A Brief Introduction to Causal Inference

Brady Neal

causalcourse.com

What is causal inference?

Inferring the effects of any treatment/policy/intervention/etc.

What is causal inference?

Inferring the effects of any treatment/policy/intervention/etc.

Examples:

- Effect of treatment on a disease

What is causal inference?

Inferring the effects of any treatment/policy/intervention/etc.

Examples:

- Effect of treatment on a disease
- Effect of climate change policy on emissions

What is causal inference?

Inferring the effects of any treatment/policy/intervention/etc.

Examples:

- Effect of treatment on a disease
- Effect of climate change policy on emissions
- Effect of social media on mental health

What is causal inference?

Inferring the effects of any treatment/policy/intervention/etc.

Examples:

- Effect of treatment on a disease
- Effect of climate change policy on emissions
- Effect of social media on mental health
- Many more (effect of X on Y)

Motivating example: Simpson's paradox

Correlation does not imply causation

Then, what does imply causation?

Causation in observational studies

Motivating example: Simpson's paradox

Correlation does not imply causation

Then, what does imply causation?

Causation in observational studies

Simpson's paradox: COVID-27

New disease: COVID-27 

Simpson's paradox: COVID-27

New disease: COVID-27 

Treatment T: A (0) and B (1)

Simpson’s paradox: COVID-27

New disease: COVID-27

Treatment T: A (0) and B (1)

Simpson's paradox: COVID-27

New disease: COVID-27

Treatment T: A (0) and B (1)

Condition C: mild (0) or severe (1)

YOU

Simpson's paradox: COVID-27

New disease: COVID-27

Treatment T: A (0) and B (1)

Condition C: mild (0) or severe (1)

Outcome Y: alive (0) or dead (1)

Simpson's paradox: mortality rate table

Treatment	Total
A	**16%** (240/1500)
B	19% (105/550)
	$\mathbb{E}[Y \mid T]$

Simpson's paradox: mortality rate table

Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)
	$\mathbb{E}[Y \mid T, C = 0]$	$\mathbb{E}[Y \mid T, C = 1]$	$\mathbb{E}[Y \mid T]$

Simpson's paradox: mortality rate table

Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)
	$\mathbb{E}[Y \mid T, C = 0]$	$\mathbb{E}[Y \mid T, C = 1]$	$\mathbb{E}[Y \mid T]$

$$\frac{1400}{1500}(0.15) + \frac{100}{1500}(0.30) = 0.16$$

$$\frac{50}{550}(0.10) + \frac{500}{550}(0.20) = 0.19$$

Simpson's paradox: mortality rate table

Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)
	$\mathbb{E}[Y \mid T, C = 0]$	$\mathbb{E}[Y \mid T, C = 1]$	$\mathbb{E}[Y \mid T]$

$$\boxed{\frac{1400}{1500}}(0.15) + \boxed{\frac{100}{1500}}(0.30) = 0.16$$

$$\frac{50}{550}(0.10) + \frac{500}{550}(0.20) = 0.19$$

Simpson's paradox: mortality rate table

Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)
	$\mathbb{E}[Y \mid T, C=0]$	$\mathbb{E}[Y \mid T, C=1]$	$\mathbb{E}[Y \mid T]$

$$\boxed{\frac{1400}{1500}}(0.15) + \boxed{\frac{100}{1500}}(0.30) = 0.16$$

$$\boxed{\frac{50}{550}}(0.10) + \boxed{\frac{500}{550}}(0.20) = 0.19$$

Simpson's paradox: mortality rate table

	Condition		
Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)
	$\mathbb{E}[Y \mid T, C = 0]$	$\mathbb{E}[Y \mid T, C = 1]$	$\mathbb{E}[Y \mid T]$

$$\boxed{\frac{1400}{1500}}(0.15) + \boxed{\frac{100}{1500}}(0.30) = 0.16$$

$$\boxed{\frac{50}{550}}(0.10) + \boxed{\frac{500}{550}}(0.20) = 0.19$$

Which treatment should you choose?

Simpson's paradox: scenario 1 (treatment B)

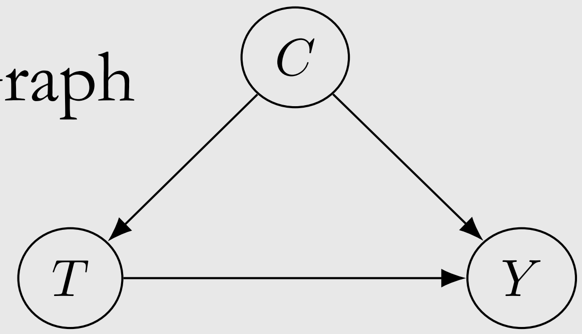
Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Simpson's paradox: scenario 1 (treatment B)

Treatment		Condition: Mild	Condition: Severe	Condition: Total
	A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
	B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

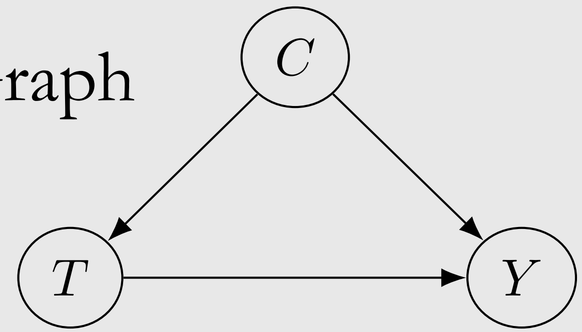
Causal Graph

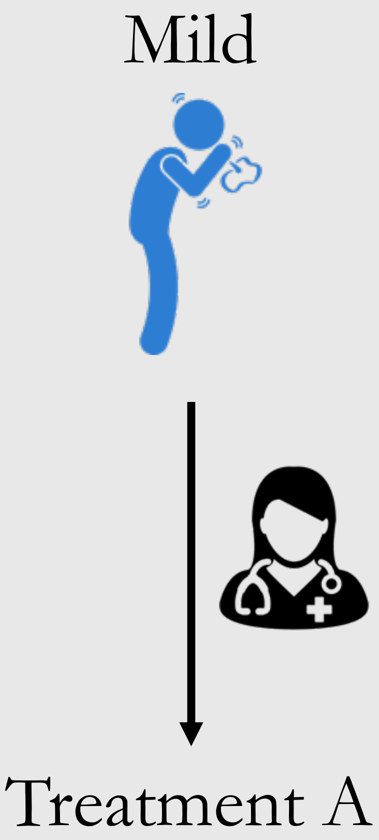


Simpson's paradox: scenario 1 (treatment B)

Treatment	Condition: Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Causal Graph

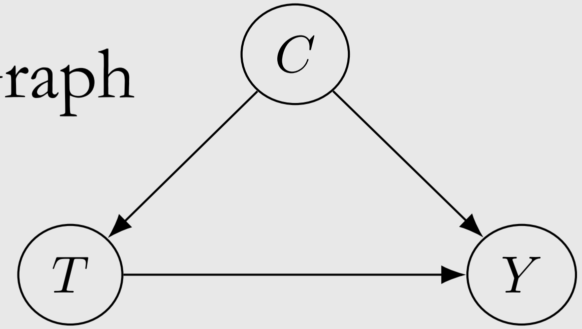


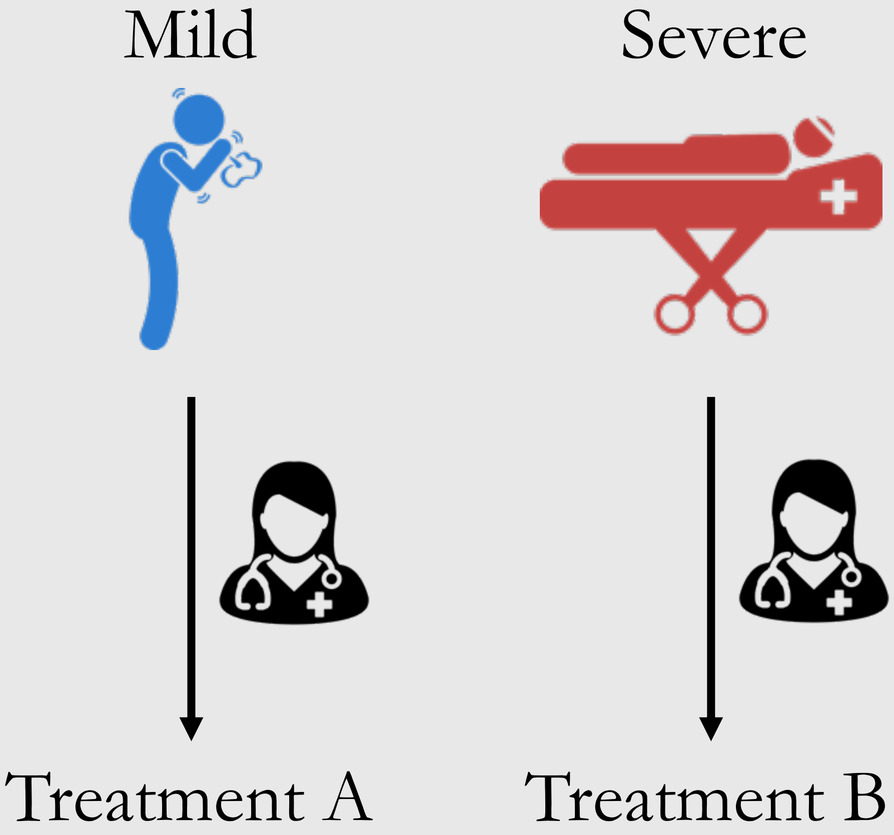


Simpson's paradox: scenario 1 (treatment B)

		Condition		
Treatment		Mild	Severe	Total
	A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
	B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Causal Graph





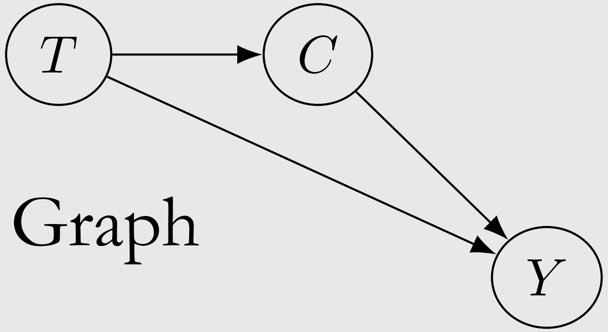
Simpson's paradox: scenario 2 (treatment A)

Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Simpson's paradox: scenario 2 (treatment A)

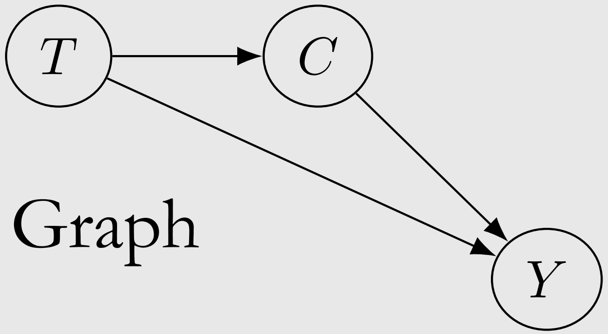
		Condition		
		Mild	Severe	Total
Treatment	A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
	B	**10%** (5/50)	**20%** (100/500)	19% (105/550)



Causal Graph

Simpson's paradox: scenario 2 (treatment A)

		Condition	
Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)



Causal Graph

Treatment B

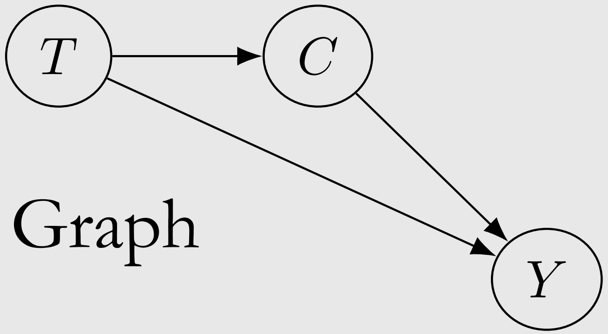
$\downarrow$

Severe

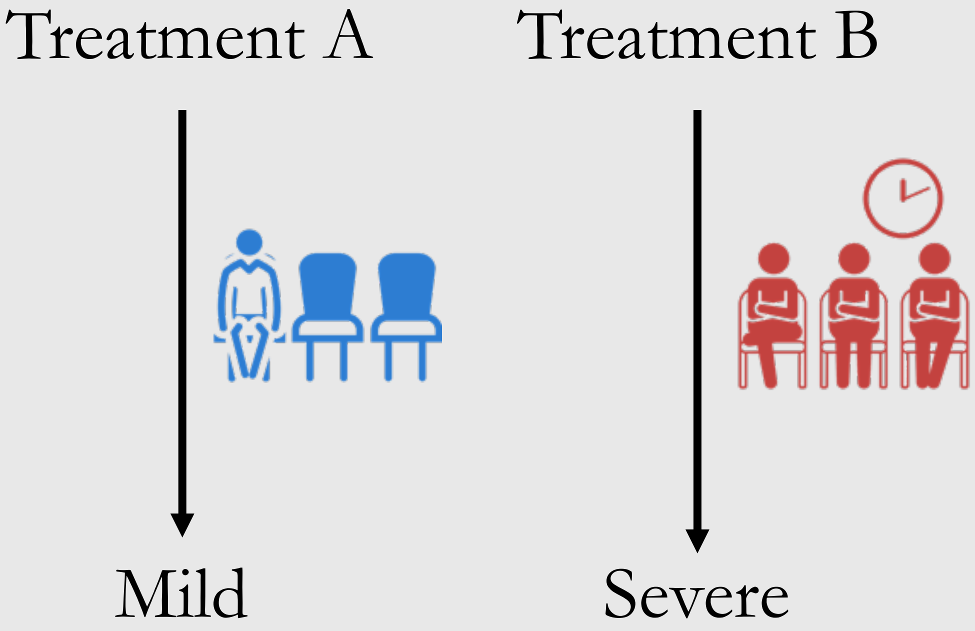
Simpson's paradox: scenario 2 (treatment A)

Condition

Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)



Causal Graph

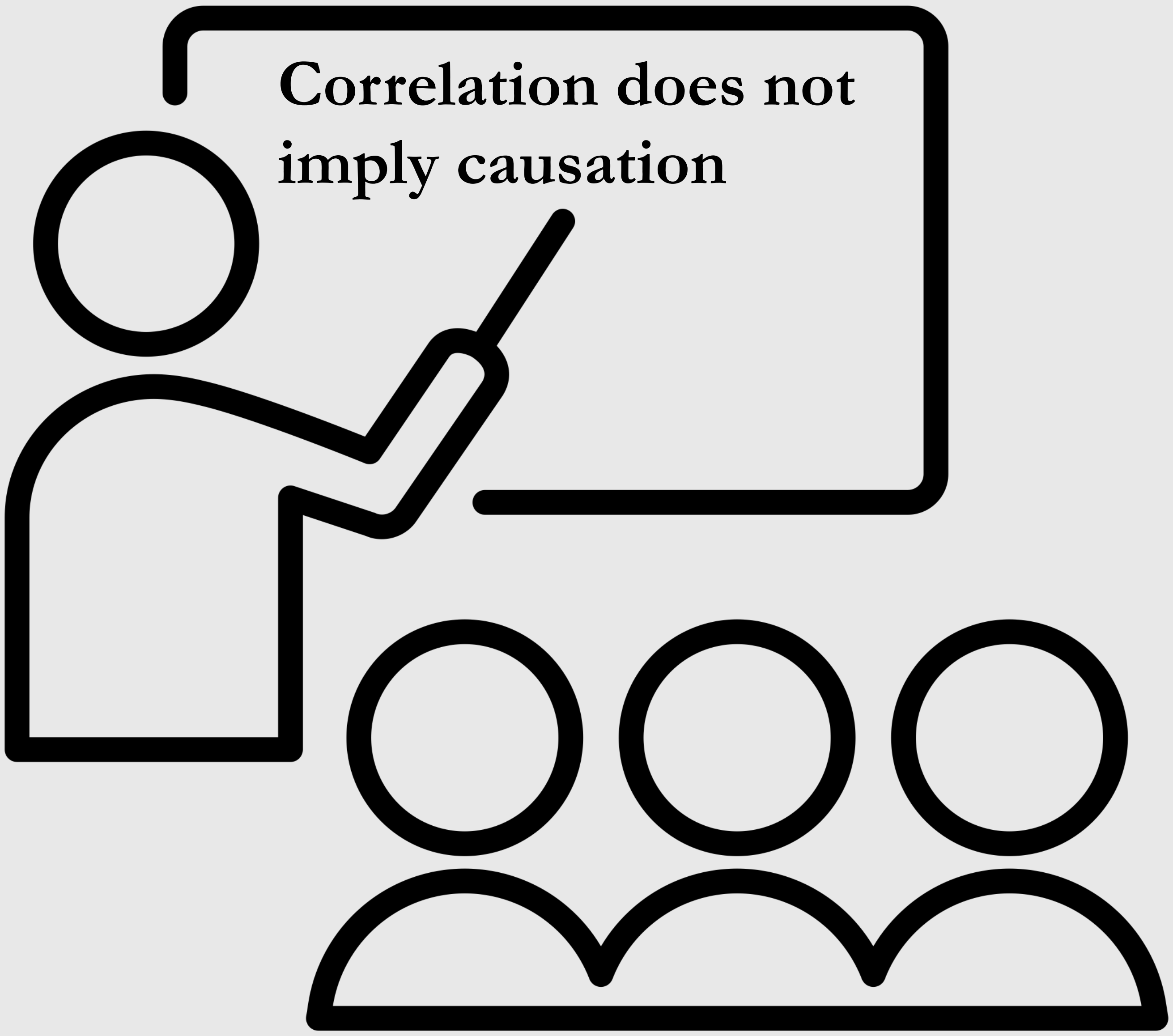


Motivating example: Simpson's paradox

Correlation does not imply causation

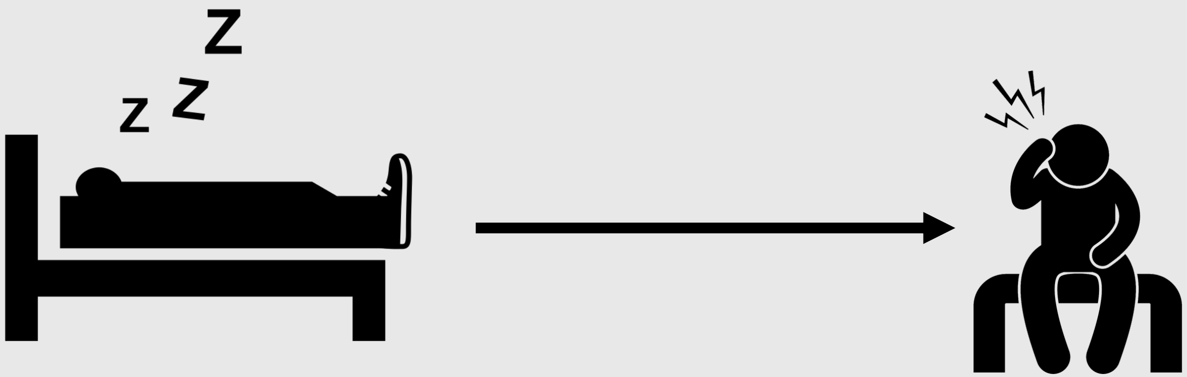
Then, what does imply causation?

Causation in observational studies


Correlation does not imply causation

Correlation does not imply causation

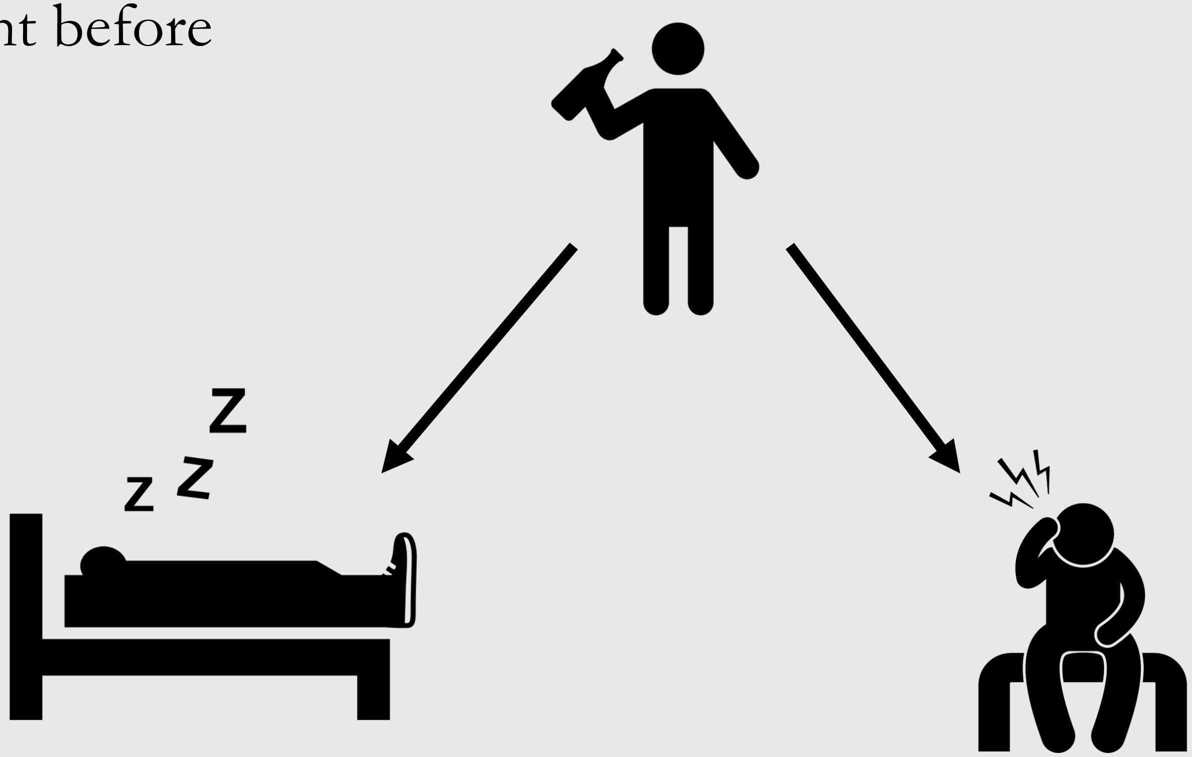
Sleeping with shoes on is strongly correlated with waking up with a headache



Correlation does not imply causation

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

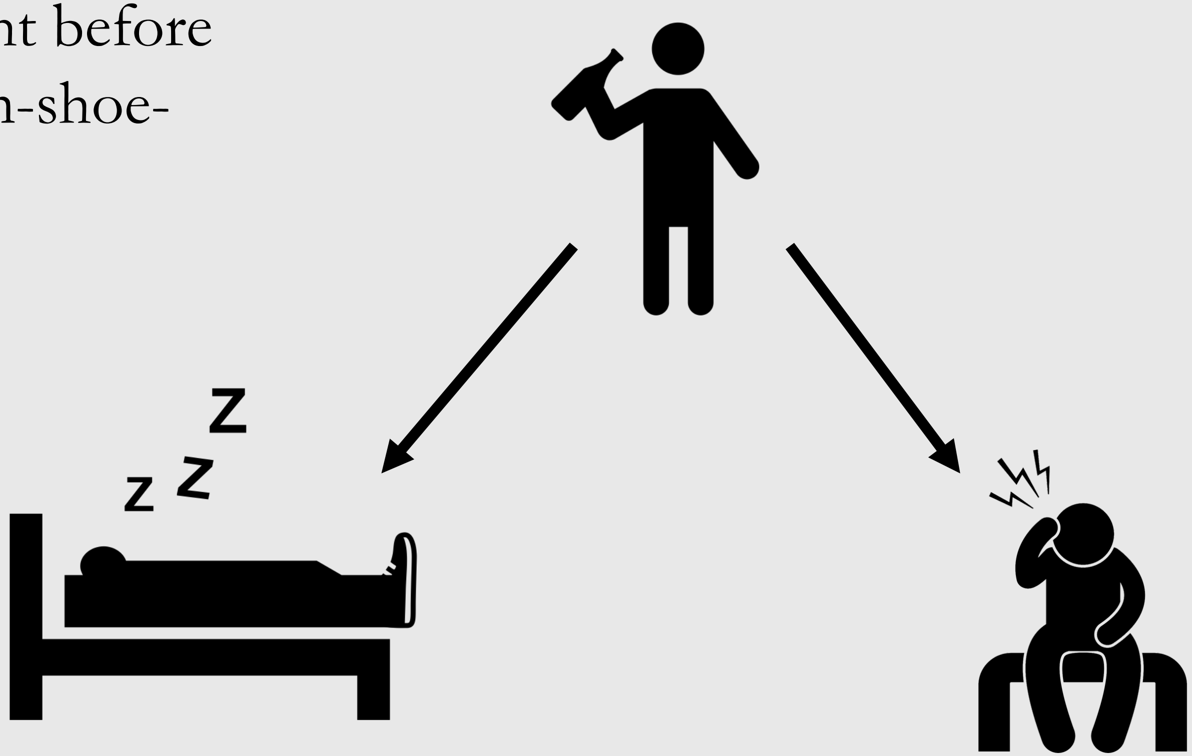


Correlation does not imply causation

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way

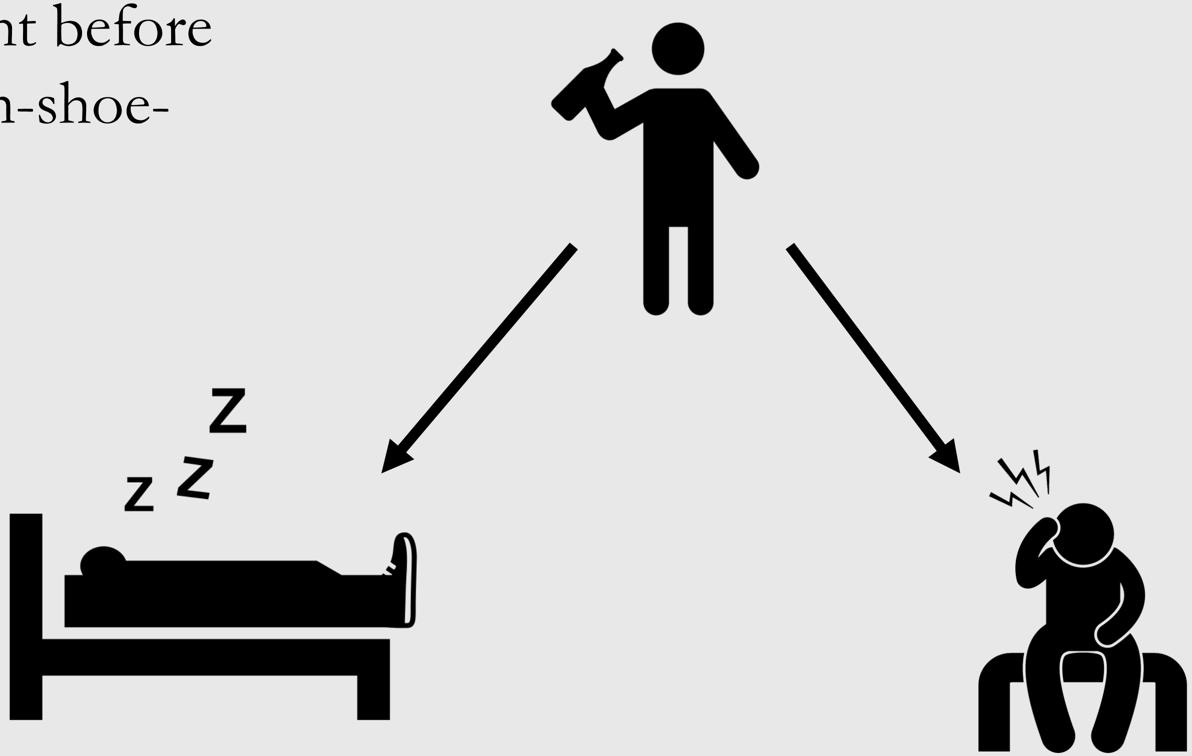


Correlation does not imply causation

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding

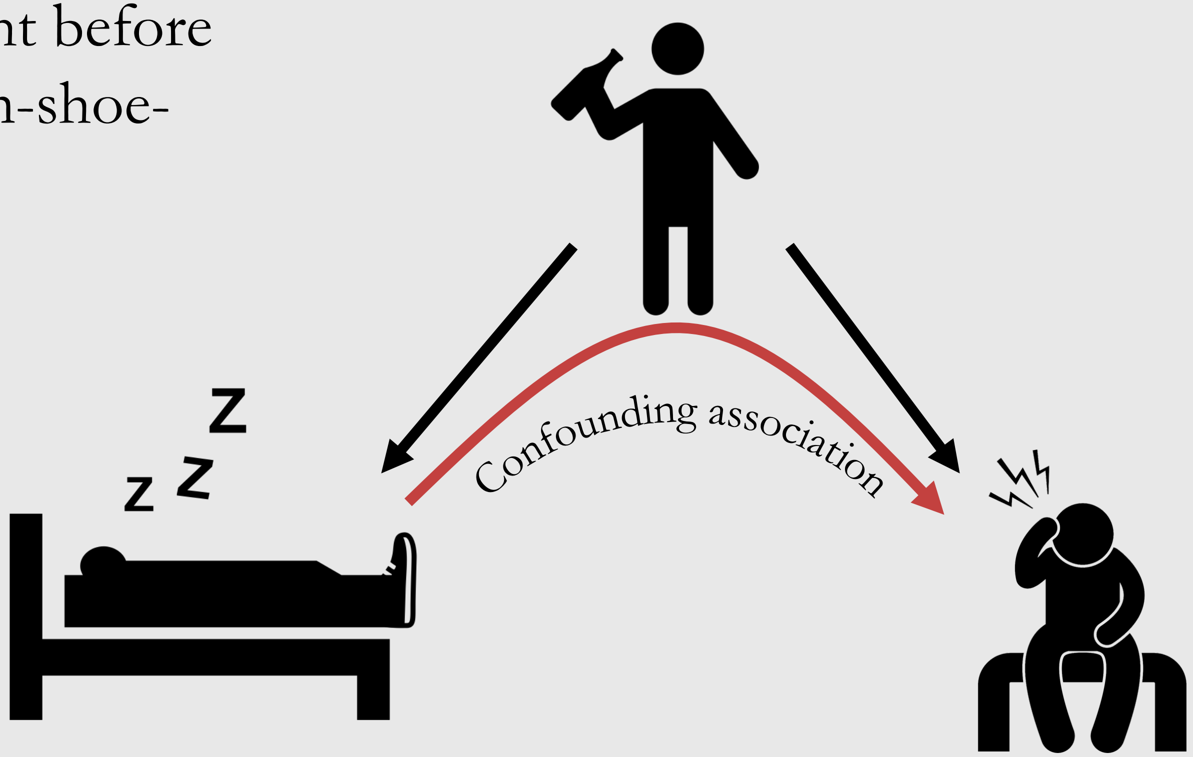


Correlation does not imply causation

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding

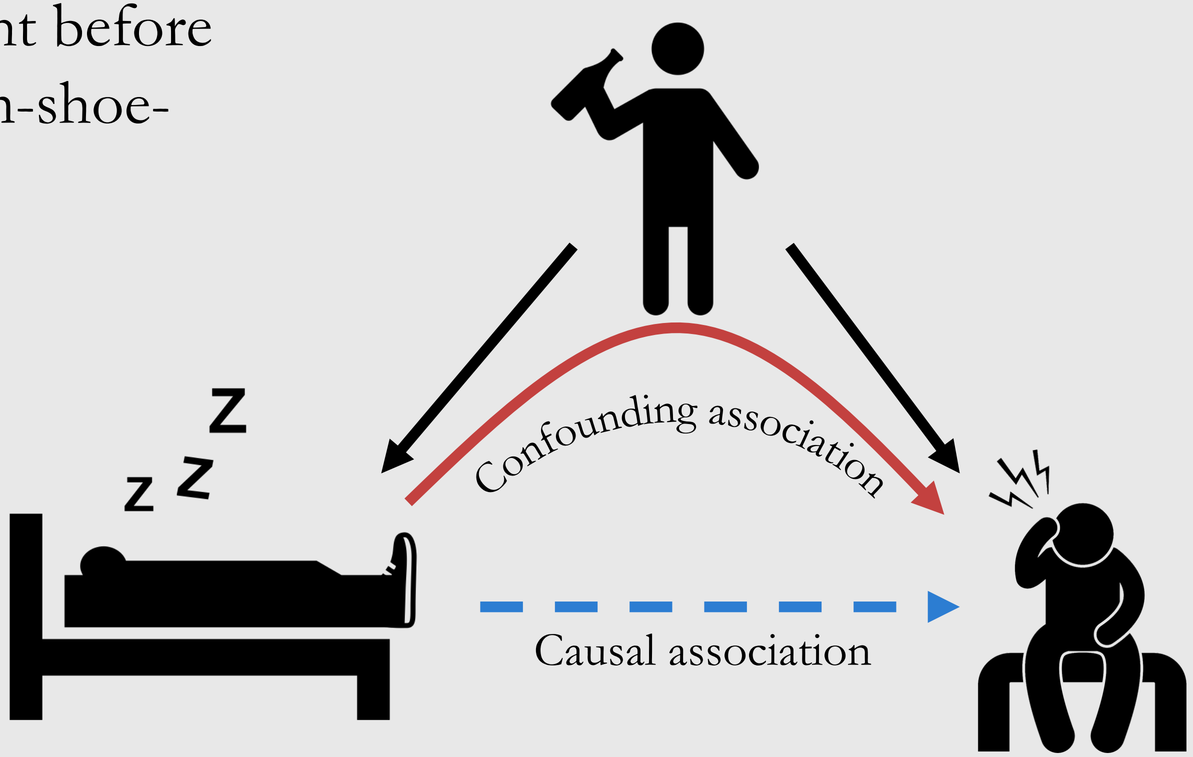


Correlation does not imply causation

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding



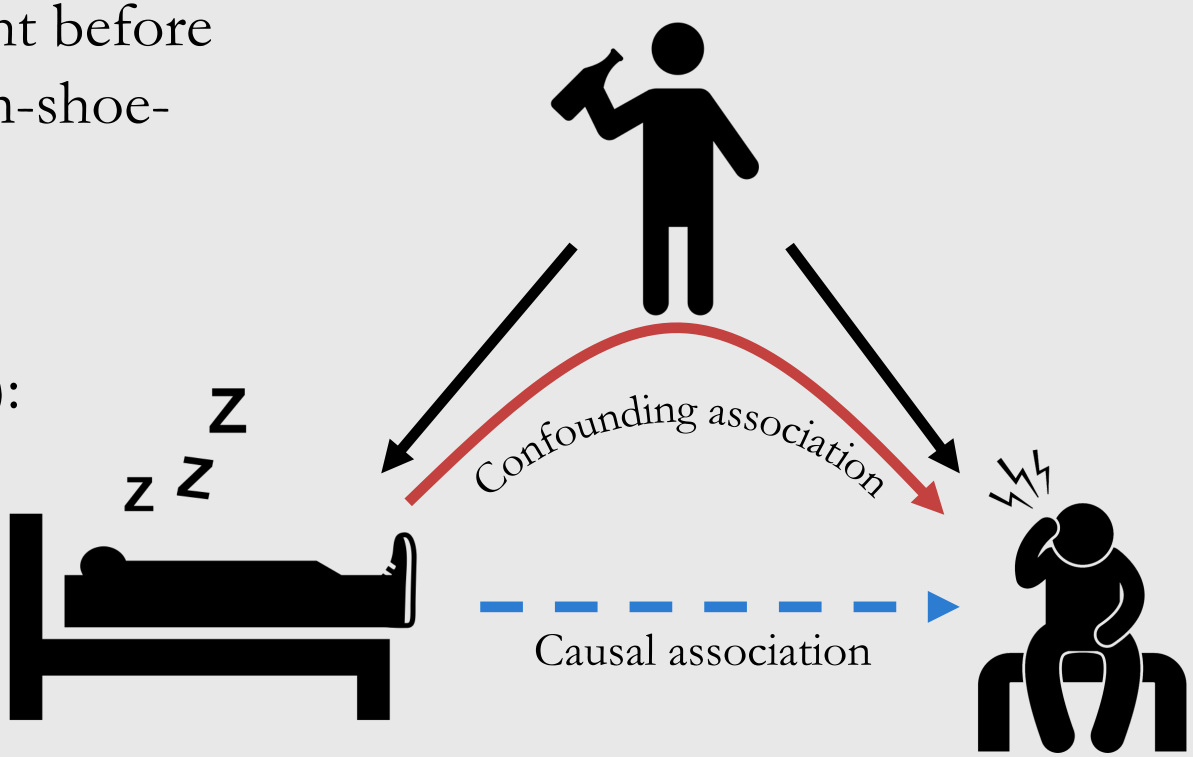
Correlation does not imply causation

Sleeping with shoes on is strongly correlated with waking up with a headache

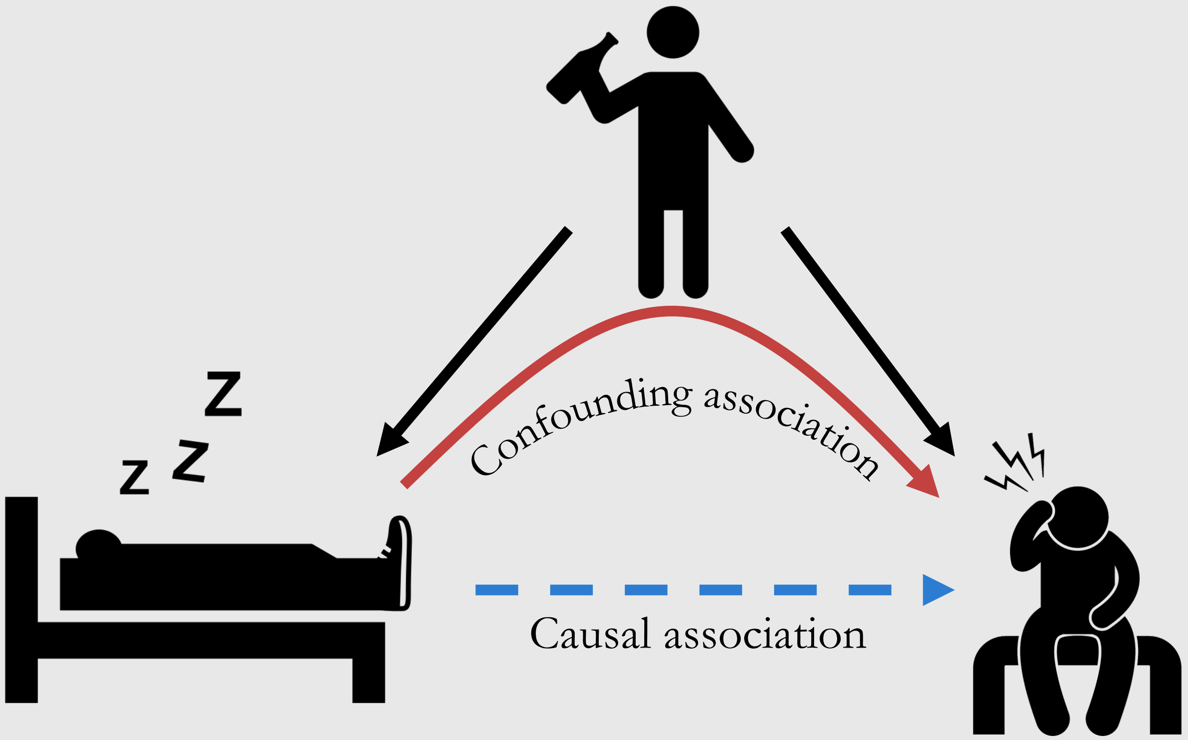
Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding

Total association (e.g. correlation): mixture of causal and confounding association

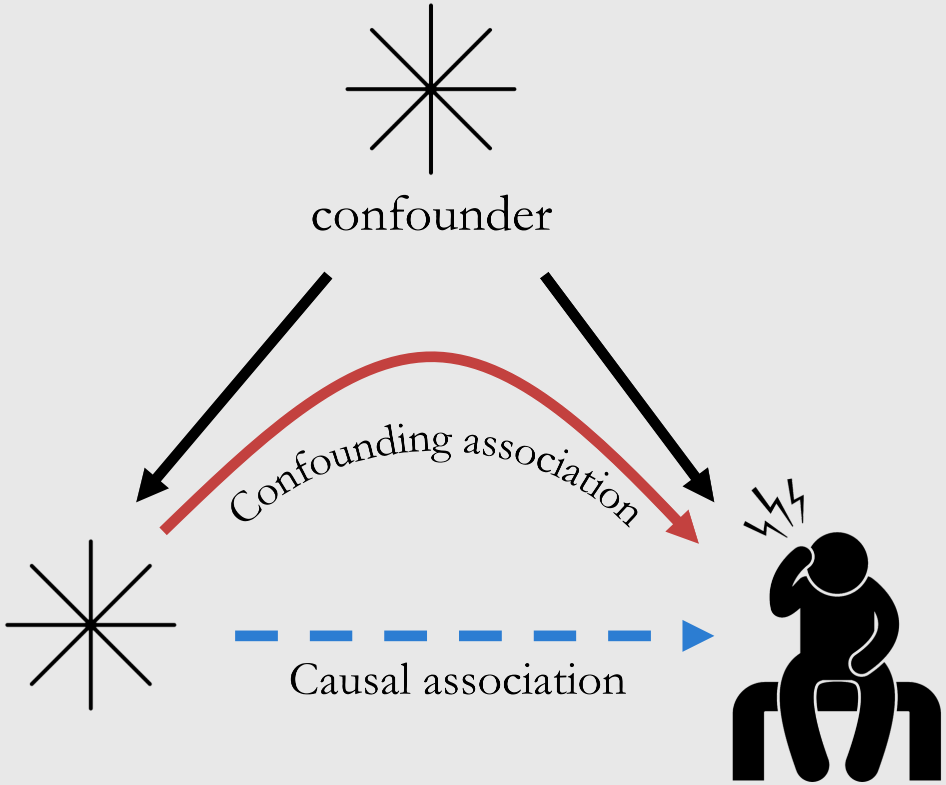


"Correlation = Causation" is a cognitive bias[1]



[1]The illusion of causality: A cognitive bias underlying pseudoscience (Blanco & Matute, 2018)

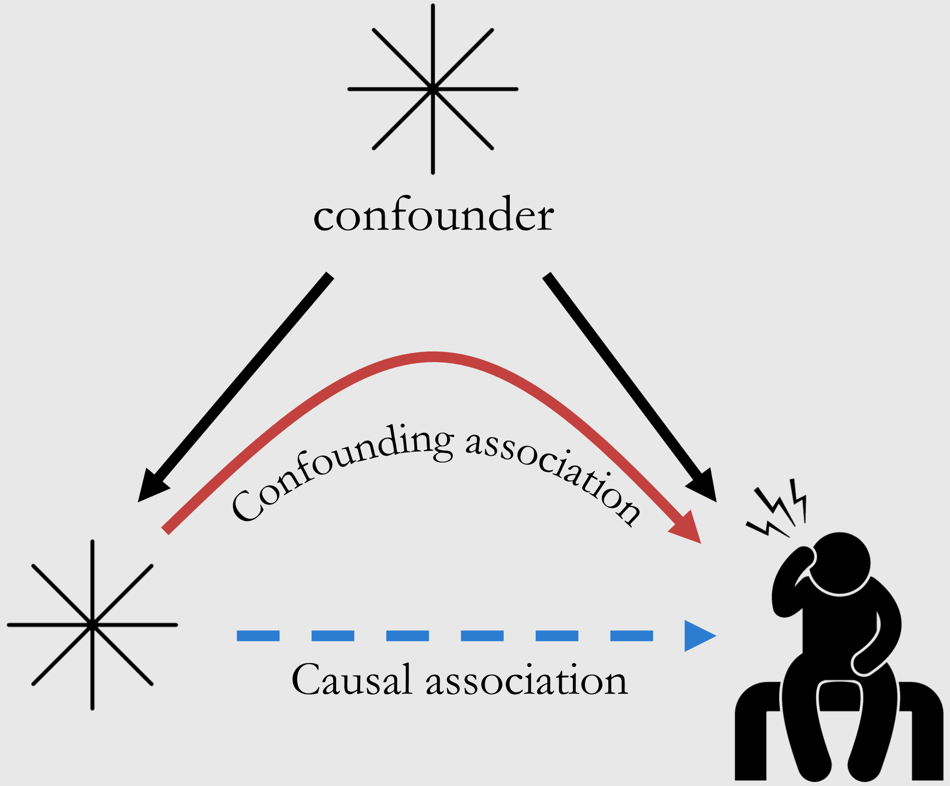
"Correlation = Causation" is a cognitive bias[1]



[1] The illusion of causality: A cognitive bias underlying pseudoscience (Blanco & Matute, 2018)

"Correlation = Causation" is a cognitive bias[1]

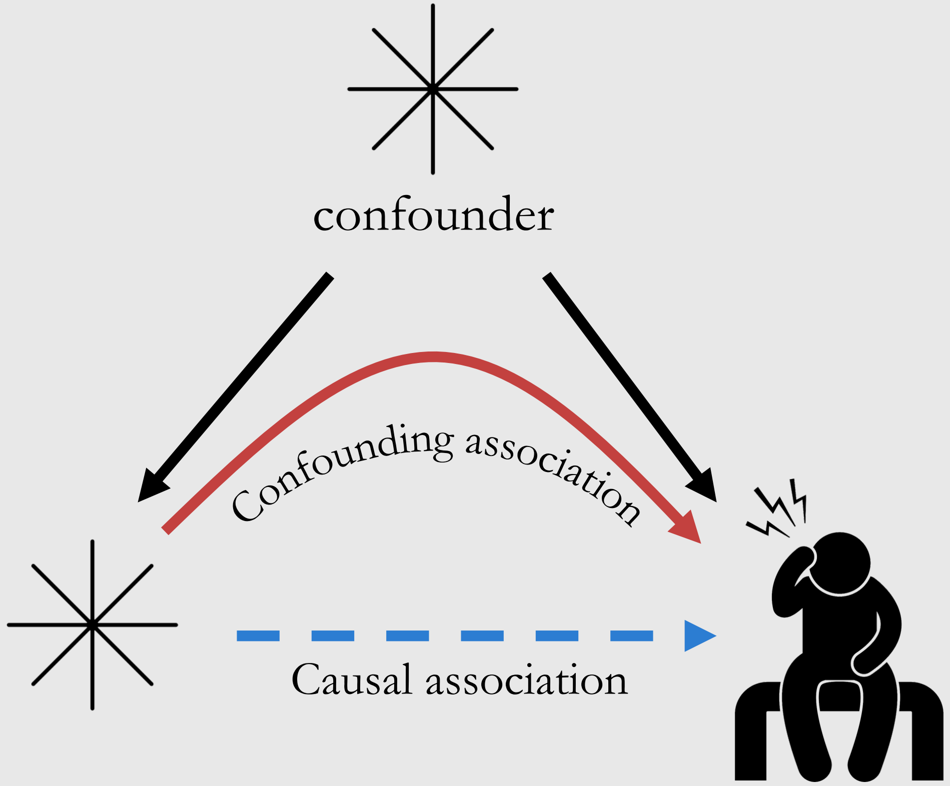
Availability heuristic (another cognitive bias) gives us ✳



[1] The illusion of causality: A cognitive bias underlying pseudoscience (Blanco & Matute, 2018)

"Correlation = Causation" is a cognitive bias[1]

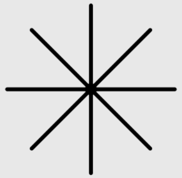
Availability heuristic (another cognitive bias) gives us ✳



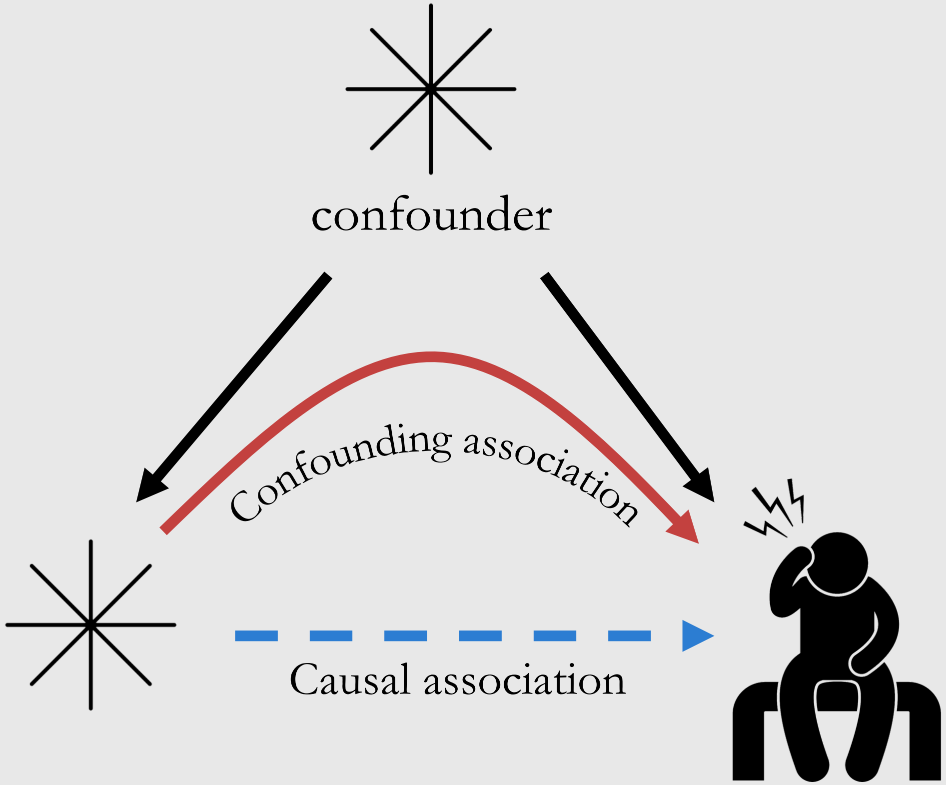
[1] The illusion of causality: A cognitive bias underlying pseudoscience (Blanco & Matute, 2018)

"Correlation = Causation" is a cognitive bias[1]

Availability heuristic (another cognitive bias) gives us ✳
Motivated reasoning (another cognitive bias)



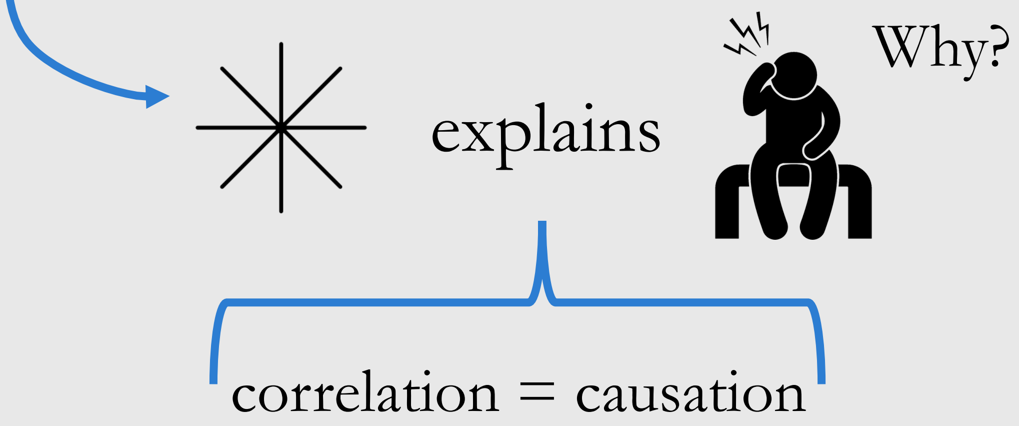
Why?

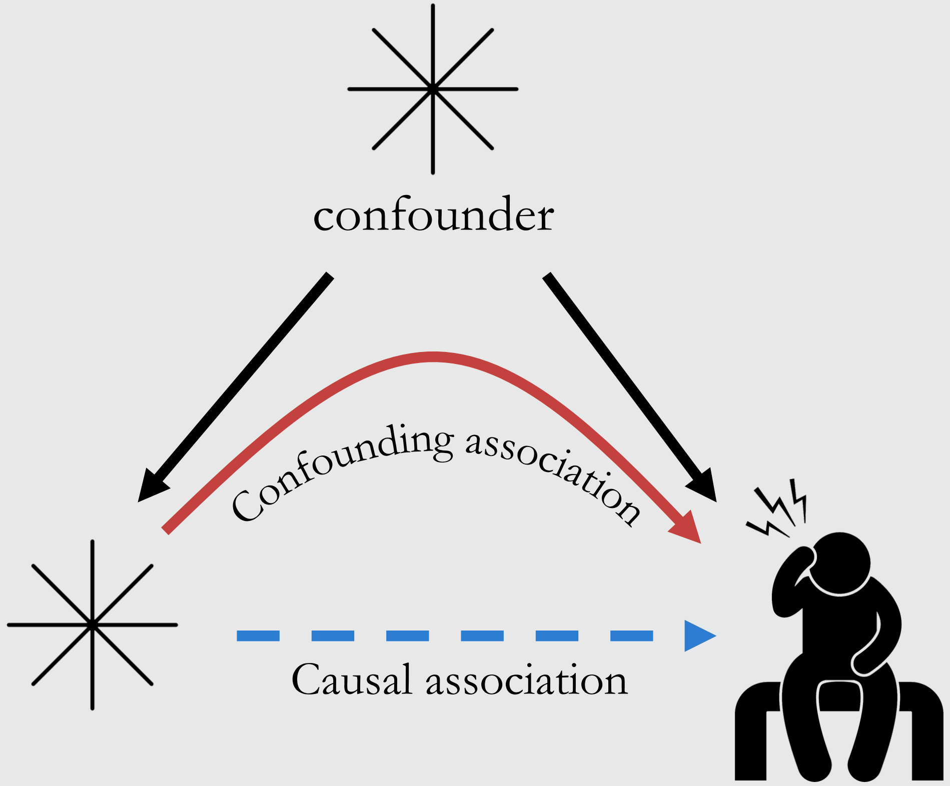


[1] The illusion of causality: A cognitive bias underlying pseudoscience (Blanco & Matute, 2018)

“Correlation = Causation” is a cognitive bias[1]

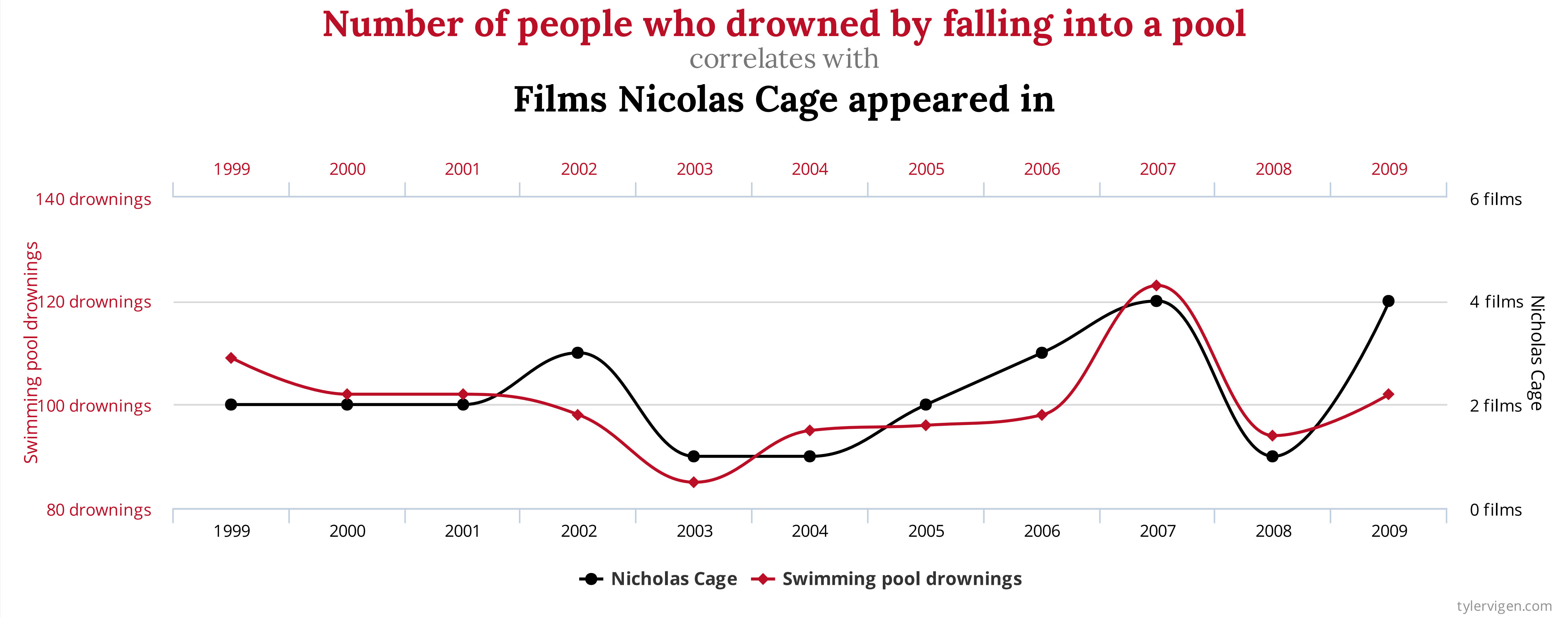
Availability heuristic (another cognitive bias) gives us ✳
Motivated reasoning (another cognitive bias)



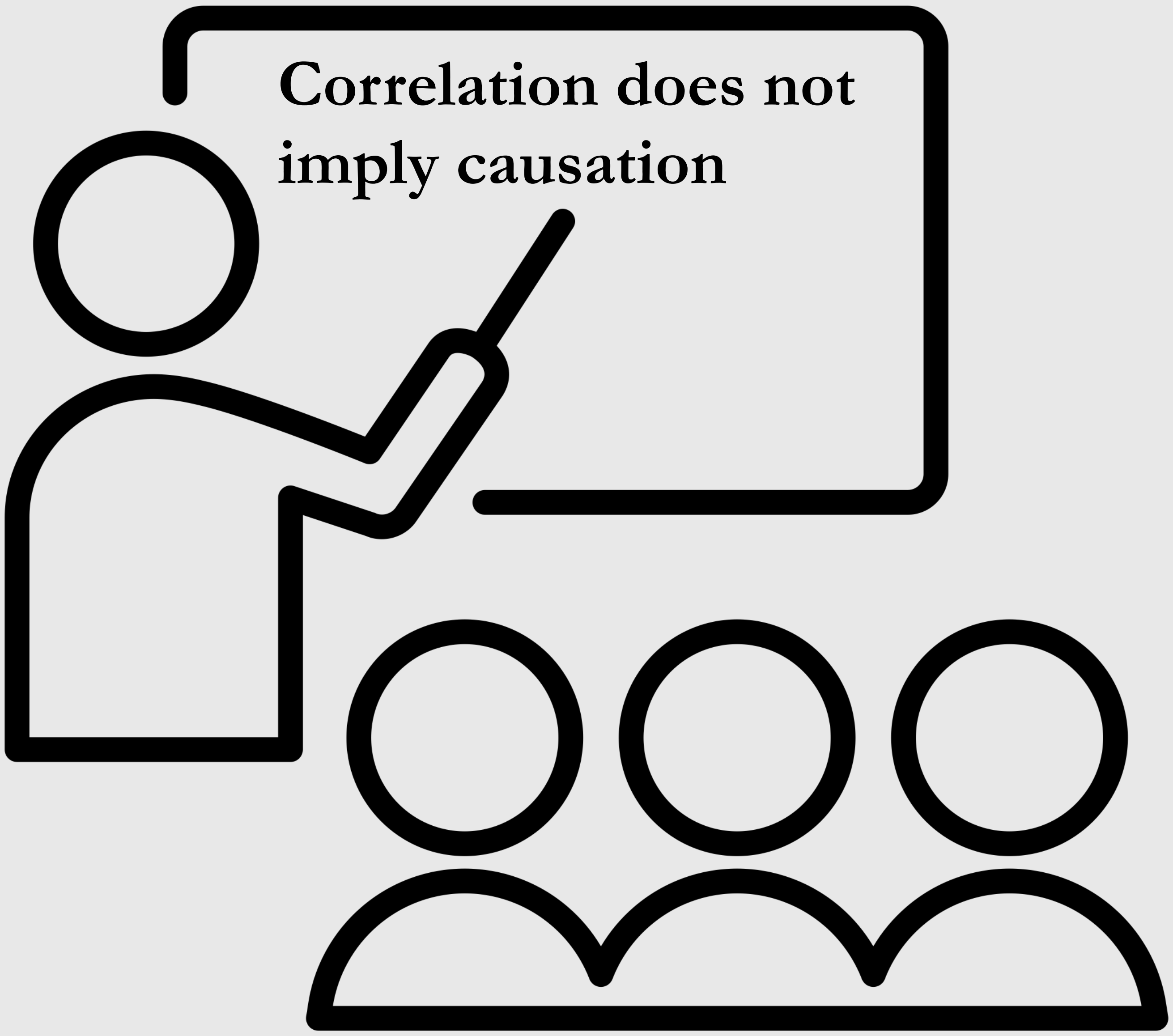


[1]The illusion of causality: A cognitive bias underlying pseudoscience (Blanco & Matute, 2018)

Nicolas Cage drives people to drown themselves



https://www.tylervigen.com/spurious-correlations


Correlation does not imply causation

Then, what does imply causation?

Motivating example: Simpson's paradox

Correlation does not imply causation

Then, what does imply causation?

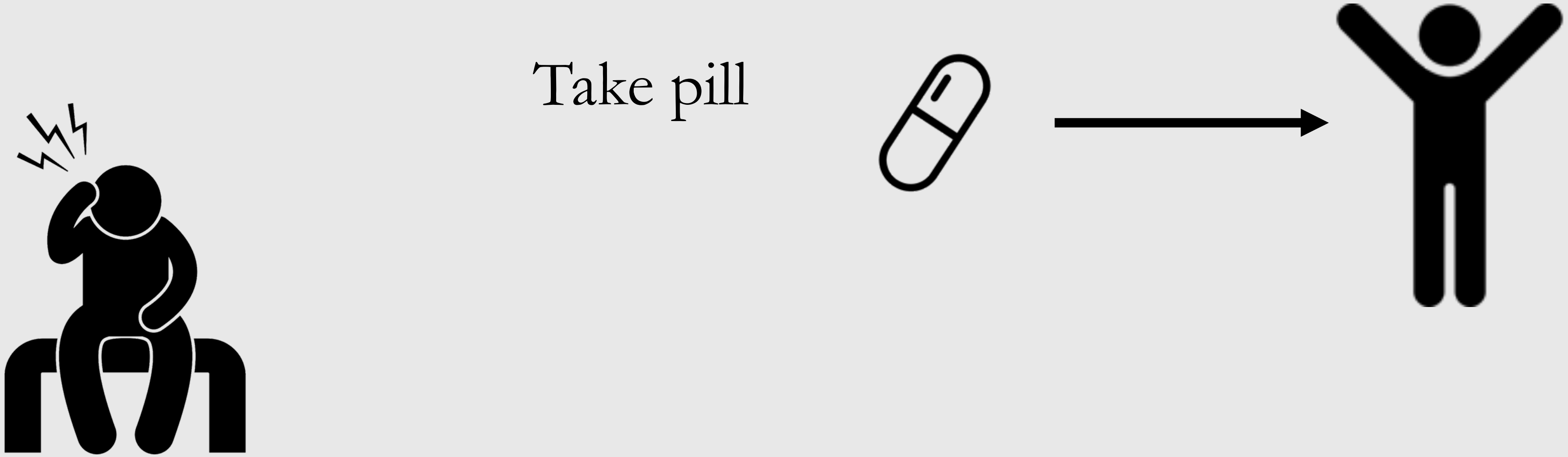
Causation in observational studies

Potential outcomes: intuition

Inferring the effect of treatment/policy on some outcome

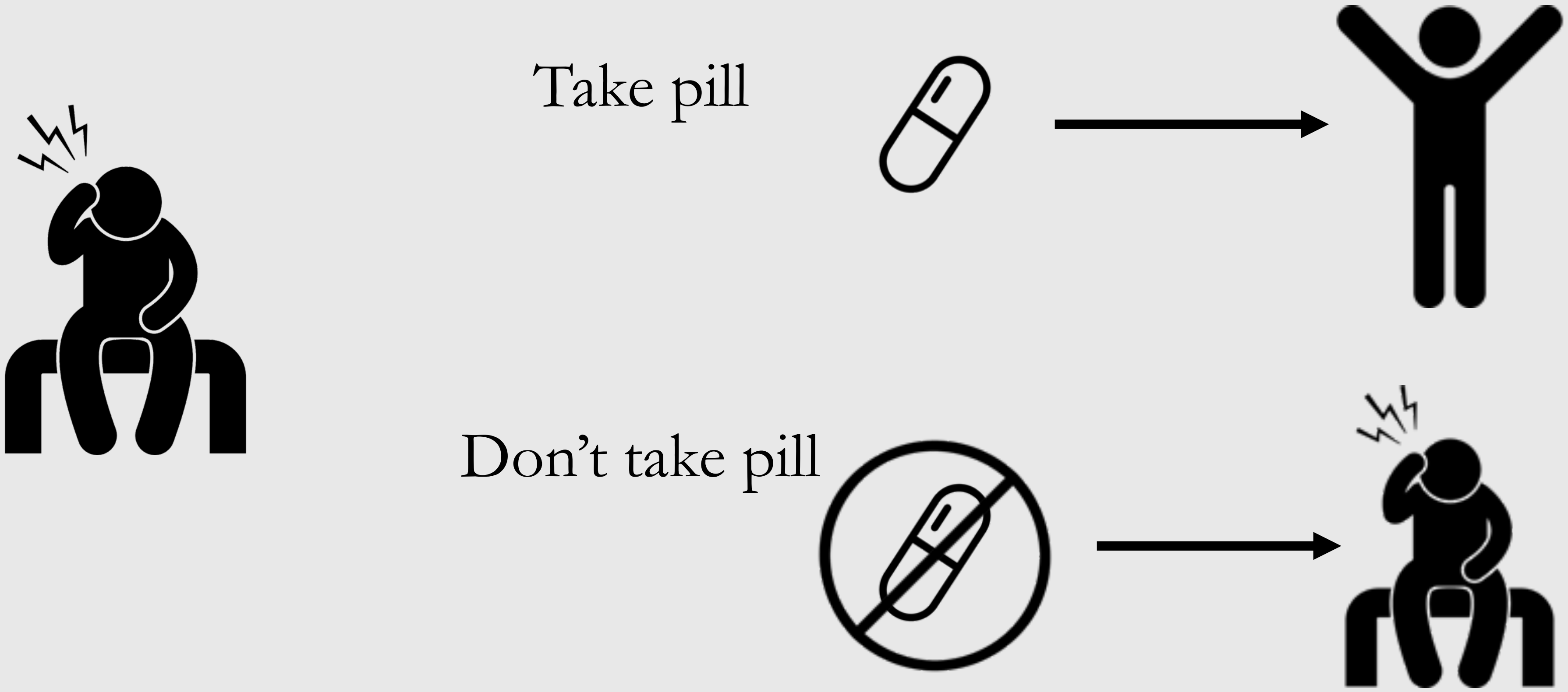
Potential outcomes: intuition

Inferring the effect of treatment/policy on some outcome



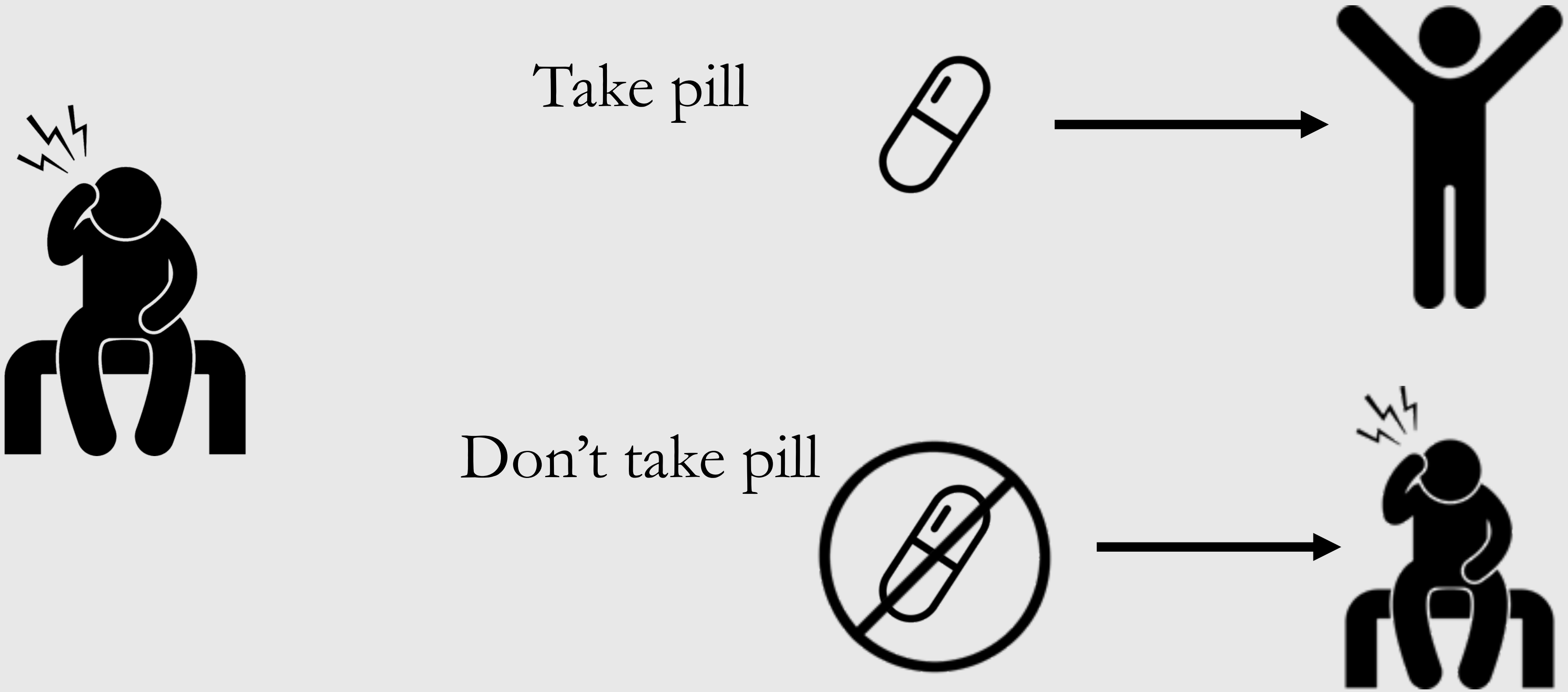
Potential outcomes: intuition

Inferring the effect of treatment/policy on some outcome



Potential outcomes: intuition

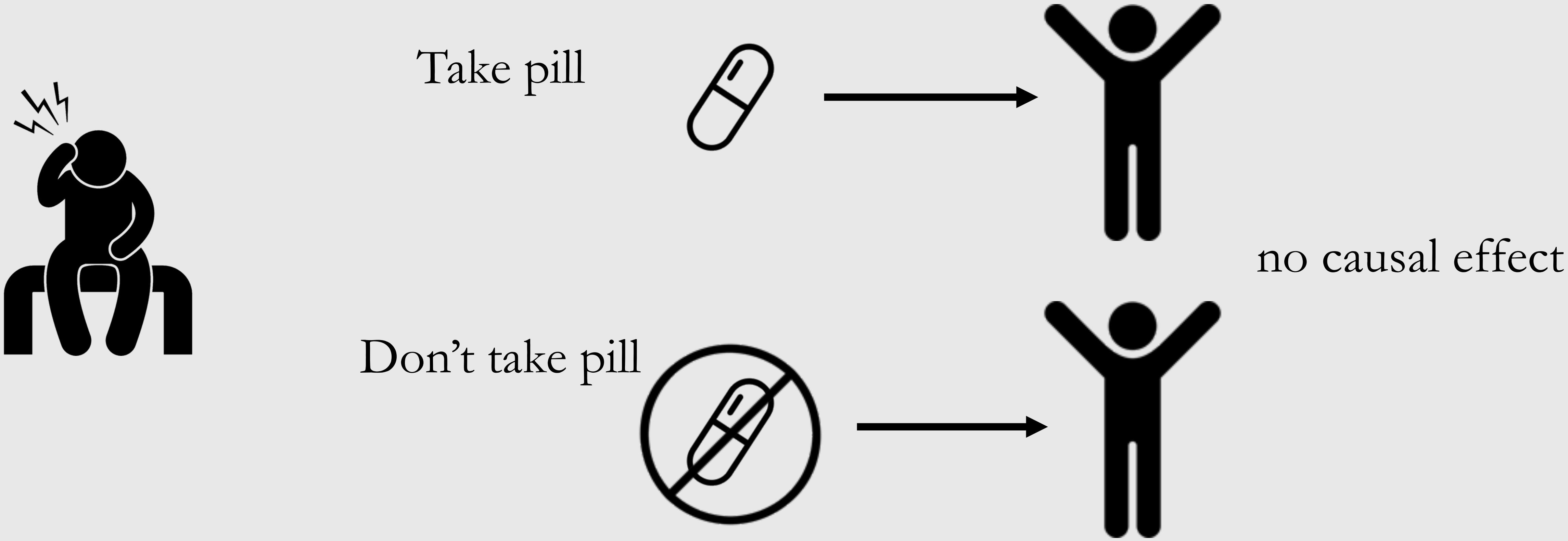
Inferring the effect of treatment/policy on some outcome



causal effect

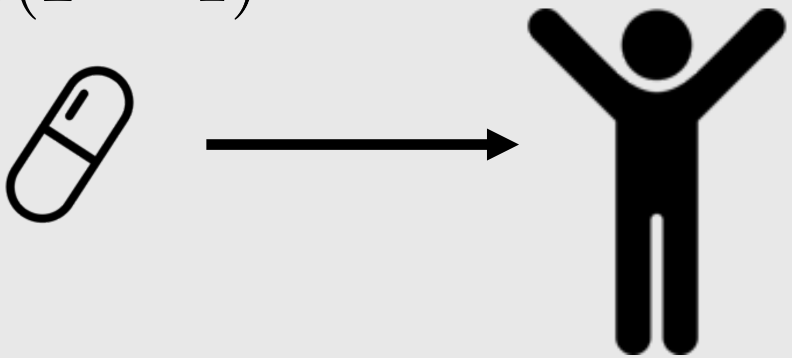
Potential outcomes: intuition

Inferring the effect of treatment/policy on some outcome

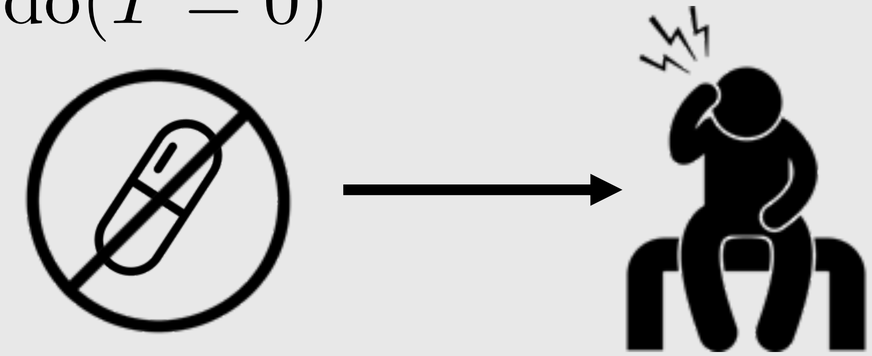


Potential outcomes: notation

$\text{do}(T = 1)$



$\text{do}(T = 0)$

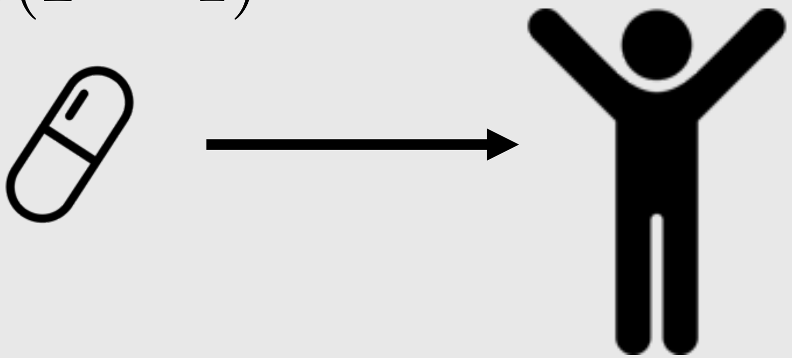


T : observed treatment
Y : observed outcome

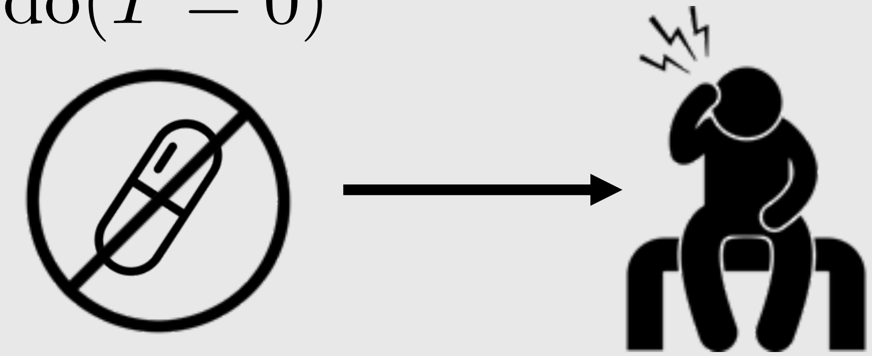
Potential outcomes: notation

$\mathrm{do}(T = 1)$

$Y_i|_{\mathrm{do}(T=1)}$



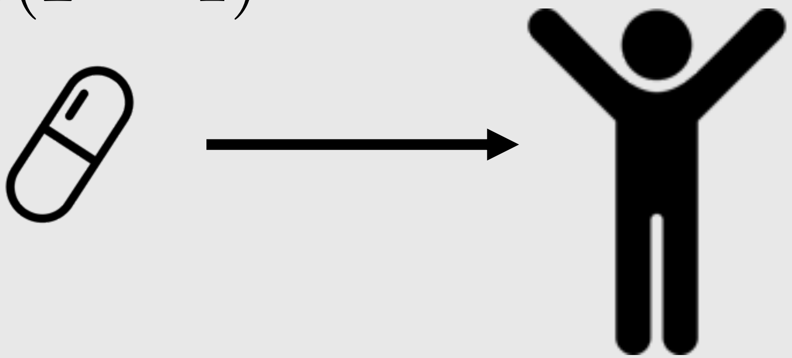
$\mathrm{do}(T = 0)$



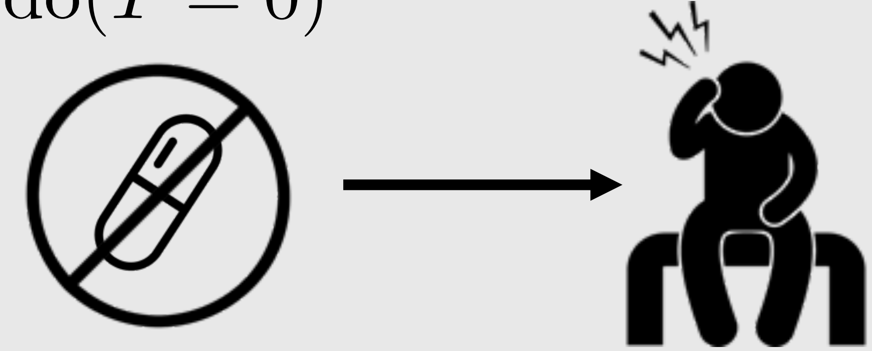
T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual

Potential outcomes: notation

$\text{do}(T = 1)$ $\qquad$ $Y_i|_{\text{do}(T=1)}$



$\text{do}(T = 0)$ $\qquad$ $Y_i|_{\text{do}(T=0)}$

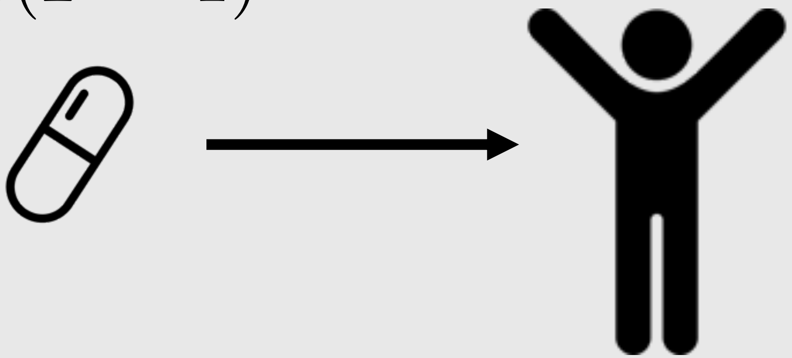


T : observed treatment
Y : observed outcome
i : used in subscript to denote a specific unit/individual

Potential outcomes: notation

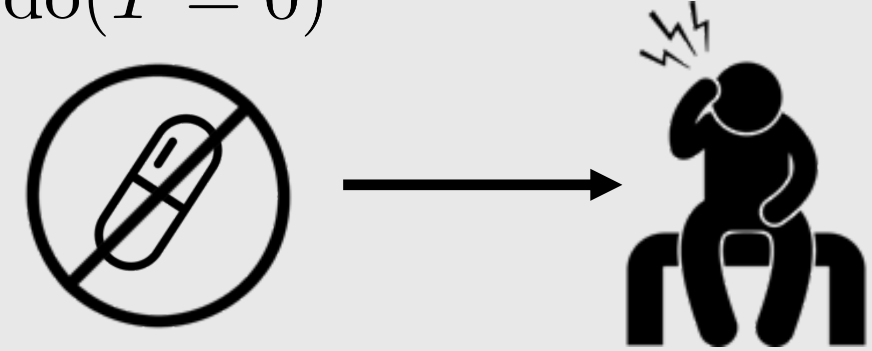
$\mathrm{do}(T = 1)$

$Y_i|_{\mathrm{do}(T=1)} \triangleq Y_i(1)$



$\mathrm{do}(T = 0)$

$Y_i|_{\mathrm{do}(T=0)}$

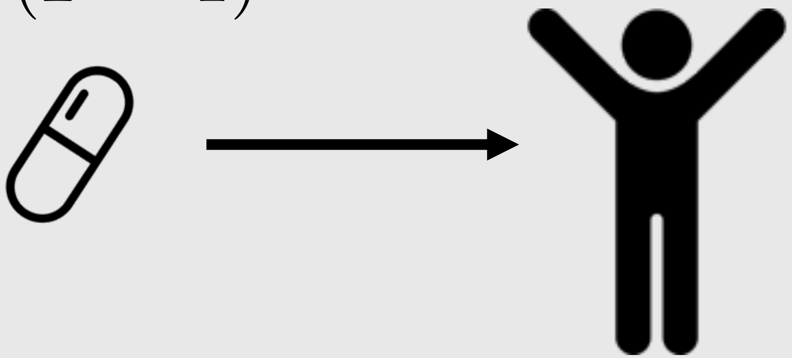


T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment

Potential outcomes: notation

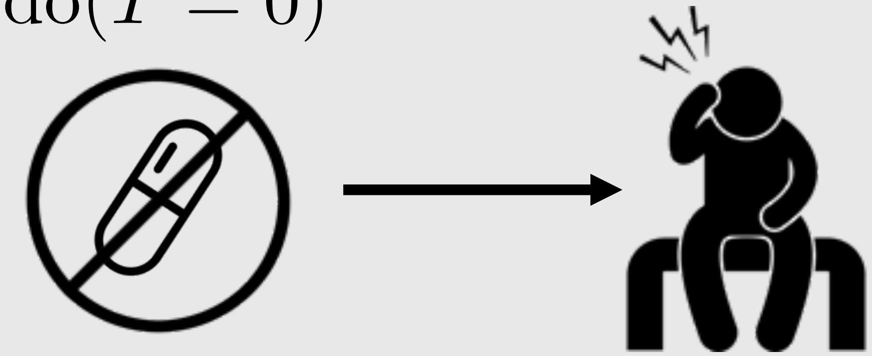
$\mathrm{do}(T = 1)$

$Y_i|_{\mathrm{do}(T=1)} \triangleq Y_i(1)$



$\mathrm{do}(T = 0)$

$Y_i|_{\mathrm{do}(T=0)} \triangleq Y_i(0)$

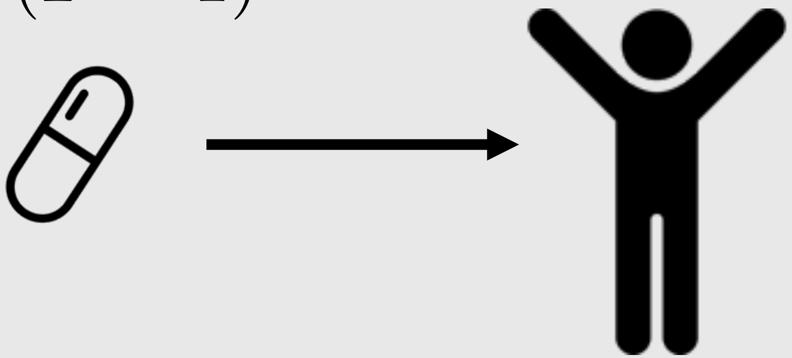


T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

Potential outcomes: notation

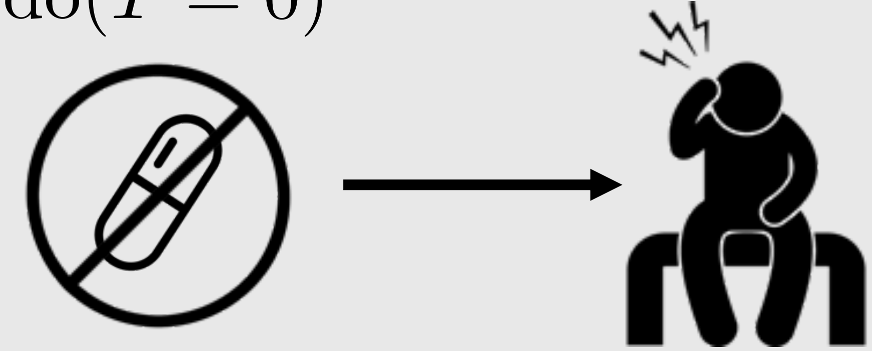
$\mathrm{do}(T = 1)$

$Y_i|_{\mathrm{do}(T=1)} \triangleq Y_i(1)$



$\mathrm{do}(T = 0)$

$Y_i|_{\mathrm{do}(T=0)} \triangleq Y_i(0)$



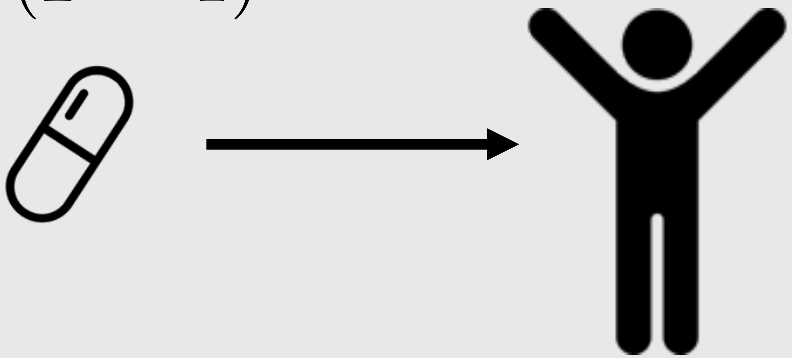
T : observed treatment
Y : observed outcome
i : used in subscript to denote a specific unit/individual
$Y_i(1)$: potential outcome under treatment
$Y_i(0)$: potential outcome under no treatment

Causal effect

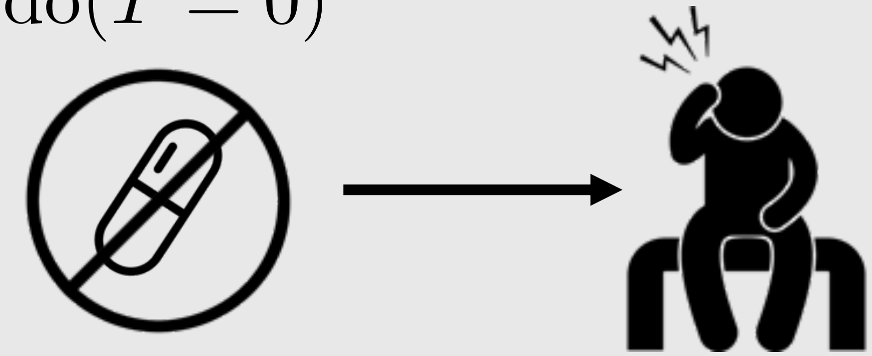
$Y_i(1) - Y_i(0)$

Fundamental problem of causal inference

$\text{do}(T = 1)$ $\quad Y_i(1) = 1$



$\text{do}(T = 0)$ $\quad Y_i(0) = 0$



T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

Causal effect

$$Y_i(1) - Y_i(0) = 1$$

Fundamental problem of causal inference

Counterfactual

$\mathrm{do}(T = 1)$

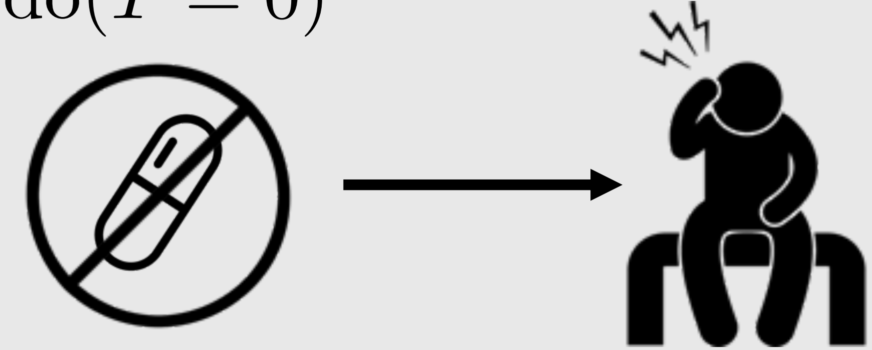
$Y_i(1) = 1$

T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

Factual

$\mathrm{do}(T = 0)$

$Y_i(0) = 0$



Causal effect

$$Y_i(1) - Y_i(0) = 1$$

Fundamental problem of causal inference

Factual

$\mathrm{do}(T = 1)$

$Y_i(1) = 1$

T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

Counterfactual

$\mathrm{do}(T = 0)$

$Y_i(0) = 0$

Causal effect

$Y_i(1) - Y_i(0) = 1$