McGill | Department of Epidemiology, Biostatistics and Occupational Health

Introduction to Target Trial Emulation and Test Negative Design Studies

Edgar Ortiz Brizuela

https://ortizbrizuela.github.io/info/

McGill

Agenda

1. A **quick refresher** on the distinction **between causation and association** and why it matters.
2. Revisiting **why randomization leads to robust causal inferences**.
3. Moving **from marginal to conditionally randomized experiments**.
4. **Observational studies** as an **alternative to randomized trials**.
5. Estimating causal effects with observational data through the **emulation of target trials**.
6. An **introduction to the test-negative study design**.
7. Potential sources of bias in **test-negative design** studies.

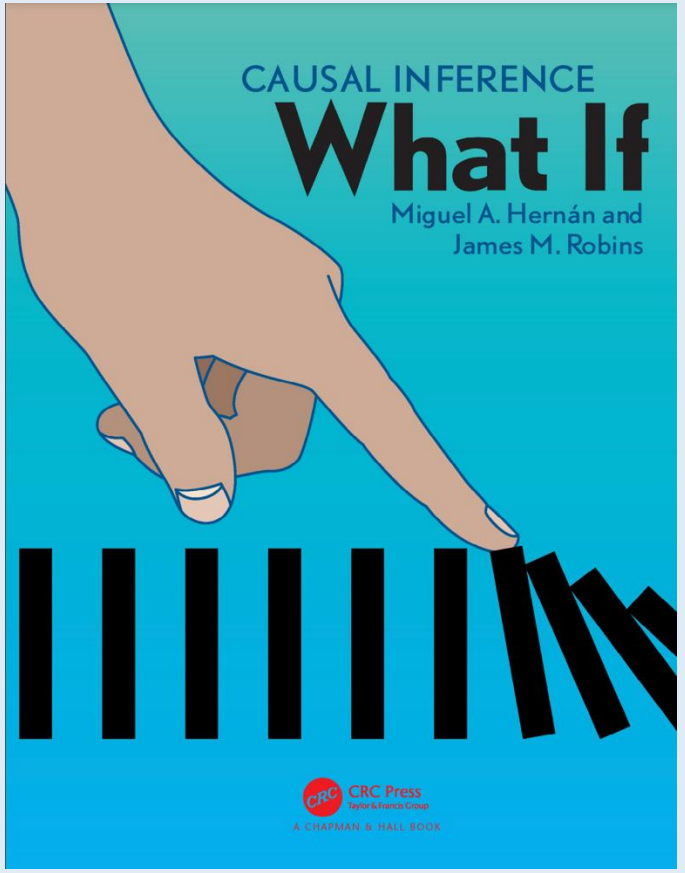
Disclaimer

Slides are based on two primary sources:

1. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

   Available at:

   https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/



Disclaimer

2. **CAUSALab Summer Courses on Causal Inference (2022 and 2023)**.

   A 5-day course in which students learn the principles of Target Trial Emulation and receive hands-on training to implement them in increasingly complex environments.

   Learn more about the 2024 courses in: https://causalab.sph.harvard.edu/courses/

1. A **quick refresher** on the distinction between **causation and association** and why it matters.

Three core tasks in health research …

a) **Description**:

b) **Prediction**:

c) **Causal inference**:

- Miguel A. Hernán, et al. (2019) CHANCE, 32:1, 42-49
- Rothman KJ, Lash TL, Haneuse S, VanderWeele TJ. Chapter 1: The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. Modern Epidemiology. 4th ed. Wolters Kluwer; 2021:3-25.

Three core tasks in health research …

a) **Description**: Using data to provide a **quantitative summary** of the characteristics of defined populations (e.g., GBD). Useful for **guiding resource allocation** and **hypothesis generation**.

b) **Prediction**:

c) **Causal inference**:

- Miguel A. Hernán, et al. (2019) CHANCE, 32:1, 42-49
- Rothman KJ, Lash TL, Haneuse S, VanderWeele TJ. Chapter 1: The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. Modern Epidemiology. 4th ed. Wolters Kluwer; 2021:3-25.

Three core tasks in health research …

a) **Description**: Using data to provide a **quantitative summary** of the characteristics of defined populations (e.g., GBD). Useful for **guiding resource allocation** and **hypothesis generation**.

b) **Prediction**: Using data to look for **associations** between different features of the population (mapping inputs and outputs). Useful for **informing some decisions** (e.g., patient monitoring).

c) **Causal inference**:

- Miguel A. Hernán, et al. (2019) CHANCE, 32:1, 42-49
- Rothman KJ, Lash TL, Haneuse S, VanderWeele TJ. Chapter 1: The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. Modern Epidemiology. 4th ed. Wolters Kluwer; 2021:3-25.

Causal inference

- It involves the **qualitative assessment** and/or **estimation of the effect of exposures**, including potential causes, on the occurrence of outcomes of interest.

- These exposures may be assessed at the **individual level** (e.g., occupations), at the **microscale** (e.g., microbiome), or at the **macroscale** (e.g., health policy).

- Miguel A. Hernán, et al. (2019) CHANCE, 32:1, 42-49
- Rothman KJ, Lash TL, Haneuse S, VanderWeele TJ. Chapter 1: The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. Modern Epidemiology. 4th ed. Wolters Kluwer; 2021:3-25.

Causal inference

- It involves the **qualitative assessment** and/or **estimation of the effect of exposures**, including potential causes, on the occurrence of outcomes of interest.

- These exposures may be assessed at the **individual level** (e.g., occupations), at the **microscale** (e.g., microbiome), or at the **macroscale** (e.g., health policy).

- They are **useful for guiding interventions**, such as deciding what treatment is appropriate for a patient with diabetes or what policy should be implemented to increase health literacy.

- Miguel A. Hernán, et al. (2019) CHANCE, 32:1, 42-49
- Rothman KJ, Lash TL, Haneuse S, VanderWeele TJ. Chapter 1: The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, eds. Modern Epidemiology. 4th ed. Wolters Kluwer; 2021:3-25.

Example

Title: **Risk Factors for** Mortality in Covid-19 Patients

Abstract: In this retrospective cohort study, we aimed to evaluate risk factors for mortality in Covid-19 patients. We used univariable logistic regression models and selected covariates for the multivariable model based on their statistical significance. Our results identified **hospitalization** (OR: 2.10, 95% CI: 1.6-2.8, p <0.001), **obesity** (OR: 1.50, 95% CI: 1.0-2.2, p = 0.03), and **age** (OR: 1.02, 95% CI: 1.01-1.04, p = 0.004) **as risk factors for Covid-19 mortality**. In contrast, vaccinated individuals had lower odds of death (OR: 0.6, 95% CI: 0.4-0.9, p = 0.002).

Multivariable analysis			
Variable	**OR**	**95% CI**	***P***
Hospitalization	2.5	1.8 - 3.4	0.001
Obesity	1.8	1.1 - 2.9	0.015
Diabetes	1.4	0.85 - 1.9	0.11
Hypertension	1.6	0.9 - 2.2	0.095
Age (years)	1.03	1.01 - 1.05	0.002
Sex (female)	0.85	0.7 - 1.03	0.08
Vaccination	0.5	0.35 - 0.7	0.0005

Example

- **Using predictions (associations) to infer causality can lead to wrong decisions about interventions**, such as advising against hospitalization for severe COVID-19 or recommending ivermectin for its treatment.

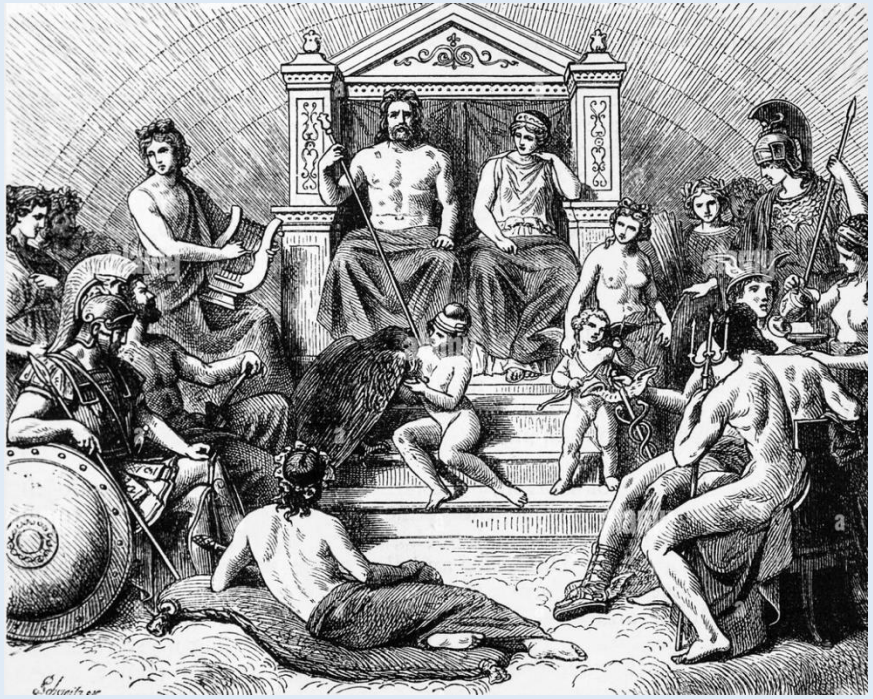
- **Causal inference is fundamental to making informed, accurate decisions about interventions**.

Multivariable analysis			
Variable	**OR**	**95% CI**	***P***
Hospitalization	2.5	1.8 - 3.4	0.001
Obesity	1.8	1.1 - 2.9	0.015
Diabetes	1.4	0.85 - 1.9	0.11
Hypertension	1.6	0.9 - 2.2	0.095
Age (years)	1.03	1.01 - 1.05	0.002
Sex (female)	0.85	0.7 - 1.03	0.08
Vaccination	0.5	0.35 - 0.7	0.0005

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

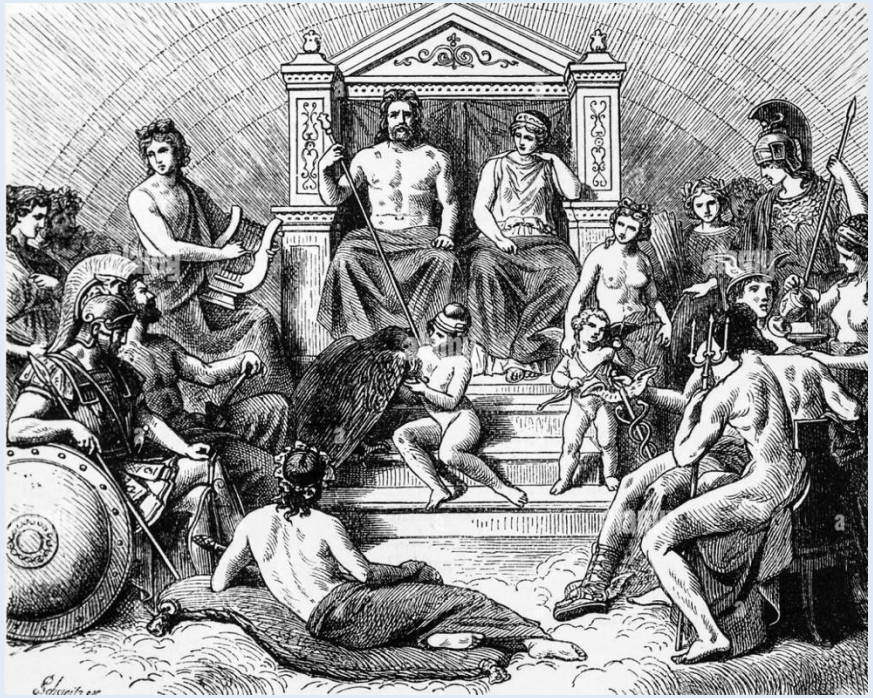


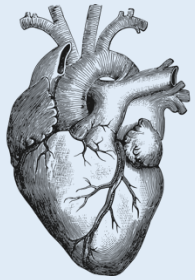
Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0





A

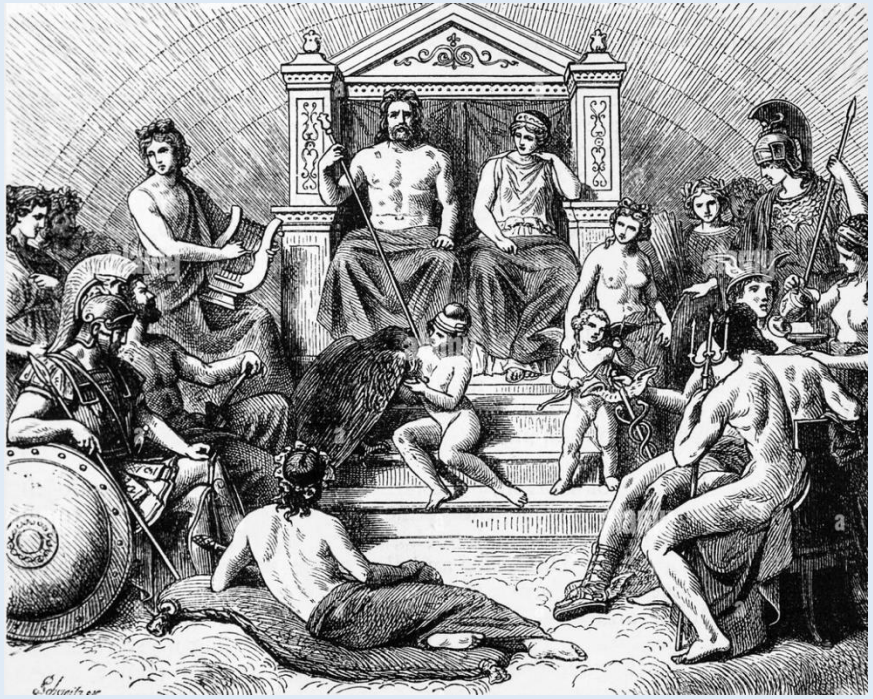
- **Intervention**
- **Exposure**
- **Policy**
- **Treatment**
- **…**

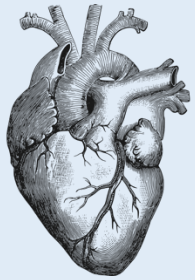
Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0





Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

The risk of death among **non-transplant recipients** is:

$$\mathbf{Pr}[\boldsymbol{Y = 1 | A = 0}] = \frac{\mathbf{3}}{\mathbf{7}} = \mathbf{0.43}$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

The risk of death among **non-transplant recipients** is:

$$\mathbf{Pr}[Y = 1 | A = 0] = \frac{3}{7} = 0.43$$

The risk of death among **transplant recipients** is:

$$\mathbf{Pr}[Y = 1 | A = 1] = \frac{7}{13} = 0.54$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

The risk of death among **non-transplant recipients** is:

$$\mathbf{Pr[Y = 1|\, A = 0] = \frac{3}{7} = 0.43}$$

The risk of death among **transplant recipients** is:

$$\mathbf{Pr[Y = 1|\, A = 1] = \frac{7}{13} = 0.54}$$

(i) $\Pr[Y = 1|A = 1] - \Pr[Y = 1|A = 0] = 0$

(ii) $\dfrac{\Pr[Y = 1|A = 1]}{\Pr[Y = 1|A = 0]} = 1$

(iii) $\dfrac{\Pr[Y = 1|A = 1] \,/\, \Pr[Y = 0|A = 1]}{\Pr[Y = 1|A = 0] \,/\, \Pr[Y = 0|A = 0]} = 1$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

The risk of death among **non-transplant recipients** is:

$$\mathbf{Pr}[\boldsymbol{Y} = \mathbf{1} | \boldsymbol{A} = \mathbf{0}] = \frac{\mathbf{3}}{\mathbf{7}} = \mathbf{0.43}$$

The risk of death among **transplant recipients** is:

$$\mathbf{Pr}[\boldsymbol{Y} = \mathbf{1} | \boldsymbol{A} = \mathbf{1}] = \frac{\mathbf{7}}{\mathbf{13}} = \mathbf{0.54}$$

Since both proportions were different, we can conclude that:

- Treatment A and outcome Y are **dependent**.
- Treatment A and outcome Y are **associated**.
- Treatment A **predicts** outcome Y.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

The risk of death among **non-transplant recipients** is:

$$\mathbf{Pr[Y = 1 | A = 0] = \frac{3}{7} = 0.43}$$

The risk of death among **transplant recipients** is:

$$\mathbf{Pr[Y = 1 | A = 1] = \frac{7}{13} = 0.54}$$

- [Again] **using predictions (associations) to infer causality can lead to wrong decisions about interventions.**
- Identifying patients with a poor prognosis is very different from identifying the best strategy to prevent or treat a disease...

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. Causation

Table 1.2

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

The risk of death among **non-transplant recipients** is:

$$\Pr[Y = 1 | A = 0] = \frac{3}{7} = 0.43$$

The risk of death among **transplant recipients** is:

$$\Pr[Y = 1 | A = 1] = \frac{7}{13} = 0.54$$

The reason for the mismatch between association and causation is simple:

- People typically receive interventions for specific reasons.
- For example, if a doctor initiates treatment A, it is because she believes it will improve the patient's prognosis…

So, what do we want to know?

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. **Causation**

Table 2.1

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1
Artemis	0	1	1	?
Apollo	0	1	1	?
Leto	0	0	0	?
Ares	1	1	?	1
Athena	1	1	?	1
Hephaestus	1	1	?	1
Aphrodite	1	1	?	1
Polyphemus	1	1	?	1
Persephone	1	1	?	1
Hermes	1	0	?	0
Hebe	1	0	?	0
Dionysus	1	0	?	0

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Association vs. **Causation**

Table 2.1

	A	Y	Y^0	Y^1	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	0	?	0	1
Kronos	0	1	1	?	1	0
Demeter	0	0	0	?	0	0
Hades	0	0	0	?	0	0
Hestia	1	0	?	0	0	0
Poseidon	1	0	?	0	1	0
Hera	1	0	?	0	0	0
Zeus	1	1	?	1	0	1
Artemis	0	1	1	?	1	1
Apollo	0	1	1	?	1	0
Leto	0	0	0	?	0	1
Ares	1	1	?	1	1	1
Athena	1	1	?	1	1	1
Hephaestus	1	1	?	1	0	1
Aphrodite	1	1	?	1	0	1
Polyphemus	1	1	?	1	0	1
Persephone	1	1	?	1	1	1
Hermes	1	0	?	0	1	0
Hebe	1	0	?	0	1	0
Dionysus	1	0	?	0	1	0

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Polyphemus	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Polyphemus	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Half would have died if none had received a transplant:

$$\Pr\left[Y^{a=0} = 1\right] = \frac{10}{20} = 0.5$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Polyphemus	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Half would have died if none had received a transplant:

$$\Pr\left[Y^{a=0} = 1\right] = \frac{10}{20} = 0.5$$

Similarly, half would have died if they had received a transplant:

$$\Pr\left[Y^{a=1} = 1\right] = \frac{10}{20} = 0.5$$

McGill

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Polyphemus	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

Half would have died if none had received a transplant:

$$\Pr[Y^{a=0} = 1] = \frac{10}{20} = 0.5$$

Similarly, half would have died if they had received a transplant:

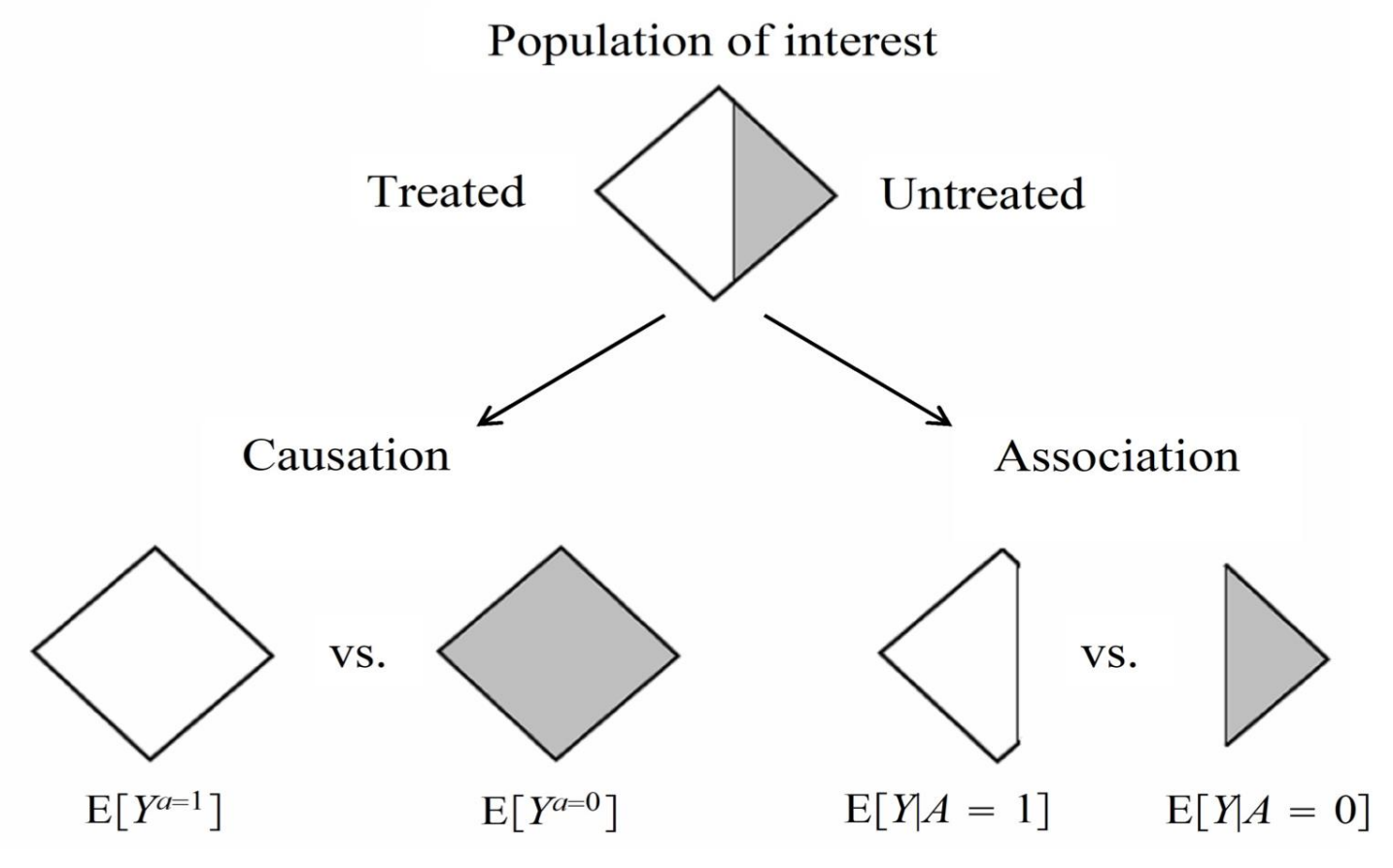
$$\Pr[Y^{a=1} = 1] = \frac{10}{20} = 0.5$$

(i) $\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1]$

(ii) $\dfrac{\Pr[Y^{a=1} = 1]}{\Pr[Y^{a=0} = 1]}$

(iii) $\dfrac{\Pr[Y^{a=1} = 1] / \Pr[Y^{a=1} = 0]}{\Pr[Y^{a=0} = 1] / \Pr[Y^{a=0} = 0]}$

Causation vs. Association (Summary)

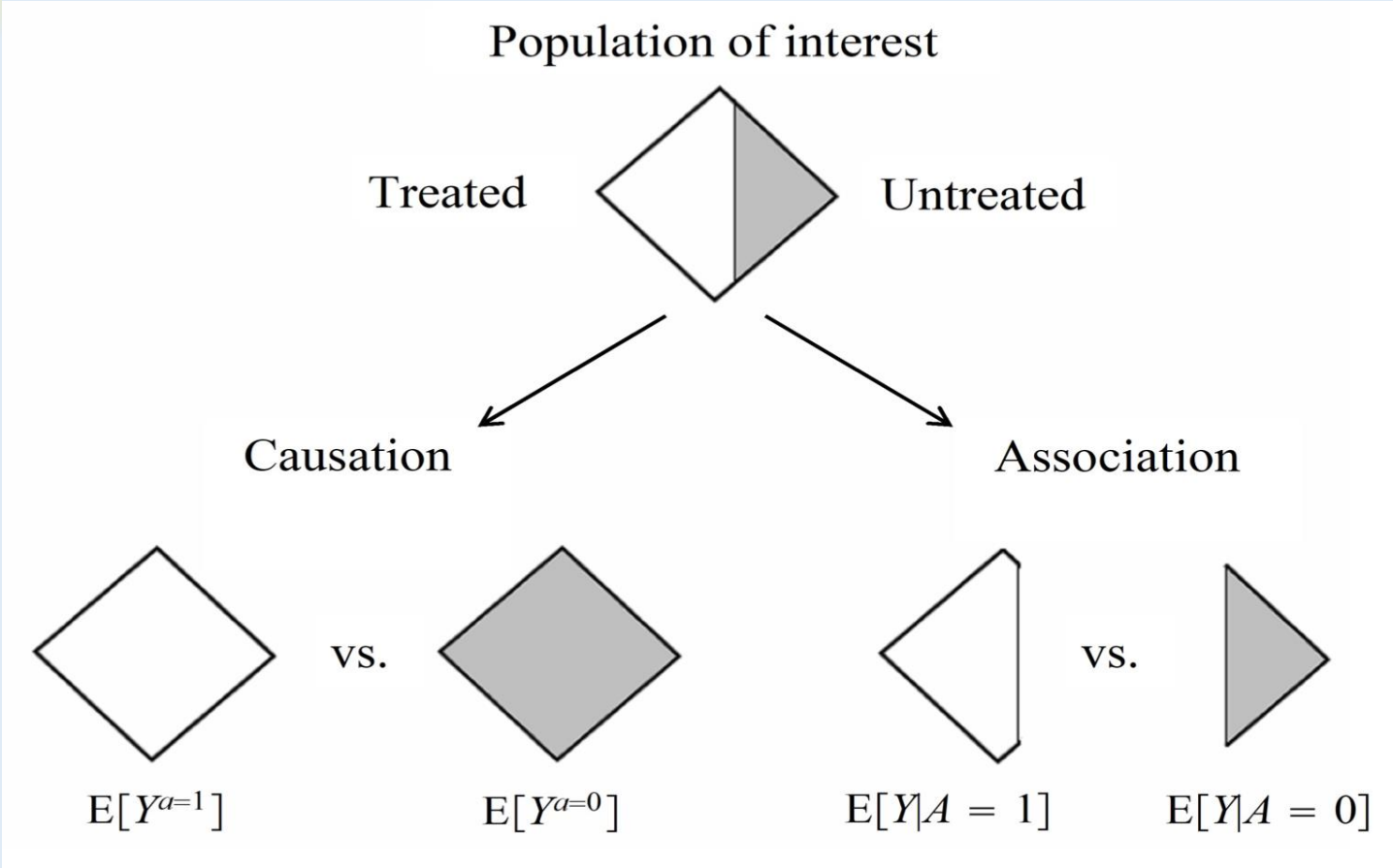


Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Causation vs. Association (Summary)

"**Inferences about causation** are concerned with **what if questions** in counterfactual worlds, such as:

- … what would be the risk if everybody had been treated?
- …"what would be the risk if everybody had been untreated?"

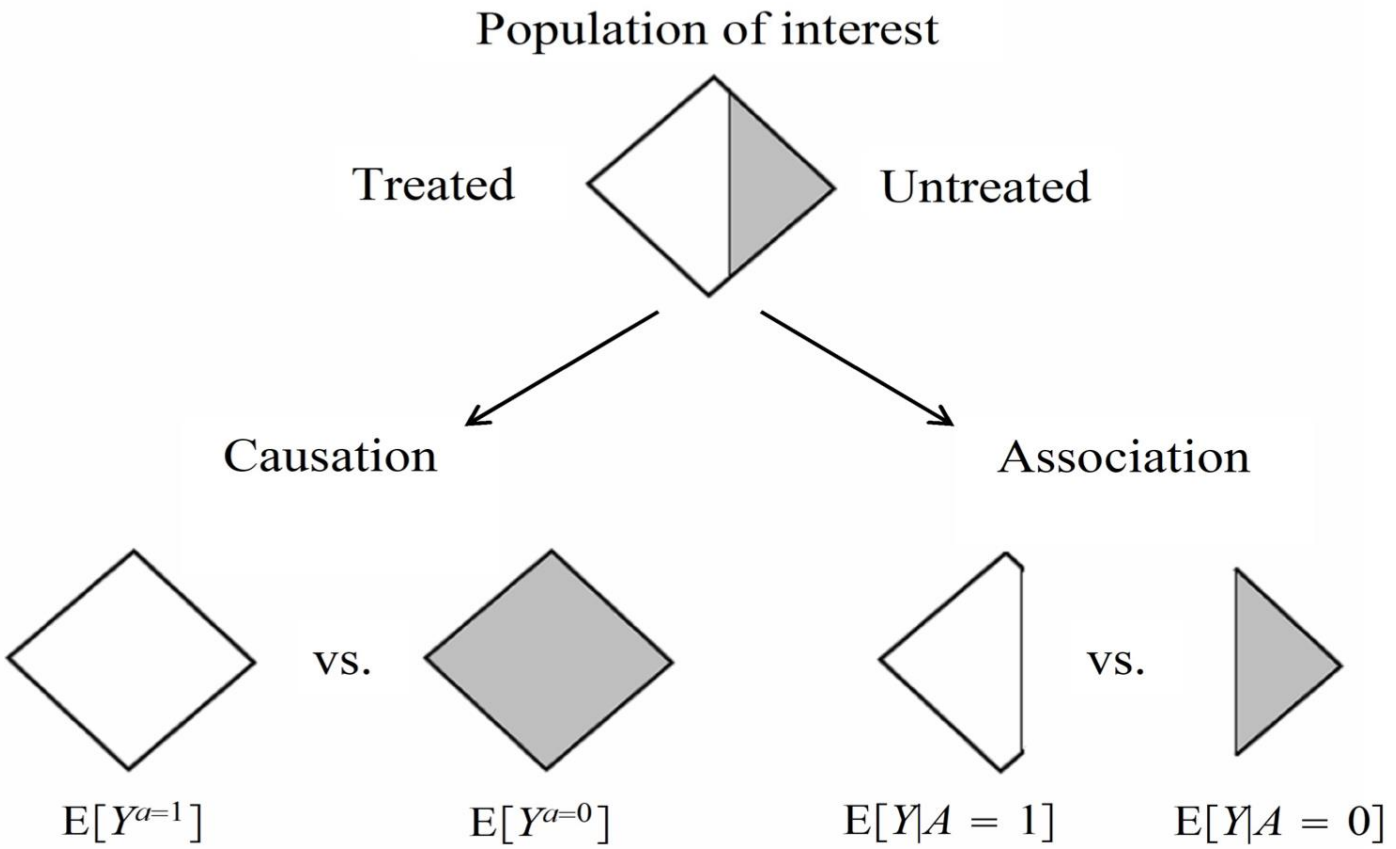


Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Causation vs. **Association** (Summary)

"**Inferences about causation** are concerned with **what if questions** in counterfactual worlds, such as:

- … what would be the risk if everybody had been treated?
- …"what would be the risk if everybody had been untreated?"



"**Inferences about association** are concerned with **questions in the actual world**, such as:

- … what is the risk in the treated?
- …"what is the risk in the untreated?"

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

2. Revisiting **why randomization leads to robust causal inferences**.

The fundamental problem of causal inference

Table 2.1

	A	Y	Y^0	Y^1	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	0	?	0	1
Kronos	0	1	1	?	1	0
Demeter	0	0	0	?	0	0
Hades	0	0	0	?	0	0
Hestia	1	0	?	0	0	0
Poseidon	1	0	?	0	1	0
Hera	1	0	?	0	0	0
Zeus	1	1	?	1	0	1
Artemis	0	1	1	?	1	1
Apollo	0	1	1	?	1	0
Leto	0	0	0	?	0	1
Ares	1	1	?	1	1	1
Athena	1	1	?	1	1	1
Hephaestus	1	1	?	1	0	1
Aphrodite	1	1	?	1	0	1
Polyphemus	1	1	?	1	0	1
Persephone	1	1	?	1	1	1
Hermes	1	0	?	0	1	0
Hebe	1	0	?	0	1	0
Dionysus	1	0	?	0	1	0

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The fundamental problem of causal inference

Table 2.1

	A	Y	Y^0	Y^1	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	0	?	0	1
Kronos	0	1	1	?	1	0
Demeter	0	0	0	?	0	0
Hades	0	0	0	?	0	0
Hestia	1	0	?	0	0	0
Poseidon	1	0	?	0	1	0
Hera	1	0	?	0	0	0
Zeus	1	1	?	1	0	1
Artemis	0	1	1	?	1	1
Apollo	0	1	1	?	1	0
Leto	0	0	0	?	0	1
Ares	1	1	?	1	1	1
Athena	1	1	?	1	1	1
Hephaestus	1	1	?	1	0	1
Aphrodite	1	1	?	1	0	1
Polyphemus	1	1	?	1	0	1
Persephone	1	1	?	1	1	1
Hermes	1	0	?	0	1	0
Hebe	1	0	?	0	1	0
Dionysus	1	0	?	0	1	0

- Because only one counterfactual is available for individuals (the one corresponding to the actual level of exposure), it is often said that **the fundamental problem of causal inference is missing data**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The fundamental problem of causal inference

Table 2.1

	A	Y	Y^0	Y^1	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	0	?	0	1
Kronos	0	1	1	?	1	0
Demeter	0	0	0	?	0	0
Hades	0	0	0	?	0	0
Hestia	1	0	?	0	0	0
Poseidon	1	0	?	0	1	0
Hera	1	0	?	0	0	0
Zeus	1	1	?	1	0	1
Artemis	0	1	1	?	1	1
Apollo	0	1	1	?	1	0
Leto	0	0	0	?	0	1
Ares	1	1	?	1	1	1
Athena	1	1	?	1	1	1
Hephaestus	1	1	?	1	0	1
Aphrodite	1	1	?	1	0	1
Polyphemus	1	1	?	1	0	1
Persephone	1	1	?	1	1	1
Hermes	1	0	?	0	1	0
Hebe	1	0	?	0	1	0
Dionysus	1	0	?	0	1	0

- Because only one counterfactual is available for individuals (the one corresponding to the actual level of exposure), it is often said that **the fundamental problem of causal inference is missing data**.

- **What would happen if counterfactuals were missing at random?**

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The fundamental problem of causal inference

Table 2.1

	A	Y	Y^0	Y^1	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	0	?	0	1
Kronos	0	1	1	?	1	0
Demeter	0	0	0	?	0	0
Hades	0	0	0	?	0	0
Hestia	1	0	?	0	0	0
Poseidon	1	0	?	0	1	0
Hera	1	0	?	0	0	0
Zeus	1	1	?	1	0	1
Artemis	0	1	1	?	1	1
Apollo	0	1	1	?	1	0
Leto	0	0	0	?	0	1
Ares	1	1	?	1	1	1
Athena	1	1	?	1	1	1
Hephaestus	1	1	?	1	0	1
Aphrodite	1	1	?	1	0	1
Polyphemus	1	1	?	1	0	1
Persephone	1	1	?	1	1	1
Hermes	1	0	?	0	1	0
Hebe	1	0	?	0	1	0
Dionysus	1	0	?	0	1	0

- Because only one counterfactual is available for individuals (the one corresponding to the actual level of exposure), it is often said that **the fundamental problem of causal inference is missing data**.
- **What would happen if counterfactuals were missing at random?**
- This is possible in an **ideal randomized experiment** (**double-blind**, **no loss to follow-up**, and **perfect adherence to the therapeutic strategy)** where **all population characteristics are expected to be evenly distributed across groups (including counterfactuals).**

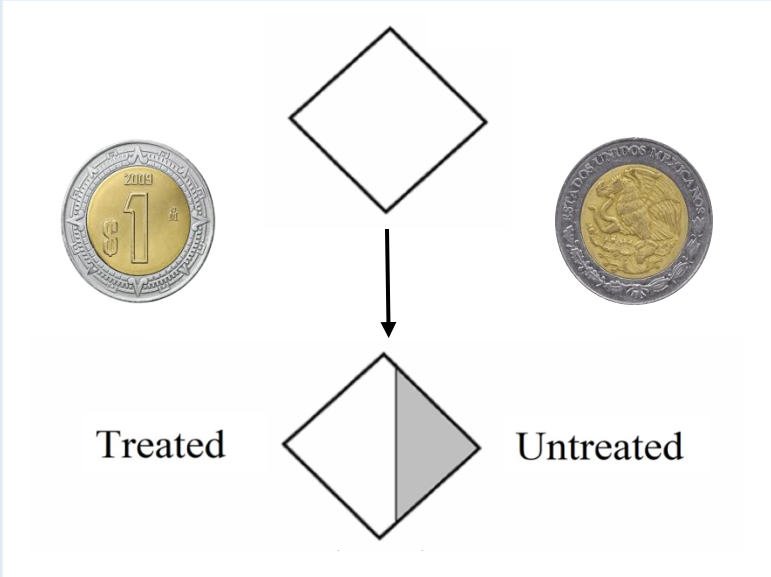
Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference

- Suppose **we randomize an almost infinite population** by **flipping a coin**.

- If the coin lands on **heads ($1)**, individuals are assigned to the **intervention** (***A* = 1** [white]); otherwise (eagle), they receive no intervention.

- We then **observe individuals for five days** and **calculate the risk of death**.



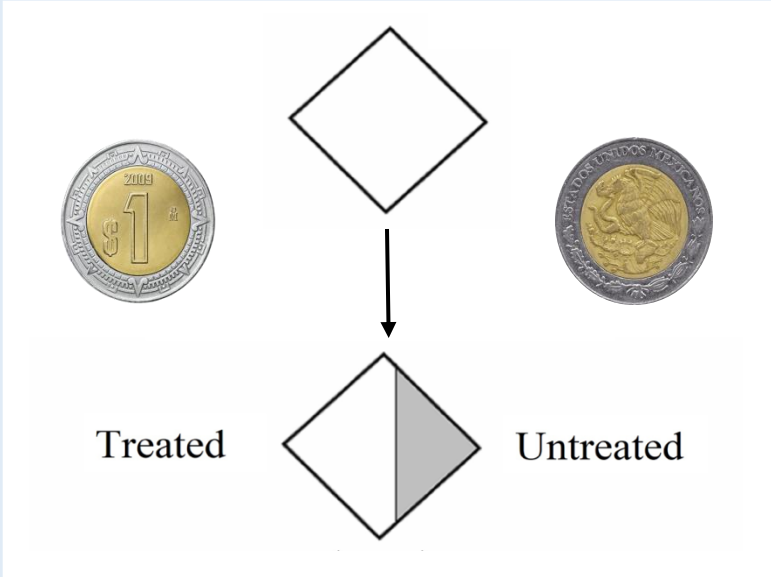
Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference

- Suppose **we randomize an almost infinite population** by **flipping a coin**.

- If the coin lands on **heads ($1)**, individuals are assigned to the **intervention** ($A = 1$ [white]); otherwise (eagle), they receive no intervention.

- We then **observe individuals for five days** and **calculate the risk of death**.



$\Pr[Y = 1|A = 1] = 0.3$ $\Pr[Y = 1|A = 0] = 0.6$

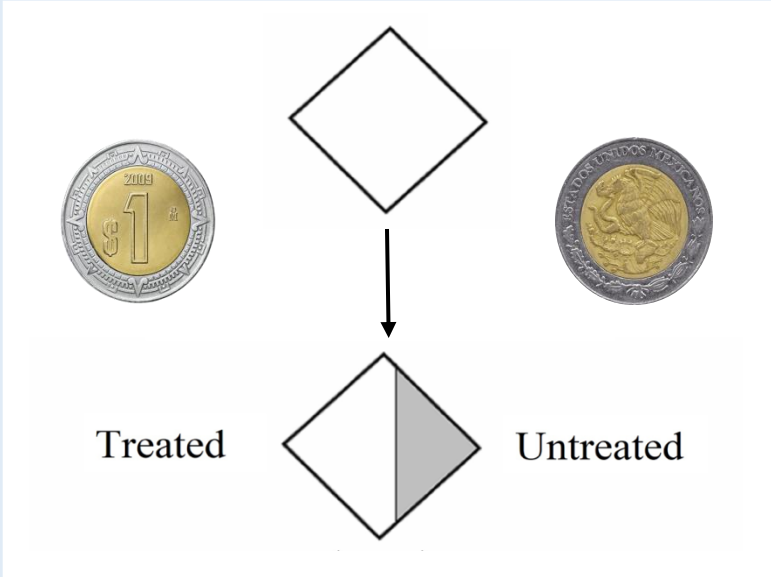
Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference

- Suppose **we randomize an almost infinite population** by **flipping a coin**.

- If the coin lands on **heads ($1)**, individuals are assigned to the **intervention** ($A = 1$ [white]); otherwise (eagle), they receive no intervention.

- We then **observe individuals for five days** and **calculate the risk of death**.



$\Pr[Y = 1|A = 1] = 0.3$ $\Pr[Y = 1|A = 0] = 0.6$

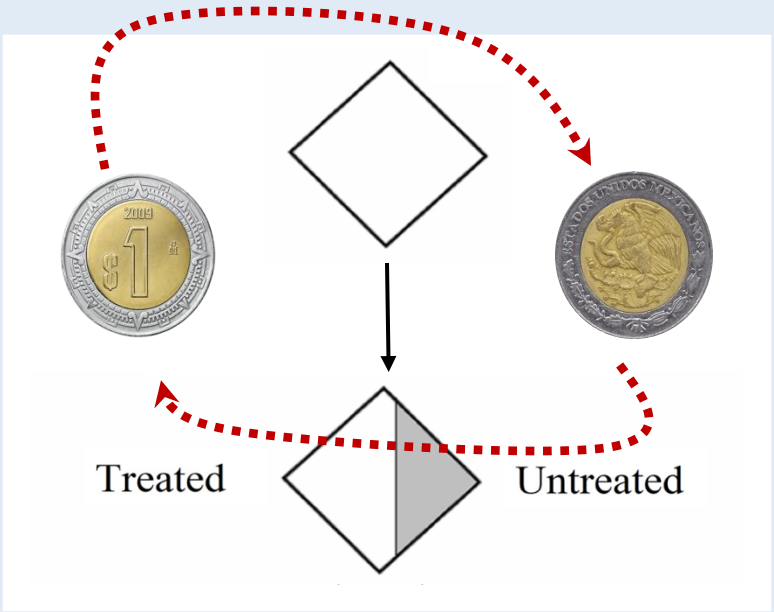
$RR = 0.3/0.6 = 0.5$
$RD = 0.3\text{-}0.6 = \text{-}0.3$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference

- **What if we had mistakenly treated the gray group instead of the white group?**

- Would our results have been affected?



$\mathbf{Pr}[Y = 1|A = 1] = ?$ $\mathbf{Pr}[Y = 1|A = 0] = ?$

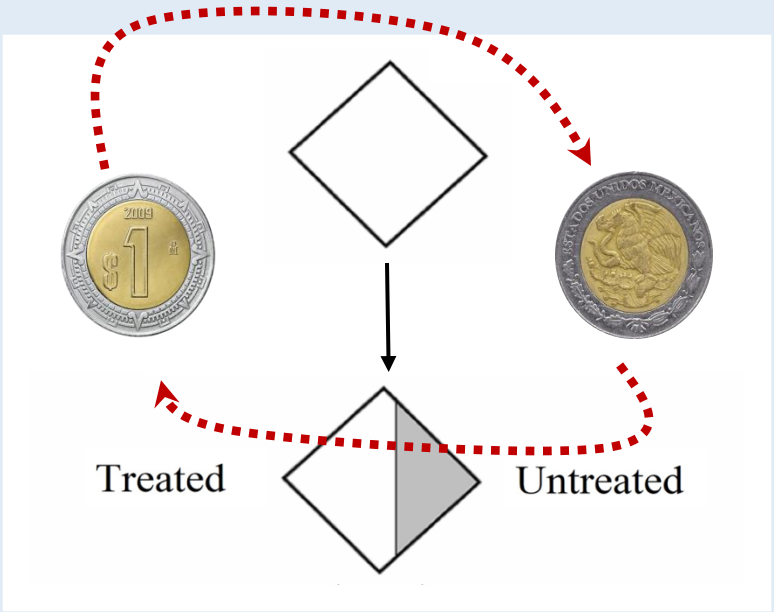
$RR = ?$ $RD = ?$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference

- **What if we had mistakenly treated the gray group instead of the white group?**

- Would our results have been affected?



$$\Pr[Y = 1|A = 1] = 0.3 \qquad \Pr[Y = 1|A = 0] = 0.6$$

$$RR = 0.3/0.6 = 0.5$$
$$RD = 0.3\text{-}0.6 = \text{-}0.3$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference: **exchangeability**

McGill

- **In summary**, **all groups** defined by treatment level **will have the same counterfactual risk of the outcome as** that observed in **the population as a whole**:

$$\Pr[Y^a = 1 \mid A = 1] =$$
$$\Pr[Y^a = 1 \mid A = 0] =$$
$$\Pr[Y^a = 1]$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference: **exchangeability**

McGill

- **In summary**, **all groups** defined by treatment level **will have the same counterfactual risk of the outcome as** that observed in **the population as a whole**:

$$\Pr[Y^a = 1 \mid A = 1] =$$
$$\Pr[Y^a = 1 \mid A = 0] =$$
$$\Pr[Y^a = 1]$$

- Because **the level of treatment actually received does not predict the counterfactual outcome**, we say that the counterfactual outcome and the level of treatment are independent (**treatment groups are exchangeable**).

Exchangeability:
$Y^a \perp\!\!\!\perp A$ for all a.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of randomization in causal inference: **exchangeability**

McGill

- **In summary**, **all groups** defined by treatment level **will have the same counterfactual risk of the outcome as** that observed in **the population as a whole**:

- **Randomization plays a key role because, in expectation, it leads to exchangeability**.

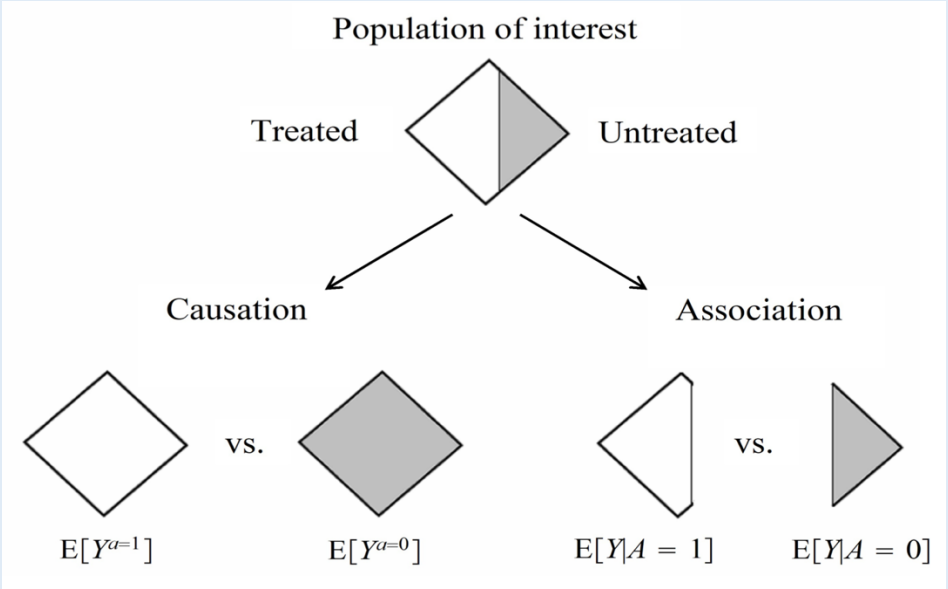
- Because **the level of treatment actually received does not predict the counterfactual outcome**, we say that the counterfactual outcome and the level of treatment are independent (**treatment groups are exchangeable**).

Exchangeability:
$Y^a \perp\!\!\!\perp A$ for all a.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

In an **ideal randomized experiment**: **association** is **causation**

- [Again] **If exchangeability holds, all treatment groups will have the same counterfactual risk as the entire population** (since *A* was randomly assigned).

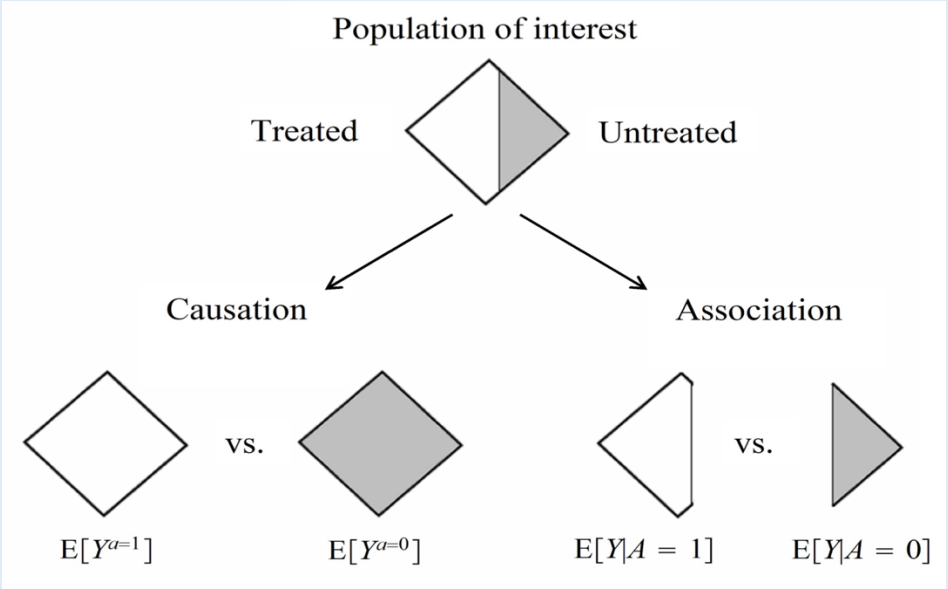


Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

In an **ideal randomized experiment**: **association** is **causation**

- [Again] **If exchangeability holds, all treatment groups will have the same counterfactual risk as the entire population** (since *A* was randomly assigned).

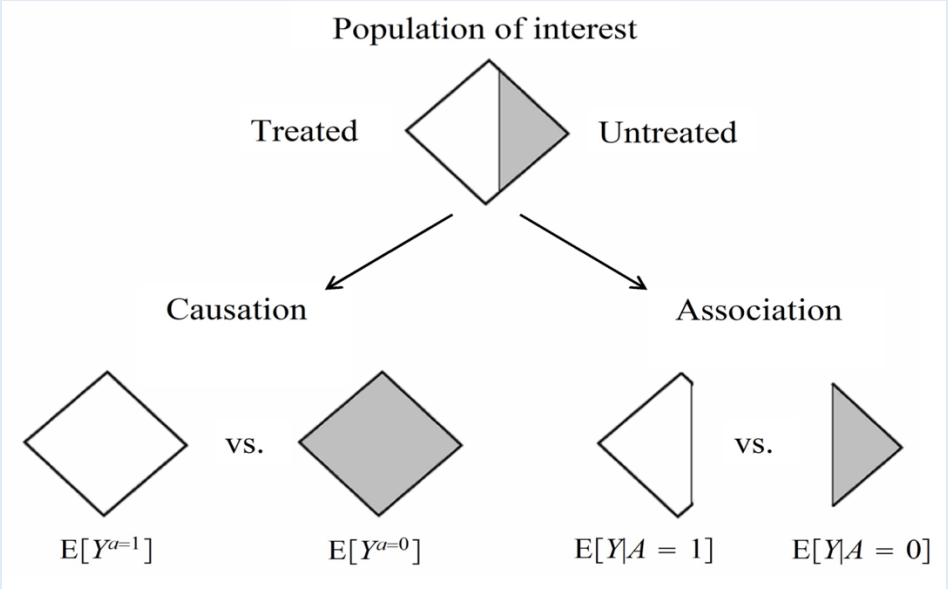
- However, **the counterfactual risk in the treated is not truly counterfactual** because they actually received the treatment (the same is true for the untreated).



Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

In an **ideal randomized experiment**: **association** is **causation**

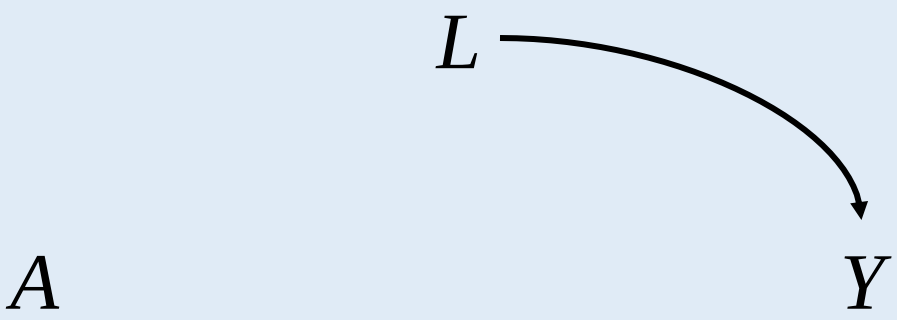
- [Again] **If exchangeability holds, all treatment groups will have the same counterfactual risk as the entire population** (since *A* was randomly assigned).
- However, **the counterfactual risk in the treated is not truly counterfactual** because they actually received the treatment (the same is true for the untreated).
- Thus, **by calculating the risk in the treated, we can estimate the counterfactual risk under treatment** (and similarly, under no treatment), **and** then comparing these two quantities allows us to **obtain causal effects**.



Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

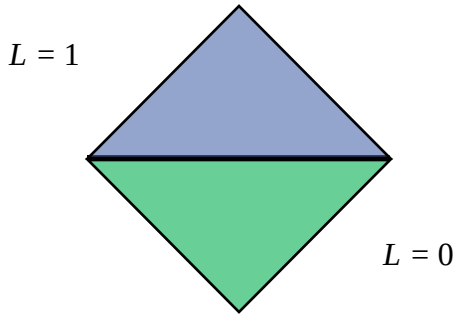
3. Moving **from marginal to conditionally randomized experiments**

The role of **conditional** randomization in causal inference

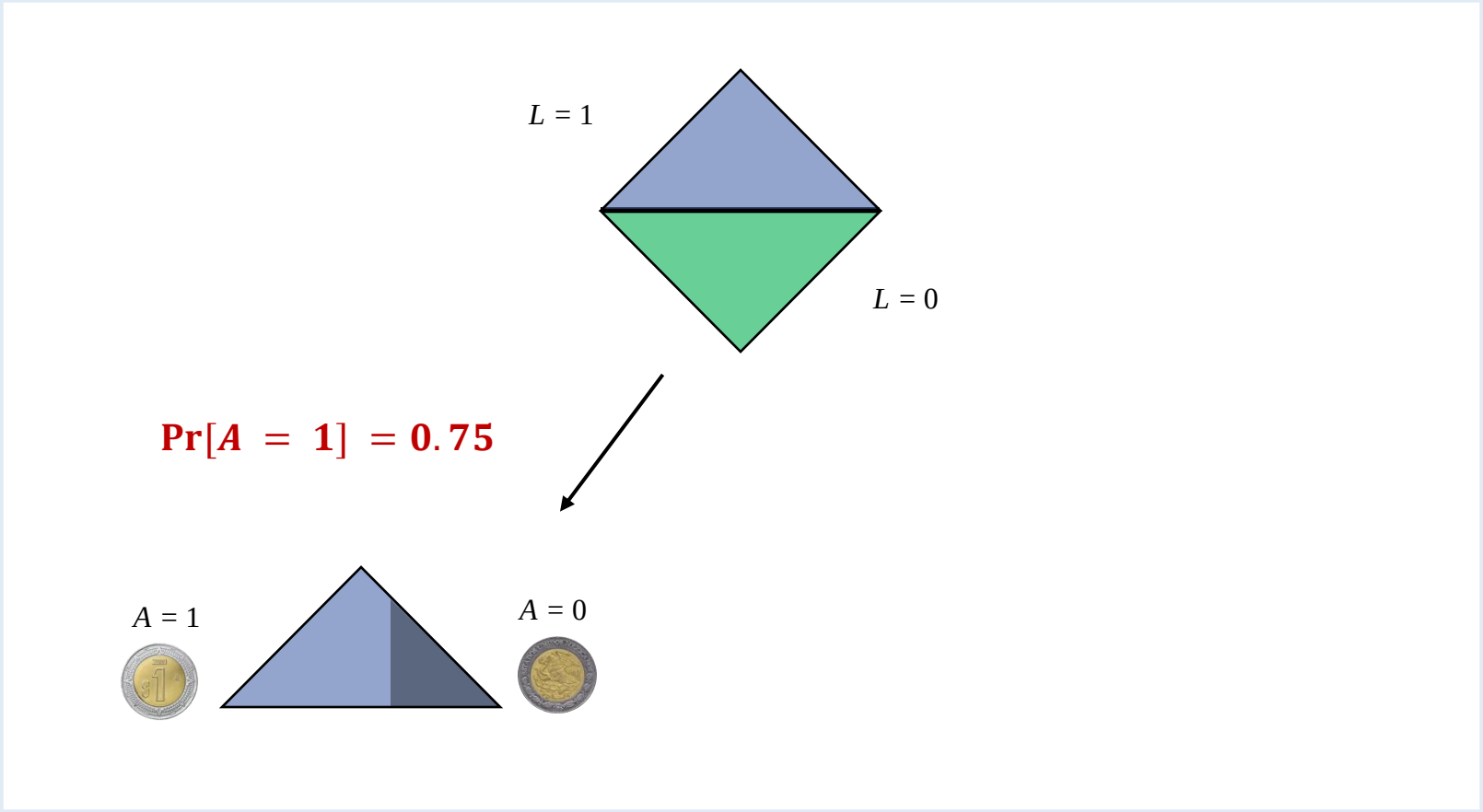


L : critical illness

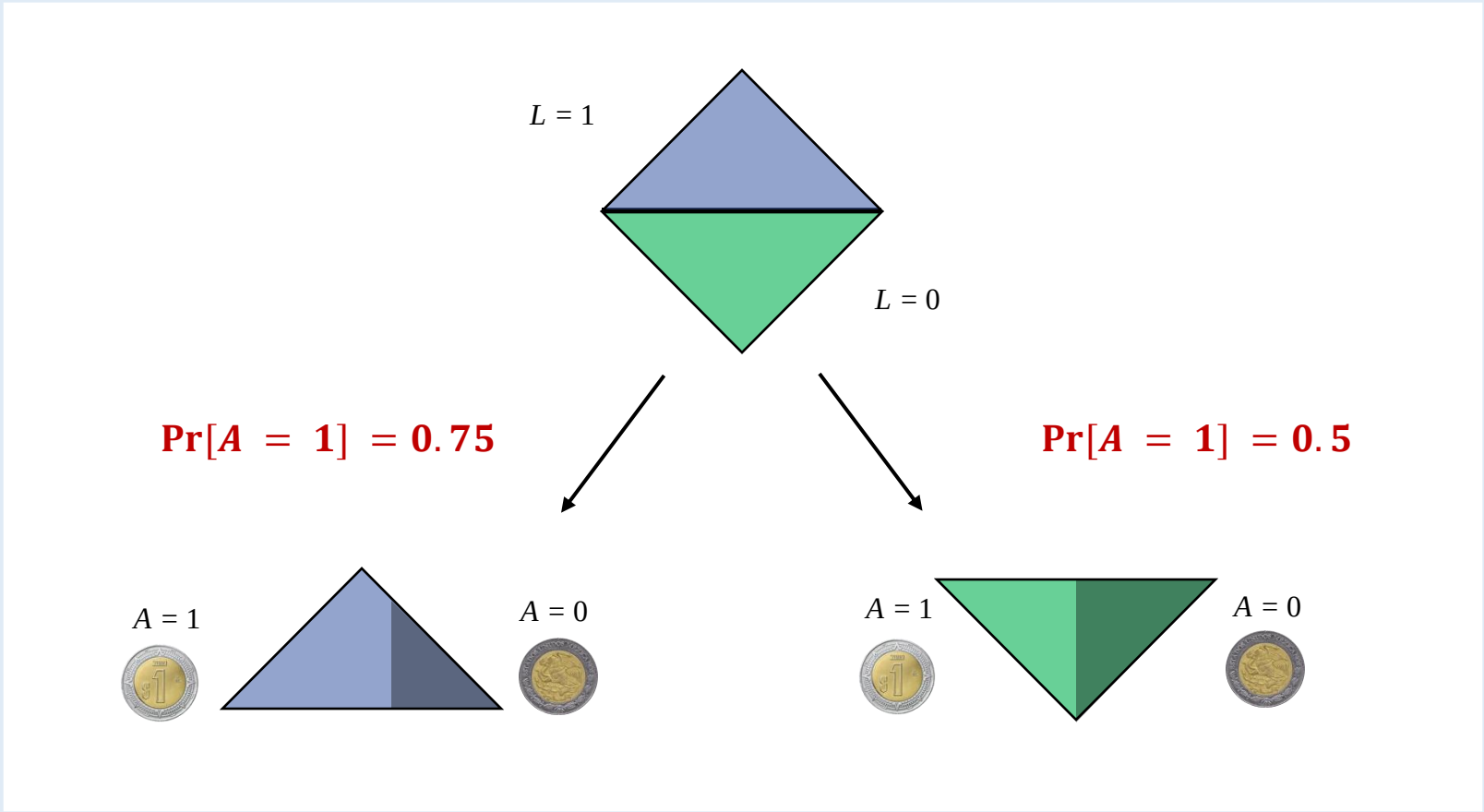
The role of **conditional** randomization in causal inference



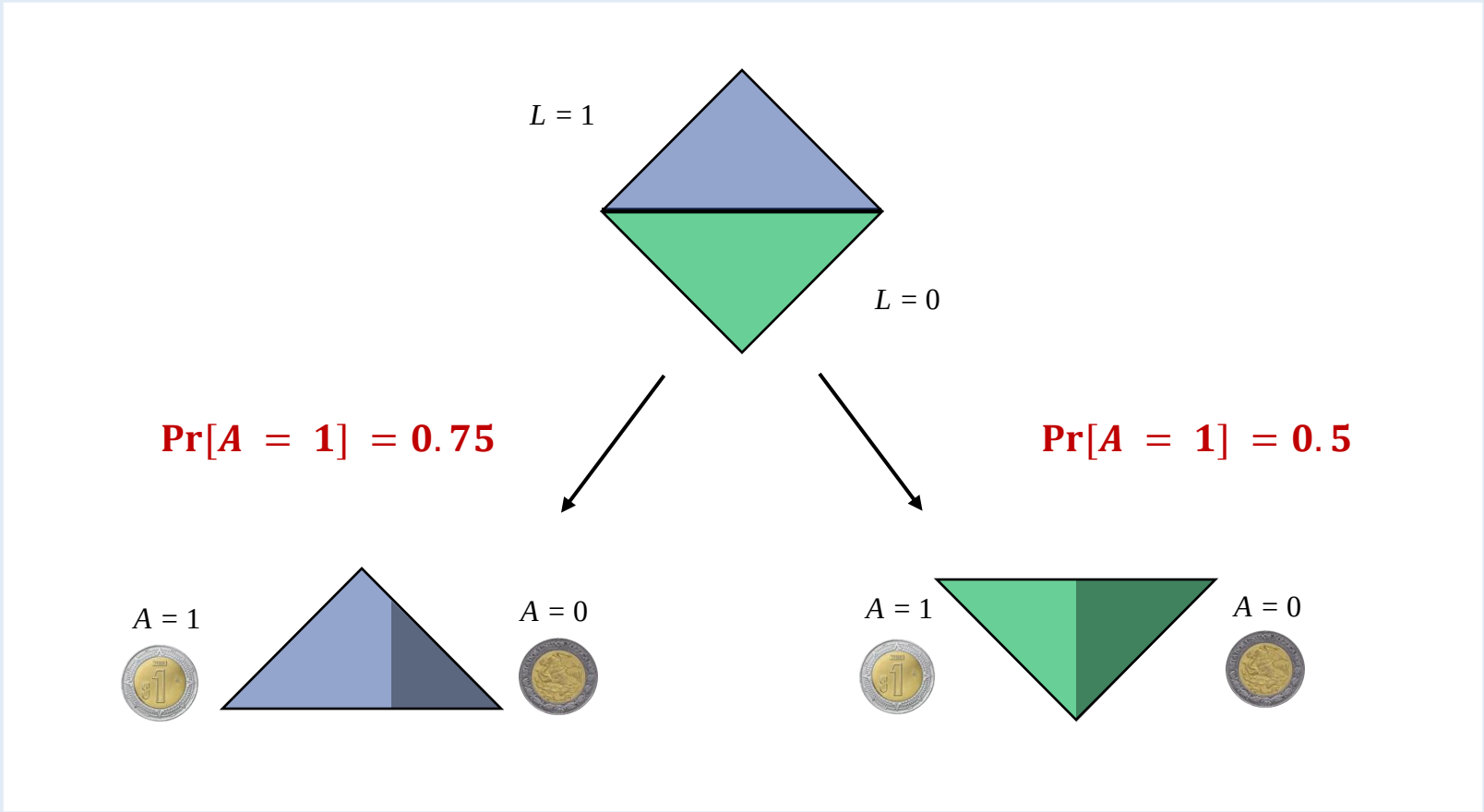
The role of **conditional** randomization in causal inference



The role of **conditional** randomization in causal inference



The role of **conditional** randomization in causal inference



The role of **conditional** randomization in causal inference

Table 2.2

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

- In this design, since critically ill individuals had a higher probability of being treated, **treated individuals will have a higher baseline risk of death**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of **conditional** randomization in causal inference

Table 2.2

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

- In this design, since critically ill individuals had a higher probability of being treated, **treated individuals will have a higher baseline risk of death**.

- Even though this is an RCT, **there is no marginal exchangeability**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

The role of **conditional** randomization in causal inference

Table 2.2

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

- In this design, since critically ill individuals had a higher probability of being treated, **treated individuals will have a higher baseline risk of death**.
- Even though this is an RCT, **there is no marginal exchangeability**.
- However, in expectation, individuals are exchangeable within levels of L (**conditional exchangeability**).

Conditional exchangeability:
$Y^a \perp\!\!\!\perp A|L$ for all a

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

We can get the **conditional causal RR** (OR, RD) …

- Since **association is causation within each subgroup defined by *L***, by calculating the **associational RR (RD, OR)** we can recover the **causal RR (RD, OR)** in each stratum:

$$\mathbf{Pr}[Y^{a=1} = 1|L = 1] / \mathbf{Pr}[Y^{a=0} = 1|L = 1]$$

And

$$\mathbf{Pr}[Y^{a=1} = 1|L = 0] / \mathbf{Pr}[Y^{a=0} = 1|L = 0]$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

We can get the **marginal causal RR** (OR, RD) …

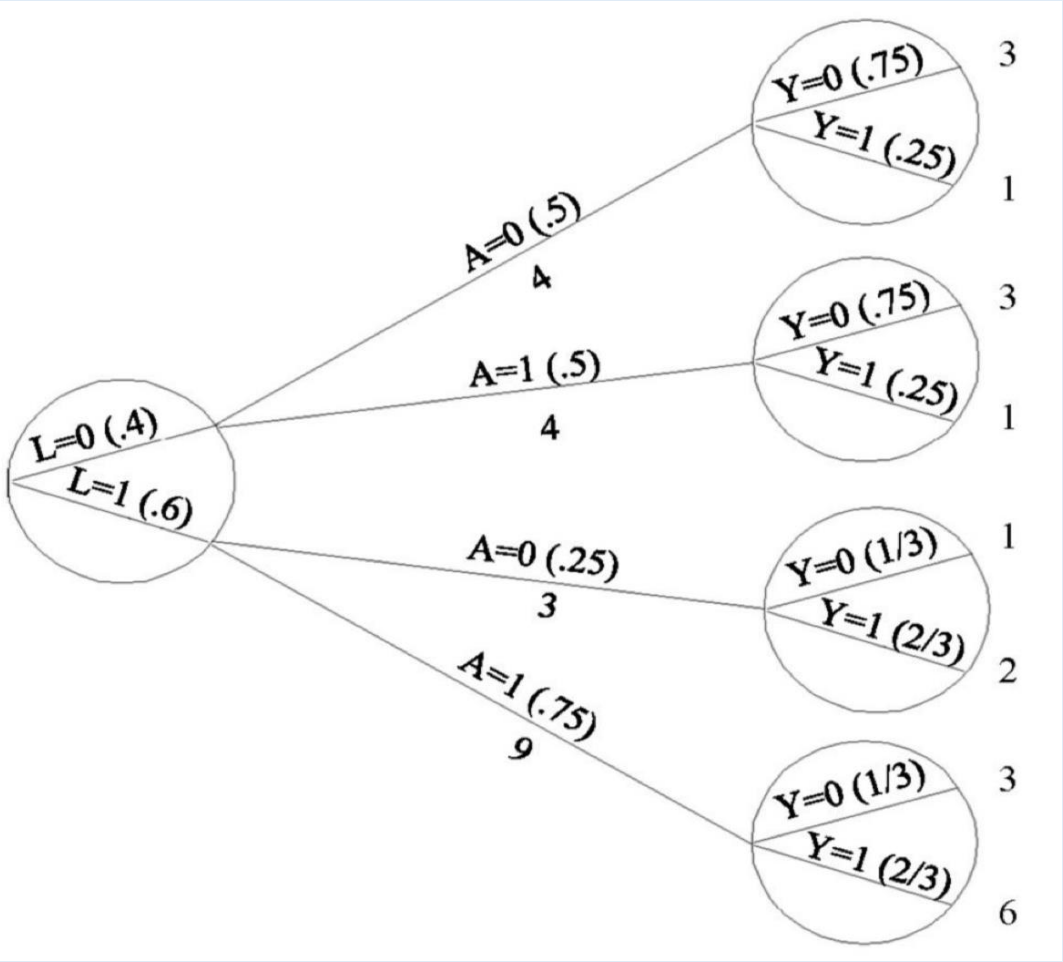
- By using either **standardization** or **inverse probability of treatment weighting (IPTW)**:

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

IPTW: observed data

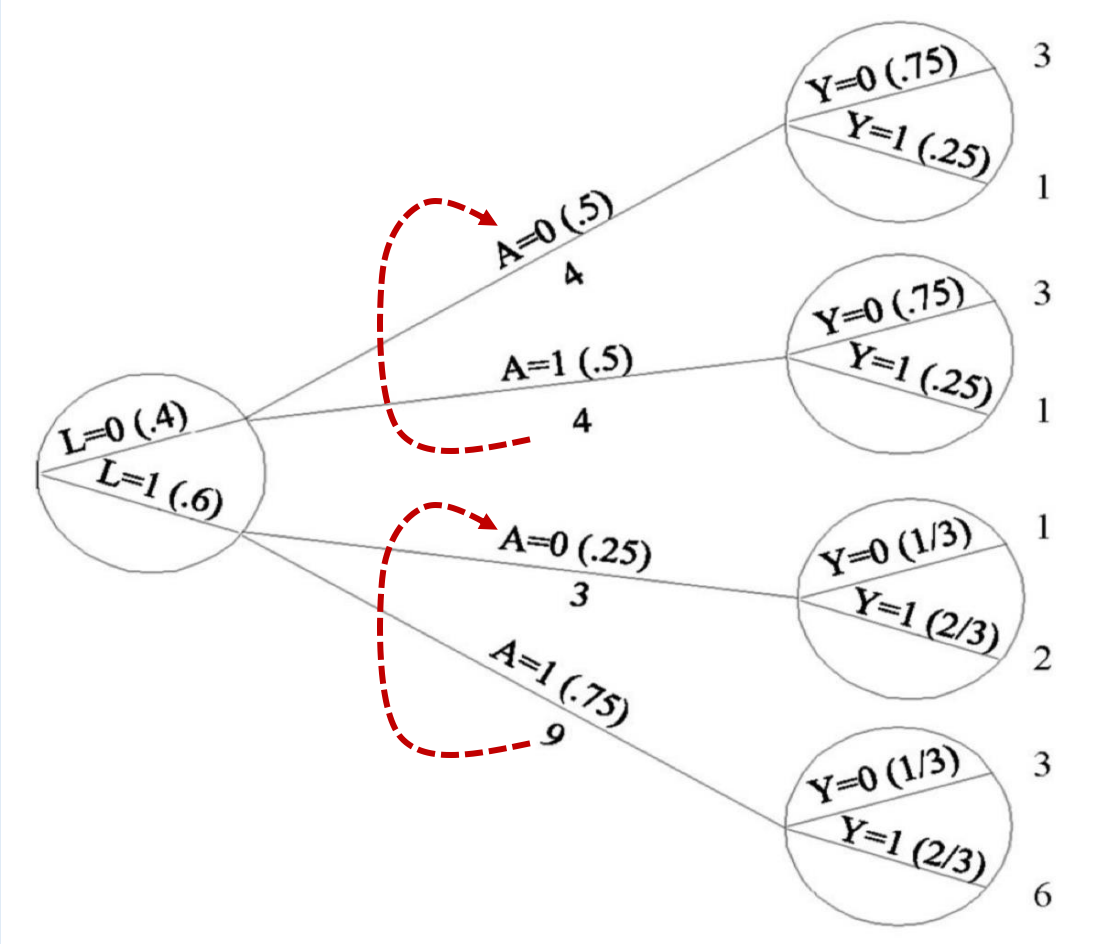
Table 2.2

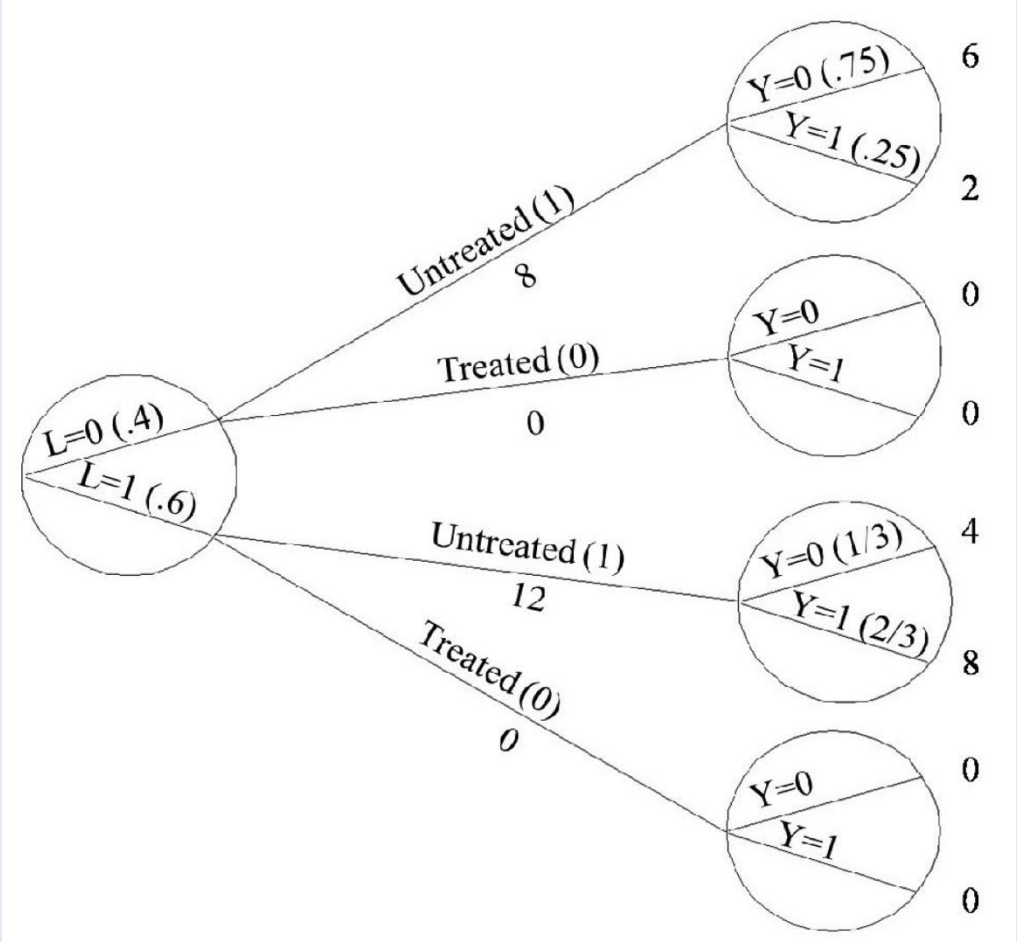
	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Polyphemus	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0



Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

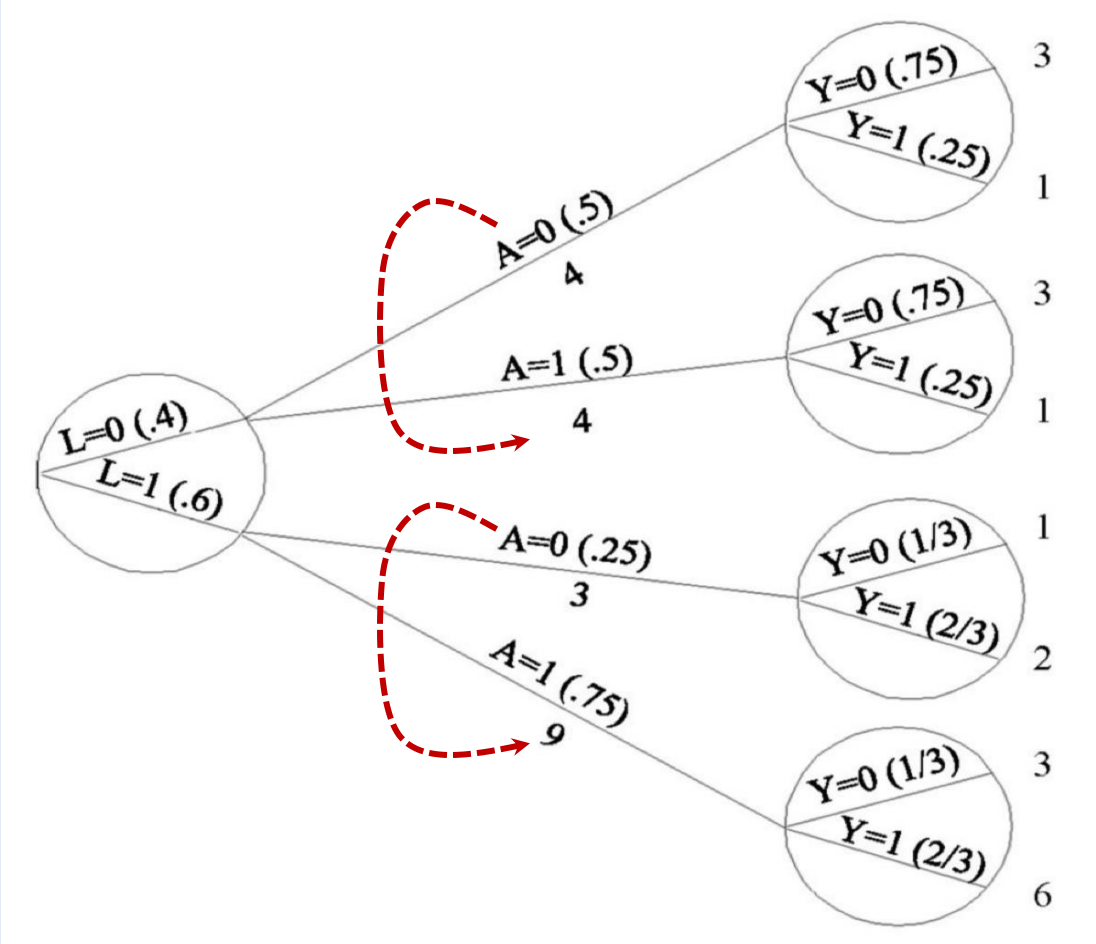
IPTW: counterfactual risk under no treatment

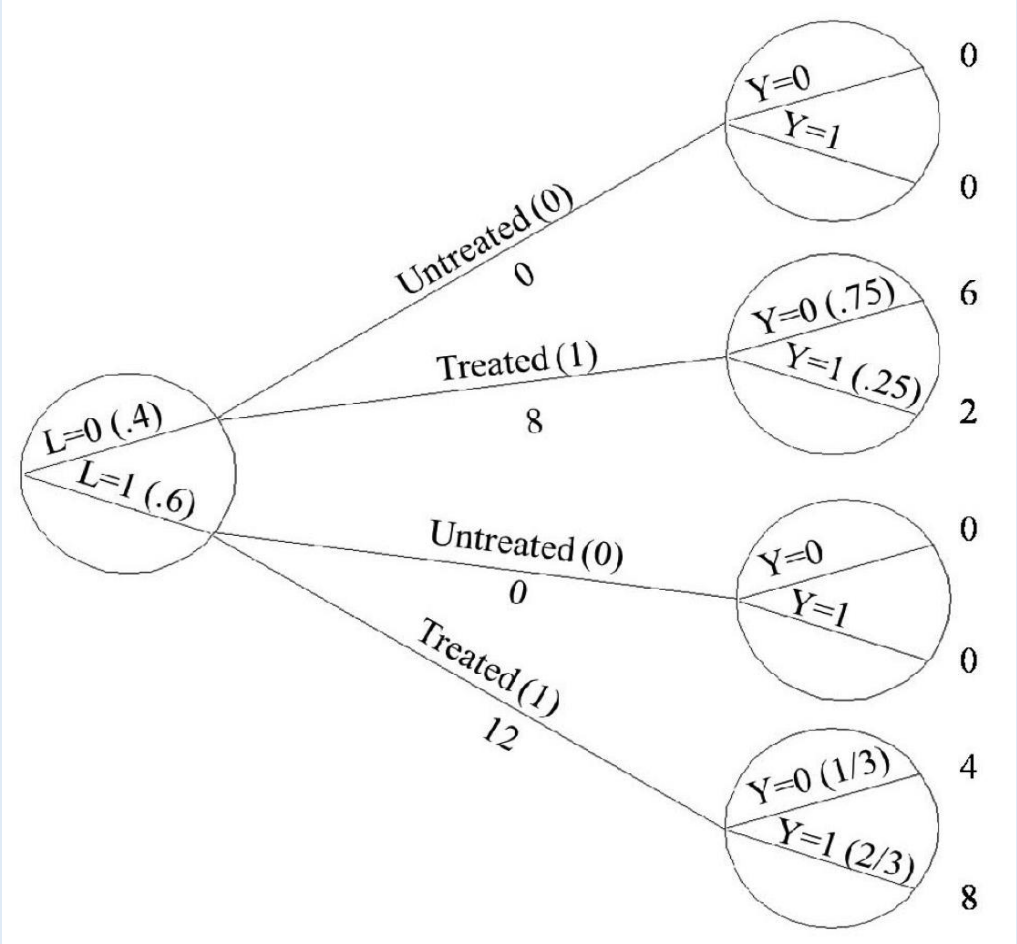




Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

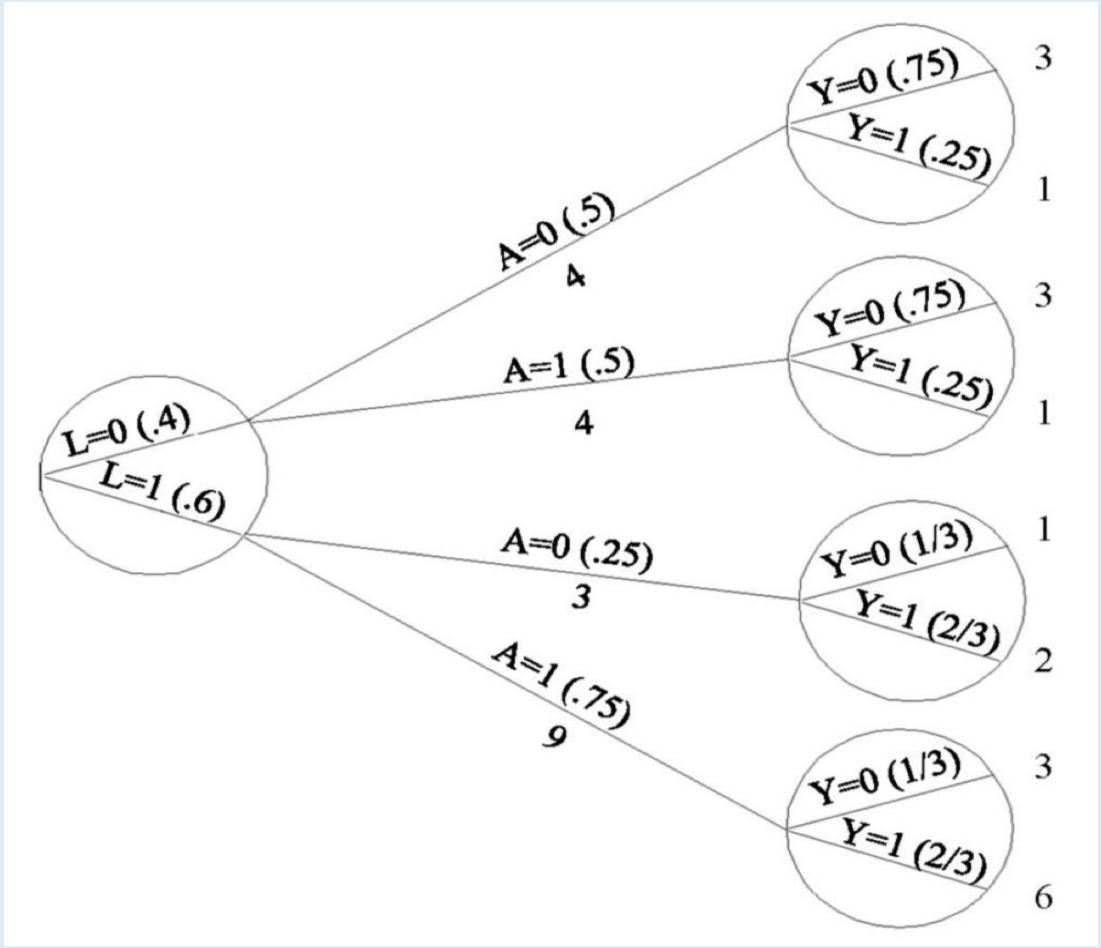
IPTW: counterfactual risk under treatment

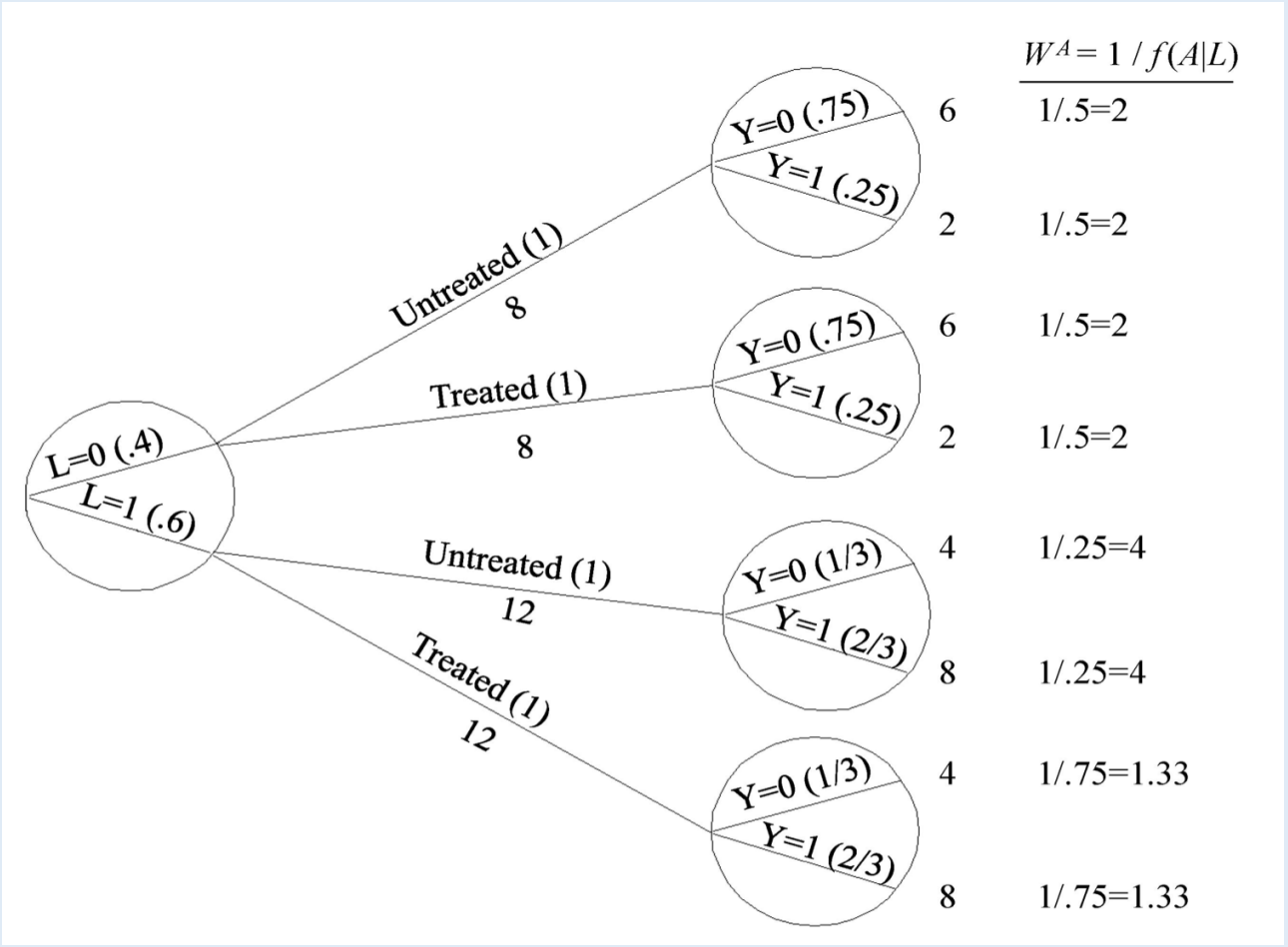




Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

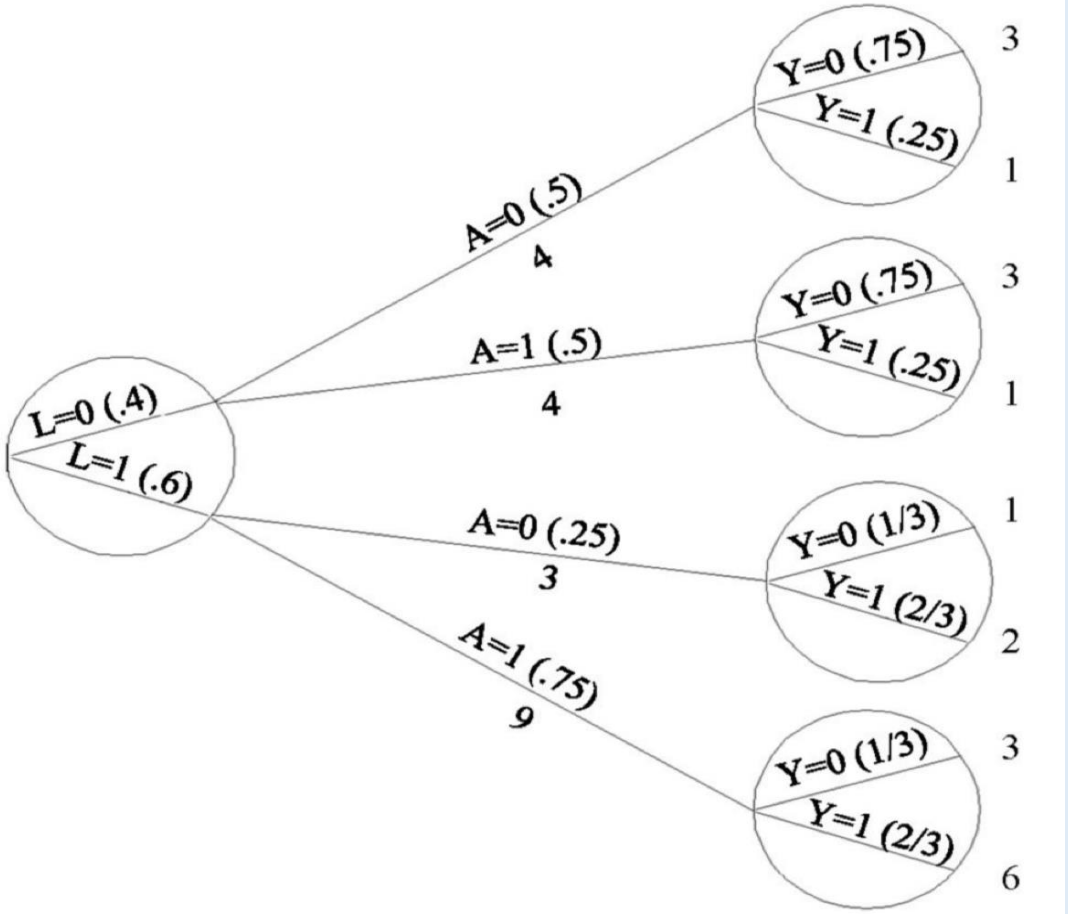
IPTW

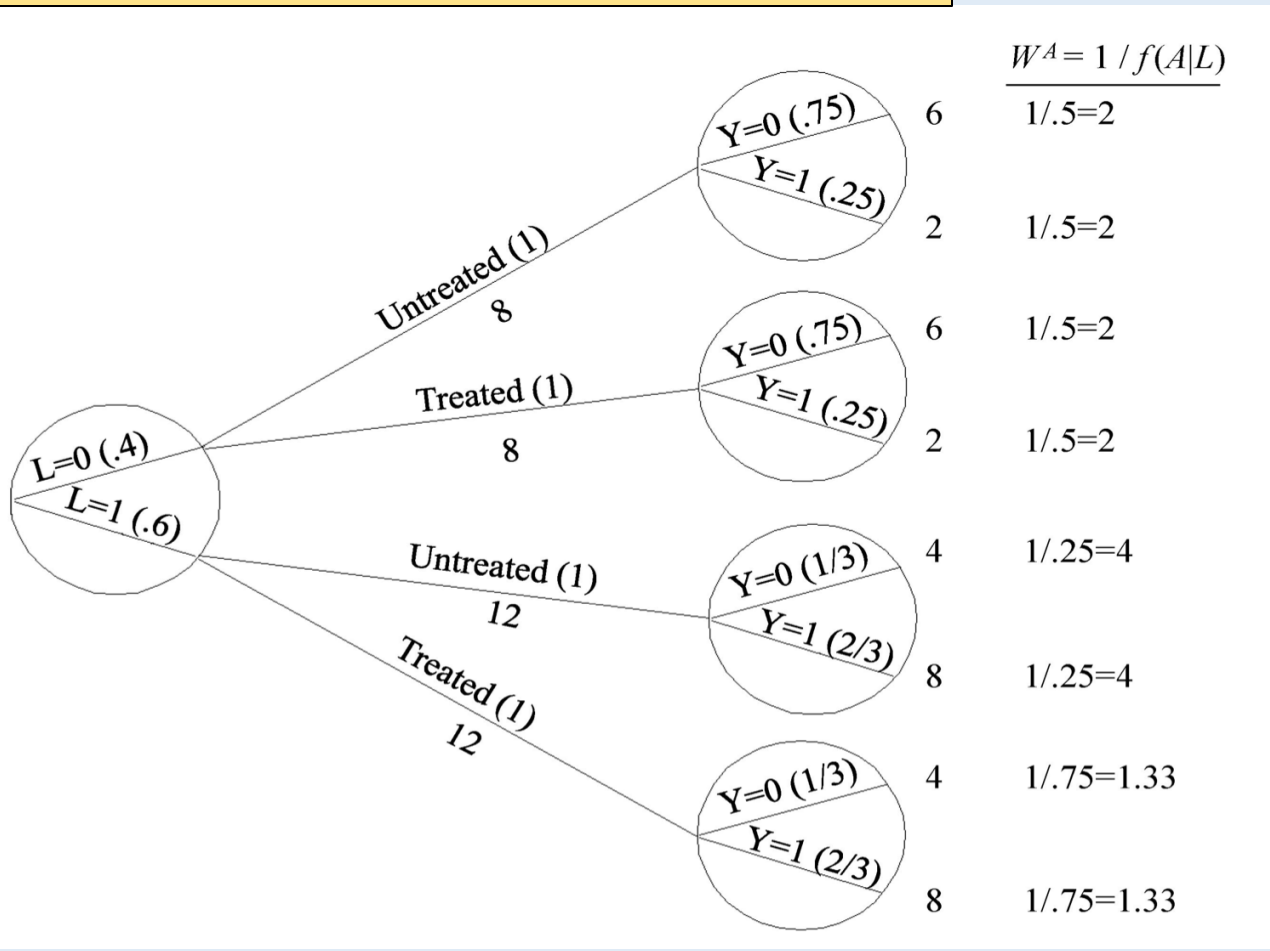




Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

- Each individual receives a **weight** equal to the **inverse of the probability** of **receiving the level of treatment they received** (**conditional on *L***).
- IPTW creates a **pseudo-population in which the distribution of confounders is balanced across treatment groups**.





Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

IPTW: R Code

```
# Estimation of IP weights via a logistic model

fit <- glm(qsmk ~ sex + race + age + I(age ^ 2) + as.factor(education) + smokeintensity +
           I(smokeintensity ^ 2) + smokeyrs + I(smokeyrs ^ 2) + as.factor(exercise) +
           as.factor(active) + wt71 + I(wt71 ^ 2),
           family = binomial(),
           data = nhefs.nmv )

p.qsmk.obs <- ifelse(nhefs.nmv$qsmk == 0,
                     1 - predict(fit, type = "response"),
                      predict(fit, type = "response"))

nhefs.nmv$w <- 1 / p.qsmk.obs
```

Shi J, McGrath S. R and Stata code for exercises accompanying: Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2024. Available at: https://remlapmot.github.io/cibookex-r/. Accessed Nov 11, 2024.

IPTW: R Code

```
# IPWeighted regression

library("geepack")
msm.w <- geeglm(wt82_71 ~ qsmk,
                data = nhefs.nmv,
                weights = w,
                id = seqn,
                corstr = "independence" )

summary(msm.w)
```

Shi J, McGrath S. R and Stata code for exercises accompanying: Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2024. Available at: https://remlapmot.github.io/cibookex-r/. Accessed Nov 11, 2024.

Standardization

- We can recover the **marginal counterfactual risk under each level of treatment** as follows:

$$\Pr[Y^a = 1] = \sum_l \Pr[Y^a = 1 | L = l] \Pr[L = l]$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Standardization

- We can recover the **marginal counterfactual risk under each level of treatment** as follows:

$$\Pr[Y^a = 1] = \sum_l \Pr[Y^a = 1 | L = l] \Pr[L = l]$$

- **Under conditional exchangeability**, within $L = l$, **the risk in the treated is equal to the risk that would have been observed if everyone had been treated** (the **same for the untreated**).

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Standardization

- We can recover the **marginal counterfactual risk under each level of treatment** as follows:

$$\Pr[Y^a = 1] = \sum_l \Pr[Y^a = 1|L = l] \Pr[L = l]$$

- **Under conditional exchangeability**, within $L = l$, **the risk in the treated is equal to the risk that would have been observed if everyone had been treated** (the **same for the untreated**).

- Unify this into **one equation**:

$$\Pr[Y^a = 1] = \sum_l \Pr[Y = 1|L = l, A = a] \Pr[L = l]$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Standardization

$$\Pr[Y^a = 1] = \sum_l \Pr[Y = 1 | L = l, A = a] \Pr[L = l]$$

- In summary:
  - ***Stratification***: Divide the population into **strata based on *L***.
  - ***Estimate stratum-specific risks***: Calculate the **outcome risk within each stratum for each treatment level**.
  - ***Weighting***: Estimate a **weighted average of the stratum-specific risks** by weighting them by the **proportion of the population in each stratum defined by *L***.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Standardization

$$\Pr[Y^a = 1] = \sum_l \Pr[Y = 1 | L = l, A = a] \Pr[L = l]$$

- The **marginal counterfactual risk** [left side of the equation] is a **weighted average of the counterfactual risks in each of the strata defined by *L*, with weights equal to the proportion of individuals in the population with *L* = 0 and *L* = 1**, respectively:

$$\mathbf{\Pr[Y^{a=1} = 1] =}$$
$$\mathbf{\Pr[Y^{a=1} = 1 | L = 0] \Pr[L = 0] + \Pr[Y^{a=1} = 1 | L = 1] \Pr[L = 1]}$$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Standardization (parametric g-formula): R Code

McGill

```
# 1) Prepare data (expansion of the dataset): create a dataset with 3 copies of each subject

## 1st copy: equal to original one
nhefs$interv <- -1

## 2nd copy: treatment set to 0, outcome to missing
interv0 <- nhefs
interv0$interv <- 0
interv0$qsmk <- 0
interv0$wt82_71 <- NA

## 3rd copy: treatment set to 1, outcome to missing
interv1 <- nhefs
interv1$interv <- 1
interv1$qsmk <- 1
interv1$wt82_71 <- NA

## combining datasets
onesample <- rbind(nhefs, interv0, interv1)
```

Shi J, McGrath S. R and Stata code for exercises accompanying: Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2024. Available at: https://remlapmot.github.io/cibookex-r/. Accessed Nov 11, 2024.

Standardization (parametric g-formula): R Code

McGill

```
# 2) Outcome modeling: Estimating the mean outcome within levels of treatment and confounders
(only original data is used for this step)

std <- glm(wt82_71 ~ qsmk + sex + race + age + I(age * age) + as.factor(education) +
           smokeintensity + I(smokeintensity * smokeintensity) + smokeyrs +
           I(smokeyrs * smokeyrs) + as.factor(exercise) +
           + as.factor(active) + wt71 + I(wt71 * wt71) + I(qsmk * smokeintensity),
           data = onesample)
```

Shi J, McGrath S. R and Stata code for exercises accompanying: Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2024. Available at: https://remlapmot.github.io/cibookex-r/. Accessed Nov 11, 2024.

Standardization (parametric g-formula): R Code

McGill

```
# 3) Outcome prediction and standardization to the baseline confounders by averaging:

onesample$predicted_meanY <- predict(std, onesample)


## estimate mean outcome in each of the groups interv=0, and interv=1
# this mean outcome is a weighted average of the mean outcomes in each combination
# of values of treatment and confounders, that is, the standardized outcome
mean(onesample[which(onesample$interv == -1), ]$predicted_meanY) #> [1] 2.56319
mean(onesample[which(onesample$interv == 0), ]$predicted_meanY) #> [1] 1.660267
mean(onesample[which(onesample$interv == 1), ]$predicted_meanY) #> [1] 5.178841
```

Shi J, McGrath S. R and Stata code for exercises accompanying: Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2024. Available at: https://remlapmot.github.io/cibookex-r/. Accessed Nov 11, 2024.

IPTW vs Standardization

- Both methods **are mathematically equivalent**: they **estimate what would happen if everyone received treatment *a***. However, **each method uses a different set of probabilities** to do so:

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

IPTW vs Standardization

- Both methods **are mathematically equivalent**: they **estimate what would happen if everyone received treatment *a***. However, **each method uses a different set of probabilities** to do so:
  - **IPTW** uses the **conditional probability of treatment *A* given covariate (s) *L***.
  - **Standardization** uses the **probability of covariate *L*** and the **conditional probability of outcome *Y* given variables *A* and *L***.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

IPTW vs Standardization

- Both methods **are mathematically equivalent**: they **estimate what would happen if everyone received treatment *a***. However, **each method uses a different set of probabilities** to do so:
  - **IPTW** uses the **conditional probability of treatment *A* given covariate (s) *L***.
  - **Standardization** uses the **probability of covariate *L*** and the **conditional probability of outcome *Y* given variables *A* and *L***.

- Since **both methods simulate what would have been observed if the variable (s) L had not influenced the probability of treatment**, these methods are often said to **adjust** (**or control**) **the analysis for *L*.**

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

4. **Observational studies** as an **alternative to randomized trials**.

When RCTs are not feasible:
The role of observational studies

McGill

- **In many cases**, **RCTs are either unethical** (e.g., when we are interested in ruling out adverse effects), **impractical** (e.g., due to high costs), or **too time-consuming**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

When RCTs are not feasible:
The role of observational studies

McGill

- **In many cases**, **RCTs are either unethical** (e.g., when we are interested in ruling out adverse effects), **impractical** (e.g., due to high costs), or **too time-consuming**.

- In these cases, **the most appropriate (or only) option** is to conduct an **observational study** (studies in which the researcher simply observes and records the relevant data).

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

But **how do we estimate causal effects** from **observational data**?

McGill

- We **analyze our data as if the treatment had been randomized conditional on the measured covariates *L*** (although this is often only an approximation).

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

But **how do we estimate causal effects** from **observational data**?

McGill

- We **analyze our data as if the treatment had been randomized conditional on the measured covariates *L*** (although this is often only an approximation).

- In other words, **causal inference from observational data is based on the hope that such studies can be considered a conditionally randomized experiment**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

But **how do we estimate causal effects** from **observational data**?

McGill

- We **analyze our data as if the treatment had been randomized conditional on the measured covariates *L*** (although this is often only an approximation).

- In other words, **causal inference from observational data is based on the hope that such studies can be considered a conditionally randomized experiment**.

- **An observational study can be conceptualized as a conditionally randomized experiment if the following conditions are met (identifiability conditions)**:

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

But **how do we estimate causal effects** from **observational data**?

McGill

- We **analyze our data as if the treatment had been randomized conditional on the measured covariates *L*** (although this is often only an approximation).

- In other words, **causal inference from observational data is based on the hope that such studies can be considered a conditionally randomized experiment**.

- **A counterfactual quantity is said to be identifiable** if it can be expressed as a function of the distribution (i.e., the probabilities) of the observed data (otherwise it is said to be unidentifiable).

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Identifiability conditions

1. **Consistency**: The treatment values to be compared correspond to **well-defined interventions (treatment variation irrelevance –** all versions of treatment have the same effects**)**, which in turn **can be mapped to** the treatment versions available in **the data**.

Consistency:
if $A_i = a$, then $Y_i^a = Y_i^A = Y_i$

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Identifiability conditions

1. **Consistency**: The treatment values to be compared correspond to **well-defined interventions (treatment variation irrelevance –** all versions of treatment have the same effects**)**, which in turn **can be mapped to** the treatment versions available in **the data**.

Consistency:
if $A_i = a$, then $Y_i^a = Y_i^A = Y_i$

2. **Exchangeability**: The **conditional probability of receiving each treatment value**, while not decided by the investigators, **depends only on the measured covariates *L***.

Conditional exchangeability:
$Y^a \perp\!\!\!\perp A|L$ for all a

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

McGill

Identifiability conditions

1. **Consistency**: The treatment values to be compared correspond to **well-defined interventions (treatment variation irrelevance –** all versions of treatment have the same effects**)**, which in turn **can be mapped to** the treatment versions available in **the data**.

Consistency:
if $A_i = a$, then $Y_i^a = Y_i^A = Y_i$

2. **Exchangeability**: The **conditional probability of receiving each treatment value**, while not decided by the investigators, **depends only on the measured covariates *L***.

Conditional exchangeability:
$Y^a \perp\!\!\!\perp A|L$ for all a

3. **Positivity**: The **probability of receiving each treatment value, conditional on *L*** (in the minimum set), **is** greater than zero (i.e., **positive**).

Positivity: $\Pr[A = a|L = l] > 0$
for all values l with $\Pr[L = l] \neq 0$
in the population of interest.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Identifiability conditions

1. **Consistency**: The treatment values to be compared correspond to **well-defined interventions (treatment variation irrelevance –** all versions of treatment have the same effects**)**, which in turn **can be mapped to** the treatment versions available in **the data**.

Consistency:
if $A_i = a$, then $Y_i^a = Y_i^A = Y_i$

2. **Exchangeability**: The **conditional probability of receiving each** decided by the investigators, **depends only on the measured cova**

Conditional exchangeability:
$Y^a \perp\!\!\!\perp A|L$ for all a

In an ideal randomized experiment, these three conditions are met by design.

3. **Positivity**: The **probability of receiving each treatment value, conditional on *L*** (in the minimum set), **is** greater than zero (i.e., **positive**).

Positivity: $\Pr[A = a|L = l] > 0$
for all values l with $\Pr[L = l] \neq 0$
in the population of interest.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Identifiability conditions

4. **No measurement error**.

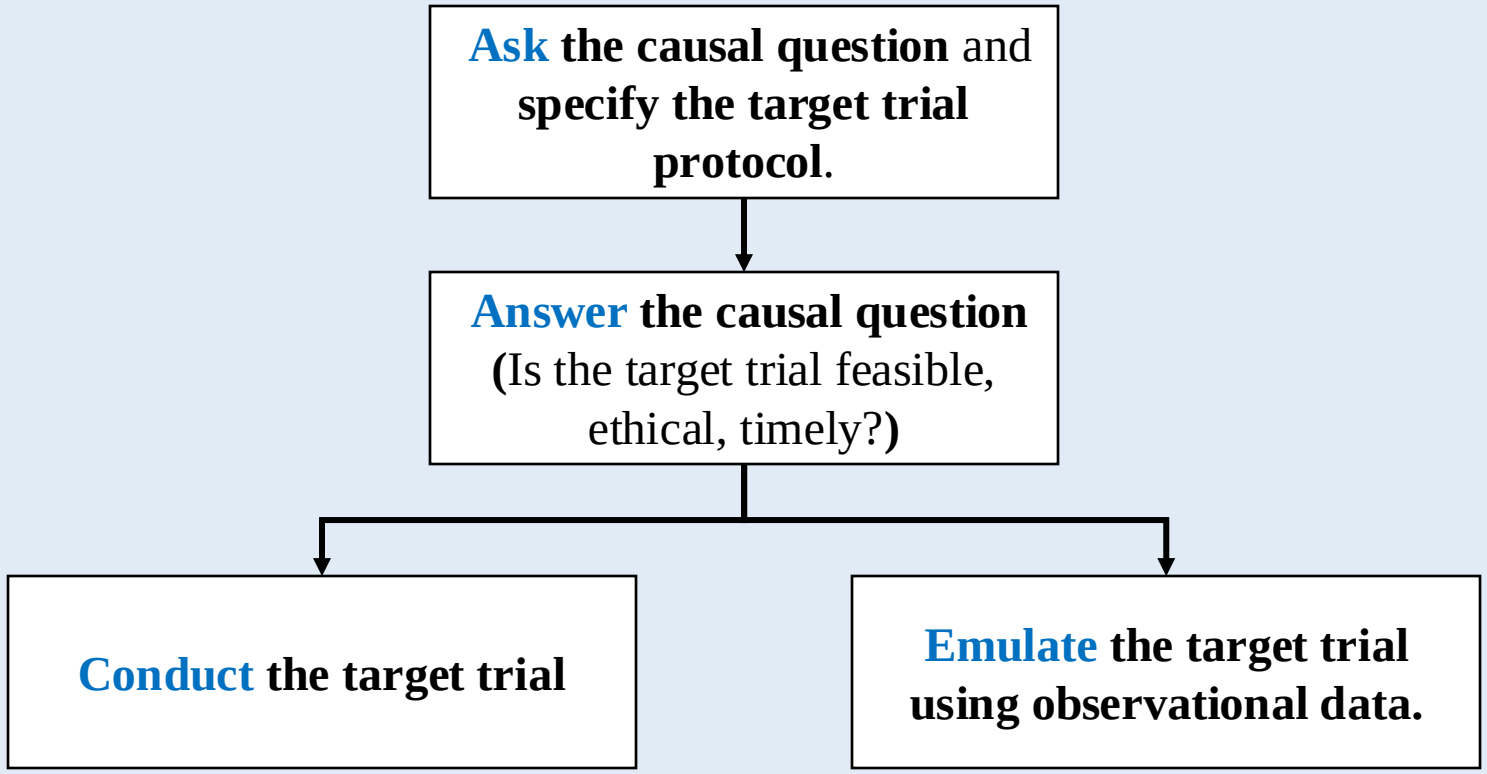
5. **No model misspecification**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

McGill | Department of Epidemiology, Biostatistics and Occupational Health

5. **Estimating causal effects with observational data** through the **emulation of target trials**.

A simple 2-step algorithm for causal inference:



CAUSALab. Material for the CAUSALab Summer Course on Causal Inference. Harvard T.H. Chan School of Public Health; 2022.

What is a **target trial**?

- **For each causal question** (and **for each average causal effect we wish to estimate** [the **estimand** or **target parameter**]), **we can imagine a hypothetical randomized trial**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

What is a **target trial**?

- **For each causal question** (and **for each average causal effect we wish to estimate** [the **estimand** or **target parameter**]), **we can imagine a hypothetical randomized trial**.

- This hypothetical trial is called the **target trial**.

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

What is a **target trial**?

- **For each causal question** (and **for each average causal effect we wish to estimate** [the **estimand** or **target parameter**]), **we can imagine a hypothetical randomized trial**.

- This hypothetical trial is called the **target trial**.

- We can then **specify and emulate the target trial protocol**:

Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Ask the causal question

- For example, we might be interested in the following question:
    - What is the effect of **Covid-19 vaccination** (completion of the 1st immunization series as specified by the manufacturer) vs. no vaccination, **on symptomatic infection and its progression to hospitalization and death** in a cohort of workers from the Mexican Social Security Institute between 2020-Dec-24 and 2021-Jun-24?

- **By defining the causal question, we avoid naive analyses**, such as assessing the effect of being a prevalent user (i.e., initiating and surviving treatment) versus never having used the drug.

- This type of analysis do not inform treatment initiation.

Target trial **protocol**

Specify the target trial protocol
Eligibility criteria
Treatment strategies
Treatment assignment
Time zero and follow-up
Outcomes
Causal contrasts
Data analysis

Target trial **protocol**

Specify the target trial protocol
Eligibility criteria
Treatment strategies
Treatment assignment
Time zero and follow-up
Outcomes
Causal contrasts
Data analysis

- Then **we will use** (and **need**) **all the data elements mentioned in the protocol** to emulate each component of the target trial.

- Therefore, we need to **specify any protocol modifications** that need to be made when using observational data.

McGill

We need to understand our data in detail

- How was data collected?
- What do variables mean?
- Are there idiosyncratic coding practices?
- Is the data accurate (e.g., are there validation studies to quantify misclassification)?
- Is there internal consistency?
- Etcetera.

Target trial **protocol**

The protocol must be for a **pragmatic trial** (a randomized trial conducted under conditions that reflect routine care, with no additional adherence promotion or monitoring):

- Participants and physicians are aware of treatments (i.e., **treatment allocation is not blinded**).
- Strategies include either active treatments or no treatments (i.e., **no one receives a placebo)**.
- **Participants are monitored as frequently as regular patients**.
- Treatment **strategies must exist in the "real world"**.

In addition, **we cannot emulate blinding of outcome ascertainment** with observational data (some exceptions, such as independent ascertainment from death registries).

Eligibility criteria

McGill

Components	TT Specification	TT Emulation
Eligibility Criteria	Age 18+; employed at IMSS; no history of vaccination; no contraindications to vaccination; no documented history of SARS-CoV-2 infection; no Covid-19-like symptoms within one week; no known pregnancy or immunodeficiency; recent health care user (at least one visit to IMSS within one year). Between 20/12/24 and 2021/06/24.	We have no information on contraindications. Two additional exclusion criteria: 1) unreliable vaccination information, 2) unknown vaccine received. (the rest is the same) Note: Eligibility criteria must be met at time zero.
Treatment Strategies		
Treatment Assignment		

Treatment strategies

Components	TT Specification	TT Emulation
Eligibility Criteria	Age 18+; employed at IMSS; no history of vaccination; no contraindications to vaccination; no documented history of SARS-CoV-2 infection; no Covid-19-like symptoms within one week; no known pregnancy or immunodeficiency; recent health care user (at least one visit to IMSS within one year). Between 20/12/24 and 2021/06/24.	We have no information on contraindications. Two additional exclusion criteria: 1) unreliable vaccination information, 2) unknown vaccine received. (the rest is the same) Note: Eligibility criteria must be met at time zero.
Treatment Strategies	1) No immediate vaccination + no LTFU. 2) Immediate vaccination (BNT162b2, AZD1222, Gam-COVID-Vac, Ad5-nCoV, CoronaVac) + no LTFU.	Same
Treatment Assignment		

Emulating **treatment assignment (randomization)**

- Conditional randomized assignment is equivalent to "confounding adjustment":
- We can use:
  - **g-methods**: Standardization, IP weighting, and g-estimation.
  - **Conventional methods for stratification-based adjustment**: stratification (including restriction) and matching.
  - **Model-based extension of conventional stratification**: outcome regression.
- **If we lack information on any relevant confounder** (as specified in our protocol), **we cannot emulate random assignment** (leading to confounding bias).

CAUSALab. Material for the CAUSALab Summer Course on Causal Inference. Harvard T.H. Chan School of Public Health; 2022.
Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2024.

Treatment assignment

Components	TT Specification	TT Emulation
Eligibility Criteria	Age 18+; employed at IMSS; no history of vaccination; no contraindications to vaccination; no documented history of SARS-CoV-2 infection; no Covid-19-like symptoms within one week; no known pregnancy or immunodeficiency; recent health care user (at least one visit to IMSS within one year). Between 20/12/24 and 2021/06/24.	We have no information on contraindications. Two additional exclusion criteria: 1) unreliable vaccination information, 2) unknown vaccine received. (the rest is the same) Note: Eligibility criteria must be met at time zero.
Treatment Strategies	1) No immediate vaccination + no LTFU. 2) Immediate vaccination (BNT162b2, AZD1222, Gam-COVID-Vac, Ad5-nCoV, CoronaVac) + no LTFU.	Same
Treatment Assignment	Unblinded randomization.	We will assume conditional randomization based on the following baseline variables: sex, comorbidities, smoking status, recent hospitalizations, prior influenza vaccination, total number of Covid-19 tests, age, Rx-Risk-ATC score, work setting, occupation, region, and healthcare level.

Follow-up

Components	TT Specification	TT Emulation
Time-zero and follow-up	Begins at treatment assignment (time zero) and continues until outcome of interest, loss to follow-up, death, treatment strategy discontinuation, or administrative censoring (2021-Jun-21), whichever first.	Same
Outcomes		
Causal Contrasts		
Statistical Analysis		

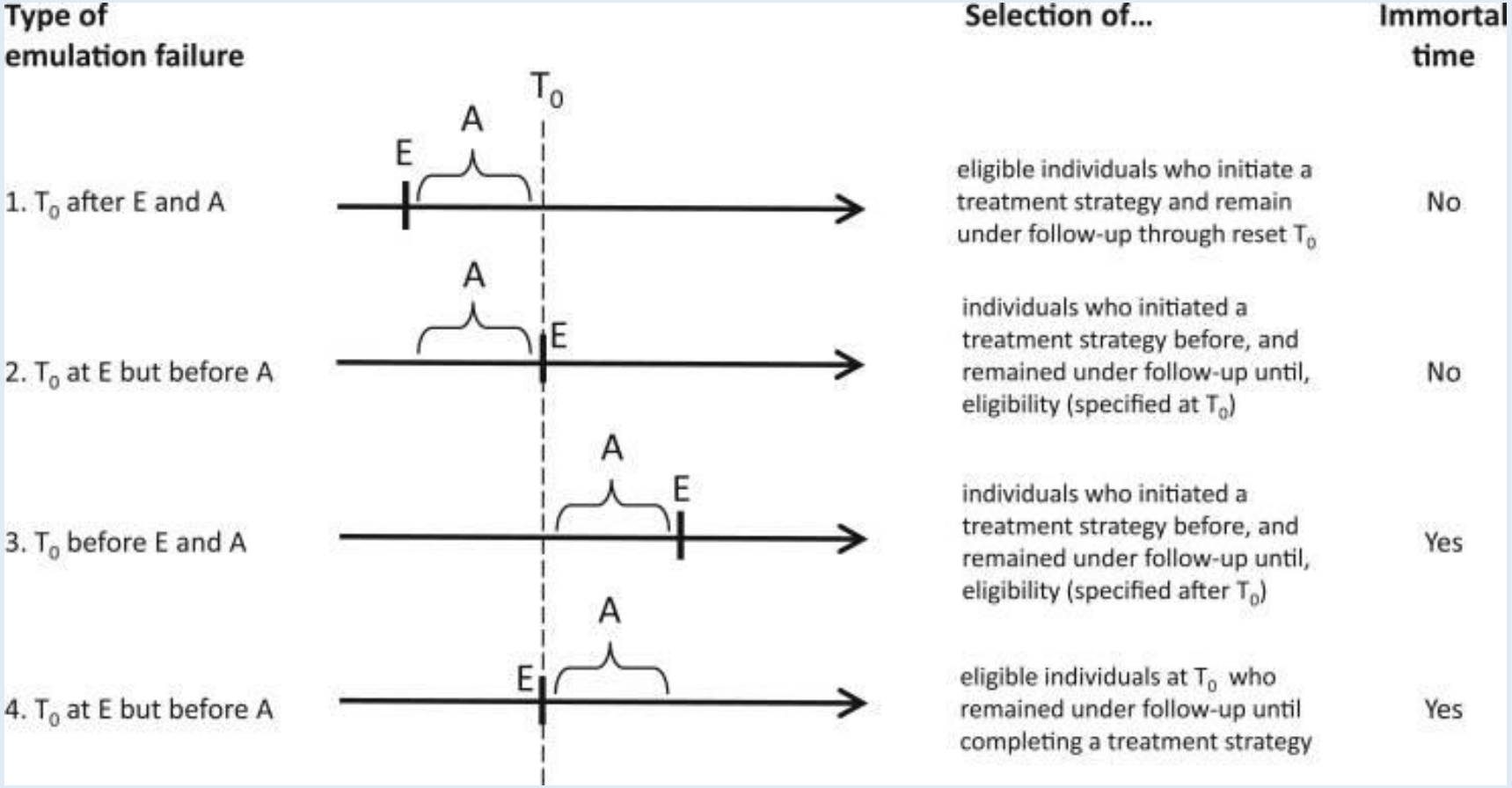
Emulating **time zero and follow-up**

- Time zero of follow-up (baseline) is **defined by the occurrence of three events**:

    1. **Meeting eligibility criteria**.
    2. **Treatment assignment**.
    3. **Start of outcome assessment**.

- **Misalignment of eligibility** and **treatment assignment** can lead to **immortal time** or **selection bias** (e.g., due to depletion of susceptibles).

Failure to emulate time zero



Hernán MA, et al. Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016;79:70-75.

What if the **eligibility criteria are met more than once**?

McGill

- It might be difficult to align eligibility and treatment assignment when eligibility criteria are met at multiple times (e.g., in our vaccine study).

- **Two potential solutions**:

    1. Choose **one eligible time at random**.

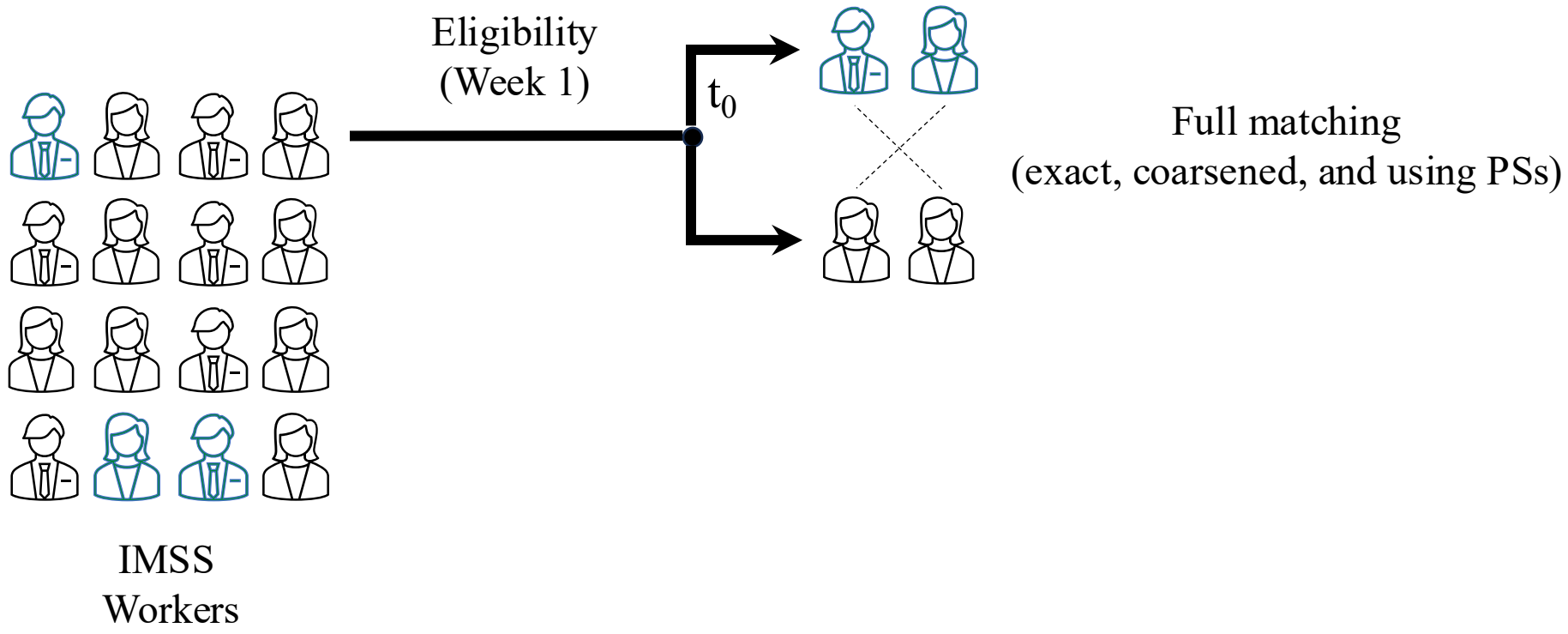
    2. **Use every eligible time** (emulate a **sequence of target trials**, each with different start of follow-up), estimate the effects in each, and then pool the results.

        ✓ Each trial has the same eligibility criteria and follow-up protocol.

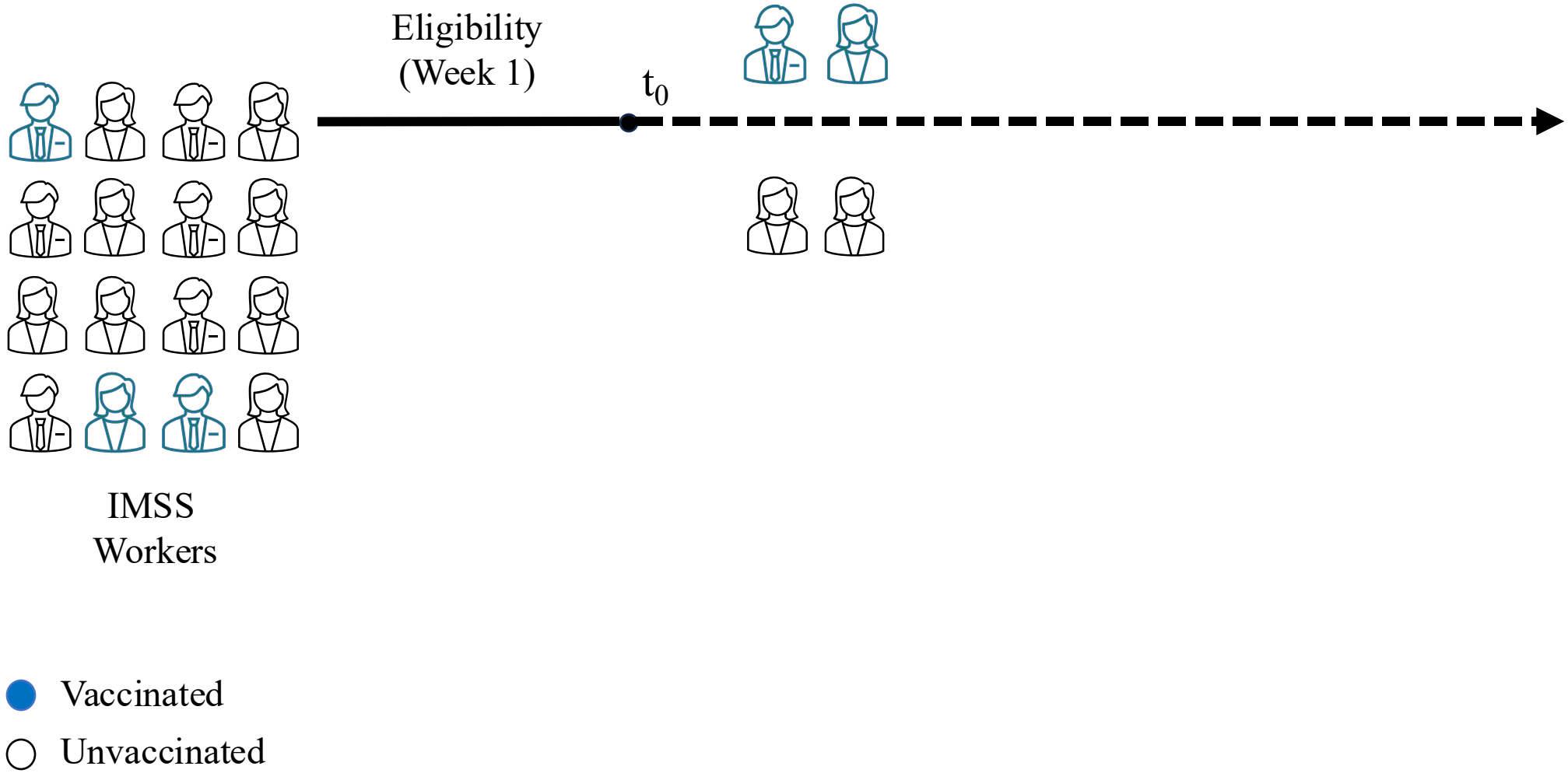
        ✓Participants can be enrolled in more than one trial and more than one treatment strategy.

        ✓Statistically efficient.

Emulating **time zero** and **follow-up**



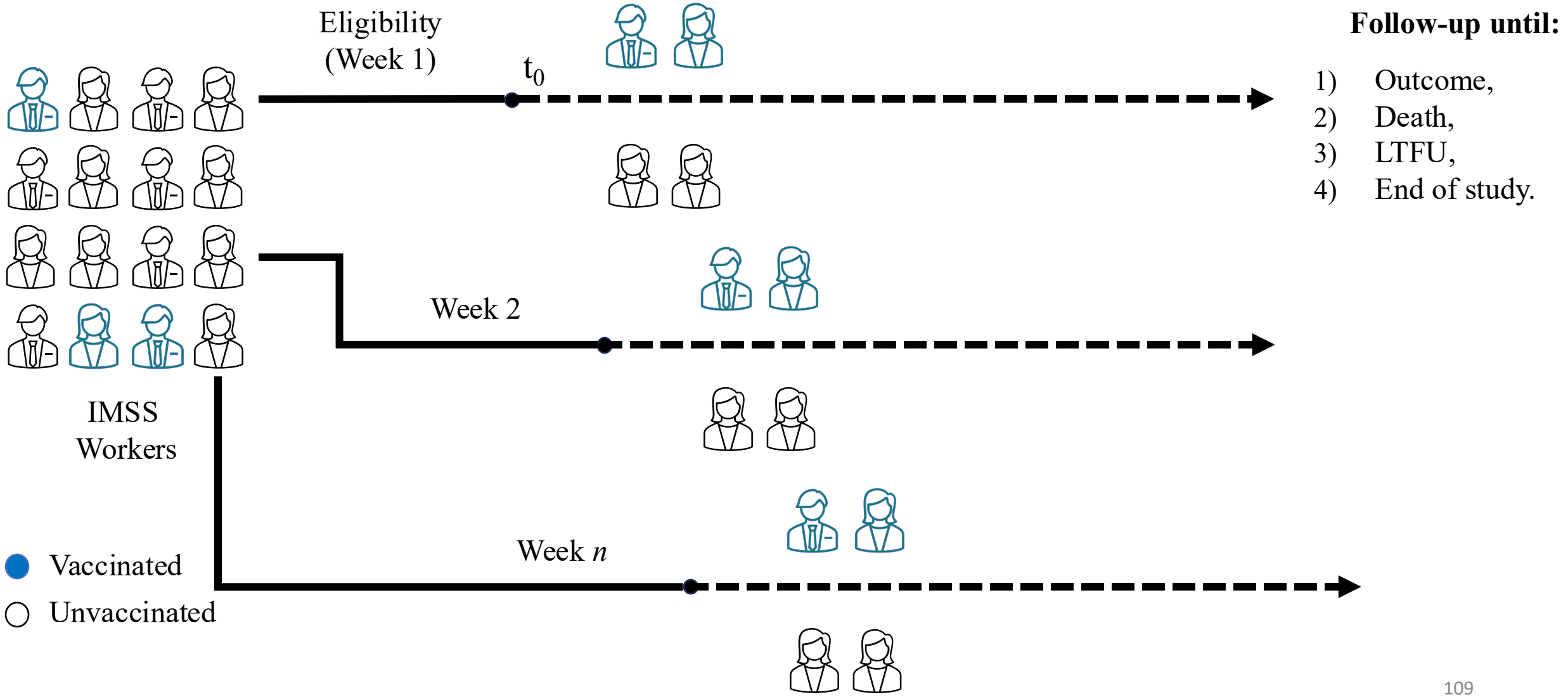
Emulating **time zero** and **follow-up**



Follow-up until:

1) Outcome,
2) Death,
3) LTFU,
4) End of study.

Emulating **time zero** and **follow-up**



Outcomes

McGill

Components	TT Specification	TT Emulation
Time-zero and follow-up	Begins at treatment assignment (time zero) and continues until outcome of interest, loss to follow-up, death, treatment strategy discontinuation, or administrative censoring (2021-Jun-21), whichever first.	Same
Outcomes	6-month risk of: a) SARS-CoV-2 infection; b) symptomatic infection; c) Covid-19-related hospitalization; d) Covid-19-related death.	Assessment of SARS-CoV-2 infection is not part of usual care. The rest is the same.
Causal Contrasts		
Statistical Analysis		

Causal contrasts

McGill

Components	TT Specification	TT Emulation
Time-zero and follow-up	Begins at treatment assignment (time zero) and continues until outcome of interest, loss to follow-up, death, treatment strategy discontinuation, or administrative censoring (2021-Jun-21), whichever first.	Same
Outcomes	6-month risk of: a) SARS-CoV-2 infection; b) symptomatic infection; c) Covid-19-related hospitalization; d) Covid-19-related death.	Assessment of SARS-CoV-2 infection is not part of usual care. The rest is the same.
Causal Contrasts	1. Intention-to-treat (ITT) effect. 2. Per-protocol (PP) effect.	The ITT may be uninformative. The rest is the same.
Statistical Analysis		

Statistical analysis

McGill

Components	TT Specification	TT Emulation
Time-zero and follow-up	Begins at treatment assignment (time zero) and continues until outcome of interest, loss to follow-up, death, treatment strategy discontinuation, or administrative censoring (2021-Jun-21), whichever first.	Same
Outcomes	6-month risk of: a) SARS-CoV-2 infection; b) symptomatic infection; c) Covid-19-related hospitalization; d) Covid-19-related death.	Assessment of SARS-CoV-2 infection is not part of usual care. The rest is the same.
Causal Contrasts	1. Intention-to-treat (ITT) effect. 2. Per-protocol (PP) effect.	The ITT may be uninformative. The rest is the same.
Statistical Analysis	ITT: Use pooled logistic regression models to estimate the 6-month risks for both groups. Then, calculate the risk differences and risk ratios along with their 95% bootstrap confidence intervals. PP: The same procedure as above, but apply time-varying inverse probability of treatment and censoring weights to the models. Individuals are censored if they do not adhere to their assignment.	Same.

Time-varying IPW for censoring due to change of treatment strategy

McGill

$$SW_k^A = \prod_{m=0}^{k} \frac{f\left(A_m | \bar{A}_{m-1}\right)}{f\left(A_m | \bar{A}_{m-1}, \bar{L}_m\right)}$$

where:

- A_m : an indicator of treatment at time m,
- $\bar{A}_{m\text{-}1}$: treatment history.
- $\bar{L}_m$: confounder history.

Hernán MA, et al. Chapter 21. G-methods for Time-Varying Treatments. Causal Inference: What If: Chapman & Hall/CRC; 2020: 269-296.

Time-varying IPW for censoring due to LTFU

McGill

$$W_{k+1}^{\bar{C}} = \begin{cases} \prod_{m=1}^{k+1} \frac{\Pr(C_m = 0 | \bar{A}_{m-1}, V, \bar{C}_{m-1} = 0)}{\Pr(C_m = 0 | \bar{A}_{m-1}, \bar{L}_{m-1}, \bar{C}_{m-1} = 0)} & \text{if } C_{k+1} = 0 \\ 0 & \text{if } C_{k+1} = 1 \end{cases}$$

where:

- $C_m = 0$: an indicator of being uncensored at time m,
- $\bar{A}_{m\text{-}1}$: treatment history.
- $\bar{L}_{m\text{-}1}$: confounder history.
- V: baseline values of covariates.

Hernán MA, et al. Chapter 21. G-methods for Time-Varying Treatments. Causal Inference: What If: Chapman & Hall/CRC; 2020: 269-296.

IPW Pooled Logistic Regression Model for the Discrete-Time Hazards

McGill

$$\text{logit Pr}\left[Y_{k+1} = 1 | \bar{A}_k, Y_k = 0, C_{k+1} = 0\right] = \theta_{0,k} + \theta_1 A + \theta_2 L + \theta_3 A \times f(k)$$

where:

- Y_{k+1}: the outcome in interval $k + 1$.
- $\bar{A}_k$: treatment history.
- $C_{k+1} = 0$: an indicator of being uncensored at time $k + 1$,
- A: a time-fixed indicator for being assigned to "always treat" ($A = 1$) or "never treat" ($A = 0$) at time zero.
- L: Baseline values of confounders.
- $A \times f(k)$: a (vector) of product terms between treatment and functions of time (allows for a time-varying hazard ratio)

Hernán MA, et al. Chapter 21. G-methods for Time-Varying Treatments. Causal Inference: What If: Chapman & Hall/CRC; 2020: 269-296.

Final considerations

- Emulating a target trial helps **prevent biases due to flawed study designs** (e.g., immortal time bias, selection bias due to prevalent users).

- However, this approach **does not eliminate bias due to data issues** (e.g., uncontrolled or unmeasured confounding or information bias).

- In these situations, **sensitivity analyses** (e.g., negative controls) and **quantitative bias analyses** may be helpful.

McGill | Department of Epidemiology, Biostatistics and Occupational Health

6. An **introduction to the test-negative study design (TND)**.

Introduction

- The TND is a widely used and increasingly popular study design for **estimating vaccine effectiveness** (i.e., the effect of vaccination on infection-related outcomes).

- **Influenza vaccine composition** is **updated twice a year** based on global surveillance data on circulating strains.

- Therefore, **influenza VE may vary** from season to season and year to year.

- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Introduction

- Researchers **needed a design to efficiently estimate influenza VE**.

- Some authors proposed using **individuals seeking medical care** for influenza-like illness (ILI) symptoms at surveillance centers or hospitals as the study population.

- **Potential advantages**: relatively **inexpensive and quick** to implement, **reduces** the possibility of **confounding by differential health-seeking behaviors** (HSBs), and **reduces the potential for misclassification** of infection status.

- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

TND theoretical basis

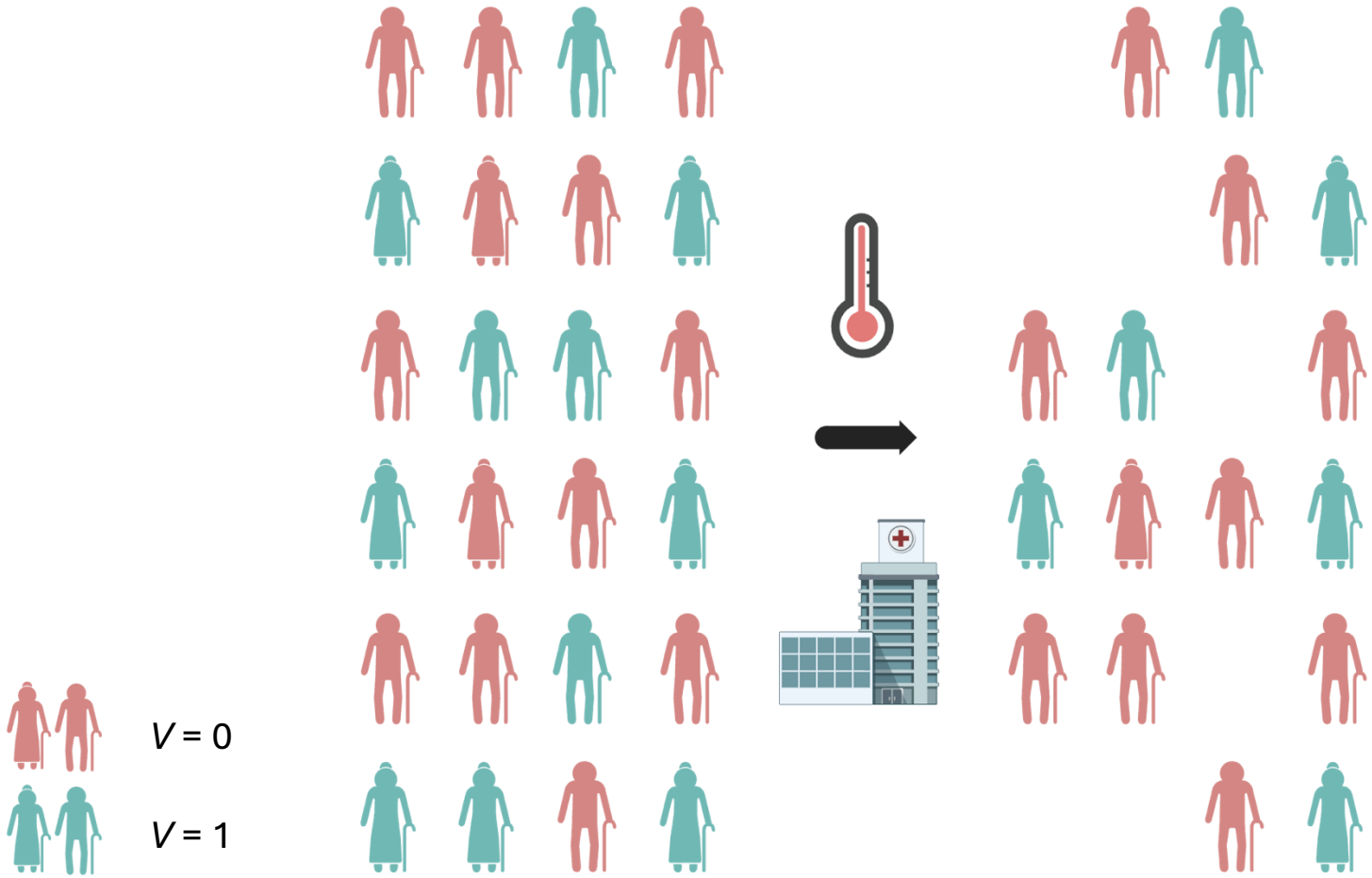
Target population (e.g., IMSS workers between 2020-Dec and 2021-Dec)

$V = 0$

$V = 1$

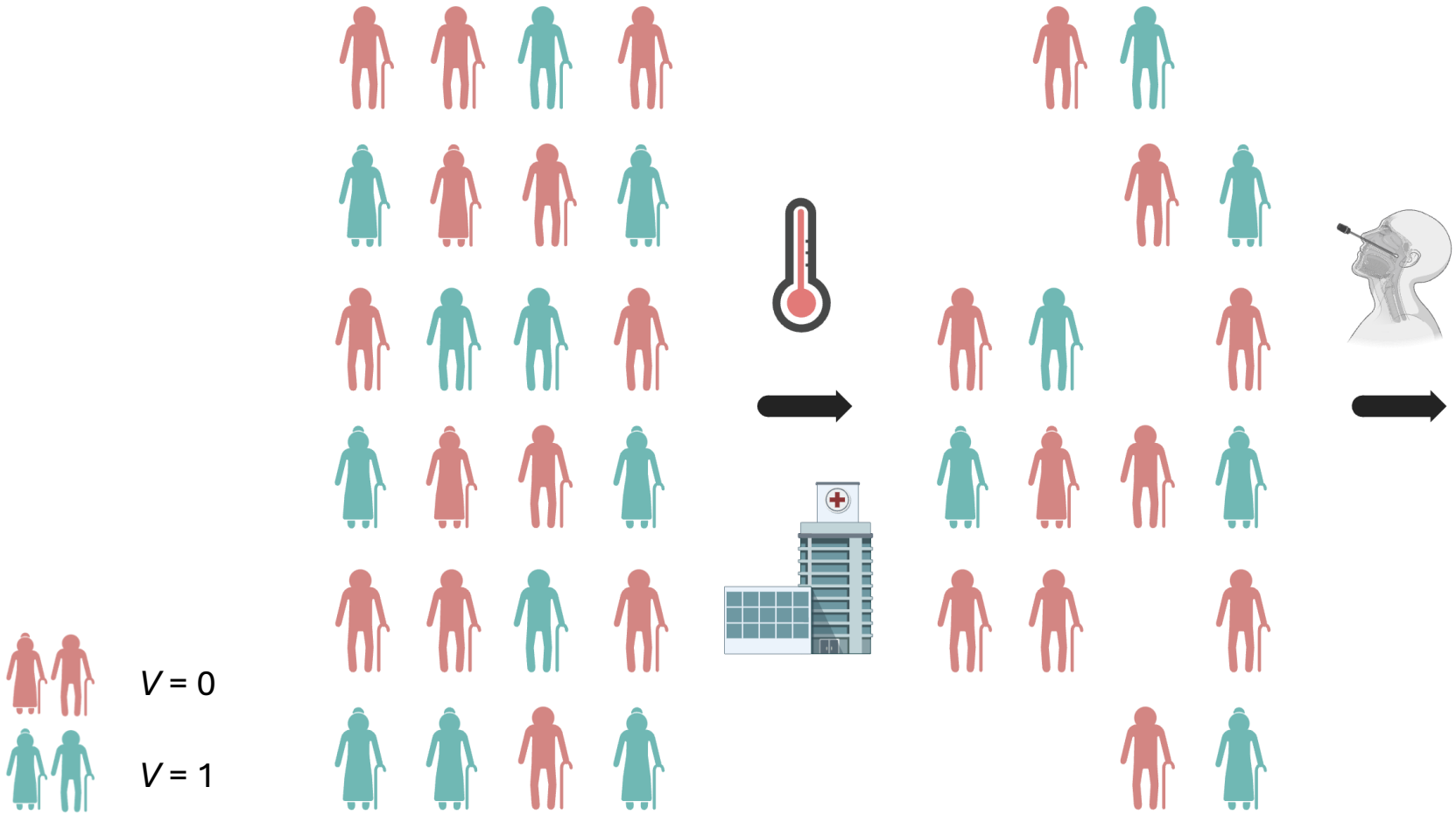
- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

TND theoretical basis



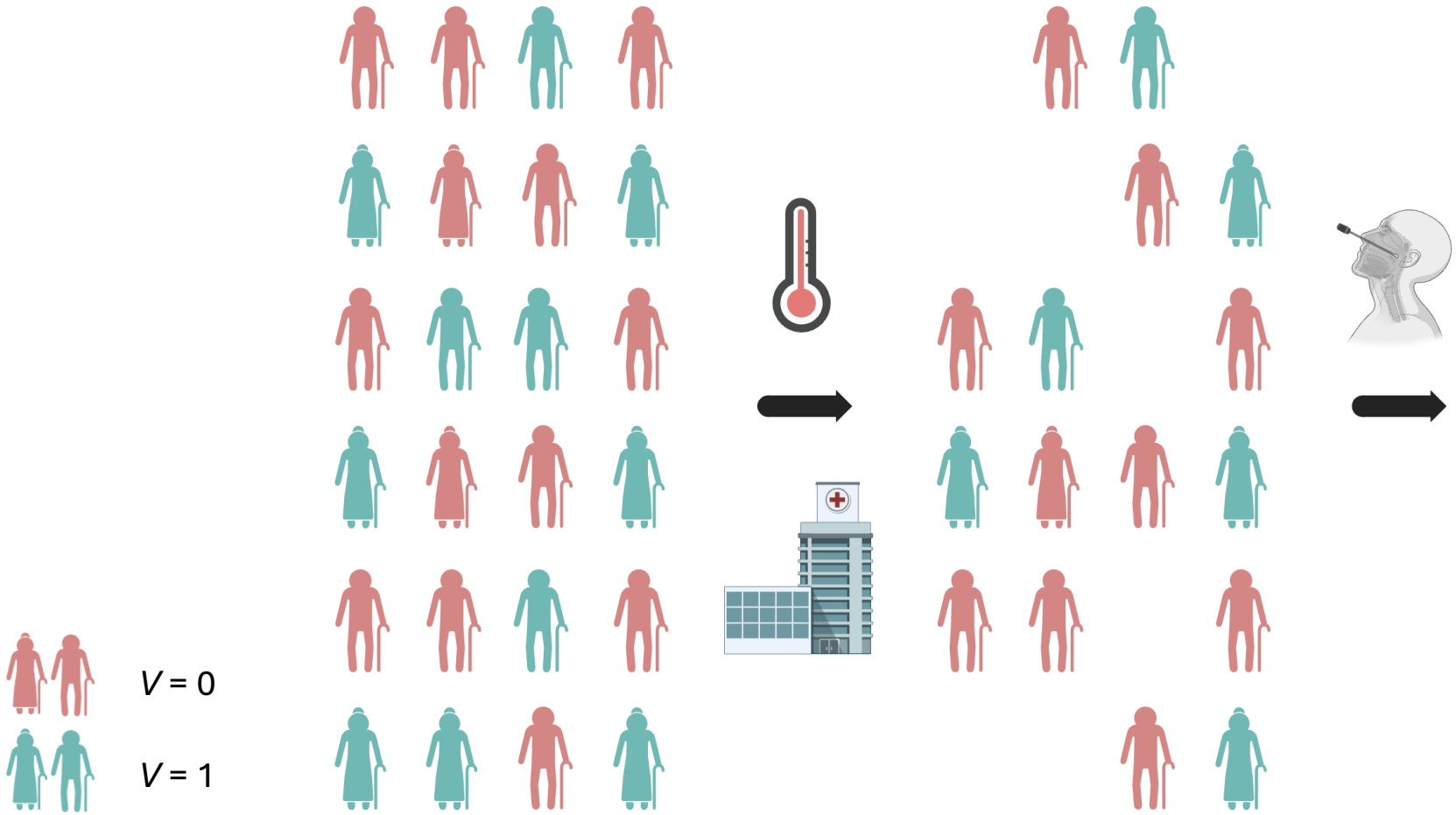
- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

TND theoretical basis



- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

TND theoretical basis

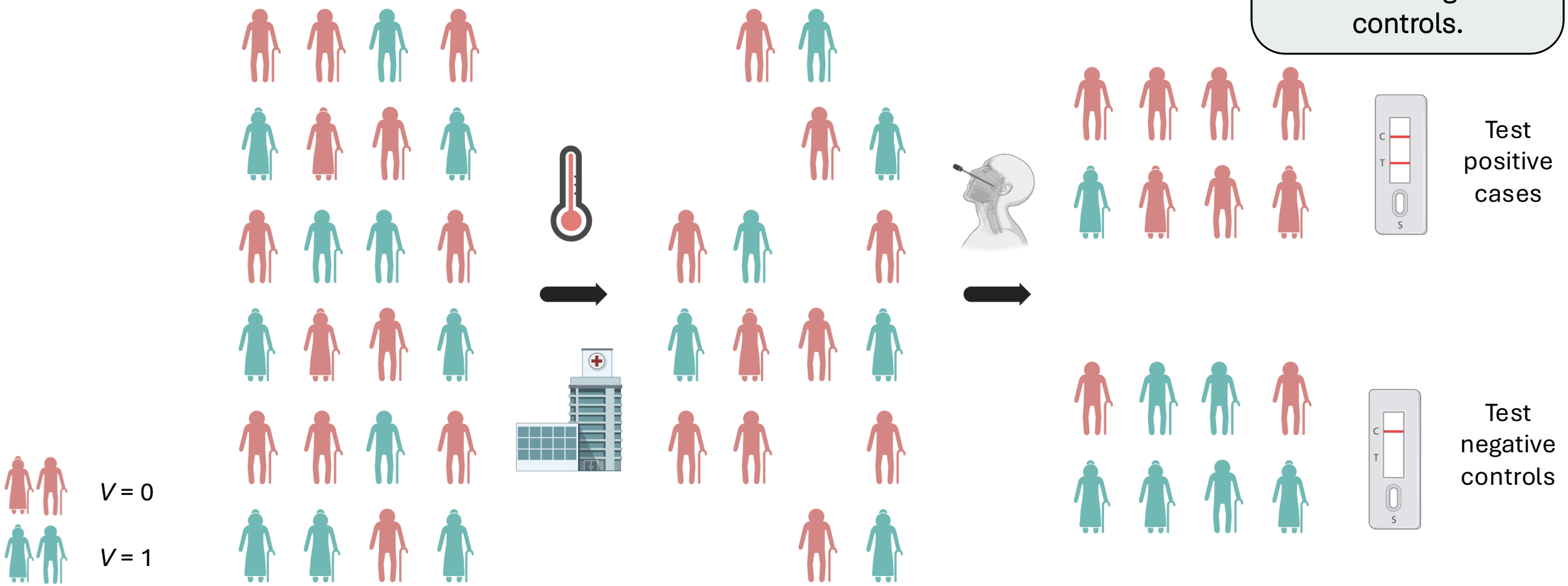


- **Selection of individuals** who underwent diagnostic testing for the vaccine-preventable disease due to symptoms of infection.
- **Collection of data** on patient vaccination history and confounder profiles.

- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

TND theoretical basis

Classification of the study population into **test-positive cases** and **test-negative controls.**

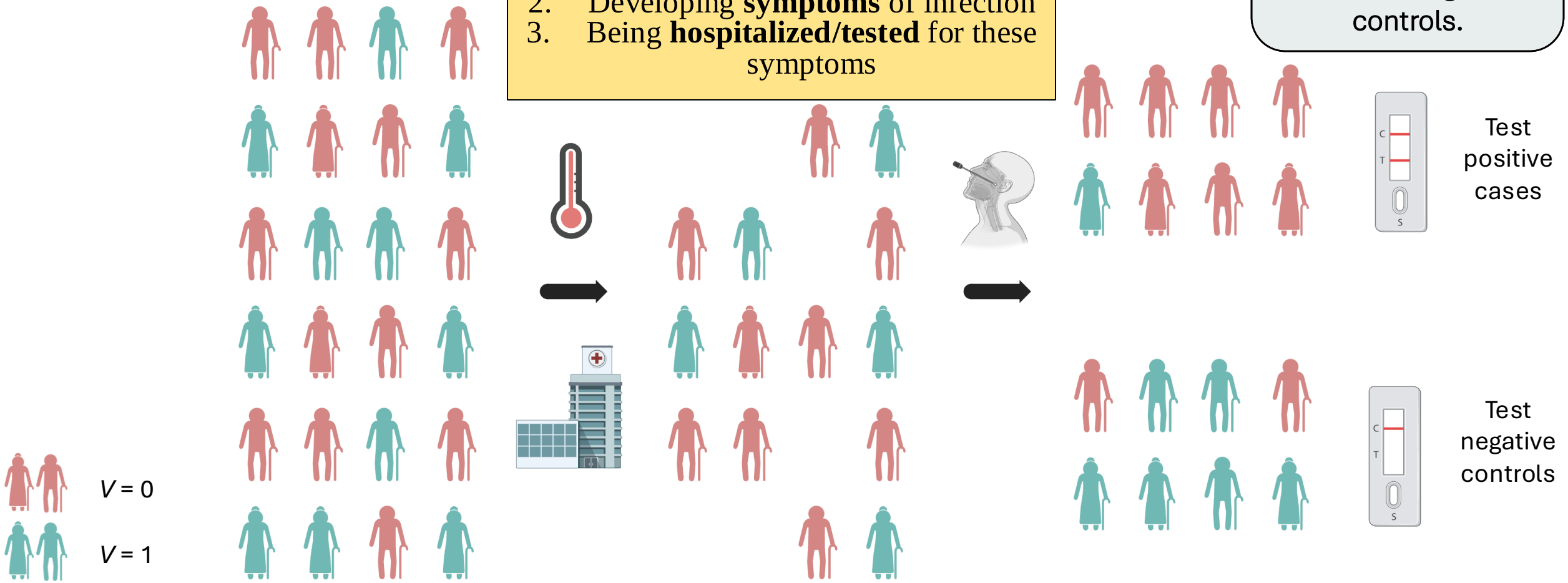


- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

In the TND, the outcome is a combination of three sequential steps:

1. Getting **infected**
2. Developing **symptoms** of infection
3. Being **hospitalized/tested** for these symptoms

Classification of the study population into **test-positive cases** and **test-negative controls.**



- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

The TND may also be implemented with EHRs

Table 1. Two Types of Studies Which Use a Test-Negative Comparison Group

Setting	Surveillance Program	Diagnostic Records/ Electronic Health Records
Data collection	Prospective	Retrospective
Sampling	Patients with a common clinical case definition recruited in a clinical setting and tested for influenza virus	All patients tested for influenza virus for diagnostic purposes
		No common clinical case definition
Case group	Patients testing positive for influenza virus	Cases testing positive for influenza virus identified from medical/laboratory records
	Case status unknown at time of recruitment	
Comparison group	Patients testing negative for influenza virus	Patients testing negative for influenza virus
Vaccination status	Prospectively ascertained and verified	Obtained from medical/registry records

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Statistical analysis

$$\text{logit}\,\Pr[Y = 1|V, L] = \theta_0 + \theta_1 V + \theta_2 L$$

where:

- Y: case status $\begin{cases} 1, & \text{test positive cases.} \\ 0, & \text{test negative controls.} \end{cases}$
- V: vaccination status.
- L: confounding variables.

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

- Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. Vaccine, 31(17), 2165–2168.
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

The control exchangeability assumption

McGill

- The probability of hospitalization/testing due to symptoms of a non-vaccine-targeted infection is independent of vaccination status conditional on covariates.

$$\{I = 1, W = 1, H = 1\} \perp\!\!\!\perp V | C$$

- This means that, in the test-negative control group, vaccinated and unvaccinated individuals are exchangeable and can therefore represent the source population for test-positive cases.

- Schnitzer ME. Estimands and Estimation of COVID-19 Vaccine Effectiveness Under the Test-Negative Design: Connections to Causal Inference. Epidemiology. 2022;33(3):325-333.

The control exchangeability assumption

McGill

$$\frac{\Pr(I=2,W=1,H=1\mid Z=z)}{\Pr(I=1,W=1,H=1\mid Z=z)} \Big/ \frac{\Pr(I=2,W=1,H=1\mid Z=z_0)}{\Pr(I=1,W=1,H=1\mid Z=z_0)} \quad (1)$$

$$\frac{\{(I=2,W=1,H=1\mid Z=z)\ *(I=1,W=1,H=1\mid Z=z_0)\}}{\{(I=2,W=1,H=1\mid Z=z_0)\ *(I=1,W=1,H=1\mid Z=z)\}} \quad (2)$$

$$\Pr(I=1,W=1,H=1\mid Z=z)=\Pr(I=1,W=1,H=1\mid Z=z_0) \quad (3)$$

$$\psi_{\text{cRR}}=\frac{\Pr(I=2,W=1,H=1\mid Z=z)}{\Pr(I=2,W=1,H=1\mid Z=z_0)} \quad (4)$$

- Schnitzer ME. Estimands and Estimation of COVID-19 Vaccine Effectiveness Under the Test-Negative Design: Connections to Causal Inference. Epidemiology. 2022;33(3):325-333.

IP weighted logistic regression model in the TND

McGill

- We can **recover the marginal RR using IPTW**.
- To obtain IP weights, the **propensity score model must be fitted to the test-negative control population**.
- This ensures that the distribution of confounders within the test-negative control group is representative of the distribution in the source population for both vaccinated and unvaccinated individuals.
- Using IPTW **helps avoiding model misspecification** due to effect measure modification.

- Schnitzer ME. Estimands and Estimation of COVID-19 Vaccine Effectiveness Under the Test-Negative Design: Connections to Causal Inference. Epidemiology. 2022;33(3):325-333.

TND IPTW: R Code

```r
# Estimation of IP weights via a logistic regression model for a TND study
## PS model
TND_ps_mod <- glm(vac ~ confounders, family = binomial(),  data = tnd_data_controls)

## IP weights
tnd_data$ps  <-  predict(TND_ps_mod, type="response", newdata= tnd_data)
tnd_data$ipw <- ifelse(dat$V == 1, 1/ps, 1 /(1 -ps))

## IP weighted LR model
fit<- glm(Y~ V, family= binomial, weights = ipw, data= tnd_data)

## Results
mRR = exp(fit$coefficients[2])
se = sqrt(vcovHC(fit)[2,2])
```

Interpretation

$$VE = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

Interpretation

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

$$\text{VE} = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$\text{VE} = \left(\frac{\text{Risk}_{v=0}}{\text{Risk}_{v=0}} - \frac{\text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

Interpretation

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

$$\text{VE} = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$\text{VE} = \left(\frac{\text{Risk}_{v=0}}{\text{Risk}_{v=0}} - \frac{\text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$\text{VE} = \left(1 - \frac{\text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

Interpretation

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

$$\text{VE} = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$\text{VE} = \left(\frac{\text{Risk}_{v=0}}{\text{Risk}_{v=0}} - \frac{\text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$\text{VE} = \left(1 - \frac{\text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

$$\text{VE} = (1 - \text{RR}) \times 100$$

Interpretation **example**

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

$$\text{VE} = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

Interpretation **example**

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

$$\text{VE} = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

- $\text{Risk}_{v=0}$: 50%
- $\text{Risk}_{v=1}$: 25%

Interpretation **example**

$$VE = (1 - (\text{adjusted OR})) \times 100$$

VE: vaccine effectiveness.
OR: odds ratio.

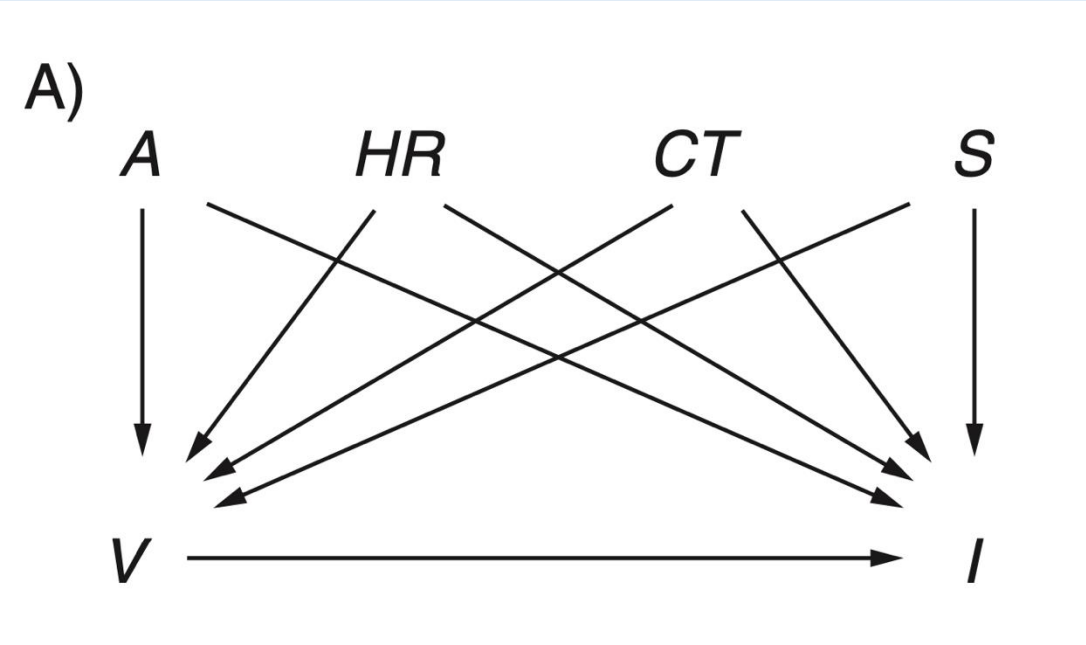
$$\text{VE} = \left(\frac{\text{Risk}_{v=0} - \text{Risk}_{v=1}}{\text{Risk}_{v=0}}\right) \times 100$$

- $\text{Risk}_{v=0}$: 50%
- $\text{Risk}_{v=1}$: 25%

$$\text{VE} = \left(\frac{50\% - 25\%}{50\%}\right) \times 100 = 50\%$$

7. Potential sources of bias in **test-negative design** studies.

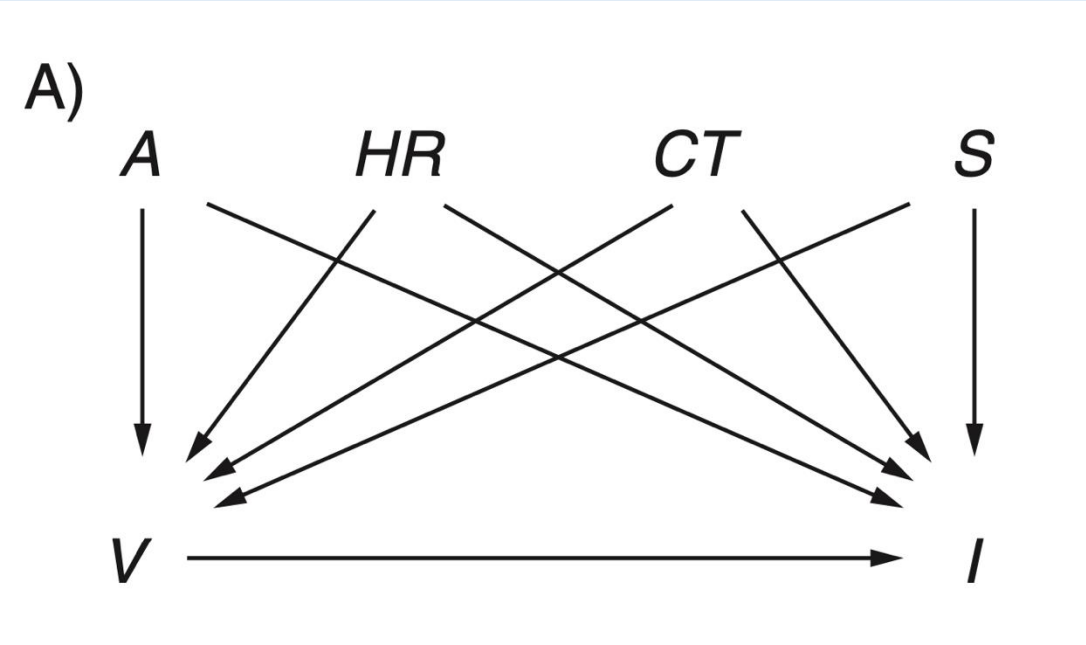
Confounding bias

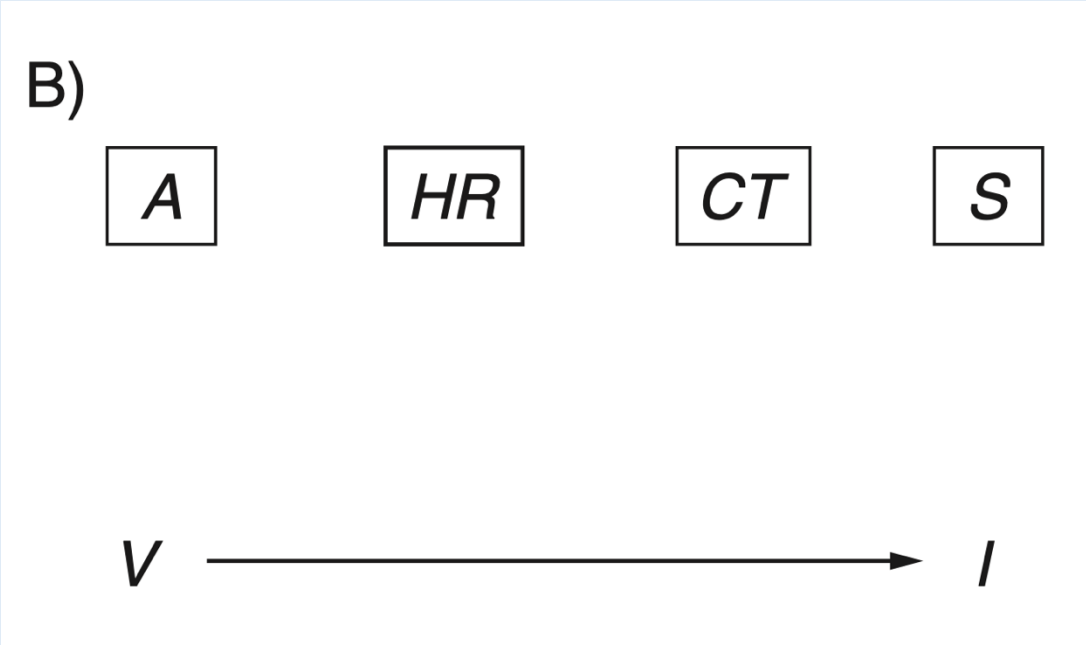


- ***A***: age.
- ***HR***: high-risk status.
- ***CT***: calendar time.
- ***S***: sex.

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

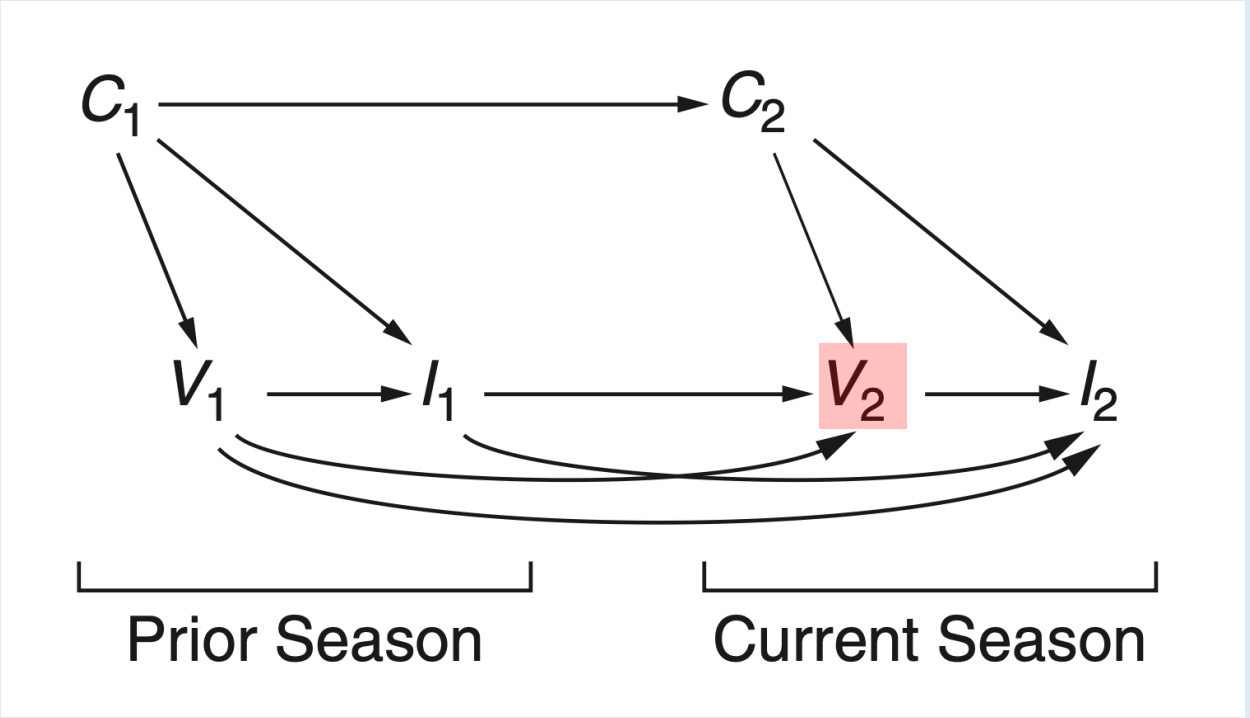
Confounding at baseline





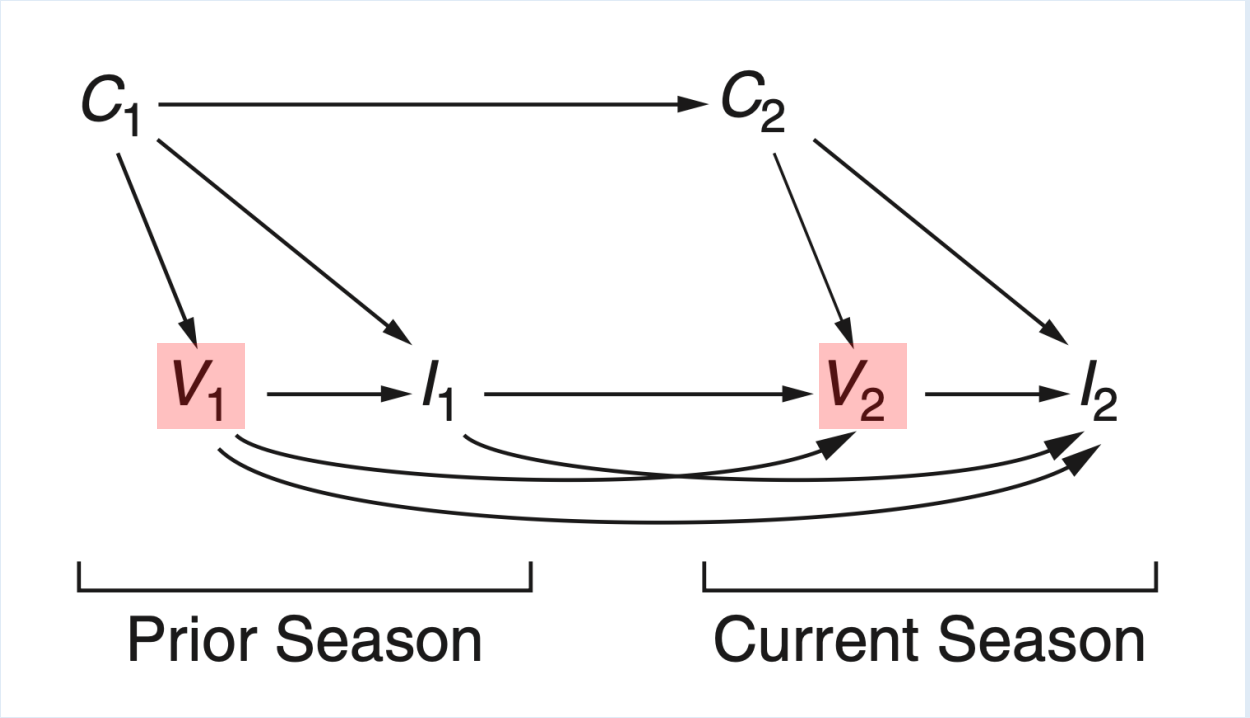
- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Confounding at baseline



- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

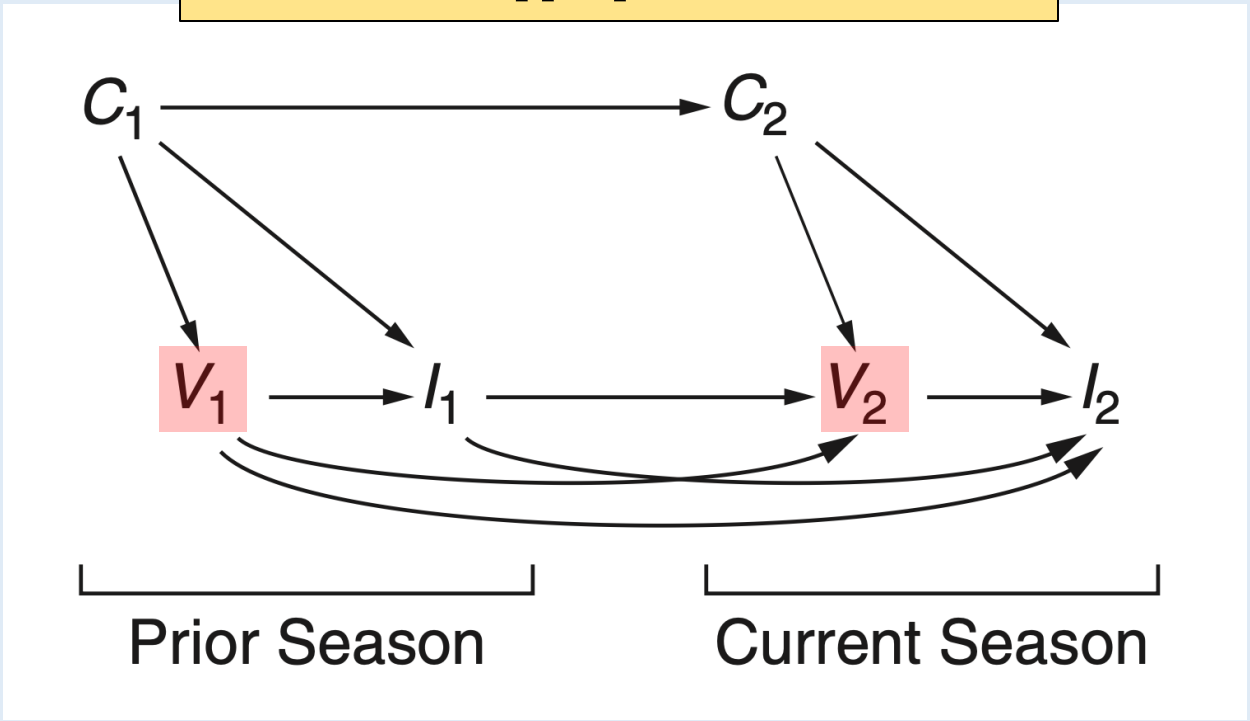
Time varying confounding



- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

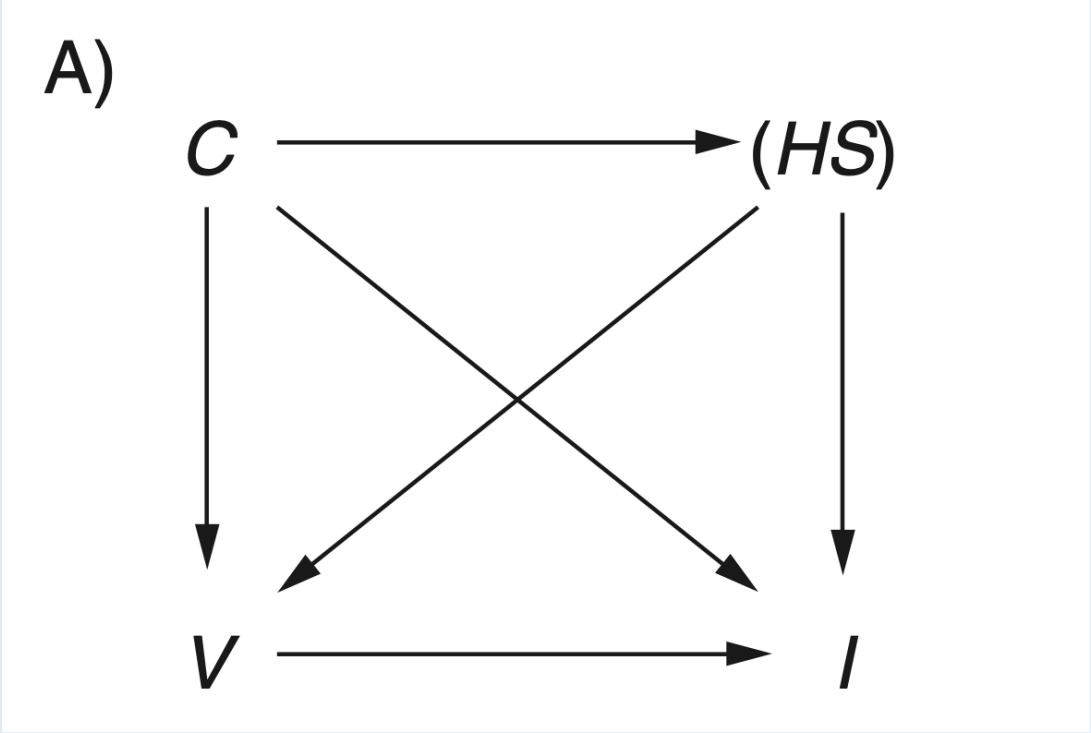
Tir ing

In the **time-varying setting** (effects of repeated vaccination), infection (*I*) acts as both a **confounder and a mediator**. This leads to "**treatment-confounder feedback**" and consequently g methods may be more appropriate.



- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

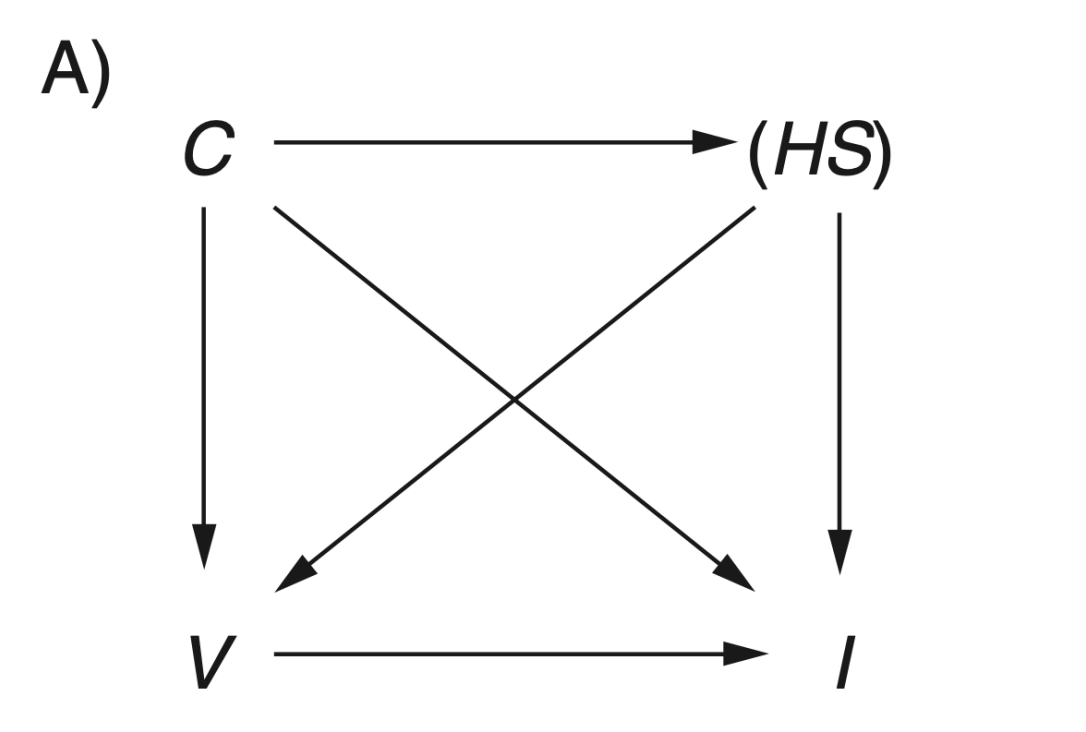
Confounding by HSBs

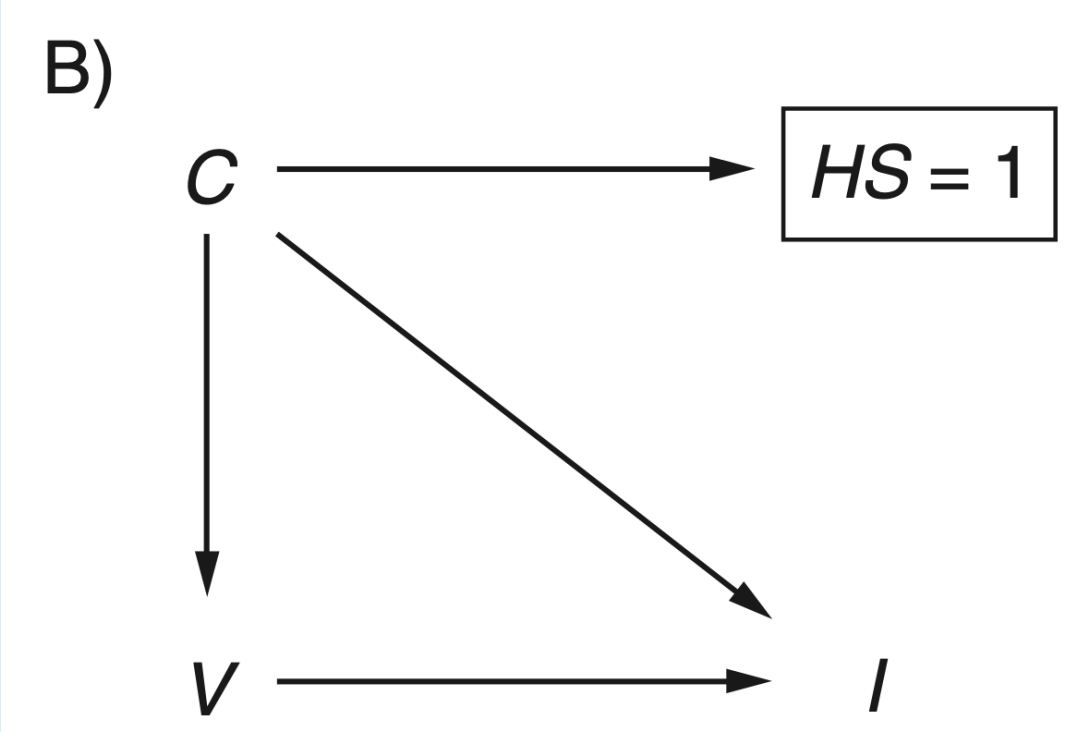


- ***C***: other confounders (i.e., age, high-risk status, sex, and calendar time).
- ***HS***: health and healthcare seeking behaviours.
- ***V***: vaccination.
- ***I***: infection.

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

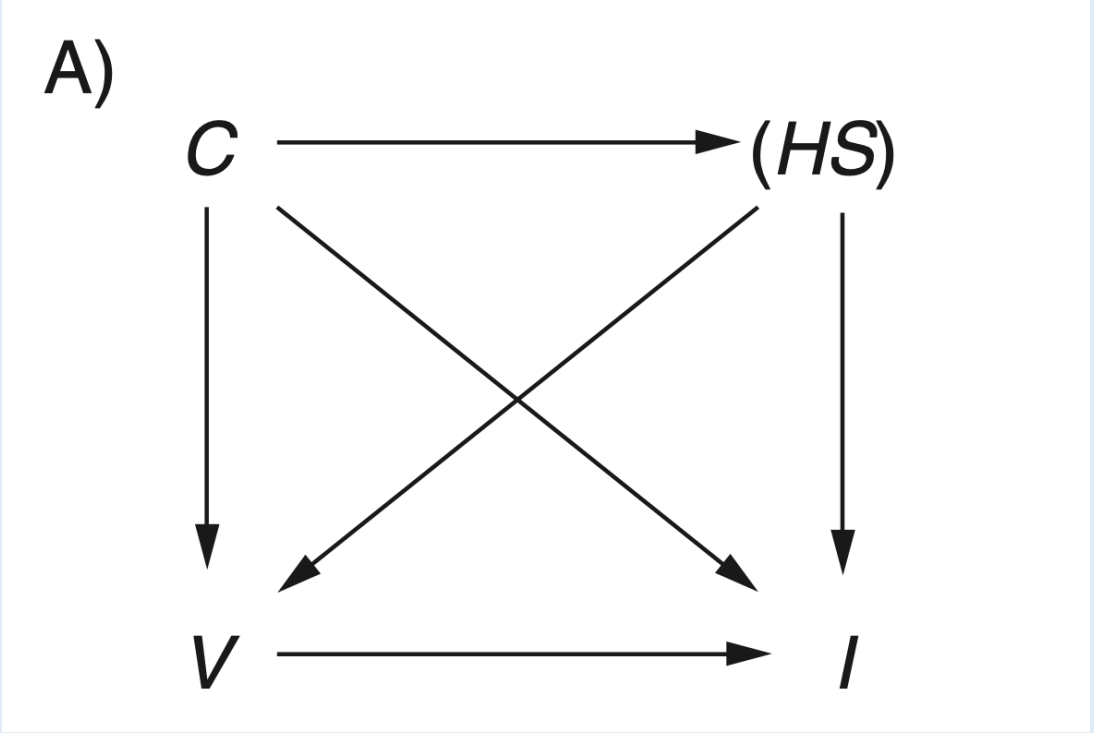
Confounding by HSBs

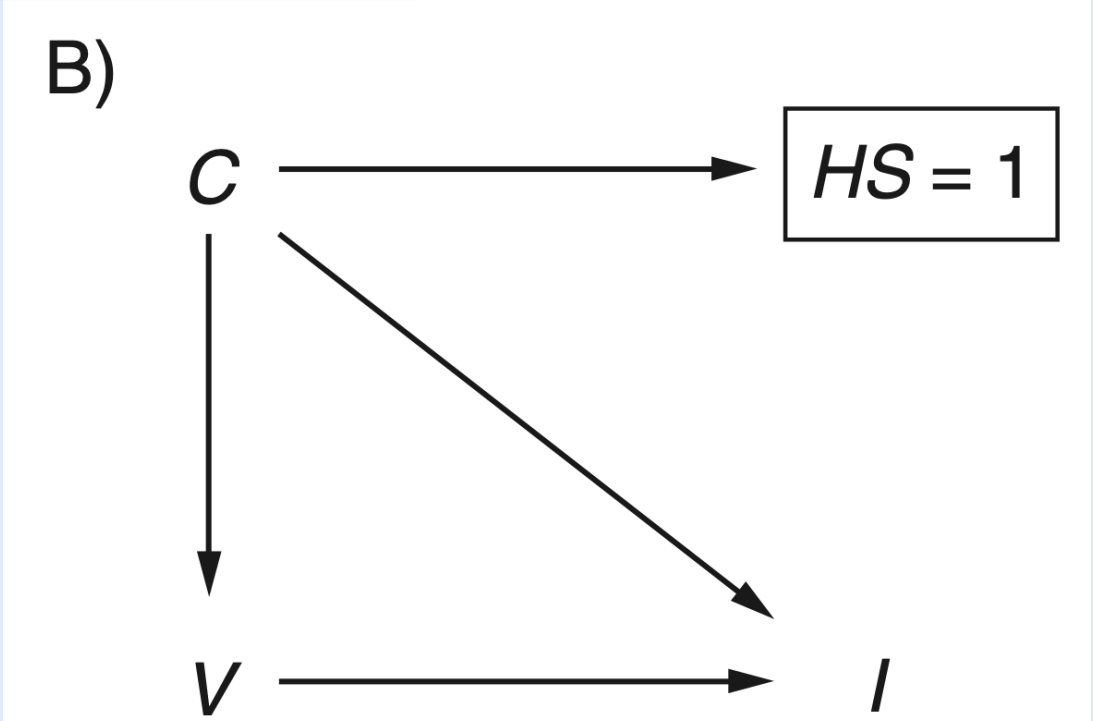




- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

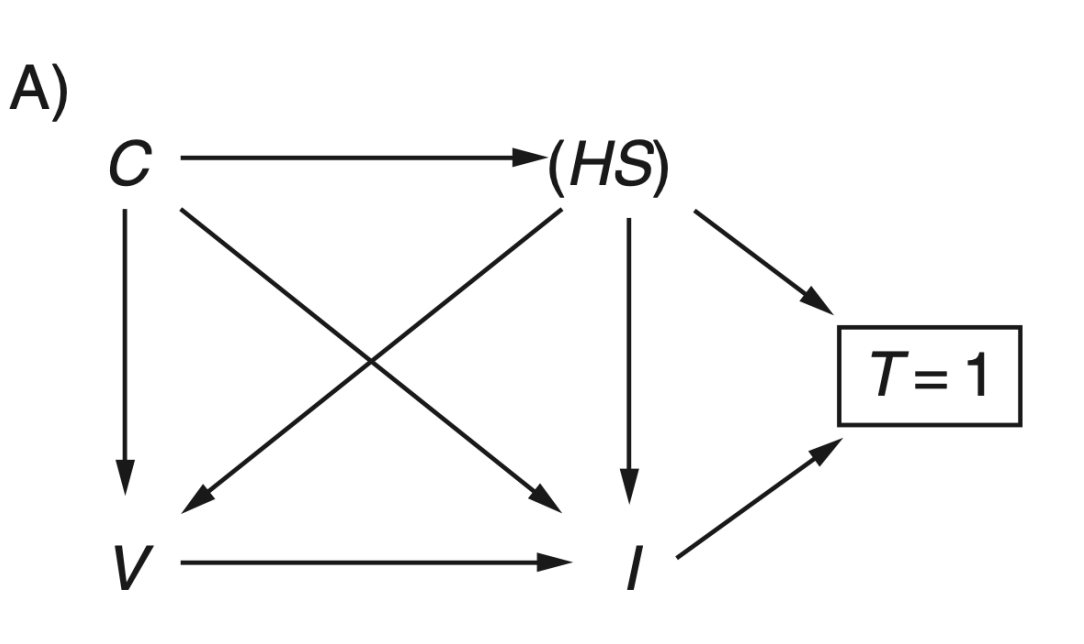
Under the assumption of **deterministic HSBs**, the TND controls for this source of confounding bias (at the **expense of reduced generalizability**).





- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

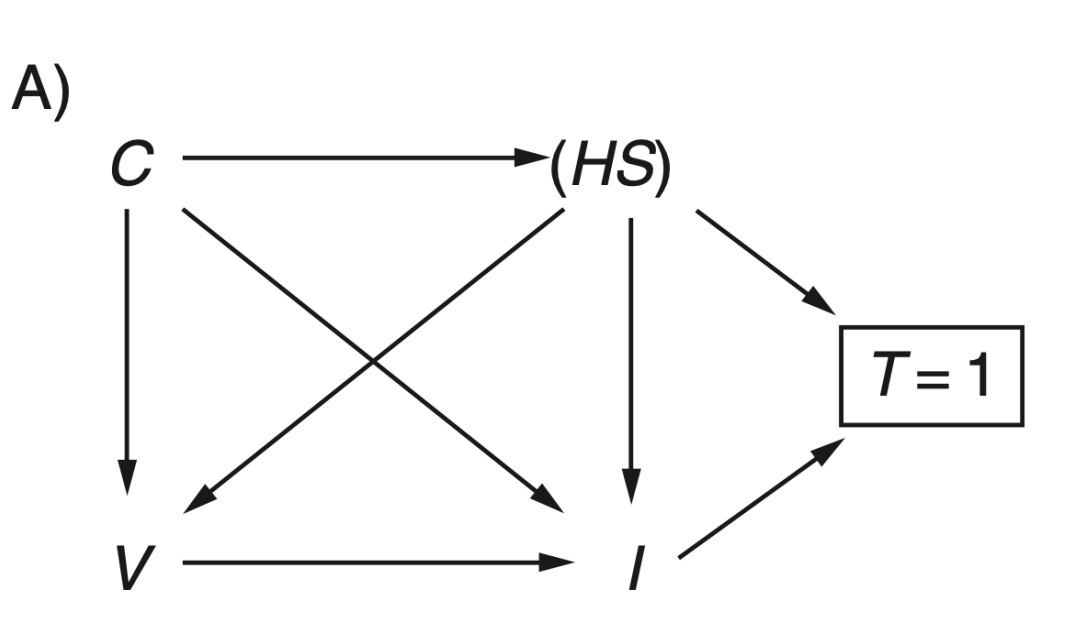
Selection bias

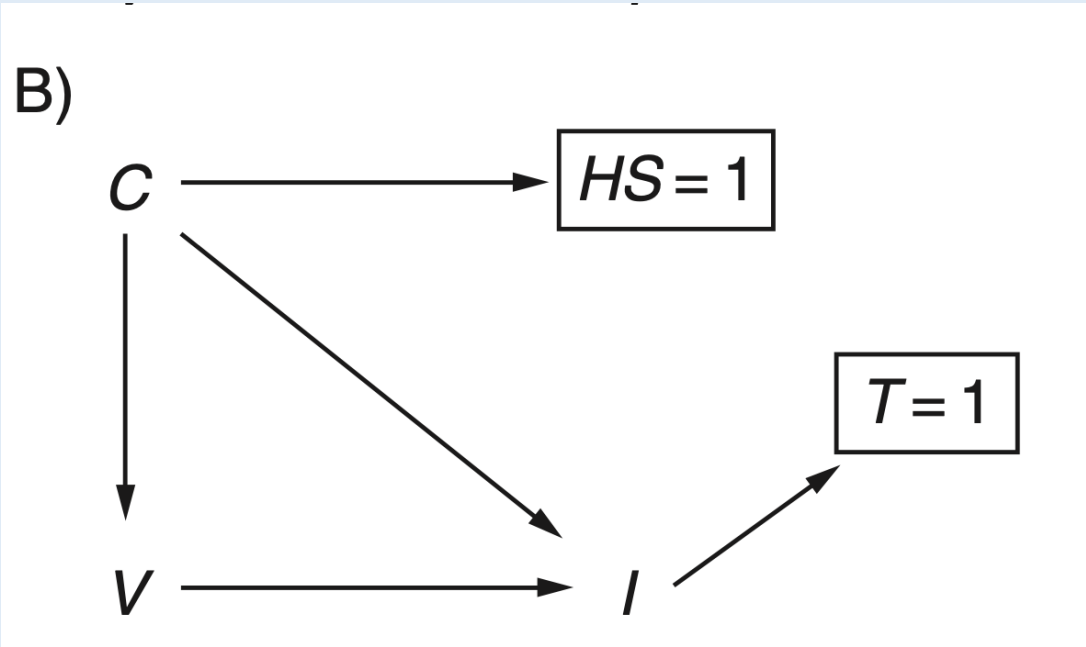


- ***C***: other confounders.
- ***HS***: HSBs.
- ***V***: vaccination.
- ***I***: infection.
- ***T***: testing.

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

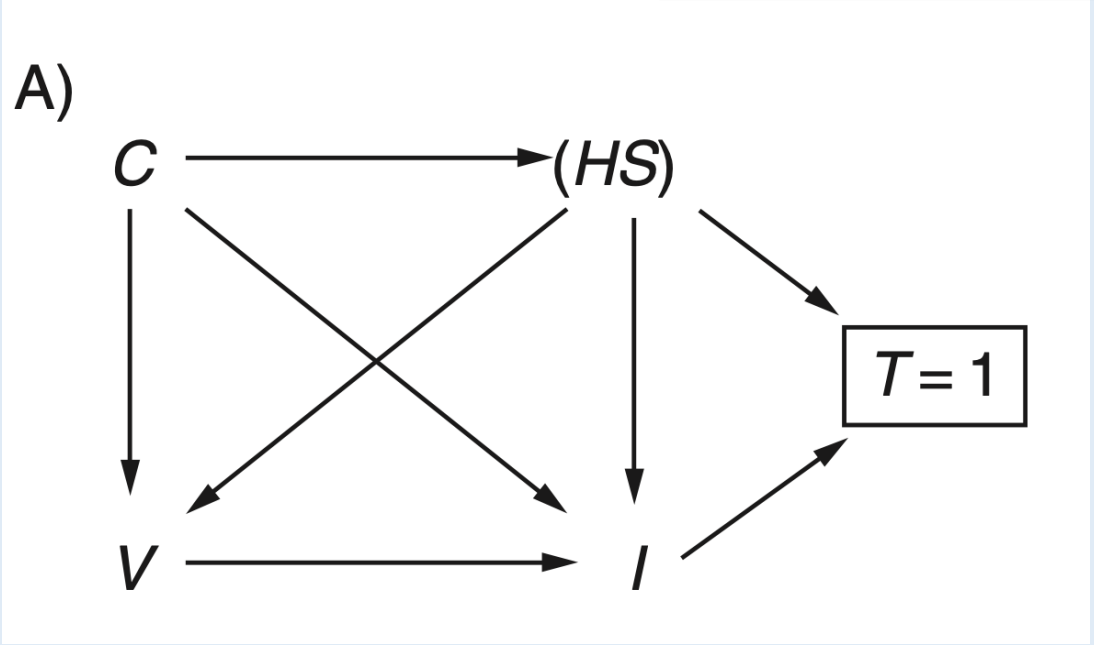
Selection bias

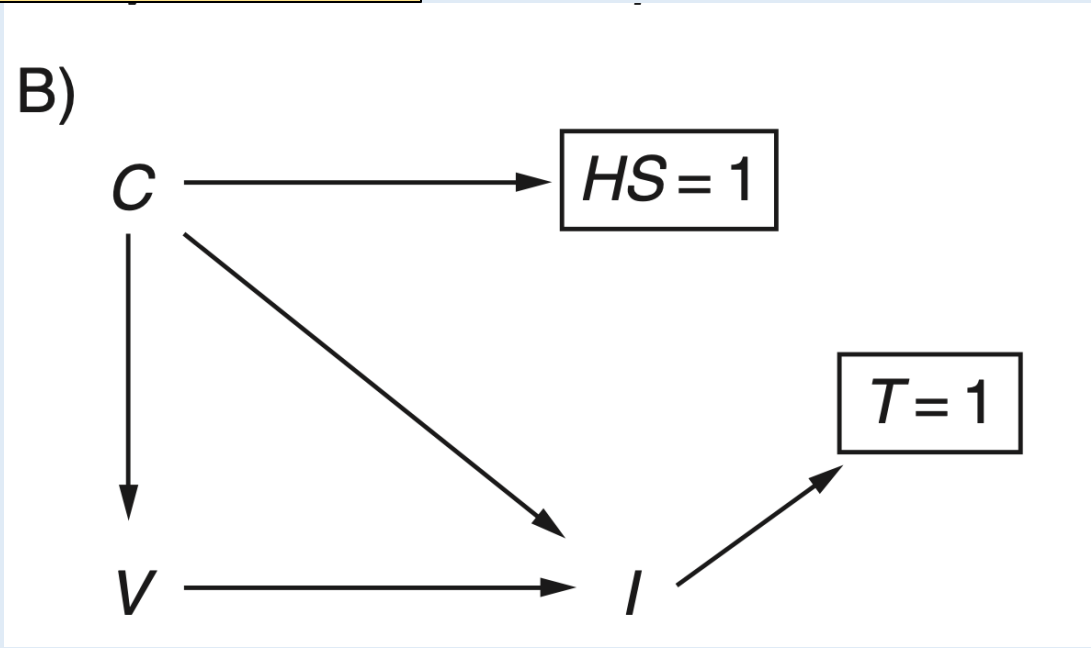




- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Under the assumption of **deterministic HSBs**, the TND controls for this source of confounding **and selection (collider)** bias.





- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Information bias

- The TND is generally **thought to reduce misclassification of infection status** by including only patients with a laboratory test result.

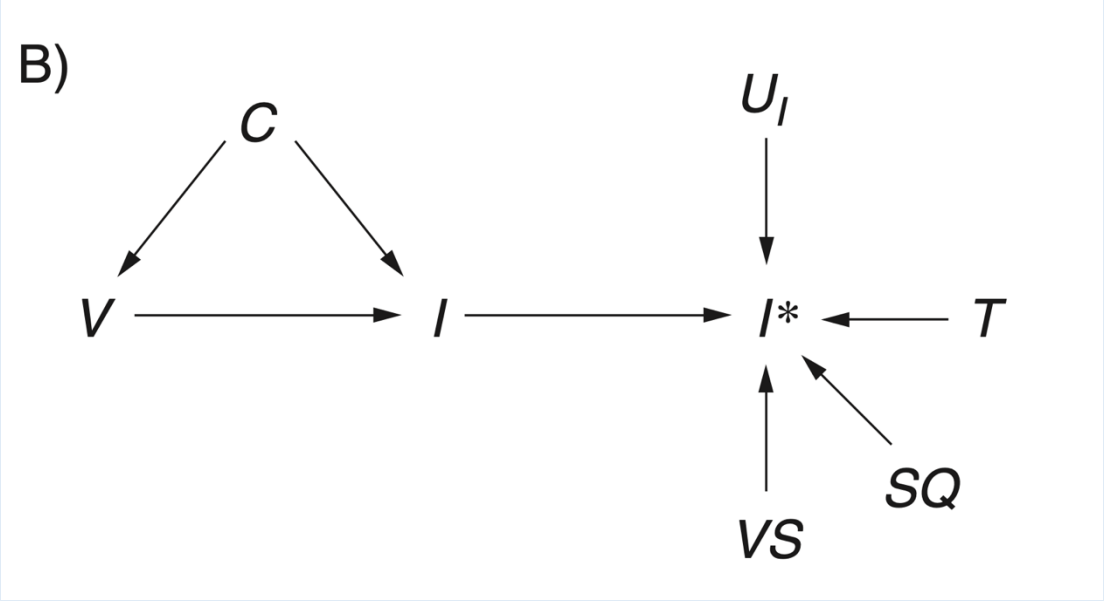
- In cohort or case-control studies, non-cases are assumed to be uninfected, but this is usually not confirmed by testing.

- **Misclassification of infected or uninfected individuals in TND is less likely** because TND studies typically use molecular tests that are highly sensitive and specific.

- However, **this may still occur with less accurate tests**, such as antigen-based tests, **or when other factors** (e.g., timing, absence of symptoms) **affect the diagnostic accuracy of the tests**.

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Outcome misclassification

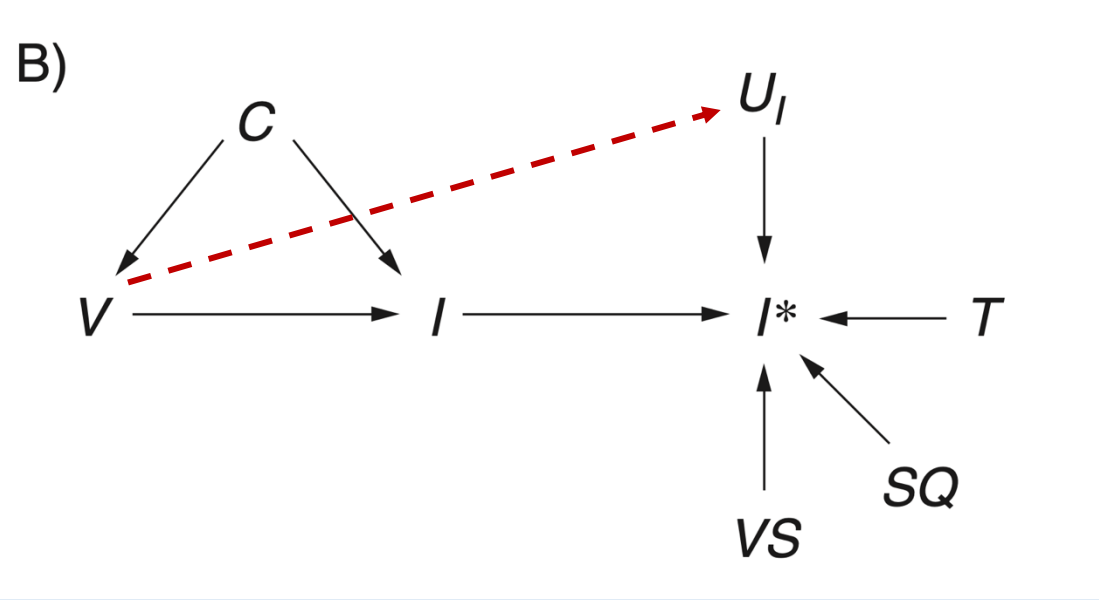


- ***C***: other confounders.
- ***V***: vaccination.
- ***I***: infection.

- ***I****: measured infection.
- ***T***: testing.
- ***SQ***: swab quality.
- ***VS***: viral shedding.
- ***U***: unmeasured factors that influence measurement of infection status

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Outcome misclassification

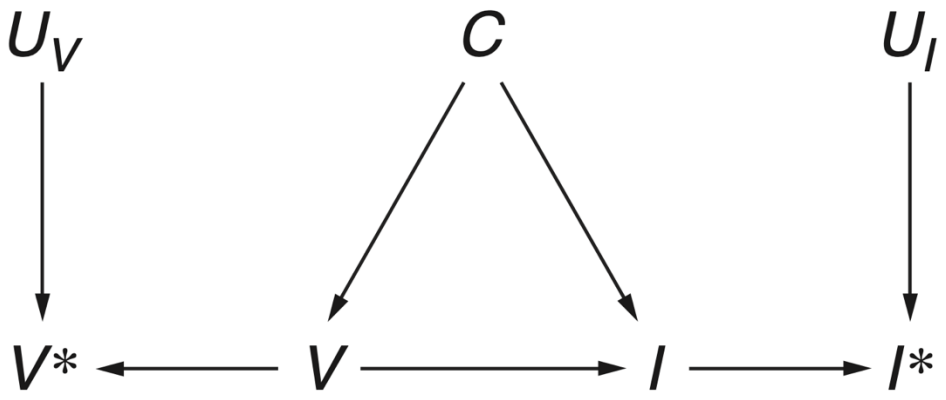


- ***C***: other confounders.
- ***V***: vaccination.
- ***I***: infection.

- ***I****: measured infection.
- ***T***: testing.
- ***SQ***: swab quality.
- ***VS***: viral shedding.
- ***U***: unmeasured factors that influence measurement of infection status

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

Exposure misclassification



- ***C***: other confounders.
- ***V***: vaccination.
- ***I***: infection.

- ***V****: measured vaccination.
- ***I****: measured infection.
- $\boldsymbol{U_I}$ and $\boldsymbol{U_v}$: unmeasured factors that influence measurement of infection/vaccination status

- Sullivan SG, Tchetgen Tchetgen EJ, Cowling BJ. Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness. Am J Epidemiol. 2016;184(5):345-353.

McGill | Department of Epidemiology, Biostatistics and Occupational Health

Thanks

https://ortizbrizuela.github.io/info/