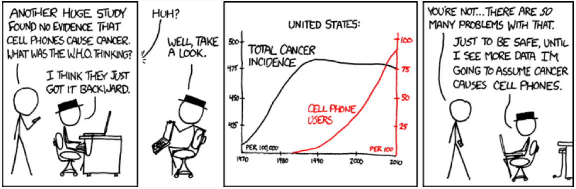
# Causal inference for combining experimental and observational data

Julie Josse (Inria) & Erwan Scornet (Polytechnique)

Ahmed BOUGHDIRI

February 16, 2024

# Causal Inference



# Causal Inference

$\Rightarrow$ **Effect of a policy/intervention/treatment $T$ on an outcome $Y$**

- What is the effect of smoking on COVID-19 mortality rate ?

# Causal Inference

$\Rightarrow$ **Effect of a policy/intervention/treatment $T$ on an outcome $Y$**

- What is the effect of smoking on COVID-19 mortality rate ?
- Does Aspirin cause my headaches to disappear ?
- What is the effect of hydrochloroquine on mortality ?

# Causal Inference

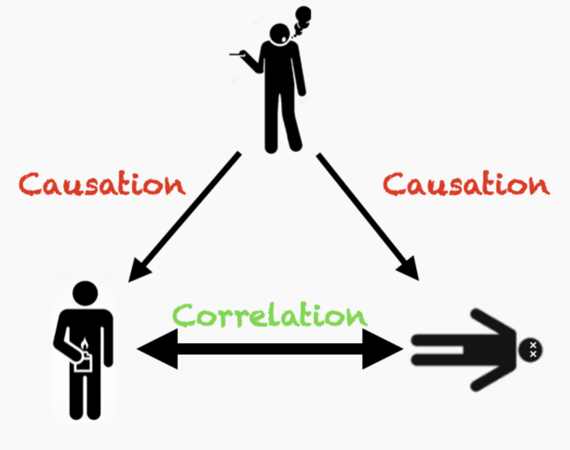
$\Rightarrow$ **Effect of a policy/intervention/treatment $T$ on an outcome $Y$**

- What is the effect of smoking on COVID-19 mortality rate ?
- Does Aspirin cause my headaches to disappear ?
- What is the effect of hydrochloroquine on mortality ?

- What is the impact of an advertising campaign ?
- What is the effect of online classes on student performance ?
- How does 4 days work week affect the economy ?

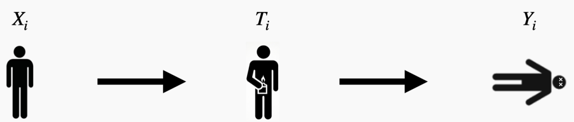
# Causal Inference

**We want to know if there is a causation and not just a correlation**



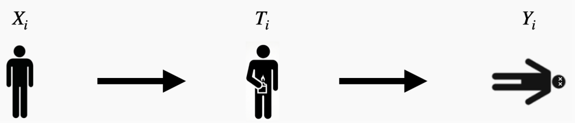

"People who have a lighter tend to have a smaller life expectancy"

# Causal Inference




"People who have a lighter tend to have a smaller life expectancy"

# Causal Inference




"People who have a lighter tend to have a smaller life expectancy"

$$\Rightarrow N \textbf{ i.i.d. } \big(\underbrace{X_i}_{\textbf{covariates}}, \overbrace{T_i}^{\textbf{treatment}}, \underbrace{Y_i}_{\text{outcome}}\big) \in \mathbb{R}^d \times \{0,1\} \times \mathbb{R} \times \mathbb{R}$$

| Covariates | | | Treatment | Outcome | Potential outcomes | |
|---|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | T | Y | Y(0) | Y(1) |
| 1.1 | 20 | F | 1 | 67 | **?** | 67 |
| 6 | 45 | F | 0 | 83 | 83 | **?** |
| 0 | 15 | M | 1 | 57 | **?** | 57 |
| | ... | | ... | ... | ... | ... |
| 12 | 52 | M | 0 | 100 | 100 | **?** |

# Potential outcomes

$$\Rightarrow N \textbf{ i.i.d. } (\underbrace{X_i}_{\textbf{covariates}}, \overbrace{T_i}^{\textbf{treatment}}, \underbrace{Y_i}_{\textit{outcome}}) \in \mathbb{R}^d \times \{0,1\} \times \mathbb{R} \times \mathbb{R}$$

| Covariates | | | Treatment | Outcome | Potential outcomes | |
|---|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | T | Y | Y(0) | Y(1) |
| 1.1 | 20 | F | 1 | 67 | **?** | 67 |
| 6 | 45 | F | 0 | 83 | 83 | **?** |
| 0 | 15 | M | 1 | 57 | **?** | 57 |
| | ... | | ... | ... | ... | ... |
| 12 | 52 | M | 0 | 100 | 100 | **?** |

Our goal is to compute the individual causal effect of the treatment:

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

However we can never observed $\Delta_i$ (only one observed outcome/indiv)

# Average Treatment Effect

Our goal is to compute the individual causal effect of the treatment:

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

In order to fixe the fundamental problem of causal inference define the Average Treatment Effect.

**Average Treatment Effect (ATE)**

$$\tau = \mathbb{E}[\Delta] = \mathbb{E}[Y(1) - Y(0)]$$

$\tau$ is also referred as the risk difference.
$\Rightarrow$ depends on the population

The ATE is the difference of the average outcome had everyone gotten treated and the average outcome had nobody gotten the treatment.

# Causal Inference

We now want to estimate $\tau$:

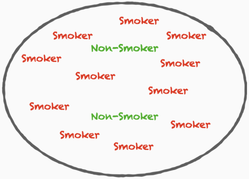
$$\begin{aligned}\tau &= \mathbb{E}[Y(1) - Y(0)] \\ &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &\stackrel{?}{=} \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]\end{aligned}$$

# Causal Inference

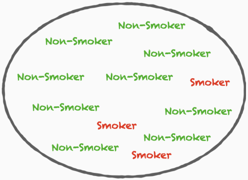
We now want to estimate $\tau$:

$$\begin{aligned}\tau &= \mathbb{E}[Y(1) - Y(0)] \\ &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &\stackrel{?}{=} \mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0]\end{aligned}$$

"People who have a lighter tend to have a smaller life expectancy"




$T = 1$



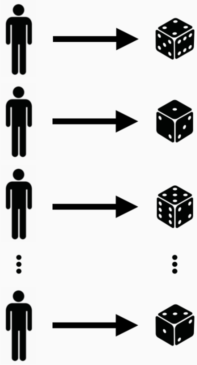

$T = 0$

# Randomized Controlled Trial

"Do people who have a lighter tend to have a smaller life expectancy ?"

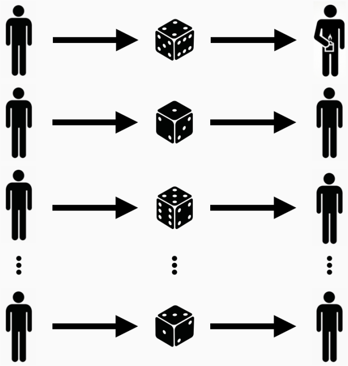
# Randomized Controlled Trial

"Do people who have a lighter tend to have a smaller life expectancy ?"



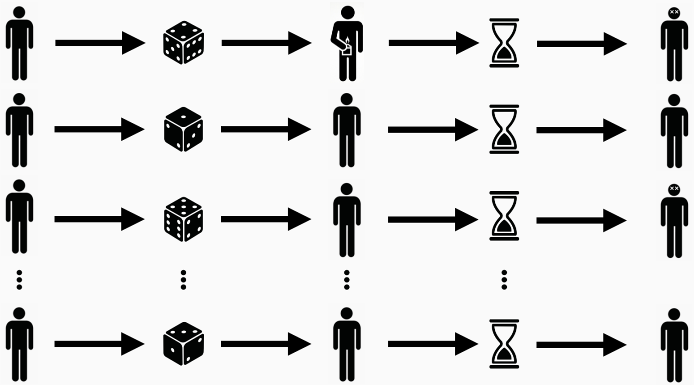
# Randomized Controlled Trial

"Do people who have a lighter tend to have a smaller life expectancy ?"



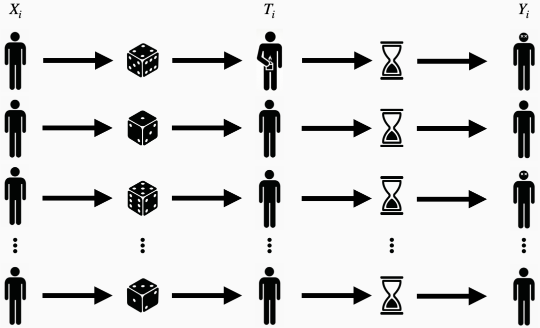
# Randomized Controlled Trial

"Do people who have a lighter tend to have a smaller life expectancy ?"



# Randomized Controlled Trial

"Do people who have a lighter tend to have a smaller life expectancy ?"




**assumptions**

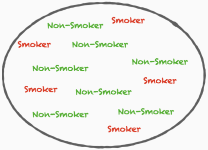
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ (consistency)
- $T_i \perp\!\!\!\perp \{Y_i(0), Y_i(1), X_i\}$ **(random treatment assignment)**
  **Flip a coin to assign the treatment**

# Randomized Controlled Trial

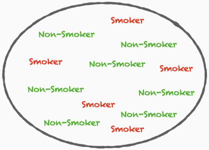
## Identifiability assumptions

- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ (consistency)
- $T_i \perp\!\!\!\perp \{Y_i(0), Y_i(1), X_i\}$ **(random treatment assignment)**

  **Flip a coin to assign the treatment**




$T = 1$




$T = 0$

We now have
$$\begin{aligned}\tau &= \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\ &= \mathbb{E}[Y_i(1) | T_i = 1] - \mathbb{E}[Y_i(0) | T_i = 0] \\ &= \mathbb{E}[Y_i | T_i = 1] - \mathbb{E}[Y_i | T_i = 0] \\ &= \frac{1}{\mathbb{P}(T_i = 1)} \mathbb{E}[Y_i T_i] - \frac{1}{\mathbb{P}(T_i = 0)} \mathbb{E}[Y_i (1 - T_i)]\end{aligned}$$

We say that $\tau$ is identifiable if it can be computed using a infinite number of observations from it.

# Randomized Controlled Trial

**Identifiability assumptions**

- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ (consistency)
- $T_i \perp\!\!\!\perp \{Y_i(0), Y_i(1), X_i\}$ **(random treatment assignment)**
  **Flip a coin to assign the treatment**

$$
\begin{aligned}
\text{We now have } \tau = \mathbb{E}[\Delta_i] &= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\
&= \frac{1}{\mathbb{P}(T_i = 1)} \mathbb{E}[Y_i T_i] - \frac{1}{\mathbb{P}(T_i = 0)} \mathbb{E}[Y_i(1 - T_i)]
\end{aligned}
$$

| Covariates | | | Treatment | Outcome | Potential outcomes | |
|---|---|---|---|---|---|---|
| $X_1$ | $X_2$ | $X_3$ | T | Y | Y(0) | Y(1) |
| 1.1 | 20 | F | 1 | 67 | **?** | 67 |
| 6 | 45 | F | 0 | 83 | 83 | **?** |
| | ... | | ... | ... | ... | ... |
| 12 | 52 | M | 0 | 100 | 100 | **?** |

$\hat{\tau}_{DM} = \frac{1}{n_1} \sum_{T_i=1} Y_i - \frac{1}{n_0} \sum_{T_i=0} Y_i;$ $\tau =$ mean(blue)-mean(red)

# Randomized Controlled Trial

### Identifiability assumptions

- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ (consistency)
- $T_i \perp\!\!\!\perp \{Y_i(0), Y_i(1), X_i\}$ **(random treatment assignment)**
  **Flip a coin to assign the treatment**

### Difference-in-means estimator

$$\hat{\tau}_{DM} = \frac{1}{n_1} \sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0} \sum_{i=1}^{n} (1 - T_i) Y_i$$

where $n_1 = \sum_{i=1}^{n} T_i$ and $n_0 = \sum_{i=1}^{n} 1 - T_i$

$\hat{\tau}_{DM}$ unbiased and $\sqrt{n}$-consistent $\sqrt{n}\,(\hat{\tau}_{DM} - \tau) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V_{DM})$
with $V_{DM} = \frac{Var(Y_i(0))}{\mathbb{P}(T_i=0)} + \frac{Var(Y_i(1))}{\mathbb{P}(T_i=1)}$.

# Data sources & evidences to estimate the treatment effect

## Randomized Controlled Trial (RCT)

- **gold standard** (allocation  )
- same covariate distributions of treated and control groups
  $\Rightarrow$ High **internal** validity

# Data sources & evidences to estimate the treatment effect

## Randomized Controlled Trial (RCT)

- **gold standard** (allocation  )
- same covariate distributions of treated and control groups
  $\Rightarrow$ High **internal** validity
- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria
  $\Rightarrow$ No personalized medicine
- **trial sample different from the population eligible for treatment**
  $\Rightarrow$ Low **external** validity

# Data sources & evidences to estimate the treatment effect

## Randomized Controlled Trial (RCT)

- **gold standard** (allocation  )
- same covariate distributions of treated and control groups
  $\Rightarrow$ High **internal** validity
- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria
  $\Rightarrow$ No personalized medicine
- **trial sample different from the population eligible for treatment**
  $\Rightarrow$ Low **external** validity

## Observational data

- low cost con
- large amounts of data (registries, biobanks, EHR, claims)
  $\Rightarrow$ patient's heterogeneity
- **representative of the target populations**
  $\Rightarrow$ High **external** validity

# Data sources & evidences to estimate the treatment effect

## Randomized Controlled Trial (RCT)

- **gold standard** (allocation )
- same covariate distributions of treated and control groups
  $\Rightarrow$ High **internal** validity
- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria
  $\Rightarrow$ No personalized medicine
- **trial sample different from the population eligible for treatment**
  $\Rightarrow$ Low **external** validity

## Observational data

- "big data": low quality
- lack of a controlled design opens the door to **confounding bias**
  $\Rightarrow$ Low **internal** validity
- low cost con
- large amounts of data (registries, biobanks, EHR, claims)
  $\Rightarrow$ patient's heterogeneity
- **representative of the target populations**
  $\Rightarrow$ High **external** validity

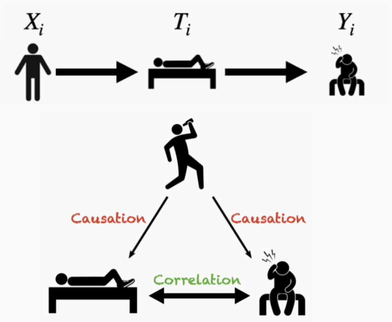
# Assumption for ATE identifiability in observational data

## Unconfoundedness

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \mid X_i$$

Measure all possible confounders

Unobserved confounders make it impossible to separate correlation and causality when correlated to both the outcome and the treatment.



# G-formula estimator

Average treatment effect (ATE): $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$

**Identifiability assumptions in observational data**

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \mid X_i$ **(Unconfoundedness)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**

Using the law of total expectation,

$$\begin{aligned}
\tau = \mathbb{E}[\Delta_i] &= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\
&= \mathbb{E}[\mathbb{E}[Y_i(1)|X]] - \mathbb{E}[\mathbb{E}[Y_i(1)|X]] \qquad \text{Law of total probability} \\
&= \mathbb{E}\left[\mathbb{E}[Y_i(1)|T_i = 1, X]\right] - \mathbb{E}\left[\mathbb{E}[Y_i(0)|T_i = 0, X]\right] \qquad \text{Unconfoundedness} \\
&= \mathbb{E}\left[\mathbb{E}[Y_i|T_i = 1, X]\right] - \mathbb{E}\left[\mathbb{E}[Y_i|T_i = 0, X]\right] \qquad \text{Consistency}
\end{aligned}$$

# G-formula estimator

Average treatment effect (ATE): $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$

**Identifiability assumptions in observational data**

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \,|\, X_i$ **(Unconfoundedness)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**

Using the law of total expectation,

$$
\begin{aligned}
\tau = \mathbb{E}[\Delta_i] &= \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\
&= \mathbb{E}[\mathbb{E}[Y_i(1)|X]] - \mathbb{E}[\mathbb{E}[Y_i(1)|X]] \qquad \text{Law of total probability} \\
&= \mathbb{E}\left[\mathbb{E}[Y_i(1)|T_i = 1, X]\right] - \mathbb{E}\left[\mathbb{E}[Y_i(0)|T_i = 0, X]\right] \qquad \text{Unconfoundedness} \\
&= \mathbb{E}\left[\mathbb{E}[Y_i|T_i = 1, X]\right] - \mathbb{E}\left[\mathbb{E}[Y_i|T_i = 0, X]\right] \qquad \text{Consistency}
\end{aligned}
$$

**G-formula estimator**

$$\hat{\tau}_G = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i)$$

where $\mu_{(t)}(X) = \mathbb{E}\left[Y|T = t, X\right]$

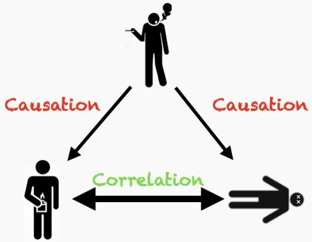
# Assumption for ATE identifiability in observational data

**Overlap**

Propensity score: probability of treatment given observed covariates.

$$e(x) \triangleq \mathbb{P}(T_i = 1 \,|\, X_i = x) \quad \forall x \in \mathcal{X}.$$

We assume overlap, i.e. $\eta < e(x) < 1 - \eta, \quad \forall x \in \mathcal{X}$ and some $\eta > 0$



# Inverse-propensity weighting estimator

Average treatment effect (ATE): $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$

**Identifiability assumptions in observational data**

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \,|\, X_i$ **(Unconfoundedness)**
- $\eta < e(x) < 1 - \eta, \quad \forall\, x \in \mathcal{X}$ and some $\eta > 0$ **(Overlap)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**

Propensity score (proba treated|covariates): $e(x) = \mathbb{P}(T_i = 1 \,|\, X_i = x)$

# Inverse-propensity weighting estimator

Average treatment effect (ATE): $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$

## Identifiability assumptions in observational data

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \,|\, X_i$ **(Unconfoundedness)**
- $\eta < e(x) < 1 - \eta, \quad \forall\, x \in \mathcal{X}$ and some $\eta > 0$ **(Overlap)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**

Propensity score (proba treated|covariates): $e(x) = \mathbb{P}(T_i = 1 \,|\, X_i = x)$

## IPW estimator

$$\hat{\tau}_{IPW} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{T_i Y_i}{\hat{e}(X_i)} - \frac{(1 - T_i) Y_i}{1 - \hat{e}(X_i)} \right)$$

$\hat{\tau}_{IPW}$ unbiased and $\sqrt{n}$-consistent $\sqrt{n}\,(\hat{\tau}_{IPW} - \tau) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, V_{IPW})$

with $V_{IPW} = \mathbb{E}\left[ \frac{\left(Y^{(0)}\right)^2}{1 - e(X)} + \frac{\left(Y^{(1)}\right)^2}{e(X)} \right] - \tau^2$ when $\hat{e}(\cdot)$ is consistent

# Augmented Inverse-propensity weighting estimator

Average treatment effect (ATE): $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$

**Identifiability assumptions in observational data**

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \mid X_i$ **(Unconfoundedness)**
- $\eta < e(x) < 1 - \eta, \quad \forall x \in \mathcal{X}$ and some $\eta > 0$ **(Overlap)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**

Model Treatment on Covariates $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x)$
Model Outcome on Covariates $\mu_{(w)}(x) = \mathbb{E}[Y_i(w) \mid X_i = x]$

# Augmented Inverse-propensity weighting estimator

Average treatment effect (ATE): $\tau = \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)]$

## Identifiability assumptions in observational data

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \mid X_i$ **(Unconfoundedness)**
- $\eta < e(x) < 1 - \eta, \quad \forall x \in \mathcal{X}$ and some $\eta > 0$ **(Overlap)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**

Model Treatment on Covariates $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x)$
Model Outcome on Covariates $\mu_{(w)}(x) = \mathbb{E}[Y_i(w) \mid X_i = x]$

## AIPW estimator

$$\hat{\tau}_{AIPW} = \frac{1}{n} \sum_{i=1}^{n} \left( \mu_{(1)}(X_i) - \mu_{(0)}(X_i) + \frac{T_i . (Y_i - \mu_{(1)}(X_i))}{e(X_i)} - \frac{(1 - T_i)(Y_i - \mu_{(0)}(X_i))}{1 - e(X_i)} \right)$$

$\hat{\tau}_{AIPW}$ unbiased and $\sqrt{n}$-consistent $\sqrt{n}\,(\hat{\tau}_{AIPW} - \tau) \xrightarrow[n \to \infty]{d} \mathcal{N}(0, V_{AIPW})$

with $V^*_{\mathrm{AIPW}} = \mathbb{E}\left[ \frac{\left(Y^{(1)} - \mu_1(X)\right)^2}{e(X)} \right] + \mathbb{E}\left[ \frac{(Y^{(0)} - \mu_0(X))^2}{1 - e(X)} \right] + \mathrm{Var}[\mu_1(X) - \mu_0(X)].$

# Augmented Inverse-propensity weighting estimator

Model Treatment on Covariates $e(x) = \mathbb{P}(W_i = 1 \,|\, X_i = x)$
Model Outcome on Covariates $\mu_{(w)}(x) = \mathbb{E}[Y_i(w) \,|\, X_i = x]$

**AIPW estimator**

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\mu_{(1)}(X_i) - \mu_{(0)}(X_i) + \frac{T_i.(Y_i - \mu_{(1)}(X_i))}{e(X_i)} - \frac{(1-T_i)(Y_i - \mu_{(0)}(X_i))}{1-e(X_i)}\right)$$

$\Rightarrow \hat{\tau}_{AIPW}$ is consistent if either the $\hat{\mu}_{(w)}(x)$ are consistent or $\hat{e}(x)$ is consistent.

# Augmented Inverse-propensity weighting estimator

Model Treatment on Covariates $e(x) = \mathbb{P}(W_i = 1 \mid X_i = x)$

Model Outcome on Covariates $\mu_{(w)}(x) = \mathbb{E}[Y_i(w) \mid X_i = x]$

**AIPW estimator**

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\mu_{(1)}(X_i) - \mu_{(0)}(X_i) + \frac{T_i.(Y_i - \mu_{(1)}(X_i))}{e(X_i)} - \frac{(1-T_i)(Y_i - \mu_{(0)}(X_i))}{1-e(X_i)}\right)$$

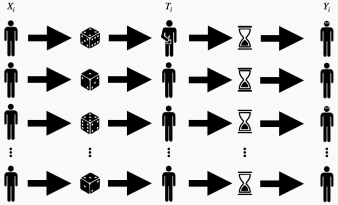
$\Rightarrow$ $\hat{\tau}_{AIPW}$ is consistent if either the $\hat{\mu}_{(w)}(x)$ are consistent or $\hat{e}(x)$ is consistent.

# Summary

When measuring a causal effect, removing all confounding bias can be done two different ways:

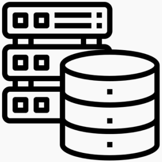
$$\tau_{RD} = \mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]$$

Randomized Controlled Trial (RCT)




$$\hat{\tau}_{DM} = \frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i)Y_i$$

Observational data



$\tau_{IPW}$ $\tau_{AIPW}$ $\tau_G$

# Other ways to measure the causal effect

$$\mathbb{E}\left[Y^{(1)}\right] \qquad \mathbb{E}\left[Y^{(0)}\right]$$

Expected outcome if treated (1) or control (0)

Count the dead

$$\tau_{RR} = \frac{\mathbb{E}\left[Y^{(1)}\right]}{\mathbb{E}\left[Y^{(0)}\right]}$$

Count the living

$$\tau_{SR} = \frac{1 - \mathbb{E}\left[Y^{(1)}\right]}{1 - \mathbb{E}\left[Y^{(0)}\right]}$$

$$\tau_{RD} = \mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]$$

Risk Difference

$$\tau_{NNT} = \tau_{RD}^{-1}$$

Number Needed to Treat

Odds Ratio

$$\tau_{OR} = \frac{\mathbb{E}[Y^{(1)}]}{1 - \mathbb{E}[Y^{(1)}]} \left( \frac{\mathbb{E}[Y^{(0)}]}{1 - \mathbb{E}[Y^{(0)}]} \right)^{-1}$$

# Estimating the risk ratio with observational data

We want to estimate the risk ratio : $\tau_{RR} = \frac{\mathbb{E}[Y_i(1)]}{\mathbb{E}[Y_i(0)]}$

**Identifiability assumptions in observational data**

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \,|\, X_i$ **(Unconfoundedness)**
- $\eta < e(x) < 1 - \eta, \quad \forall x \in \mathcal{X}$ and some $\eta > 0$ **(Overlap)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**
- $\forall i, j \quad Y_i Y_j \geq 0$ **(Name ?)**

# Estimating the risk ratio with observational data

We want to estimate the risk ratio : $\tau_{RR} = \frac{\mathbb{E}[Y_i(1)]}{\mathbb{E}[Y_i(0)]}$

**Identifiability assumptions in observational data**

- $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp T_i \mid X_i$ **(Unconfoundedness)**
- $\eta < e(x) < 1 - \eta, \quad \forall x \in \mathcal{X}$ and some $\eta > 0$ **(Overlap)**
- $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$ **(Consistency)**
- $\forall i, j \quad Y_i Y_j \geq 0$ **(Name ?)**

Estimating the ratio is harder:

$$\tau_{RR} = \frac{\mathbb{E}[Y_i(1)]}{\mathbb{E}[Y_i(0)]} \neq \mathbb{E}\left[\frac{Y_i(1)}{Y_i(0)}\right]$$

# Estimating the risk ratio with observational data

$\Rightarrow$ Using M-estimation, we can solve this issue and get asymptotical normality

# Estimating the risk ratio with observational data

$\Rightarrow$ Using M-estimation, we can solve this issue and get asymptotical normality

$$\hat{\tau}_G = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i)$$

$$\downarrow$$

$$\tau_{R-G} = \frac{\sum_{i=1}^{n} \hat{\mu}_1(X_i)}{\sum_{i=1}^{n} \hat{\mu}_0(X_i)}$$

$$\hat{\tau}_{IPW} = \frac{1}{n}\sum_{i=1}^{n} \left( \frac{T_i Y_i}{\hat{e}(X_i)} - \frac{(1-T_i)Y_i}{1-\hat{e}(X_i)} \right)$$

$$\downarrow$$

$$\tau_{R-IPW} = \frac{\sum_{i=1}^{n} \frac{T_i Y_i}{\hat{e}(X_i)}}{\sum_{i=1}^{n} \frac{(1-T_i)Y_i}{1-\hat{e}(X_i)}}$$

⇒ Using M-estimation, we can solve this issue and get asymptotical normality

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + \frac{T_i\,.\,(Y_i - \hat{\mu}_{(1)}(X_i))}{\hat{e}(X_i)} - \frac{(1-T_i)(Y_i - \hat{\mu}_{(0)}(X_i))}{1-\hat{e}(X_i)}\right)$$

$$\hat{\tau}_{R-AIPW} = \frac{\sum_{i=1}^{n}\hat{\mu}_1(X_i) + \frac{T_i(Y_i - \hat{\mu}_1(X_i))}{\hat{e}(X_i)}}{\sum_{i=1}^{n}\hat{\mu}_0(X_i) + \frac{(1-T_i)(Y_i - \hat{\mu}_0(X_i))}{1-\hat{e}(X_i)}}$$

$\hat{\tau}_{R-X}$ unbiased and $\sqrt{n}$-consistent

$$\sqrt{n}\,(\hat{\tau}_{R-X} - \tau_{RR}) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V_X)$$

# Estimating the risk ratio with observational data

$\Rightarrow$ Using M-estimation, we can solve this issue and get asymptotical normality

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + \frac{T_i \,.\, (Y_i - \hat{\mu}_{(1)}(X_i))}{\hat{e}(X_i)} - \frac{(1-T_i)(Y_i - \hat{\mu}_{(0)}(X_i))}{1-\hat{e}(X_i)}\right)$$

$$\hat{\tau}_{R-AIPW} = \frac{\sum_{i=1}^{n} \hat{\mu}_1(X_i) + \frac{T_i(Y_i - \hat{\mu}_1(X_i))}{\hat{e}(X_i)}}{\sum_{i=1}^{n} \hat{\mu}_0(X_i) + \frac{(1-T_i)(Y_i - \hat{\mu}_0(X_i))}{1-\hat{e}(X_i)}}$$

$\hat{\tau}_{R-X}$ unbiased and $\sqrt{n}$-consistent

$$\sqrt{n}\,(\hat{\tau}_{R-X} - \tau_{RR}) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, V_X)$$

Note that $V_X$ is not symmetrical anymore: Estimating $\frac{\mathbb{E}[Y_i(1)]}{\mathbb{E}[Y_i(0)]}$ or $\frac{\mathbb{E}[Y_i(0)]}{\mathbb{E}[Y_i(1)]}$ will not give the same confidence intervals!

# Different treatment measures give different impressions

Let's suppose an RCT was conducted on a given population:

- $Y = 1$ stroke in 5 years and $Y = 0$ no stroke

| | $\tau_{RD}$ | $\tau_{RR}$ | $\tau_{SR}$ | $\tau_{NNT}$ | $\tau_{OR}$ |
|---|---|---|---|---|---|
| *All* $(P_r)$ | -0.06 | 0.93 | 0.27 | -17 | 0.26 |

# Different treatment measures give different impressions

Let's suppose an RCT was conducted on a given population:

- $Y = 1$ stroke in 5 years and $Y = 0$ no stroke

| | $\tau_{RD}$ | $\tau_{RR}$ | $\tau_{SR}$ | $\tau_{NNT}$ | $\tau_{OR}$ |
|---|---|---|---|---|---|
| *All* ($P_r$) | -0.06 | 0.93 | 0.27 | -17 | 0.26 |

- RD: treatment reduces by 0.06 the probability to suffer from a stroke
- RR: the treated has 0.93 $\times$ the risk of having a stroke comp. with the control
- SR: there is an increased chance of not having a stroke when treated compared to the control by a factor 0.27.
- NNT: one has to treat 17 people to prevent one additional stroke
- OR: people who had a stroke are 0.26 less likely to be treated

# Different treatment measures give different impressions

Let's suppose two RCTs were conducted on these two subpopulations:

- $X = 1$ high blood pressure, $X = 0$ moderate blood pressure.
- $Y = 1$ stroke in 5 years and $Y = 0$ no stroke

| | $\tau_{RD}$ | $\tau_{RR}$ | $\tau_{SR}$ | $\tau_{NNT}$ | $\tau_{OR}$ |
|---|---|---|---|---|---|
| *All* ($P_r$) | -0.06 | 0.93 | 0.27 | -17 | 0.26 |
| $X = 1$ | -0.4 | 0.5 | 0.333 | -2.5 | 0.167 |
| $X = 0$ | -0.029 | 0.97 | 0.17 | -34 | 0.166 |

# Measures' properties

We define $\tau(X) := \mathbb{E}[Y^{(1)} - Y^{(0)} | X]$

**Direct collapsibility**

$$\mathbb{E}\left[\tau(X)\right] = \tau$$

$\Rightarrow$ Only $\tau^{RD} = \mathbb{E}[Y^{(1)} - Y^{(0)}]$ is directly collapsible:

$$\tau^{\mathrm{RD}} = p_{\mathrm{R}}(X = 1) \times \tau_{\mathrm{R}}^{\mathrm{RD}}(X = 1) + p_{\mathrm{R}}(X = 0) \times \tau_{\mathrm{R}}^{\mathrm{RD}}(X = 0)$$

$$-0.06 = -0.4 \times 0.091 - 0.029 \times 0.909$$

| | $\tau_{RD}$ | $\tau_{RR}$ | $\tau_{SR}$ | $\tau_{NNT}$ | $\tau_{OR}$ |
|---|---|---|---|---|---|
| *All* $(P_r)$ | -0.06 | 0.93 | 0.27 | -17 | 0.26 |
| $X = 1$ | -0.4 | 0.5 | 0.333 | -2.5 | 0.167 |
| $X = 0$ | -0.029 | 0.97 | 0.17 | -34 | 0.166 |

# Measures' properties

**Collapsibility (require knowing $Y(0)$)**

$$\mathbb{E}\left[w(X, P(X, Y(0)))\,\tau(X)\right] = \tau \quad \text{with } w \geq 0,\ \mathbb{E}\left[w(X, P(X, Y(0)))\right] = 1$$

Where $\tau(X) := \mathbb{E}[Y^{(1)} - Y^{(0)}|X]$

# Measures' properties

## Collapsibility (require knowing $Y(0)$)

$$\mathbb{E}\left[w(X, P(X, Y(0)))\,\tau(X)\right] = \tau \quad \text{with } w \geq 0,\ \mathbb{E}\left[w(X, P(X, Y(0)))\right] = 1$$

Where $\tau(X) := \mathbb{E}[Y^{(1)} - Y^{(0)} | X]$

| | $\tau_{RD}$ | $\tau_{RR}$ | $\tau_{SR}$ | $\tau_{NNT}$ | $\tau_{OR}$ |
|---|---|---|---|---|---|
| *All* $(P_r)$ | -0.06 | 0.93 | 0.27 | -17 | 0.26 |
| $X = 1$ | -0.4 | 0.5 | 0.333 | -2.5 | 0.167 |
| $X = 0$ | -0.029 | 0.97 | 0.17 | -34 | 0.166 |

## Logic respecting

$$\tau \in \left[\min_x(\tau(x)), \max_x(\tau(x))\right].$$

# Measures' properties

## Collapsibility (require knowing $Y(0)$)

$$\mathbb{E}\left[w(X, P(X, Y(0)))\,\tau(X)\right] = \tau \quad \text{with } w \geq 0,\ \mathbb{E}\left[w(X, P(X, Y(0)))\right] = 1$$

## Logic respecting

$$\tau \in \left[\min_x(\tau(x)), \max_x(\tau(x))\right].$$

| Measure | Collapsible | Logic-respecting |
|---|---|---|
| Risk Difference (RD) | Yes | Yes |
| Number Neeeded to Treat (NNT) | No | Yes |
| Risk Ratio (RR) | Yes | Yes |
| Survival Ratio (SR) | Yes | Yes |
| Odds Ratio (OR) | No | No |

# Leverage both RCT and observational data

**Average Treatment Effect (ATE)**

$$\tau = \mathbb{E}[\Delta] = \mathbb{E}[Y(1) - Y(0)]$$

$\tau$ is also referred as the risk difference.

$\Rightarrow$ depends on the population

The ATE is the difference of the average outcome had everyone gotten treated and the average outcome had nobody gotten the treatment.

Covariates distribution not the same in the RCT & target pop:

$$p_{\mathrm{R}}(x) \neq p_{\mathrm{T}}(x) \Rightarrow \quad \underbrace{\tau_{\mathrm{R}} := \mathbb{E}_{\mathrm{R}}[Y(1) - Y(0)]}_{\text{ATE in the RCT}} \neq \underbrace{\mathbb{E}_{\mathrm{T}}[Y(1) - Y(0)] := \tau_{\mathrm{T}}}_{\text{Target ATE}}$$

# Leverage both RCT and observational data

**RCT**

- $+$ No confounding
- $-$ Trial sample different from the population eligible for treatment

**(big) Observational data**

- $-$ Confounding
- $+$ Representative of the target population

# Leverage both RCT and observational data

**RCT**

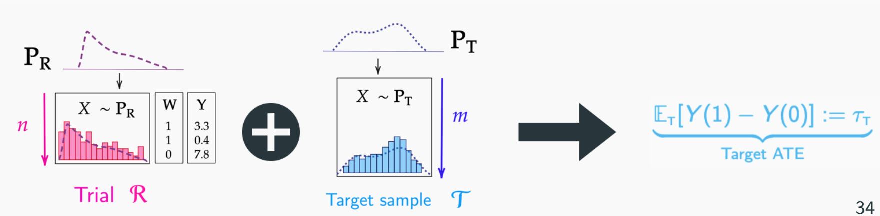
- $+$ No confounding
- $-$ Trial sample different from the population eligible for treatment

**(big) Observational data**

- $-$ Confounding
- $+$ Representative of the target population

$$\tau_R^{\mathrm{RD}} = p_{\mathrm{R}}(X=1) \times \tau_{\mathrm{R}}^{\mathrm{RD}}(X=1) + p_{\mathrm{R}}(X=0) \times \tau_{\mathrm{R}}^{\mathrm{RD}}(X=0)$$

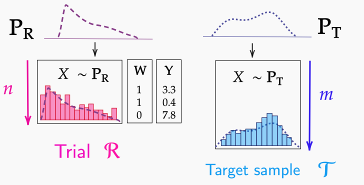
$$\tau_T^{\mathrm{RD}} = p_{\mathrm{T}}(X=1) \times \tau_{\mathrm{R}}^{\mathrm{RD}}(X=1) + p_{\mathrm{T}}(X=0) \times \tau_{\mathrm{R}}^{\mathrm{RD}}(X=0)$$



# Generalization task from a RCT to a target population

**Two data sources**:

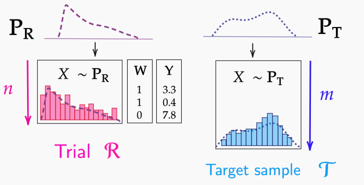
- A trial of size $n$ with $p_R(x)$ the probability of observing individual with $X = x$,
- A sample of the target population of interest – for e.g. a national cohort (resp. $m$ and $p_T(x)$).



# Generalization task from a RCT to a target population

**Two data sources**:

- A trial of size $n$ with $p_R(x)$ the probability of observing individual with $X = x$,
- A sample of the target population of interest – for e.g. a national cohort (resp. $m$ and $p_T(x)$).




**Transportability (Ignorability on trial participation)**

$$\forall w \in \{0,1\} \quad \mathbb{E}_R[Y(w) \mid X] = \mathbb{E}_T[Y(w) \mid X]$$

**Overlap assumption**

$$\forall x \in \mathbb{X}, p_R(x) > 0 \text{ and } \operatorname{supp}(P_T(X)) \subset \operatorname{supp}(P_R(X))$$

# Generalization of conditional outcome: identifiability

Average Treatment Effect: $\tau_{\mathrm{T}} = \mathbb{E}_{\mathrm{T}}[Y_i(1) - Y_i(0)], \forall t \in \{0, 1\}$

$$
\begin{aligned}
\mathbb{E}_{\mathrm{T}}[Y(t)] &= \mathbb{E}_{\mathrm{T}}[\mathbb{E}_{\mathrm{T}}[Y(t) \mid X]] && \text{Law of total expectation} \\
&= \mathbb{E}_{\mathrm{T}}[\mathbb{E}_{\mathrm{R}}[Y(t) \mid X]] && \text{Ignorability} \\
&= \mathbb{E}_{\mathrm{T}}[\mathbb{E}_{\mathrm{R}}[Y(t) \mid X = x, T = t]] && \text{Random treatment} \\
&= \mathbb{E}_{\mathrm{T}}[\underbrace{\mathbb{E}_{\mathrm{R}}[Y \mid X = x, T = t]}_{\mu_t(x)}] && \text{Consistency}
\end{aligned}
$$

# Generalization of conditional outcome: identifiability

Average Treatment Effect: $\tau_{\text{T}} = \mathbb{E}_{\text{T}}[Y_i(1) - Y_i(0)], \forall t \in \{0,1\}$

$$
\begin{aligned}
\mathbb{E}_{\text{T}}[Y(t)] &= \mathbb{E}_{\text{T}}[\mathbb{E}_{\text{T}}[Y(t) \mid X]] && \text{Law of total expectation} \\
&= \mathbb{E}_{\text{T}}[\mathbb{E}_{\text{R}}[Y(t) \mid X]] && \text{Ignorability} \\
&= \mathbb{E}_{\text{T}}[\mathbb{E}_{\text{R}}[Y(t) \mid X = x, T = t]] && \text{Random treatment} \\
&= \mathbb{E}_{\text{T}}\underbrace{[\mathbb{E}_{\text{R}}[Y \mid X = x, T = t]]}_{\mu_t(x)} && \text{Consistency}
\end{aligned}
$$

**Regression adjustment - plug-in gformula**

$$\hat{\tau}_{g,n,m} = \frac{1}{m} \sum_{i \in \mathcal{T}} \left(\hat{\mu}_{1,n}(X_i) - \hat{\mu}_{0,n}(X_i)\right)$$

# Plug-in gformula: difference between conditional mean

## Plug-in gformula

$$\hat{\tau}_{g,n,m} = \frac{1}{m} \sum_{i=n+1}^{n+m} \left( \hat{\mu}_{1,n}(X_i) - \hat{\mu}_{0,n}(X_i) \right),$$

$$\mu_t(x) = \mathbb{E}_{\mathrm{R}}[Y \mid X = x, T = t]$$

| | | | Covariates | | | Treat | Outcomes |
|---|---|---|---|---|---|---|---|
| | Set | $S$ | $X_1$ | $X_2$ | $X_3$ | $T$ | $Y$ |
| 1 | $\mathcal{R}$ | 1 | 1.1 | 20 | 9.4 | 1 | 24.1 |
| | $\mathcal{R}$ | 1 | -6 | 45 | 8.3 | 0 | 26.3 |
| $n$ | $\mathcal{R}$ | 1 | 0 | 15 | 6.2 | 1 | 23.5 |
| $n+1$ | $\mathcal{O}$ | ? | -1 | 35 | 7.1 | | |
| $n+2$ | $\mathcal{O}$ | ? | -2 | 52 | 2.4 | | |
| | $\mathcal{O}$ | ? | | ... | | | |
| $n+m$ | $\mathcal{O}$ | ? | -2 | 22 | 3.4 | | |

- Fit two models of the outcome ($Y$) on covariates ($X$) among trial participants ($\mathcal{R}$) for treated and for control to get $\widehat{\mu}_{1,n}$ & $\widehat{\mu}_{0,n}$

# Plug-in gformula: difference between conditional mean

## Plug-in gformula

$$\hat{\tau}_{g,n,m} = \frac{1}{m} \sum_{i=n+1}^{n+m} \left(\hat{\mu}_{1,n}(X_i) - \hat{\mu}_{0,n}(X_i)\right),$$

$\mu_t(x) = \mathbb{E}_{\text{R}}[Y \mid X = x, T = t]$

| | | | Covariates | | | Treat | Outcomes | |
|---|---|---|---|---|---|---|---|---|
| | Set | $S$ | $X_1$ | $X_2$ | $X_3$ | $T$ | $Y(0)$ | Y(1) |
| 1 | $\mathcal{R}$ | 1 | 1.1 | 20 | 9.4 | 1 | | 24.1 |
| | $\mathcal{R}$ | 1 | -6 | 45 | 8.3 | 0 | 26.3 | |
| $n$ | $\mathcal{R}$ | 1 | 0 | 15 | 6.2 | 1 | | 23.5 |
| $n+1$ | $\mathcal{O}$ | ? | -1 | 35 | 7.1 | | $\hat{\mu}_0(X_{n+1})$ | $\hat{\mu}_1(X_{n+1})$ |
| $n+2$ | $\mathcal{O}$ | ? | -2 | 52 | 2.4 | | $\hat{\mu}_0(X_{n+2})$ | $\hat{\mu}_1(X_{n+2})$ |
| | $\mathcal{O}$ | ? | | ... | | ... | ... | ... |
| $n+m$ | $\mathcal{O}$ | ? | -2 | 22 | 3.4 | | $\hat{\mu}_0(X_{n+m})$ | $\hat{\mu}_1(X_{n+m})$ |

- Fit two models of the outcome ($Y$) on covariates ($X$) among trial participants ($\mathcal{R}$) for treated and for control to get $\widehat{\mu}_{1,n}$ & $\widehat{\mu}_{0,n}$
- Apply these models to the covariates in the target pop, i.e., marginalize over the covariate distribution of the target pop, gives the expected outcomes
- Compute the differences between the expected outcomes on the target population $\overline{\hat{\mu}_{1,n}(\cdot)}$ - $\overline{\hat{\mu}_{0,n}(\cdot)}$

## Assumptions for identifiability with fewer covariates

**Transportability (Ignorability on trial participation)**

$$\forall t \in \{0,1\} \quad \mathbb{E}_{\mathrm{R}}[Y(t) \mid X] = \mathbb{E}_{\mathrm{T}}[Y(t) \mid X]$$

# Assumptions for identifiability with fewer covariates

**Transportability (Ignorability on trial participation)**

$$\forall t \in \{0,1\} \quad \mathbb{E}_{\mathrm{R}}[Y(t) \mid X] = \mathbb{E}_{\mathrm{T}}[Y(t) \mid X]$$

**Transportability of the CATE**

$$\underbrace{\tau_R(X)}_{\mathbb{E}_{\mathrm{R}}[Y(1)-Y(0)|X]} = \underbrace{\tau_{\mathrm{T}}(X)}_{\mathbb{E}_{\mathrm{T}}[Y(1)-Y(0)|X]}$$

# Identifiability and estimation for generalization: weighting

$$\begin{aligned}\tau_{\mathrm{T}} &= \mathbb{E}_{\mathrm{T}}[Y_i(1) - Y_i(0)] = \mathbb{E}_{\mathrm{T}}[\mathbb{E}_{\mathrm{T}}[Y_i(1) - Y_i(0) | X]] \\ &= \mathbb{E}_{\mathrm{T}}\left[\tau_{\mathrm{T}}(X)\right] = \mathbb{E}_{\mathrm{T}}\left[\tau_{\mathrm{R}}(X)\right] \text{ Transportability CATE} \\ &= \mathbb{E}_{\mathrm{R}}\left[\frac{p_{\mathrm{T}}(X)}{p_{\mathrm{R}}(X)}\tau_{\mathrm{R}}(X)\right]\end{aligned}$$

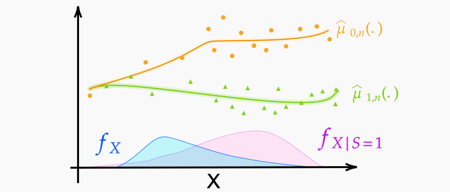
**IPSW: inverse propensity sampling weighting**

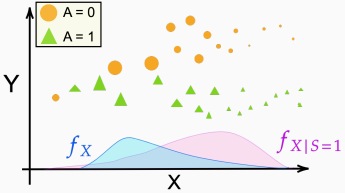
$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n}\sum_{i\in\mathcal{R}} \quad \frac{\hat{p}_{\mathrm{T}}(X_i)}{\hat{p}_{\mathrm{R}}(X_i)} \quad \left(\frac{TY}{e_{\mathrm{R}}(x)} - \frac{(1-T)Y}{1-e_{\mathrm{R}}(x)}\right),$$

$e_{\mathrm{R}}(x) = P(T=1 \mid X=x) = 0.5.$

Re-weight, so that the trial follows the target sample's distribution: if proba to be in trial when old is small, then up-weight old in trial

# Generalization estimators: illustrative schematics




The trial findings $\hat{\tau}_{1,n}$ would over-estimate the target treatment effect $\tau_T$
Left: the plug-in G-formula model the response using the RCT observation
Right: IPSW weight the RCT observations
$f_X$ ($f_{X|S=1}$) density of the target (resp. trial) pop., $\hat{\mu}_{a,n}(\cdot)$ fitted response surface with $n$ trial obs.

## Theorem - consistency[1]

Under assumptions, $\hat{\tau}_{\mathrm{IPSW},n,m}$ and $\hat{\tau}_{g,n,m}$ converges toward $\tau_{\mathrm{T}}$ in $L^1$ norm,

$$\hat{\tau}_{\mathrm{IPSW},n,m} \xrightarrow[n,m\to\infty]{L^1} \tau_{\mathrm{T}}$$

$$\hat{\tau}_{g,n,m} \xrightarrow[n,m\to\infty]{L^1} \tau_{\mathrm{T}}$$

[1] Colnet, J.J et al. 2022. Generalizing a causal effect: sensitivity analysis and missing covariates. *Journal of Causal Inference*.

# Conclusion

| Generalizing | Conditional Outcome | Local effects |
|---|---|---|
| Assumption | $\mathbb{E}_{\mathrm{R}}[Y(w) \mid X] = \mathbb{E}_{\mathrm{T}}[Y(w) \mid X = x]$ | $\tau_R(X) = \tau_{\mathrm{T}}(X)$ |
| Identification | $\mathbb{E}_{\mathrm{T}}\left[Y(w)\right] = \mathbb{E}_{\mathrm{T}}\left[\mathbb{E}_{\mathrm{R}}\left[Y(w) \mid X\right]\right]$ | $\mathbb{E}_{\mathrm{R}}\left[\frac{p_{\mathrm{T}}(X)}{p_{\mathrm{R}}(X)}\, w_{\mathrm{T}}(Y(0), X)\, \tau_{\mathrm{R}}(X)\right]$ |
| Estimator | $\frac{1}{m}\sum_{i=n+1}^{n+m}\left(\hat{\mu}_{1,n}(X_i) - \hat{\mu}_{0,n}(X_i)\right)$ | $\frac{1}{n}\sum_{i\in\mathcal{R}} \frac{\hat{p}_{\mathrm{T}}(X_i)}{\hat{p}_{\mathrm{R}}(X_i)}\left(\frac{TY}{e_{\mathrm{R}}(x)} - \frac{(1-T)Y}{1-e_{\mathrm{R}}(x)}\right)$ |

- Depending on the assumptions, either conditional outcome or local treatment effect can be generalised
- Generalizing local effects only for collapsible measure, information on $Y^{(0)}$ with weights required

$\Rightarrow$ My goal is to do the same for the risk ratio!

Thanks for your attention!

# Appendix

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| | $Y = 1$ | $Y = 0$ |
|---|---|---|
| $T = 1$ | $A$ | $B$ |
| $T = 0$ | $C$ | $D$ |

**Risk difference:**

$$\begin{aligned}\tau_{RD} &= \mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right] \\ &= \frac{A}{A+B} - \frac{C}{C+D}\end{aligned}$$

How much higher is the risk of the outcome among people who are exposed to the risk factor?

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0,1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| Toxic pollutant level | Death | |
|---|---|---|
| | Yes | No |
| High | 50 | 50 |
| Low | 25 | 75 |

**Risk difference:**

$$\tau_{RD} = \mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]$$
$$= \frac{50}{50+50} - \frac{25}{25+75} = 0.25$$

People exposed to high levels of the toxic pollutant had a 25 percentage point higher chance of dying within the next 20 years

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| | $Y = 1$ | $Y = 0$ |
|---|---|---|
| $T = 1$ | $A$ | $B$ |
| $T = 0$ | $C$ | $D$ |

**Risk ratio:**

$$\tau_{RR} = \frac{\mathbb{E}\left[Y^{(1)}\right]}{\mathbb{E}\left[Y^{(0)}\right]} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

How many times higher is the risk of the outcome among people who are exposed to the risk factor?

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| Toxic pollutant level | Death | |
|---|---|---|
| | Yes | No |
| High | 50 | 50 |
| Low | 25 | 75 |

**Risk ratio:**

$$\tau_{RR} = \frac{\mathbb{E}\left[Y^{(1)}\right]}{\mathbb{E}\left[Y^{(0)}\right]} = \frac{\frac{50}{50+50}}{\frac{25}{25+75}} = 2$$

People exposed to high levels of toxic pollutant were twice as likely to die within the next 20 years.

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| | $Y = 1$ | $Y = 0$ |
|---|---|---|
| $T = 1$ | $A$ | $B$ |
| $T = 0$ | $C$ | $D$ |

**Survival ratio:**

$$\tau_{SR} = \frac{1 - \mathbb{E}\left[Y^{(1)}\right]}{1 - \mathbb{E}\left[Y^{(0)}\right]} = \frac{\frac{D}{C+D}}{\frac{B}{A+B}}$$

how many times higher is the chance of avoiding the outcome, among people not exposed to the risk factor?

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| Toxic pollutant level | Death | |
|---|---|---|
| | Yes | No |
| High | 50 | 50 |
| Low | 25 | 75 |

**Survival ratio:**

$$\tau_{SR} = \frac{1 - \mathbb{E}\left[Y^{(1)}\right]}{1 - \mathbb{E}\left[Y^{(0)}\right]}$$

$$= \frac{\frac{75}{25+75}}{\frac{50}{50+50}} = 1.5$$

People only exposed to low levels of this toxic pollutant were 1.5 times as likely to survive the next 20 years.

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| | $Y = 1$ | $Y = 0$ |
|---|---|---|
| $T = 1$ | $A$ | $B$ |
| $T = 0$ | $C$ | $D$ |

**Number needed to harm/treat:**

$$\tau_{NNH} = \frac{1}{\mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]} = \tau_{RD}^{-1}$$
$$= \frac{1}{\frac{A}{A+B} - \frac{C}{C+D}}$$

How many people would need to be exposed to the risk factor, to see the outcome in one of them?

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| Toxic pollutant level | Death | |
|---|---|---|
| | Yes | No |
| High | 50 | 50 |
| Low | 25 | 75 |

**Number needed to harm/treat:**

$$\tau_{NNH} = \frac{1}{\mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]} = \tau_{RD}^{-1}$$

$$= \frac{1}{\frac{50}{50+50} - \frac{25}{25+75}} = 4$$

Four people would need to be exposed to high levels of the toxic pollutant for one to die within the next 20 years, on average.

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| | $Y = 1$ | $Y = 0$ |
|---|---|---|
| $T = 1$ | $A$ | $B$ |
| $T = 0$ | $C$ | $D$ |

**Odds ratio:**

$$\begin{aligned}\tau_{OR} &= \frac{\mathbb{E}\left[Y^{(1)}\right]}{1 - \mathbb{E}\left[Y^{(1)}\right]} \left(\frac{\mathbb{E}\left[Y^{(0)}\right]}{1 - \mathbb{E}\left[Y^{(0)}\right]}\right)^{-1} \\ &= \frac{A}{B}\left(\frac{C}{D}\right)^{-1}\end{aligned}$$

How many times higher were the odds of the outcome, in people exposed to the risk factor?

# Other ways to measure the causal effect

To simplify things, suppose $Y \in \{0, 1\}$ and that our data is from an RCT:

$$\mathbb{E}\left[Y^{(1)}\right] = \mathbb{P}(Y^{(1)} = 1) \qquad \mathbb{E}\left[Y^{(0)}\right] = \mathbb{P}(Y^{(0)} = 1)$$

| Toxic pollutant level | Death | |
|---|---|---|
| | Yes | No |
| High | 50 | 50 |
| Low | 25 | 75 |

**Odds ratio:**

$$\tau_{OR} = \frac{\mathbb{E}\left[Y^{(1)}\right]}{1 - \mathbb{E}\left[Y^{(1)}\right]} \left( \frac{\mathbb{E}\left[Y^{(0)}\right]}{1 - \mathbb{E}\left[Y^{(0)}\right]} \right)^{-1}$$

$$= \frac{50}{50} \left( \frac{25}{75} \right)^{-1} = 3$$

People who died had 3 times the odds of having been exposed to high levels of the toxic pollutant during the past 20 years.

# Causal Inference

$$Y(t) \textbf{ Vs } Y|T = t$$