise

If the estimand is causal, you have to tell the reader:

what is the intervention?

> "If poor women would marry,
> they would no longer be poor."
>
> — A policymaker

Group Exercise

If the estimand is causal, you have to tell the reader:

what is the intervention?

> "If poor women would marry,
> they would no longer be poor."
>
> — A policymaker

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\text{Poverty}_i(\text{Married}) - \text{Poverty}_i(\text{Unmarried})\Big)$$

Group Exercise

If the estimand is causal, you have to tell the reader:

what is the intervention?

> "If poor women would marry,
> they would no longer be poor."
>
> — A policymaker

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\text{Poverty}_i(\text{Married}) - \text{Poverty}_i(\text{Unmarried})\Big)$$

Is this policymaker being sufficiently precise about the intervention?

Group Exercise

If the estimand is causal, you have to tell the reader:

what is the intervention?

> "If poor women would marry,
> they would no longer be poor."
>
> — A policymaker

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\text{Poverty}_i(\text{Married}) - \text{Poverty}_i(\text{Unmarried})\Big)$$

Is this policymaker being sufficiently precise about the intervention?

Suggestion: For a given woman i, are there versions of "married" and "unmarried" that might lead to distinct outcomes? Does it matter who you marry?

Regardless of descriptive or causal,
you have to tell the reader the **target population**

Regardless of descriptive or causal,
you have to tell the reader the **target population**

Example: The unemployment rate

Regardless of descriptive or causal,
you have to tell the reader the **target population**

Example: The unemployment rate

U.S. non-institutionalized civilians age 16+

actively looking for work in past 4 weeks

Group Exercise

Regardless of descriptive or causal,
you have to tell the reader the **target population**

Group Exercise

Regardless of descriptive or causal,
you have to tell the reader the **target population**

> "My online sample is nationally representative"
>
> — An academic

Group Exercise

Regardless of descriptive or causal,
you have to tell the reader the **target population**

> "My online sample is nationally representative"
>
> — An academic

What might the target population be if the question is:

Group Exercise

Regardless of descriptive or causal,
you have to tell the reader the **target population**

> "My online sample is nationally representative"
>
> — An academic

What might the target population be if the question is:

1) Do you plan to vote for Biden?

Group Exercise

Regardless of descriptive or causal,
you have to tell the reader the **target population**

> "My online sample is nationally representative"
>
> — An academic

What might the target population be if the question is:

1) Do you plan to vote for Biden?

2) What is your hourly wage?

Group Exercise

Regardless of descriptive or causal,
you have to tell the reader the **target population**

> "My online sample is nationally representative"
>
> — An academic

What might the target population be if the question is:

1) Do you plan to vote for Biden?
2) What is your hourly wage?
3) Would you recommend this hypothetical resume for an interview?

Policy implications depend on the type of claim.

Policy implications depend on the type of claim.

Descriptive Claim

"People who go to Disneyland twice are more likely to buy an annual pass"

Policy implications depend on the type of claim.

Descriptive Claim

"People who go to Disneyland twice are more likely to buy an annual pass"

Causal Claim

"A second trip to Disneyland causes people to decide they need an annual pass"

Policy implications depend on the type of claim.

Descriptive Claim

"People who go to Disneyland twice are more likely to buy an annual pass"

Causal Claim

"A second trip to Disneyland causes people to decide they need an annual pass"

Policy implications for Disney

Policy implications depend on the type of claim.

Descriptive Claim

"People who go to Disneyland twice are more likely to buy an annual pass"

Causal Claim

"A second trip to Disneyland causes people to decide they need an annual pass"

Target annual pass ads at the return visitors

Policy implications for Disney

Policy implications depend on the type of claim.

Descriptive Claim

"People who go to Disneyland twice are more likely to buy an annual pass"

Target annual pass ads at the return visitors

Causal Claim

"A second trip to Disneyland causes people to decide they need an annual pass"

Offer discounts to induce a second visit

Policy implications for Disney

Group Exercise

Policy implications depend on the type of claim.

Descriptive Claim

Causal Claim

Group Exercise

Policy implications depend on the type of claim.

Descriptive Claim

"College graduates are more financially stable than non-graduates"

Causal Claim

Group Exercise

Policy implications depend on the type of claim.

Descriptive Claim

"College graduates are more financially stable than non-graduates"

Causal Claim

"College completion causes greater financial stability"

Group Exercise

Policy implications depend on the type of claim.

Descriptive Claim

"College graduates are more financially stable than non-graduates"

Causal Claim

"College completion causes greater financial stability"

What kinds of policies would these claims suggest?

Professional development

- Social media
- Professional networking
- Anything else you want to discuss!