How can we account for sampling bias in randomized trials using observational data?

Bénédicte Colnet, PhD student at Inria (Soda & PreMeDICaL teams)

Trevor Hastie, Jonathan Taylor, and Rob Tibshirani's research group for students, Wednesday, May 18th, 2022

My PhD advisors

Julie Josse
Senior Researcher

Inria

Missing values, causal inference

Erwan Scornet

Associate professor

École Polytechnique

Random forests, missing values

Gaël Varoquaux

Research director

Inria

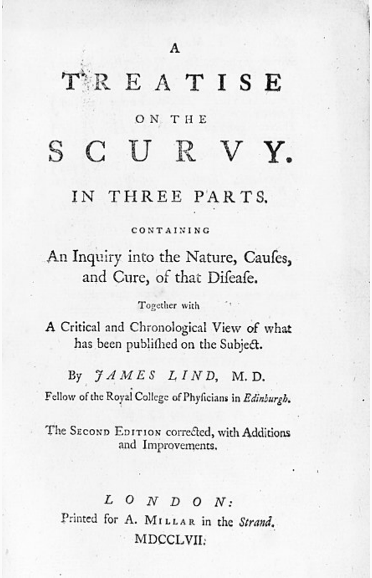
Co-founder of scikit-learn, Machine-Learning

Todays' presentation[1],[2]

[1] Colnet & Mayer et al. (2020) Causal inference methods for combining randomized trials and observational studies: a review. *Under revisions.*

[2] Colnet et al. (2021) Causal effect on a target population: a sensitivity analysis to handle missing covariates. *Under revisions for Journal of Causal Inference.*

A rather old question



A

TREATISE

ON THE

SCURVY.

IN THREE PARTS.

CONTAINING

An Inquiry into the Nature, Cauſes,
and Cure, of that Diſeaſe.

Together with

A Critical and Chronological View of what
has been publiſhed on the Subject.

By *JAMES LIND*, M. D.
Fellow of the Royal College of Phyſicians in *Edinburgh.*

The SECOND EDITION corrected, with Additions
and Improvements.

LONDON:
Printed for A. MILLAR in the *Strand.*
MDCCLVII.

Source: Wikipedia

James Lind’s experiment formalization

This slide is an introduction to the Potential Outcome framework.

Assume your goal is to measure the effect of a drug on an outcome.

James Lind's experiment formalization

This slide is an introduction to the Potential Outcome framework.

Assume your goal is to measure the effect of a drug on an outcome.

Using the potential outcome framework (Neyman, 1923), we denote

- 💊 A the treatment,
- 🩺 X the covariates,
- 🌡️ Y the **observed** outcome.

For each individual i, consider each of the possible outcomes, as if we consider counterfactual worlds, $Y_i^{(1)}$ **(treated)**, and $Y_i^{(0)}$ **(untreated)**.

James Lind's experiment formalization

This slide is an introduction to the Potential Outcome framework.

Assume your goal is to **measure the effect** of a drug on an outcome.

Using the potential outcome framework (Neyman, 1923), we denote

- 💊 A the treatment,
- 🩺 X the covariates,
- 🌡️ Y the **observed** outcome.

For each individual i, consider each of the possible outcomes, as if we consider counterfactual worlds, $Y_i^{(1)}$ **(treated)**, and $Y_i^{(0)}$ **(untreated)**.

Question: $Y_i^{(1)} \overset{?}{=} Y_i^{(0)}$

Data at hand

Individual causal effect of the treatment: $\Delta_i = Y_i(1) - Y_i(0)$

Problem: Δ_i never observed (only observe one outcome/indiv). Causal inference as a missing value problem?

Covariates			Treatment	Outcome(s)		Observed outcome
X_1	X_2	X_3	A	Y(0)	Y(1)	Y(A)
1.1	20	F	1	NA	T	T
-6	45	F	0	F	NA	F
0	15	M	1	NA	F	F
	...		...	...	...	...
-2	52	M	0	T	NA	T

Data at hand

Individual causal effect of the treatment: $\Delta_i = Y_i(1) - Y_i(0)$

Problem: Δ_i never observed (only observe one outcome/indiv). Causal inference as a missing value problem?

Covariates			Treatment	Outcome(s)		Observed outcome
X_1	X_2	X_3	A	Y(0)	Y(1)	Y(A)
1.1	20	F	1	NA	T	T
-6	45	F	0	F	NA	F
0	15	M	1	NA	F	F
	...		...	...	...	...
-2	52	M	0	T	NA	T

💡 Two sources of randomness in this data set:

- Treatment assignment allocation,
- Sampling individuals in a wider population.

Randomized Controlled Trial: an empirical trick to measure the causal effect

Statistical trick: Inference on potential outcomes' distributions.

$$\mathbb{E}\left[Y^{(1)}\right] \overset{?}{=} \mathbb{E}\left[Y^{(0)}\right].$$

Randomized Controlled Trial: an empirical trick to measure the causal effect

Statistical trick: Inference on potential outcomes' distributions.

$$\mathbb{E}\left[Y^{(1)}\right] \overset{?}{=} \mathbb{E}\left[Y^{(0)}\right].$$

More precisely people often target the so-called Average Treatment Effect (ATE),

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

Randomized Controlled Trial: an empirical trick to measure the causal effect

Statistical trick: Inference on potential outcomes' distributions.

$$\mathbb{E}\left[Y^{(1)}\right] \overset{?}{=} \mathbb{E}\left[Y^{(0)}\right].$$

More precisely people often target the so-called Average Treatment Effect (ATE),

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

Running a randomized controlled trial corresponds to:

Randomized Controlled Trial: an empirical trick to measure the causal effect

Statistical trick: Inference on potential outcomes' distributions.

$$\mathbb{E}\left[Y^{(1)}\right] \overset{?}{=} \mathbb{E}\left[Y^{(0)}\right].$$

More precisely people often target the so-called Average Treatment Effect (ATE),

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

Running a randomized controlled trial is a way to ensure,

Assumption - Treatment assignment exchangeability

$$\forall i, \quad Y_i^{(1)}, Y_i^{(0)} \perp\!\!\!\perp A_i,$$

💡 *Treated and control groups differ only with respect to treatment allocation.*

Another assumption we will assume today is the SUTVA assumption: no interference and consistency $Y_i(A_1, A_2, \ldots, A_n) = Y_i(A_i)$.

Statistical properties of the difference-in-means

Suppose we have access to n independent and identically distributed examples labeled $i = 1, \ldots, n$, a response $Y_i \in \mathcal{Y}$, and a binary treatment indicator $A_i \in \{0, 1\}$ assigned randomly.

Definition - Difference in means

$$\hat{\tau}_{\mathrm{DM}} = \frac{1}{n_1} \sum_{A_i=1} Y_i - \frac{1}{n_0} \sum_{A_i=0} Y_i \quad , \text{where } n_a = \left|\{i : A_i = a\}\right| ,$$

Proposition - Asymptotically normal estimator

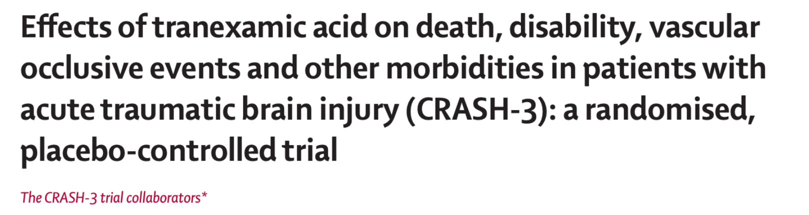
The difference-in-means estimator is asymptotically normal,

$$\sqrt{n} \left(\hat{\tau}_{\mathrm{DM}} - \tau\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\mathrm{DM}}\right) ,$$

where $\sigma^2_{\mathrm{DM}} = \frac{1}{n_0} \operatorname{Var}\left[Y(0)\right] + \frac{1}{n_1} \operatorname{Var}\left[Y(1)\right]$.

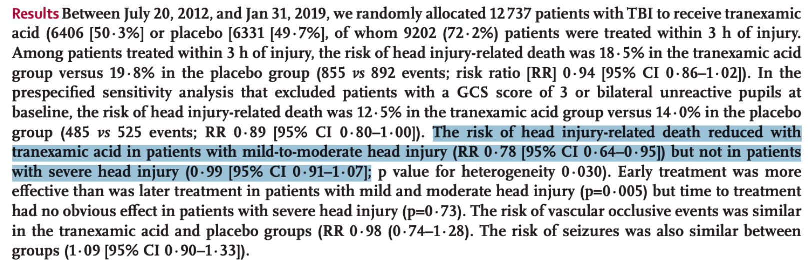
Bonus: $\hat{\tau}_{\mathrm{DM}}$ is an unbiased estimator.

Example of Tranexamic Acid (TXA) on brain injury related death



Effects of tranexamic acid on death, disability, vascular occlusive events and other morbidities in patients with acute traumatic brain injury (CRASH-3): a randomised, placebo-controlled trial

*The CRASH-3 trial collaborators**

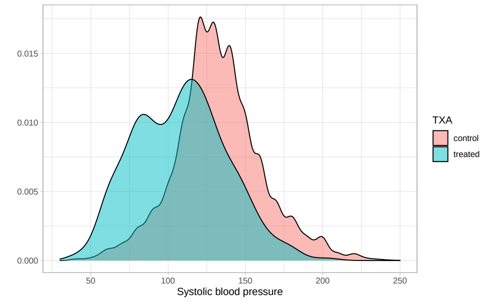


Results Between July 20, 2012, and Jan 31, 2019, we randomly allocated 12 737 patients with TBI to receive tranexamic acid (6406 [50·3%] or placebo [6331 [49·7%], of whom 9202 (72·2%) patients were treated within 3 h of injury. Among patients treated within 3 h of injury, the risk of head injury-related death was 18·5% in the tranexamic acid group versus 19·8% in the placebo group (855 *vs* 892 events; risk ratio [RR] 0·94 [95% CI 0·86–1·02]). In the prespecified sensitivity analysis that excluded patients with a GCS score of 3 or bilateral unreactive pupils at baseline, the risk of head injury-related death was 12·5% in the tranexamic acid group versus 14·0% in the placebo group (485 *vs* 525 events; RR 0·89 [95% CI 0·80–1·00]). The risk of head injury-related death reduced with tranexamic acid in patients with mild-to-moderate head injury (RR 0·78 [95% CI 0·64–0·95]) but not in patients with severe head injury (0·99 [95% CI 0·91–1·07]; p value for heterogeneity 0·030). Early treatment was more effective than was later treatment in patients with mild and moderate head injury (p=0·005) but time to treatment had no obvious effect in patients with severe head injury (p=0·73). The risk of vascular occlusive events was similar in the tranexamic acid and placebo groups (RR 0·98 (0·74–1·28). The risk of seizures was also similar between groups (1·09 [95% CI 0·90–1·33]).

Source: www.thelancet.com Vol 394, 2019

Non-randomized data

Non-experimental studies – called Observational data – are often **confounded**, meaning that treated patients are not exactly like untreated ones.



In other words, the conditional independence does no longer hold,

$$\mathbb{E}\left[Y \mid A = a\right] \neq \mathbb{E}[Y^{(a)}]$$

Source: Traumabase data, own plot.

Motivation

Question from clinicians[a]

[a] www.traumabase.eu

Can we estimate the average effect of Tranexamic Acid (TXA) on brain-injured death (TBI) on the French population in trauma centers?

Motivation

Question from clinicians[a]

[a] www.traumabase.eu

Can we estimate the average effect of Tranexamic Acid (TXA) on brain-injured death (TBI) on the French population in trauma centers?

Data sources and evidence at hand:

CRASH3

- Multi-centric RCT over 29 counties,
- $\sim$ 9 000 individuals,
- High **internal** validity
- Measured a positive effect of TXA on moderate injured patients

Traumabase

- Observational sample,
- $\sim$ 30 000 individuals,
- High **external** validity
- Observational analysis can not reject the null hypothesis of no effect (and pushing toward negative effect).

Motivation

Question from clinicians[a]

[a] www.traumabase.eu

Can we estimate the average effect of Tranexamic Acid (TXA) on brain-injured death (TBI) on the French population in trauma centers?

Data sources and evidence at hand:

CRASH3

- Multi-centric RCT over 29 counties,
- $\sim$ 9 000 individuals,
- High **internal** validity
- Measured a positive effect of TXA on moderate injured patients

Traumabase

- Observational sample,
- $\sim$ 30 000 individuals,
- High **external** validity
- Observational analysis can not reject the null hypothesis of no effect (and pushing toward negative effect).

Is there a paradox?

Possible explanations

- Treatment and outcome are not exactly the same[3],

[3] Sara Lodi & Miguel A Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

Possible explanations

- Treatment and outcome are not exactly the same[3],

[3] Sara Lodi & Miguel A Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

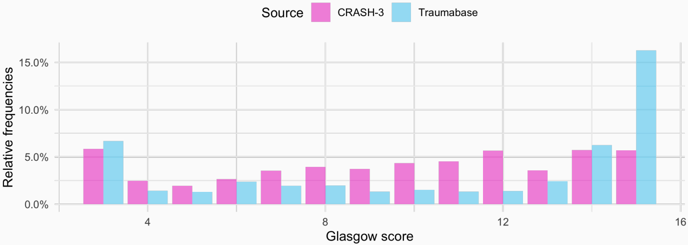
Possible explanations

- Treatment and outcome are not exactly the same[3],
- Traumabase's analysis suffers from unobserved confounding,

[3] Sara Lodi & Miguel A Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

Possible explanations

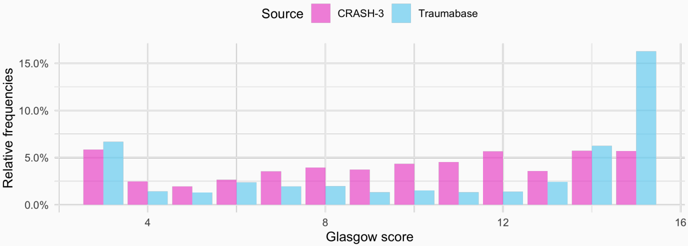
- Treatment and outcome are not exactly the same[3],
- Traumabase's analysis suffers from unobserved confounding,
- Populations are different.



[3] Sara Lodi & Miguel A Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

Possible explanations

- Treatment and outcome are not exactly the same[3],
- Traumabase's analysis suffers from unobserved confounding,
- Populations are different.



Could we generalize the evidence from the trial to the Traumabase?
Would a trial directly conducted on the Traumabase's individuals had found the same effect?

[3] Sara Lodi & Miguel A Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

This topic seems to be a burning question

Within the last 7 days at Stanford:

- Last Thursday, in the Biostatistic seminar, talk about eligibility criteria in oncology, distributional shifts, and validity of trials,
- Yesterday in stat seminar ”*Is empirical medical research doomed? Generalizability of predictions and treatment effect estimates*”,

[4] Stephen R. Cole, Elizabeth A. Stuart. (2010) Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial, *American Journal of Epidemiology*

[5] Elias Bareinboim & Judea Pearl. (2016). Causal inference & the data-fusion problem. *PNAS*.

[6] Rothman & Greenland, *Modern Epidemiology*

This topic seems to be a burning question

Within the last 7 days at Stanford:

- Last Thursday, in the Biostatistic seminar, talk about eligibility criteria in oncology, distributional shifts, and validity of trials,
- Yesterday in stat seminar ”*Is empirical medical research doomed? Generalizability of predictions and treatment effect estimates*”,

This question is found under many names in literature,

- Generalization[4],
- Transportability, data fusion, or recoverability[5],
- External validity,
- Standardization[6],
- . . .

[4] Stephen R. Cole, Elizabeth A. Stuart. (2010) Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial, *American Journal of Epidemiology*

[5] Elias Bareinboim & Judea Pearl. (2016). Causal inference & the data-fusion problem. *PNAS*.

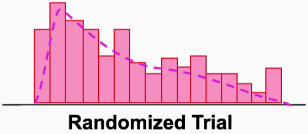
[6] Rothman & Greenland, *Modern Epidemiology*

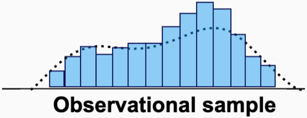
Combining data for generalizability or transportability

Context

Consider that a policy maker has at hand:

- an already conducted trial about a treatment or policy ($\rightarrow \hat{\tau}_1$),
- and a sample of the target population of interest ($\hat{\tau}$?).

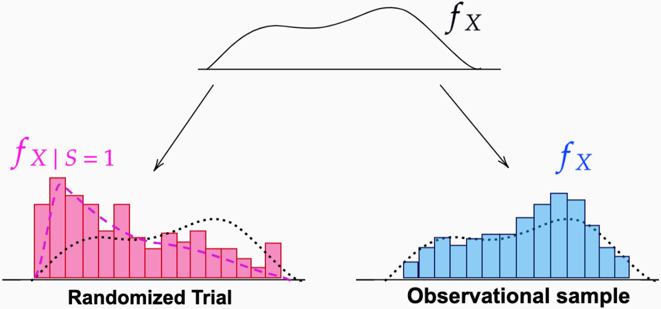




Context

Consider that a policy maker has at hand:

- an already conducted trial about a treatment or policy ($\rightarrow \hat{\tau}_1$),
- and a sample of the target population of interest ($\hat{\tau}$?).



Notations

Notations

Using the potential outcome framework (Neyman, 1923), we denote

- 💊 A the treatment,
- 🩺 X the covariates,
- 🌡️ Y the observed outcome,
- 🙋‍♀️ S trial selection or eligibility.

Identical to the classical framework

Notations

Notations

Using the potential outcome framework (Neyman, 1923), we denote

- 💊 A the treatment,
- 🩺 X the covariates,
- 🌡️ Y the observed outcome,
- 🙋‍♀️ S trial selection or eligibility.

Identical to the classical framework

	Set	S	X_1	X_2	X_3	A	$Y(0)$	Y(1)
1	$\mathcal{R}$	1	1.1	20	5.4	1	?	24.1
...	$\mathcal{R}$	1		...		...	...	
$n-1$	$\mathcal{R}$	1	-6	45	8.3	0	26.3	?
n	$\mathcal{R}$	1	0	15	6.2	1	?	23.5
$n+1$	$\mathcal{O}$	?(0)	-2	52	7.1	NA	NA	NA
$n+2$	$\mathcal{O}$	?(1)	-1	35	2.4	NA	NA	NA
...	$\mathcal{O}$	?(0)		...		NA	NA	NA
$n+m$	$\mathcal{O}$	?(1)	-2	22	3.4	NA	NA	NA

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

$$f_{X|S=1} \neq f_X$$

$$\Rightarrow \underbrace{\tau_1 = \mathbb{E}[Y(1) - Y(0)|S=1]}_{\text{ATE in the RCT}} \neq \underbrace{\mathbb{E}[Y(1) - Y(0)] = \tau}_{\text{Target ATE}}$$

⚠️ We consider a non-nested design.

Generalization's *causal* assumptions.

Ignorability on trial participation

$$\{Y(0), Y(1)\} \perp S \mid X$$

- Transportability[7] of the CATE $\implies \underbrace{\mathbb{E}\left[Y(1) - Y(0) \mid X = x, S = 1\right]}_{:=\tau_1(X)} = \underbrace{\mathbb{E}\left[Y(1) - Y(0) \mid X = x\right]}_{:=\tau(X)},$
- Corresponding to shifted treatment effect modifier.

Sampling score overlap

$$\mathbb{P}(S_i = 1 \mid X_i = x) \quad \forall x \in \mathcal{X}.$$

Assume overlap, i.e. $\mathbb{P}(S_i = 1 \mid X_i = x) \geq c > 0, \quad \forall x \in \mathcal{X}$ and some constant c.

- Every individuals in the target population could have been recruited,
- Similar to ATT or ATC assumptions (asymetric).

[7] Depend on the treatment effect metric

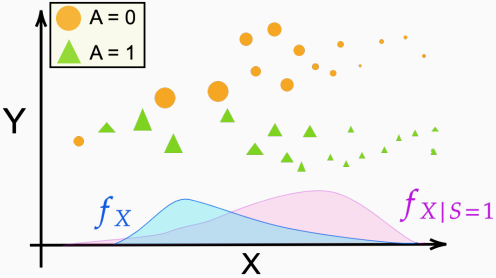
Reweighting

Identifiability

$$\tau = \mathbb{E}\left[\frac{f(X)}{f(X \mid S=1)}\left(\frac{AY}{e_1(X)} - \frac{(1-A)Y}{1-e_1(X)}\right) \mid S=1\right],$$

where $e_1(X) = \mathbb{P}(A = 1 \mid X, S = 1)$.

Intuition

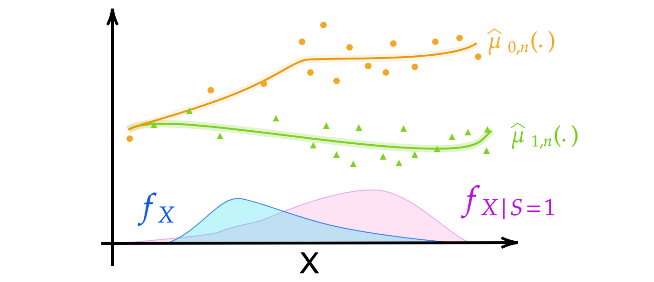


Outcome modeling

Identifiability

$$\tau = \mathbb{E}\left[\underbrace{\mathbb{E}\left[Y(1) \mid X, A=1, S=1\right]}_{:=\mu_1(X)} - \underbrace{\mathbb{E}\left[Y(0) \mid X, A=0, S=1\right]}_{:=\mu_0(X)}\right],$$

Intuition



Estimators and consistency

Inverse probability of sampling weighting (IPSW)

Definition - Stuart et al. (2011); Buchanan et al. (2018)

The IPSW estimator is denoted $\hat{\tau}_{IPSW,n,m}$, and defined as

$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n}\sum_{i=1}^{n} \frac{n}{m}\frac{Y_i}{\hat{\alpha}_{n,m}(X_i)}\left(\frac{A_i}{e_1(X_i)} - \frac{1-A_i}{1-e_1(X_i)}\right),$$

where $\hat{\alpha}_{n,m}$ is an estimate of the odd ratio of the indicatrix of being in the RCT:

Sampling bias or two populations point of view?

$$\text{Odds } \alpha(x) = \frac{\mathbb{P}(i \in \mathcal{R} \mid \exists i \in \mathcal{R} \cup \mathcal{O}, X_i = x)}{\mathbb{P}(i \in \mathcal{O} \mid \exists i \in \mathcal{R} \cup \mathcal{O}, X_i = x)} = \underbrace{\frac{\mathbb{P}(i \in \mathcal{R})}{\mathbb{P}(i \in \mathcal{O})}}_{\sim \frac{n}{m}} \times \underbrace{\frac{\mathbb{P}(X_i = x \mid i \in \mathcal{R})}{\mathbb{P}(X_i = x \mid i \in \mathcal{O})}}_{\frac{f(x|S=1)}{f(x)} = \frac{\mathbb{P}(S=1)}{\mathbb{P}(S=1|X=x)}}$$

where $\alpha(.)$ is the odds ratio of being in the RCT versus observational data conditioned to the covariates.

IPSW consistency

IPSW nuisance parameters consistency's assumption

- $\sup_{x \in \mathcal{X}} \left| \frac{n}{m \hat{\alpha}_{n,m}(x)} - \frac{f_X(x)}{f_{X|S=1}(x)} \right| = \varepsilon_{n,m} \xrightarrow{a.s.} 0$, when $n, m \to \infty$,
- for all n, m large enough $\mathbb{E}[\varepsilon_{n,m}^2]$ exists and $\mathbb{E}[\varepsilon_{n,m}^2] \xrightarrow{a.s.} 0$, when $n, m \to \infty$.

Theorem - IPSW consistency and asymptotic normality

Under causal and consistency assumption, $\hat{\tau}_{\text{IPSW},n,m}$ converges toward τ in L^1 norm,

$$\hat{\tau}_{\text{IPSW},n,m} \xrightarrow[n,m \to \infty]{L^1} \tau.$$

Providing that the potential outcomes are square integrable,

$$\sqrt{n} \left(\hat{\tau}_{\text{IPSW},n,m} - \tau \right) \xrightarrow{d} \mathcal{N}\left(0, V_{\text{IPSW}}\right),$$

where

$$V_{\text{IPSW}} = \frac{1}{n} \left(\mathbb{E}\left[\left(\frac{f_X(x)}{f_{X|S=1}(x)} \right)^2 \left(\frac{(Y(0))^2}{1 - e(X)} + \frac{(Y(1))^2}{e(X)} \right) \mid S = 1 \right] - \tau^2 \right).$$

Outcome regression (G-formula)

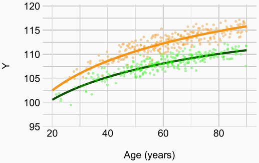
Definition

The G-formula is denoted $\hat{\tau}_{G,n,m}$, and defined as

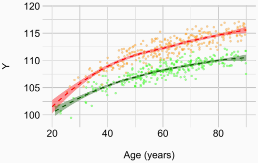
$$\hat{\tau}_{G,n,m} = \frac{1}{m} \sum_{i=n+1}^{n+m} \left(\hat{\mu}_{1,n}(X_i) - \hat{\mu}_{0,n}(X_i) \right),$$

where $\hat{\mu}_{a,n}(X_i)$ is an estimator of $\mu_a(X_i)$ obtained on the RCT sample.

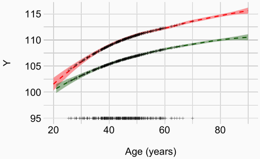
1. Consider RCT data



2. Estimate $\hat{\mu}_a(.)$



3. Marginalize



G-formula consistency

G-formula nuisance parameters consistency's assumption

Denoting $\hat{\mu}_{0,n}$ and $\hat{\mu}_{1,n}$ estimators of μ_0 and μ_1 respectively, and $\mathcal{D}_n$ the RCT sample,

(H1-G) For $a \in \{0,1\}$, $\mathbb{E}\left[|\hat{\mu}_{a,n}(X) - \mu_a(X)| \mid \mathcal{D}_n\right] \xrightarrow{p} 0$ when $n \to \infty$,

(H2-G) For $a \in \{0,1\}$, there exist C_1, N_1 so that for all $n \geqslant N_1$, almost surely, $\mathbb{E}[\hat{\mu}_{a,n}^2(X) \mid \mathcal{D}_n] \leqslant C_1$.

Theorem - G-formula consistency and asymptotic normality

Under causal and consistency assumption, $\hat{\tau}_{\mathrm{G},n,m}$ converges toward τ in L^1 norm,

$$\hat{\tau}_{\mathrm{G},n,m} \underset{n,m \to \infty}{\xrightarrow{L^1}} \tau.$$

Augmented IPSW - AIPSW

Definition

The AIPSW estimator is denoted $\hat{\tau}_{AIPSW,n,m}$, and defined as

$$\hat{\tau}_{AIPSW,n,m} = \frac{1}{n}\sum_{i=1}^{n}\frac{n}{m\,\hat{\alpha}_{n,m}(X_i)}\left[\frac{A_i\left(Y_i-\hat{\mu}_{1,n}(X_i)\right)}{e_1(X_i)}-\frac{(1-A_i)\left(Y_i-\hat{\mu}_{0,n}(X_i)\right)}{1-e_1(X_i)}\right]$$

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

On-working consistency proof,

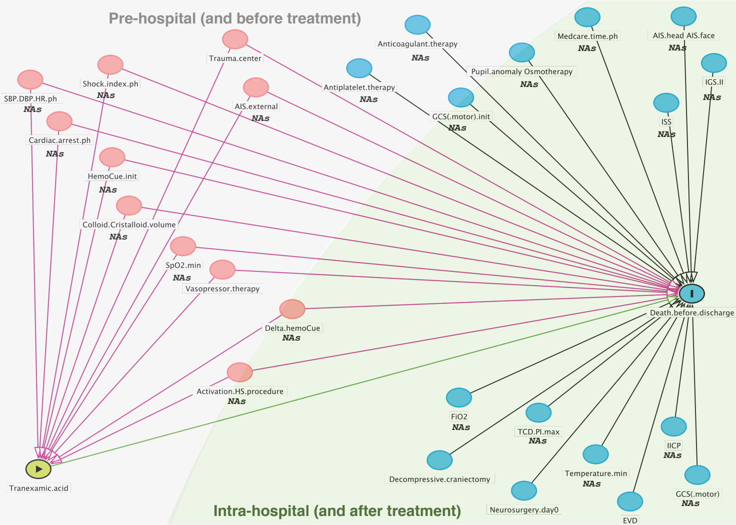
- Require surface-response cross-fitting estimation,
- Asymptotic normality achieved under sufficient convergence rates,
- Probable asymptotic variance being:

$$V_{\text{AIPW}} = \mathbb{E}\left[\left(\frac{f(X\mid S=1)}{f(X)}\right)^2\left(\frac{(Y(1)-\mu_1(X))^2}{e(X)}+\frac{(Y(0)-\mu_0(X))^2}{1-e(X)}\right)\mid S=1\right]+\text{Var}[\tau(X)].$$

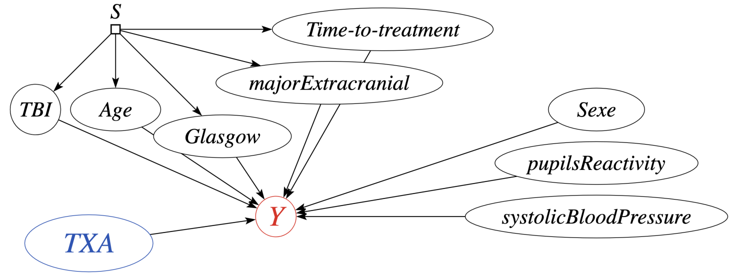
Toward the application

Covariate selection in causal inference

With my advisors and collaborators we currently apply the Delphi method.



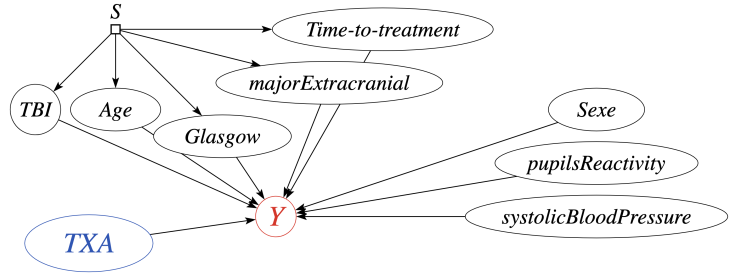
Covariate selection in generalization[8]



Structural causal model representing treatment, outcome, inclusion criteria with S and other predictors of outcome.

[8] and a SCM comment 🎁

Covariate selection in generalization[8]



Structural causal model representing treatment, outcome, inclusion criteria with S and other predictors of outcome.

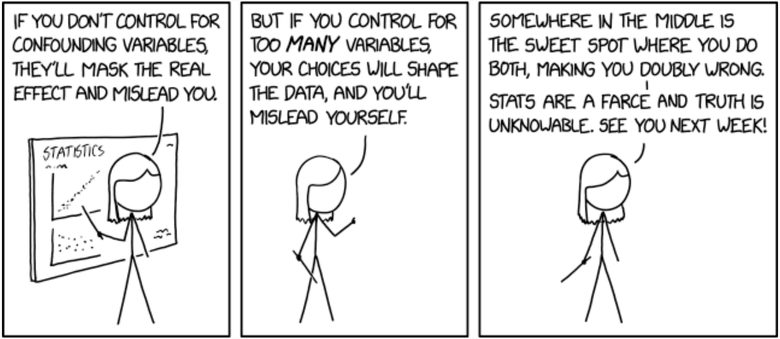
Selecting covariates in any application with a causal question is a challenge for:

- Identification,
- Statistical efficiency.

$\Longrightarrow$ ongoing work...

[8] and a SCM comment 🎁

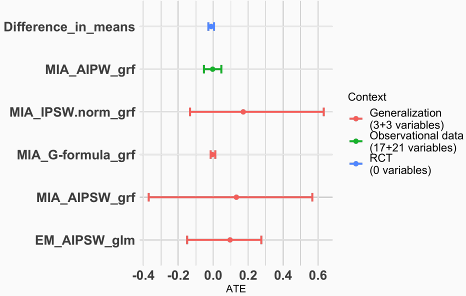
N.B.: To find X *really* is a tricky task!


IF YOU DON'T CONTROL FOR CONFOUNDING VARIABLES, THEY'LL MASK THE REAL EFFECT AND MISLEAD YOU.
STATISTICS
BUT IF YOU CONTROL FOR TOO MANY VARIABLES, YOUR CHOICES WILL SHAPE THE DATA, AND YOU'LL MISLEAD YOURSELF.
SOMEWHERE IN THE MIDDLE IS THE SWEET SPOT WHERE YOU DO BOTH, MAKING YOU DOUBLY WRONG.
STATS ARE A FARCE AND TRUTH IS UNKNOWABLE. SEE YOU NEXT WEEK!

Source: xkcd.com/2560/

On the Traumabase

Comparison with trials and observational data results[9][10]



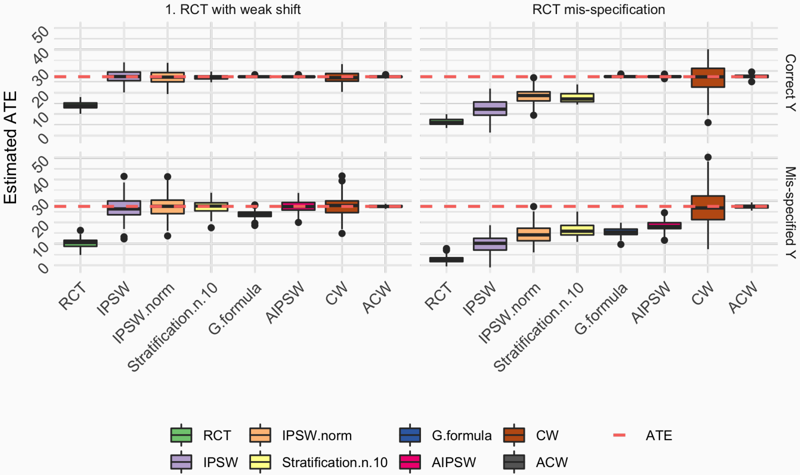
Issues:

- Heterogeneous point estimates,
- (Very) High variance,
- Heterogeneous missing values patterns.

[9] MIA = Missing Incorporated in Attributes (MIA, Twala et al. 2008; implemented in `grf`); EM, Jiang et al. (2018)

[10] Mayer et al. (2020) Doubly Robust Treatment Effect Estimation with Missing Attributes. *Annals of Applied Statistics.*

Applications with simulated data



Additional estimators are represented in these simulations, namely CW and ACW. See Yang et al. (2020) Improving trial generalizability using observational studies, *Biometrics*.

Sensitivity analysis

What if a covariate is missing / not observed?

	Set	S	X_1	X_2	X_3	A	$Y(0)$	Y(1)
1	$\mathcal{R}$	1	NA	NA	5.4	1	?	24.1
...	$\mathcal{R}$	1		...		...	...	
$n-1$	$\mathcal{R}$	1	NA	NA	8.3	0	26.3	?
n	$\mathcal{R}$	1	NA	NA	6.2	1	?	23.5
$n+1$	$\mathcal{O}$	?(0)	NA	52	NA	NA	NA	NA
$n+2$	$\mathcal{O}$	?(1)	NA	35	NA	NA	NA	NA
...	$\mathcal{O}$	?(0)	NA	...		NA	NA	NA
$n+m$	$\mathcal{O}$	?(1)	NA	22	NA	NA	NA	NA

X_1 totally missing, while X_2, X_3 are partially observed.

$$X = X_{\text{mis}} \cup X_{obs}$$

What if a covariate is missing / not observed?

	Set	S	X_1	X_2	X_3	A	$Y(0)$	Y(1)
1	$\mathcal{R}$	1	NA	NA	5.4	1	?	24.1
...	$\mathcal{R}$	1		...		...	...	
$n-1$	$\mathcal{R}$	1	NA	NA	8.3	0	26.3	?
n	$\mathcal{R}$	1	NA	NA	6.2	1	?	23.5
$n+1$	$\mathcal{O}$	?(0)	NA	52	NA	NA	NA	NA
$n+2$	$\mathcal{O}$	?(1)	NA	35	NA	NA	NA	NA
...	$\mathcal{O}$	?(0)	NA	...		NA	NA	NA
$n+m$	$\mathcal{O}$	?(1)	NA	22	NA	NA	NA	NA

X_1 totally missing, while X_2, X_3 are partially observed.

$$X = X_{\text{mis}} \cup X_{obs}$$

$$\{Y(1), Y(0)\} \not\perp\!\!\!\perp S \mid X_{obs}$$

What if a covariate is missing / not observed?

	Set	S	X_1	X_2	X_3	A	$Y(0)$	Y(1)
1	$\mathcal{R}$	1	NA	NA	5.4	1	?	24.1
...	$\mathcal{R}$	1		...		...	...	
$n-1$	$\mathcal{R}$	1	NA	NA	8.3	0	26.3	?
n	$\mathcal{R}$	1	NA	NA	6.2	1	?	23.5
$n+1$	$\mathcal{O}$	?(0)	NA	52	NA	NA	NA	NA
$n+2$	$\mathcal{O}$	?(1)	NA	35	NA	NA	NA	NA
...	$\mathcal{O}$	?(0)	NA	...		NA	NA	NA
$n+m$	$\mathcal{O}$	?(1)	NA	22	NA	NA	NA	NA

X_1 totally missing, while X_2, X_3 are partially observed.

$$X = X_{\text{mis}} \cup X_{obs}$$

$$\{Y(1), Y(0)\} \not\perp S \mid X_{obs}$$

Is there a way to assess how dramatic the situation is?

- Andrews and Oster (2019) consider a totally unobserved covariate;
- Nguyen et al. (2018) study a missing covariate in observational;
- Practitioners sometimes rely on imputation, see Lesko et al. (2016);
- Pearl and Bareinboim (2011) propose a proxy (though not in the generalization set-up);
- Nie et al. (2021) considers a totally unobserved covariate with an approach inspired from Rosenbaum.

Sensitivity analysis in a nutshell

Source: YouTube's screenshot.

How strong should you push the man before he falls?

Generalization’s case and model chosen

Intuition

A poorly shifted missing covariate and/or a weak treatment effect missing covariate will lead to a small bias.

Generalization's case and model chosen

Intuition

A poorly shifted missing covariate and/or a weak treatment effect missing covariate will lead to a small bias.

Assumption on the generative model

Assume that $X, Y^{(0)}, Y^{(1)} \in \mathbb{R}^{p+2}$, along with assuming there exist $\delta \in \mathbb{R}^{p}$, $\sigma \in \mathbb{R}^{+}$, any function $g \in \mathsf{L}^2(\mathcal{X} \to \mathbb{R})$ such that:

$$\begin{aligned} Y &= g(X) + A\langle X, \delta\rangle + \varepsilon \\ &= g(X) + A\left(\langle X_{obs}, \delta_{obs}\rangle + \langle X_{mis}, \delta_{mis}\rangle\right) + \varepsilon \end{aligned}$$

where $\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right), \mathbb{E}[\varepsilon \mid X] = 0$.

Generalization's case and model chosen

Intuition

A poorly shifted missing covariate and/or a weak treatment effect missing covariate will lead to a small bias.

Assumption on the generative model

Assume that $X, Y^{(0)}, Y^{(1)} \in \mathbb{R}^{p+2}$, along with assuming there exist $\delta \in \mathbb{R}^p$, $\sigma \in \mathbb{R}^+$, any function $g \in \mathrm{L}^2(\mathcal{X} \to \mathbb{R})$ such that:

$$\begin{aligned} Y &= g(X) + A\langle X, \delta\rangle + \varepsilon \\ &= g(X) + A\left(\langle X_{obs}, \delta_{obs}\rangle + \langle X_{mis}, \delta_{mis}\rangle\right) + \varepsilon \end{aligned}$$

where $\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right), \mathbb{E}[\varepsilon \mid X] = 0$.

Is it a strong assumption?

When assuming $Y^{(0)}, Y^{(1)} \in \mathbb{R}^{p+2}$ the treatment is automatically additively separable,

$$Y(A) = g(X) + A\,\tau(X) + \varepsilon.$$

Note that if $\tau(X)$ is a constant, then $\tau_1 = \tau$.

Transportability of covariates relationship

Assumption on covariates

The distribution of X is Gaussian, that is, $X \sim \mathcal{N}(\mu, \Sigma)$, and transportability of Σ is true, that is, $X \mid S = 1 \sim \mathcal{N}(\mu_{RCT}, \Sigma)$.

- Relation between covariates are preserved in the sources, while the expectancy can be different explaining the bias,
- Allows to prevent from assuming independence.

The plausibility of this assumption can be partially-assessed through a statistical test on $\mathbf{\Sigma}_{obs,obs}$ for example Box's M test (Box, 1949), supported with vizualizations (Friendly and Sigal, 2020)[a].

[a] This part will be illustrated on the application.

Asymptotic bias

Theorem

Assume that the partially linear generative model holds, along with the transportability of covariates relationship. Let B be the following quantity:

$$B = \sum_{j \in mis} \delta_j \left(\mathbb{E}[X_j] - \mathbb{E}[X_j \mid S = 1] - \boldsymbol{\Sigma}_{j,obs} \boldsymbol{\Sigma}_{obs,obs}^{-1} (\mathbb{E}[X_{obs}] - \mathbb{E}[X_{obs} \mid S = 1]) \right),$$

Consider a procedure $\hat{\tau}_{n,m}$ that estimates τ with no asymptotic bias. Let $\hat{\tau}_{n,m,obs}$ be the same procedure but trained on observed data only, then

$$\tau - \lim_{n,m \to \infty} \mathbb{E}[\hat{\tau}_{n,m,obs}] = B.$$

where $\boldsymbol{\Sigma}_{obs,obs}$ is the sub matrix of $\boldsymbol{\Sigma}$ corresponding to observed index rows and columns, and $\boldsymbol{\Sigma}_{j,obs}$ is the row j with column corresponding to observed index of $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{mis,mis} & \boldsymbol{\Sigma}_{mis,obs} \\ \hline \boldsymbol{\Sigma}_{mis,obs} & \boldsymbol{\Sigma}_{obs,obs} \end{array} \right)$$

Toward sensitivity analysis

"Translating expert judgments into a bias."

Assume the covariate is missing in the RCT

$$B = \underbrace{\delta_{mis}}_{X_{mis}\text{'s strength}} \left(\underbrace{\mathbb{E}[X_{mis}] - \mathbb{E}[X_{mis} \mid S = 1]}_{\text{Shift of } X_{mis}:\ \Delta_m} - \underbrace{\Sigma_{mis,obs}\Sigma^{-1}_{obs,obs}(\mathbb{E}[X_{obs}] - \mathbb{E}[X_{obs} \mid S = 1])}_{\text{Can be estimated from the data}} \right)$$

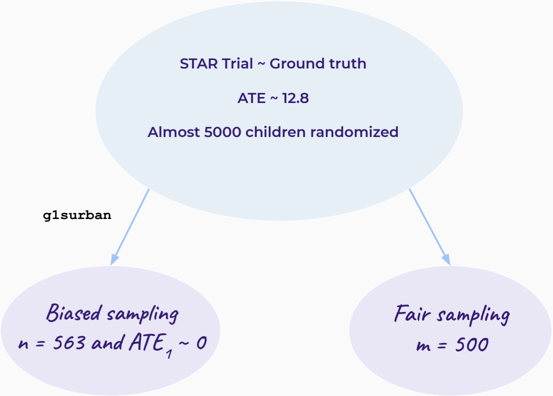
The sensitivity parameters are from two natures:

- δ_{mis}
  CATE coefficient $\sim$ Treatment effect modifier's strength
  $\implies$ ⚠ Complicated to translate,
- $\mathbb{E}[X_{mis}] - \mathbb{E}[X_{mis} \mid S = 1]$
  Covariate shift's strength
  $\implies$ Straightforward to translate.

Semi synthetic simulation

Using the data from the Tennessee Student/Teacher Achievement Ratio (STAR) study (Finn and Achilles, 1990).

We generate a biased RCT sample based on covariate `g1surban` and a representative sample.



STAR Trial ~ Ground truth

ATE ~ 12.8

Almost 5000 children randomized

g1surban

Biased sampling

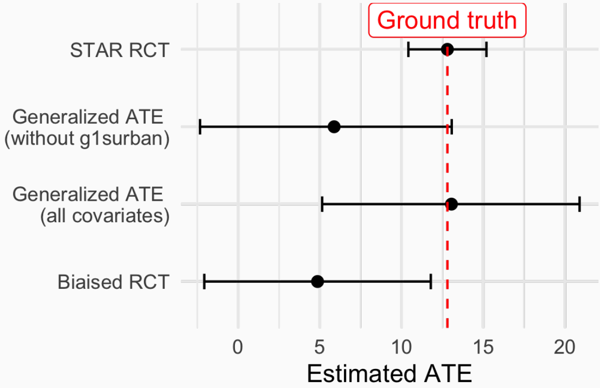
$n = 563$ and $ATE_1 \sim 0$

Fair sampling

$m = 500$

Semi synthetic simulation - Generalization with missing covariate

Bias induced is around 7 points when omitting `g1surban`.



Can the sensitivity analysis estimates the bias when `g1surban` is missing in the observational data but not the RCT?

Semi synthetic simulation - Sensitivity analysis

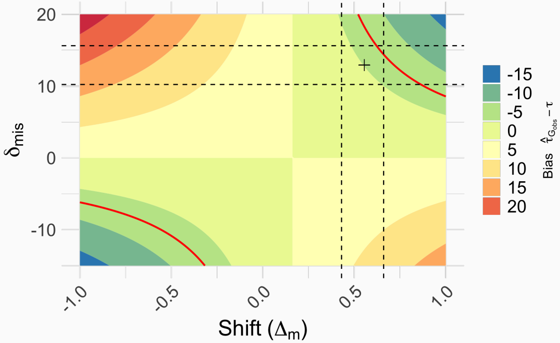
- Δ_m can be proposed by domain expert (interpretable quantity, here the shift in children proportion leaving in suburbs versus city center),
- To estimate δ_{mis}:
  - Learn a model on the observational data,
  - Impute X_{mis} in the RCT,
  - Estimate δ_{mis} with a Robinson procedure.

	Set	S	X_1	X_2	X_3	A	$Y(0)$	$Y(1)$
1	$\mathcal{R}$	1	NA	20	5.4	1	?	24.1
...	$\mathcal{R}$	1		...		...	...	
$n-1$	$\mathcal{R}$	1	NA	45	8.3	0	26.3	?
n	$\mathcal{R}$	1	NA	15	6.2	1	?	23.5
$n+1$	$\mathcal{O}$	?(0)	-2	52	7.1	NA	NA	NA
$n+2$	$\mathcal{O}$	?(1)	-1	35	2.4	NA	NA	NA
...	$\mathcal{O}$	?(0)		...		NA	NA	NA
$n+m$	$\mathcal{O}$	?(1)	-2	22	3.4	NA	NA	NA

Semi synthetic simulation - Sensitivity analysis

- Δ_m can be proposed by domain expert (interpretable quantity, here the shift in children proportion leaving in suburbs versus city center),
- To estimate δ_{mis}:
  - Learn a model on the observational data,
  - Impute X_{mis} in the RCT,
  - Estimate δ_{mis} with a Robinson procedure.

$\Longrightarrow$ then plot a sensitivity map!



Other results - Sketches

Linear imputation?

- Assuming the true linear relation between X_{mis} as a function of X_{obs}, which leads to the optimal imputation $\hat{X}_{mis}$,
- and denoting the oracle estimator $\hat{\tau}_{\infty,\infty,imp}$ aware of these linear model imputation,

Then,

$$\mathbb{E}[\hat{\tau}_{\infty,\infty,imp}] - \tau = \lim_{n,m \to \infty} \mathbb{E}[\hat{\tau}_{n,m,obs}] - \tau$$

Relying on a proxy?

Assume that $X_{mis} \perp\!\!\!\perp X_{obs}$, and that there exist a proxy variable X_{prox} such that,

$$X_{prox} = X_{mis} + \eta$$

where $\mathbb{E}[\eta] = 0$, $\mathbf{Var}[\eta] = \sigma^2_{prox}$, and $\mathbf{Cov}\left(\eta, X_{mis}\right) = 0$,

$$\implies B = \delta_{mis}\, \boldsymbol{\Delta}_m \left(1 - \frac{\sigma^2_{mis}}{\sigma^2_{mis} + \sigma^2_{prox}}\right),$$

where $\boldsymbol{\Delta}_{mis} = \mathbb{E}[X_{mis}] - \mathbb{E}[X_{mis} \mid S = 1]$

Conclusion, open questions & remarks

- This method relies on two key assumptions
  $\Longrightarrow$ *CATE linearity* & **Σ** *transportability*,

- Currently applying generalization to other data,
  $\Longrightarrow$ *Confront statistical assumptions with reality*
  $\Longrightarrow$ *Quantify with several trials the effective external validity bias*

Conclusion, open questions & remarks

- This method relies on two key assumptions
  $\implies$ *CATE linearity* & Σ *transportability*,
- Currently applying generalization to other data,
  $\implies$ *Confront statistical assumptions with reality*
  $\implies$ *Quantify with several trials the effective external validity bias*
- Working on covariate selection and variance
  $\implies$ Extensions of Lunceford and Davidian (2004)
  $\implies$ How non-parametric estimation affects convergence?
- Which covariates for generalization?
  heterogeneities depends on the causal scale chosen

Binary outcome and heterogeneities?

- Physicians usually face binary outcome and are interested in ratio,
- Treatment effect heterogeneity has different meaning depending whether people are interested in the ratio, absolute difference, else.

Sensitivity analysis transposed for binary outcome could be,

$$\ln\left(\frac{\mathbb{P}(Y^{(a)}=1\mid X)}{\mathbb{P}(Y^{(a)}=0\mid X)}\right)=f(X)+a\,\tau(X),$$

such that,

$$\tau_{\text{log-OR}}:=\mathbb{E}\left[\ln\left(\frac{\mathbb{P}(Y^{(1)}=1\mid X)}{\mathbb{P}(Y^{(1)}=0\mid X)}\left(\frac{\mathbb{P}(Y^{(0)}=1\mid X)}{\mathbb{P}(Y^{(0)}=0\mid X)}\right)^{-1}\right)\right]=\mathbb{E}\left[\tau(X)\right]=\sum_{j=1}^{p}\beta_j\mathbb{E}\left[X_j\right].$$

DINA[11] on CRASH3 data

Zijun's work could be applied in this situation, targeting natural parameters.

Covariates	$\hat{\beta}$
Age	0.022
Glasgow	-0.05
Time to treatment	0.05

Many questions:

- Is there a better causal measure for RCT's generalizability?
- How different are the necessary sets to transport a difference versus a ratio?

[11] Zijun Gao & Trevor Hastie, *Estimating Heterogeneous Treatment Effects for General Responses*

Thank you very much for your attention!! 🌹

References i

Andrews, I. and Oster, E. (2019). A simple approximation for evaluating external validity bias. Economics Letters, 178:58–62. Working Paper.

Angrist, J. D. and Pischke, J.-S. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36(3-4):317–346.

Buchanan, A. L., Hudgens, M. G., Cole, S. R., Mollan, K. R., Sax, P. E., Daar, E. S., Adimora, A. A., Eron, J. J., and Mugavero, M. J. (2018). Generalizing evidence from randomized trials using inverse probability of sampling weights. Journal of the Royal Statistical Society: Series A (Statistics in Society), 181:1193–1209.

Cinelli, C. and Pearl, J. (2020). Generalizing experimental results by leveraging knowledge of mechanisms. European journal of epidemiology.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959). Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions. JNCI: Journal of the National Cancer Institute, 22(1):173–203.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (2009). Smoking and lung cancer: recent evidence and a discussion of some questions*. International Journal of Epidemiology, 38(5):1175–1191.

Finn, J. D. and Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. American Educational Research Journal, 27(3):557–577.

Franks, A., D'Amour, A., and Feller, A. (2019). Flexible sensitivity analysis for observational studies without observable implications. Journal of the American Statistical Association, pages 1–38.

Friendly, M. and Sigal, M. (2020). Visualizing tests for equality of covariance matrices. The American Statistician, 74(2):144–155.

Greenhouse, J. B. (2009). Commentary: Cornfield, Epidemiology and Causality. International Journal of Epidemiology, 38(5):1199–1201.

Imbens, G. (2003). Sensitivity to exogeneity assumptions in program evaluation. The American Economic Review.

Kallus, N., Puli, A. M., and Shalit, U. (2018). Removing hidden confounding by experimental grounding. In Advances in neural information processing systems, pages 10888–10897.

Lesko, C. R., Cole, S. R., Hall, H. I., Westreich, D., Miller, W. C., Eron, J. J., Li, J., Mugavero, M. J., and for the CNICS Investigators (2016). The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009–11. International Journal of Epidemiology, 45(1):140–150.

References ii

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. In Statistics in Medicine, pages 2937–2960.

Nguyen, T., Ackerman, B., Schmid, I., Cole, S., and Stuart, E. (2018). Sensitivity analyses for effect modifiers not observed in the target population when generalizing treatment effects from a randomized controlled trial: Assumptions, models, effect scales, data scenarios, and implementation details. PLOS ONE, 13:e0208795.

Nguyen, T. Q., Ebnesajjad, C., Cole, S. R., Stuart, E. A., et al. (2017). Sensitivity analysis for an unobserved moderator in rct-to-target-population generalization of treatment effects. The Annals of Applied Statistics, 11(1):225–247.

Nie, X., Imbens, G., and Wager, S. (2021). Covariate balancing sensitivity analysis for extrapolating randomized trials across locations.

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. Proceedings of the AAAI Conference on Artificial Intelligence, 25(1).

Rosenbaum, J. P. R., o-c oi, D. B. R. D., Rosenbaum, P. R., Rubin, D. B., and Sei, D. (1983). Assessing the sensitivity to an unobserved binary covariate in an observational study with binary outcome. In Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," JASA, pages 212–218.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., and Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. Journal of the Royal Statistical Society: Series A (Statistics in Society), 174:369–386.

Veitch, V. and Zaveri, A. (2020). Sense and sensitivity analysis: Simple post-hoc analysis of bias due to unobserved confounding.

When covariate is partially observed in RCT

$$
\begin{aligned}
\tau &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\
&= \mathbb{E}[g(X) + W\langle X, \delta\rangle \mid W = 1] - \mathbb{E}[g(X) + W\langle X, \delta\rangle \mid W = 0] \\
&= \langle \delta, \mathbb{E}[X]\rangle = \langle \delta_{obs}, \mathbb{E}\left[X_{obs}\right]\rangle + \langle \delta_{mis}, \underbrace{\mathbb{E}\left[X_{mis}\right]}_{\text{Unknown}}\rangle
\end{aligned}
$$

Extension of (Nguyen et al., 2017): $\mathbb{E}\left[Y \mid A, X\right] = \underbrace{g(X)}_{\text{non-linear}} + A\langle \delta, X\rangle$

- Define range for plausible $\mathbb{E}[X_{mis}]$ values
- Estimate δ with Robinson procedure (residuals on residuals) on the RCT [12] [13] that is:
  - Estimate $m(x) = \mathbb{E}[Y \mid X = x, S = 1]$ with non parametric regression,
  - Define transformed features $\tilde{Y} = Y - \hat{m}_n(X)$ and $\tilde{Z} = (W - e_1(X))X$,
  - Estimate $\hat{\delta}$ with OLS regression: $\tilde{Y} \sim \tilde{Z}$.
- Estimate $\mathbb{E}[X_{obs}]$ on the observational dataset
- Compute all possible bias for range of $\mathbb{E}[X_{mis}]$ and return austen plot

[12] Robinson, P. 1988, Root- N-Consistent Semiparametric Regression, *Econometrica*

[13] Nie, X & Wager, S. 2020, Quasi-Oracle Estimation of Heterogeneous Treatment, *Biometrika*

Historical background

In fact, the fear of missing covariate or missing confounder is a central issue in causal inference.

Several methods have been developed so far including:

- **Sensitivity analysis**,
  A well-known example dating back from Cornfield et al. (1959), followed by Rosenbaum et al. (1983); Imbens (2003) and more recently Franks et al. (2019); Veitch and Zaveri (2020); Cinelli and Pearl (2020)
- **Instrumental variables**,
  For example Angrist and Pischke (2008)
- **Experimental grounding**,
  For example Kallus et al. (2018)

Smoking and lung cancer[14]

Formally, suppose that a true causal agent exist, for example hormone producer with a specific gene, and this is denoted B. If people have B, then their disease rate is r_1. If not, their disease rate is r_2 (and we suppose a lower prevalence).

[14] This derivations were inspired from reprint of the original discussion (Greenhouse, 2009; Cornfield et al., 2009).

Smoking and lung cancer[14]

Formally, suppose that a true causal agent exist, for example hormone producer with a specific gene, and this is denoted B. If people have B, then their disease rate is r_1. If not, their disease rate is r_2 (and we suppose a lower prevalence). But instead of B, we observe A, for example the smoking status. Suppose now that, $p(B \mid A) = p_1$ and $p(B \mid \bar{A}) = p_2$, such that the presence of B is correlated with A, so $p_1 > p_2$.

[14] This derivations were inspired from reprint of the original discussion (Greenhouse, 2009; Cornfield et al., 2009).

Smoking and lung cancer[14]

Formally, suppose that a true causal agent exist, for example hormone producer with a specific gene, and this is denoted B. If people have B, then their disease rate is r_1. If not, their disease rate is r_2 (and we suppose a lower prevalence). But instead of B, we observe A, for example the smoking status. Suppose now that, $p(B \mid A) = p_1$ and $p(B \mid \bar{A}) = p_2$, such that the presence of B is correlated with A, so $p_1 > p_2$. In practice, when observing A, then an apparent rate of disease is observed in association. We denote R_A this rate, and we can write is as $p_1 r_1 + (1 - p_1)\, r_2 = R_A$.

Because $R_A > R_{\bar{A}}$, and doing a bit of computation gives ...

$$\frac{p_1}{p_2} = \frac{R_A}{R_{\bar{A}}} + \frac{r_2}{p_2 r_1}\left(\frac{R_A}{R_{\bar{A}}}(1 - p_2) - (1 - p_1)\right).$$

Because $p_1 > p_2$ and $R_A > R_{\bar{A}}$, the third term is positive, therefore, $\frac{R_A}{R_{\bar{A}}} < \frac{p_1}{p_2}$.

[14] This derivations were inspired from reprint of the original discussion (Greenhouse, 2009; Cornfield et al., 2009).

Smoking and lung cancer[14]

Formally, suppose that a true causal agent exist, for example hormone producer with a specific gene, and this is denoted B. If people have B, then their disease rate is r_1. If not, their disease rate is r_2 (and we suppose a lower prevalence). But instead of B, we observe A, for example the smoking status. Suppose now that, $p(B \mid A) = p_1$ and $p(B \mid \bar{A}) = p_2$, such that the presence of B is correlated with A, so $p_1 > p_2$. In practice, when observing A, then an apparent rate of disease is observed in association. We denote R_A this rate, and we can write is as $p_1 r_1 + (1 - p_1)\, r_2 = R_A$.

Because $R_A > R_{\bar{A}}$, and doing a bit of computation gives ...

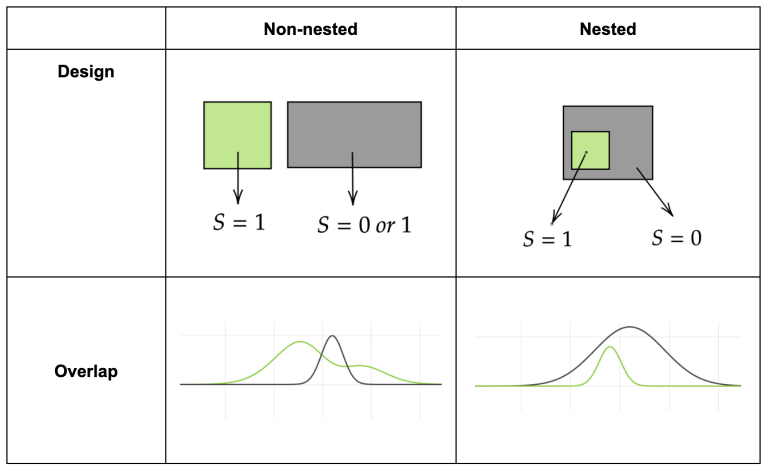
$$\frac{p_1}{p_2} = \frac{R_A}{R_{\bar{A}}} + \frac{r_2}{p_2 r_1}\left(\frac{R_A}{R_{\bar{A}}}(1 - p_2) - (1 - p_1)\right).$$

Because $p_1 > p_2$ and $R_A > R_{\bar{A}}$, the third term is positive, therefore, $\frac{R_A}{R_{\bar{A}}} < \frac{p_1}{p_2}$.

🗨 *If cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer (i.e. $\frac{R_A}{R_{\bar{A}}} = 9$), and this is not because cigarette smoke is a causal agent, but only because cigarette smokers produce hormone X, then the proportion of hormone-X producers among cigarette smokers must be at least 9 times greater than nonsmokers (i.e. $\frac{p_1}{p_2} > 9$).* – Cornfield, 1956

[14] This derivations were inspired from reprint of the original discussion (Greenhouse, 2009; Cornfield et al., 2009).

Nested and non-nested



	Non-nested	Nested
Design	$S = 1$ $S = 0 \; or \; 1$	$S = 1$ $S = 0$
Overlap		