Generalizing a causal effect from a trial to a target population

Bénédicte Colnet — Wednesday, 28 June 2023 — Ph.D. Defense

Jury members

Ph.D. advisors	Reviewers	Examiners
- Julie Josse (Inria) - Erwan Scornet (École polytechnique) - Gaël Varoquaux (Inria)	- Nicolai Meinshausen (ETH Zürich) - Stinj Vansteelandt (Ghent University)	- Trevor Hastie (Stanford) - Erwan Le Pennec (École polytechnique) - Elizabeth Ogburn (John Hopkins) - Philippe Ravaud (Paris' hospitals)

ECOLE DOCTORALE DE MATHEMATIQUES HADAMARD

Outline

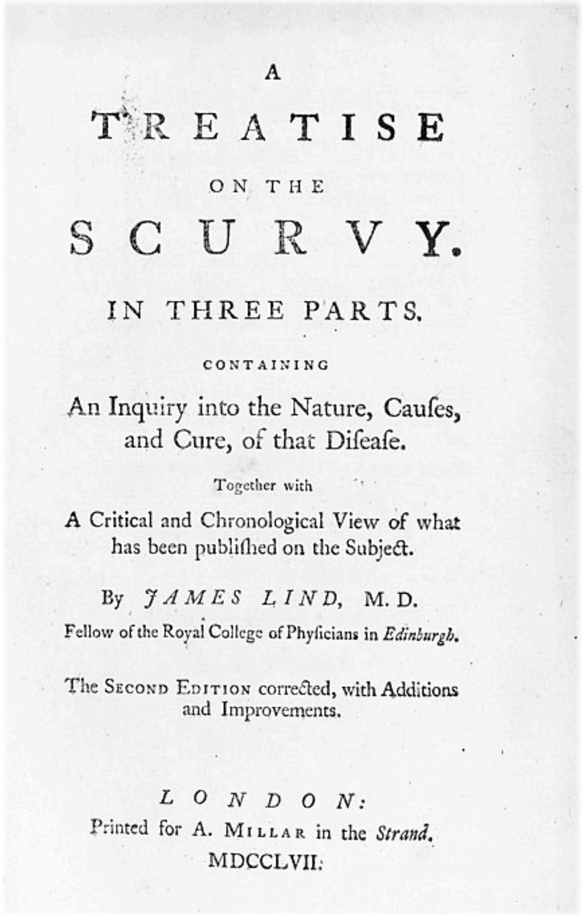
1. **Introduction**
   A. Motivating example from critical care medicine
   B. State-of-the-art

— Focus on two contributions —

2. **Finite and large sample analysis of the IPSW estimator**

3. **Extension to different causal measures**

A longstanding presence of Randomized Controlled Trials (RCTs)



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:

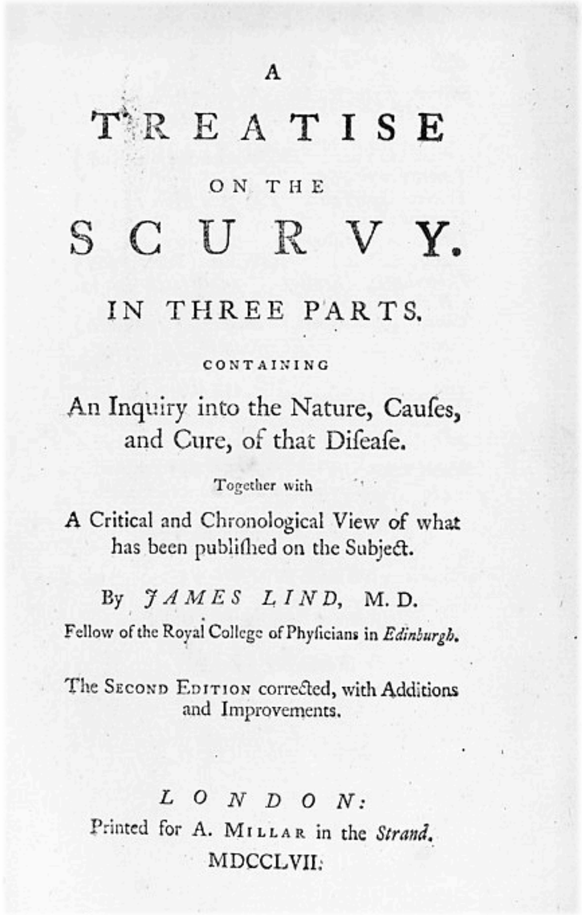
Printed for A. MILLAR in the *Strand*.

MDCCLVII.

James Lind experiment on scorbut in **1757**

Source: Wikipedia

A longstanding presence of Randomized Controlled Trials (RCTs) … now being the gold-standard



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.

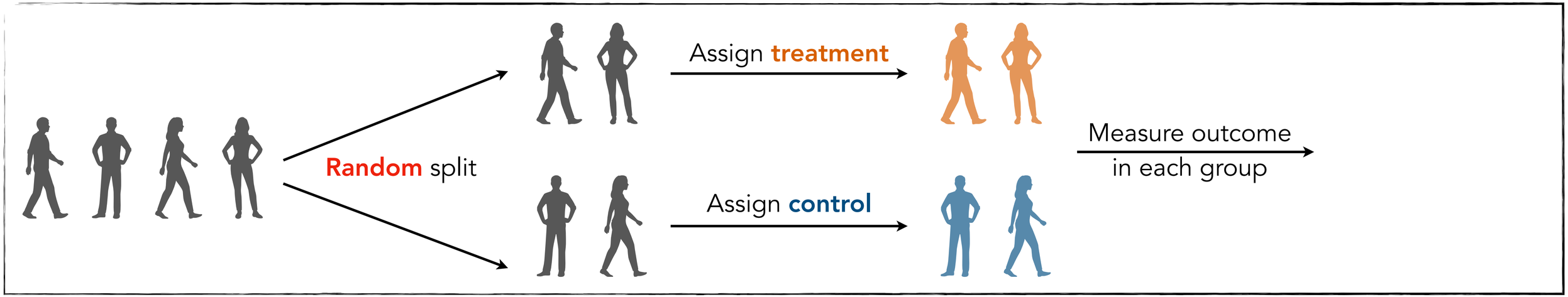
James Lind experiment on scorbut in **1757**
Source: Wikipedia

Drug Trials Snapshot	Active Ingredient	Date of FDA Approval	What is it Approved For
CABENUVA	cabotegravir and rilpivirine	January 20, 2021	Treatment of HIV-1 infection.
LUPKYNIS	voclosporin	January 22, 2021	Treatment of lupus nephritis
VERQUVO	vericiguat	January 19, 2021	Treatment of chronic heart failure
GEMTESA	vibegron	December 23, 2020	Treatment of symptoms of overactive bladder
EBANGA	ansuvimab-zykl	December 21, 2020	Treatment of Zaire ebolavirus infection
ORGOVYX	relugolix	December 18, 2020	Treatment of advanced prostate cancer

Recently approved drugs by the Food and Drug Administration (FDA), all with their corresponding RCT snapshot and information.
Source: www.fda.gov - **2022**

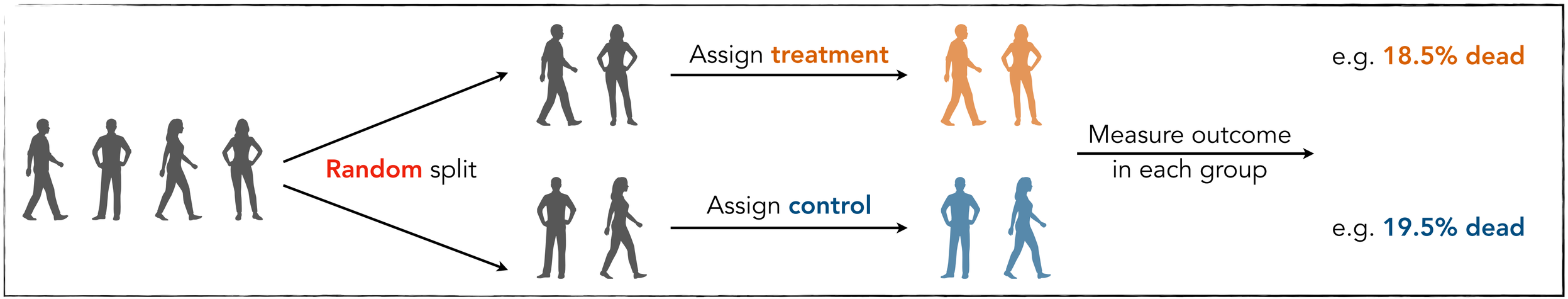
RCTs' principle : estimating a causal effect

Principle



RCTs' principle : estimating a causal effect

Principle

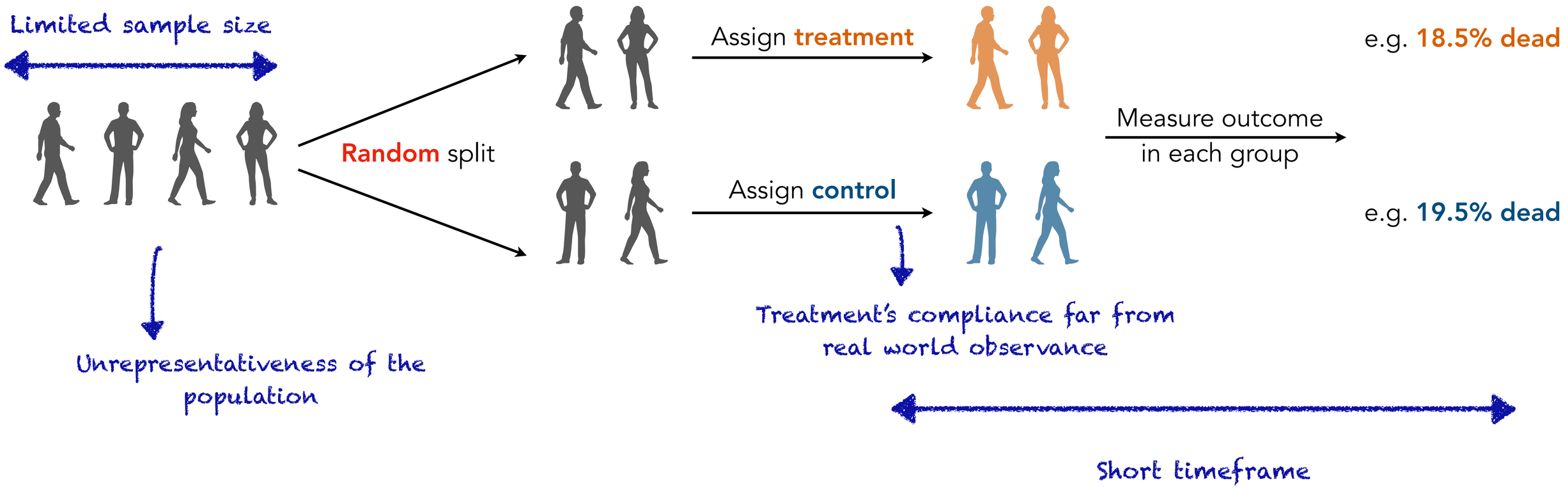


In practice : the CRASH-3 trial investigating Tranexamic Acid effect on brain injured (TBI) related death

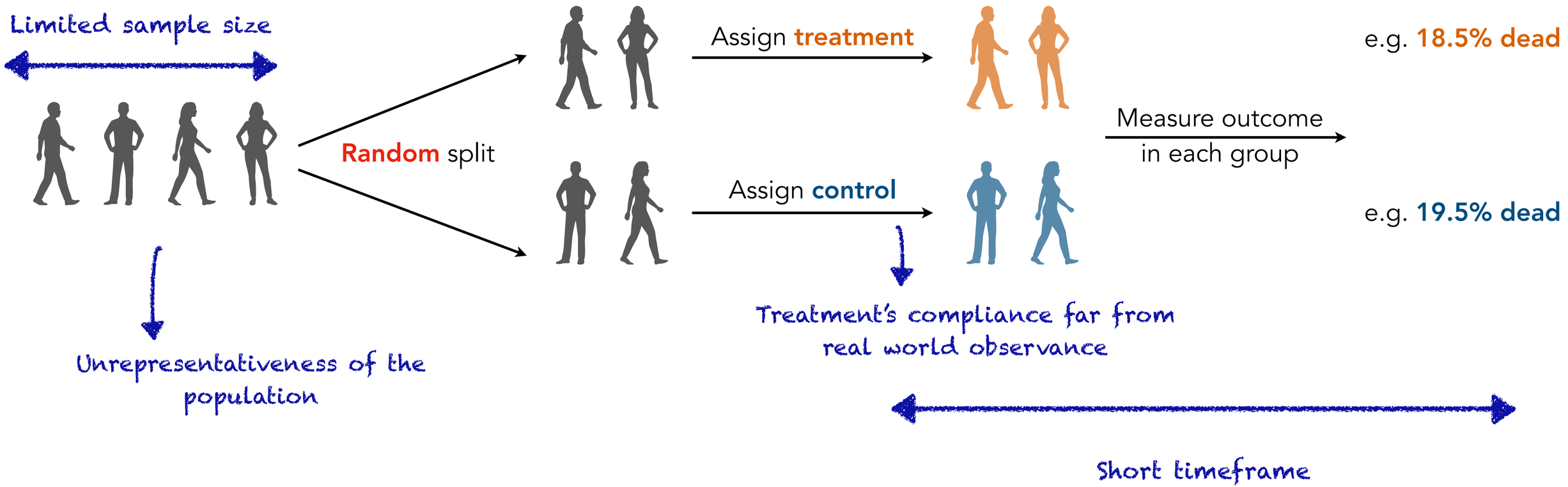
> **Results** Between July 20, 2012, and Jan 31, 2019, we randomly allocated 12 737 patients with TBI to receive tranexamic acid (6406 [50·3%] or placebo [6331 [49·7%], of whom 9202 (72·2%) patients were treated within 3 h of injury. Among patients treated within 3 h of injury, the risk of head injury-related 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]).

Source: Screenshot from the Lancet (CRASH-3 main report)

The scope of RCTs is increasingly under scrutiny



The scope of RCTs is increasingly under scrutiny



"'*External validity*' asks the question of generalizability: to what populations, settings, treatment variables, and measurement variables can this effect be generalized?" — Campbell and Stanley (1963), p. 5

The promise of detailed and larger observational or *real world* data sets

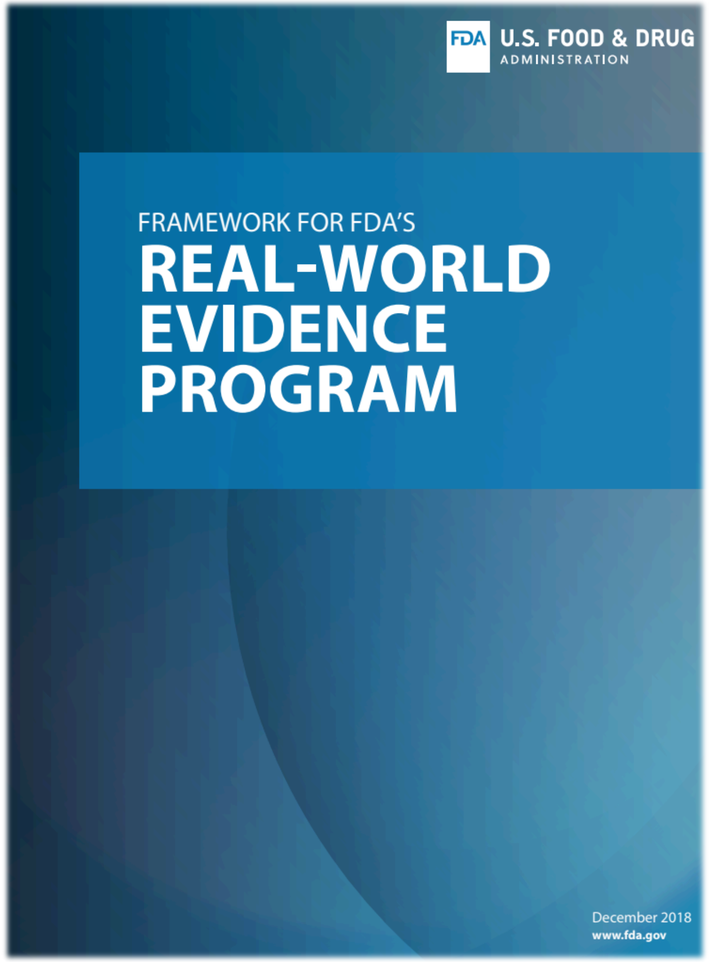
Estimate the efficacy in real-world conditions

- Using large cohorts like hospital data bases

- To emulate a target trial[1] leveraging observed confounding variables

- Solving both representativity and effective treatment given

🎁 *Large sample enabling more personalization (i.e stratified effects)*



Source: FDA's website

(1) Hernán and Robins, Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available, *Am J Epidemiol*, 2016

The example of a large French national cohort — The Traumabase

- 30,000 patients of unique size and granularity in Europe (~9,000 suffering from TBI)
- But randomisation does not hold, e.g. severe trauma are more likely to be treated

Among control	Among treated
16% dead	**38% dead**

Confusion problem

The example of a large French national cohort — The Traumabase

- 30,000 patients of unique size and granularity in Europe (~9,000 suffering from TBI)
- But randomisation does not hold, e.g. severe trauma are more likely to be treated

Among control	Among treated
16% dead	**38% dead**

← Confusion problem

After adjustment on confounding covariates (Glasgow score, age, blood pressure, …), the null hypothesis of no effect can not be rejected[2].

Is there a paradox?

CRASH-3 key results

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]

(2) Mayer et al., Doubly robust treatment effect estimation with missing attributes, *Annals of Applied Statistics* 2019

Idea — Using both types of data : experimental and observational

Fear of **unobserved confounding** in the observational sample.

Idea — Using both types of data : experimental and observational

Fear of **unobserved confounding** in the observational sample.

Both Randomized Controlled Trial (RCT) data and observational data have limitations and advantages.

The idea is to **combine** them to get the **best of both worlds**.

Causal inference methods for combining randomized trials and observational studies: a review

Bénédicte Colnet[1], Imke Mayer[1], Guanhua Chen, Awa Dieng, Ruohong Li, Gaël Varoquaux, Jean-Philippe Vert, Julie Josse[2], Shu Yang[2]

Accepted for publication in Statistical Science

Idea — Using both types of data

Fear of **unobserved confounding** in the observational sample.

Both Randomized Controlled Trial (RCT) data and observational data have limitations and advantages.

The idea is to **combine** them to get the **best of both worlds**.

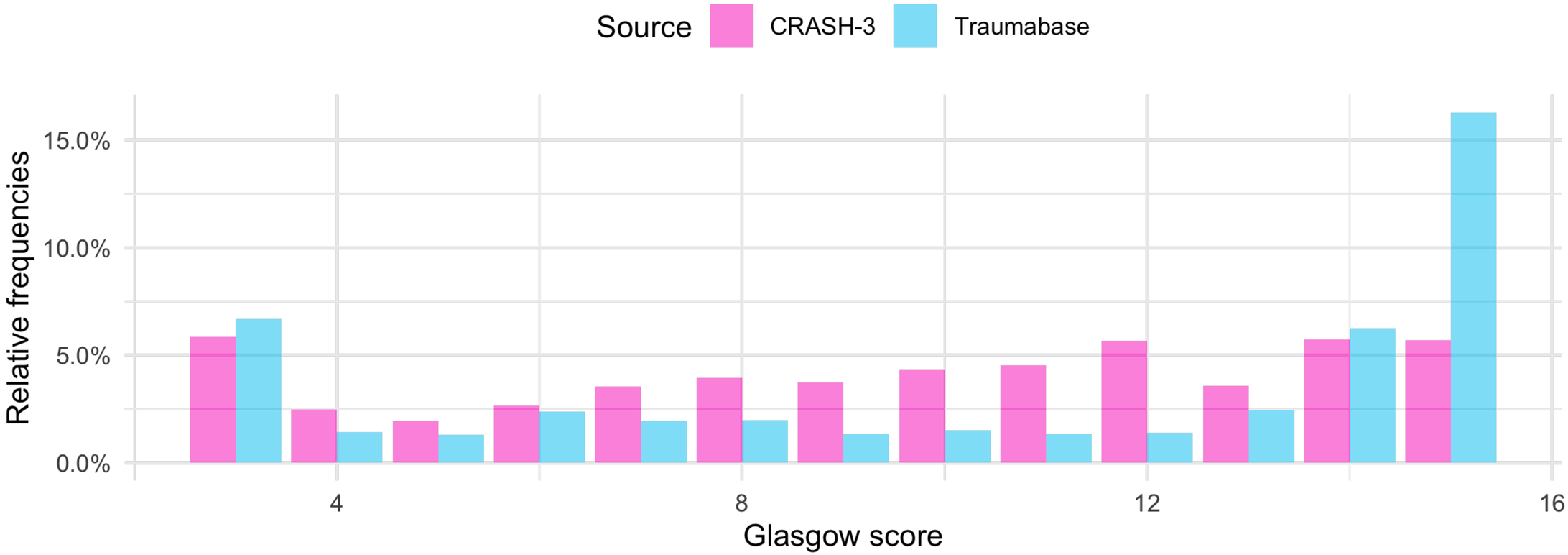
Causal inference methods for combining randomized trials and observational studies: a review

Bénédicte Colnet[1], Imke Mayer[1], Guanhua Chen, Awa Dieng, Ruohong Li, Gaël Varoquaux, Jean-Philippe Vert, Julie Josse[2], Shu Yang[2]

Accepted for publication in Statistical Science

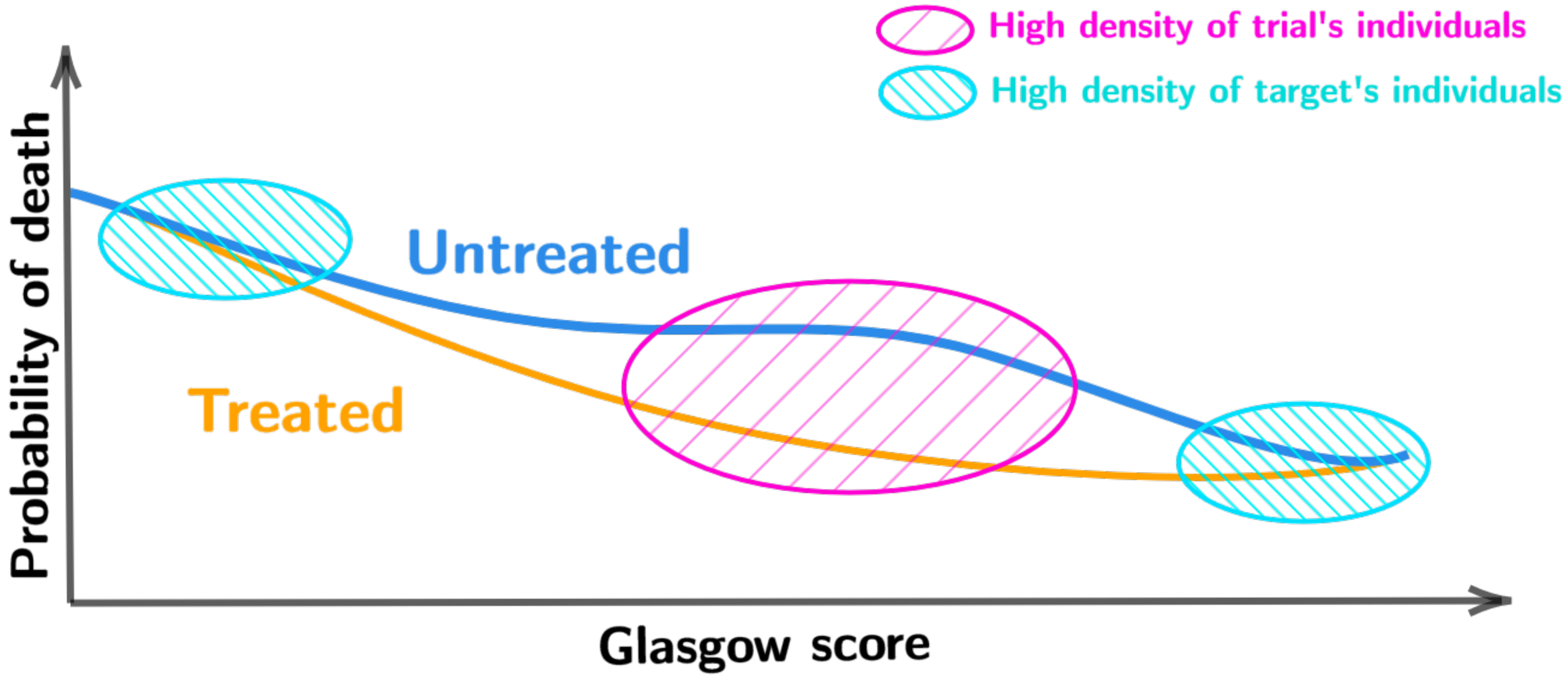
— Using observational data to improve trial's representativity

Generalizing or *transporting* CRASH-3 findings to the Traumabase population



👩‍⚕️ ``*What would have been measured as an effect in CRASH-3 if the trial was sampled in the Traumabase'?"*

Generalizing or *transporting* CRASH-3 findings to the Traumabase population



Hypothetical drawing of how the Glasgow score could modulate treatment effect

State-of-the art in a Nutshell

- Foundational work in epidemiological books (Rothman & Greenland, 2000)

- Idea of using two data sets (Stuart et al. 2010 and Pearl & Barenboim 2011)

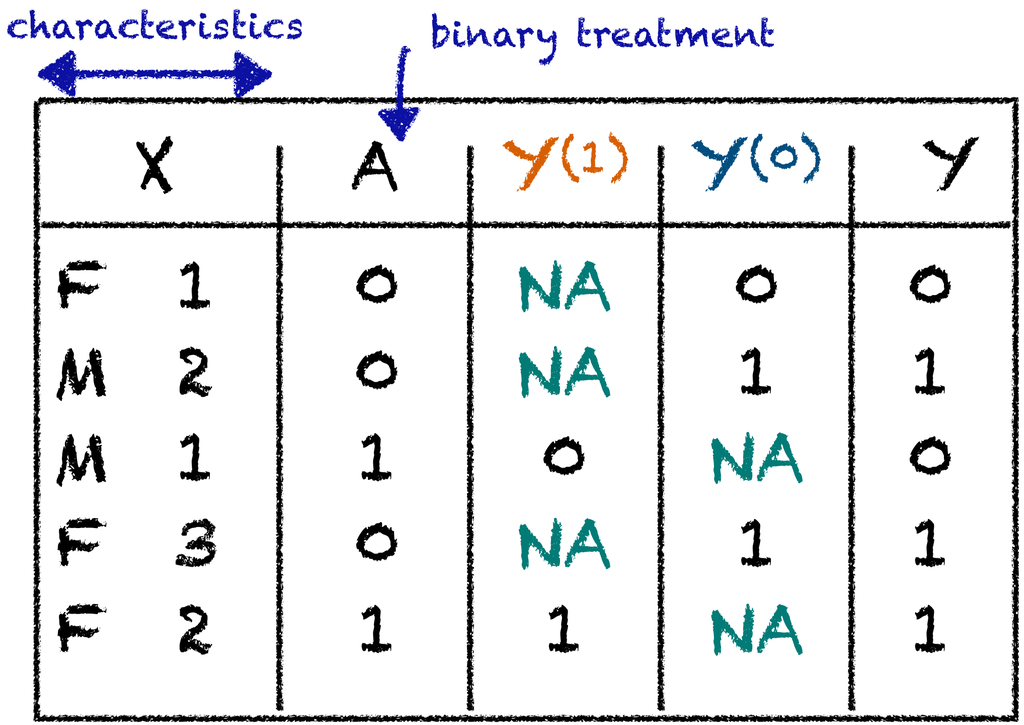
- Flourishing field …. in statistics!

- Usually clinical papers focus on characterising the lack of representativeness

  - Comparison of Table 1
  - % of patients actually treated that would have been eligible

Notations

For each individual ***i***, consider each of the possible outcomes for treated $Y^{(1)}$, and control $Y^{(0)}$.



X		A	Y(1)	Y(0)	Y
F	1	0	NA	0	0
M	2	0	NA	1	1
M	1	1	0	NA	0
F	3	0	NA	1	1
F	2	1	1	NA	1

Comparison of two potential outcomes

Individual effect $\Delta_i := Y_i^{(1)} - Y_i^{(0)}$

Detailed introduction to potential outcomes framework from Imbens and Rubin, Causal Inference for Statistics, Social, and Biomedical Sciences, *2015*

Notations

For each individual ***i***, consider each of the possible outcomes for treated $Y^{(1)}$, and control $Y^{(0)}$.

characteristics ↔ | binary treatment ↓

X		A	Y(1)	Y(0)	Y
F	1	0	NA	0	0
M	2	0	NA	1	1
M	1	1	0	NA	0
F	3	0	NA	1	1
F	2	1	1	NA	1

Individual effect $\Delta_i := Y_i^{(1)} - Y_i^{(0)}$

Can not be observed!

Average effect $ATE \equiv \tau := \mathbb{E}\left[\Delta_i\right]$

Detailed introduction to potential outcomes framework from Imbens and Rubin, Causal Inference for Statistics, Social, and Biomedical Sciences, *2015*

The potential outcomes framework for generalization

Denoting,

- **A** the binary treatment
- **X** the covariates
- **Y** the observed outcome

Two samples,

- A **trial** of size **n** sampled from a population $p_R(X)$,
- A data set of size **m** sampled from $p_T(X)$ the **target** population of interest.

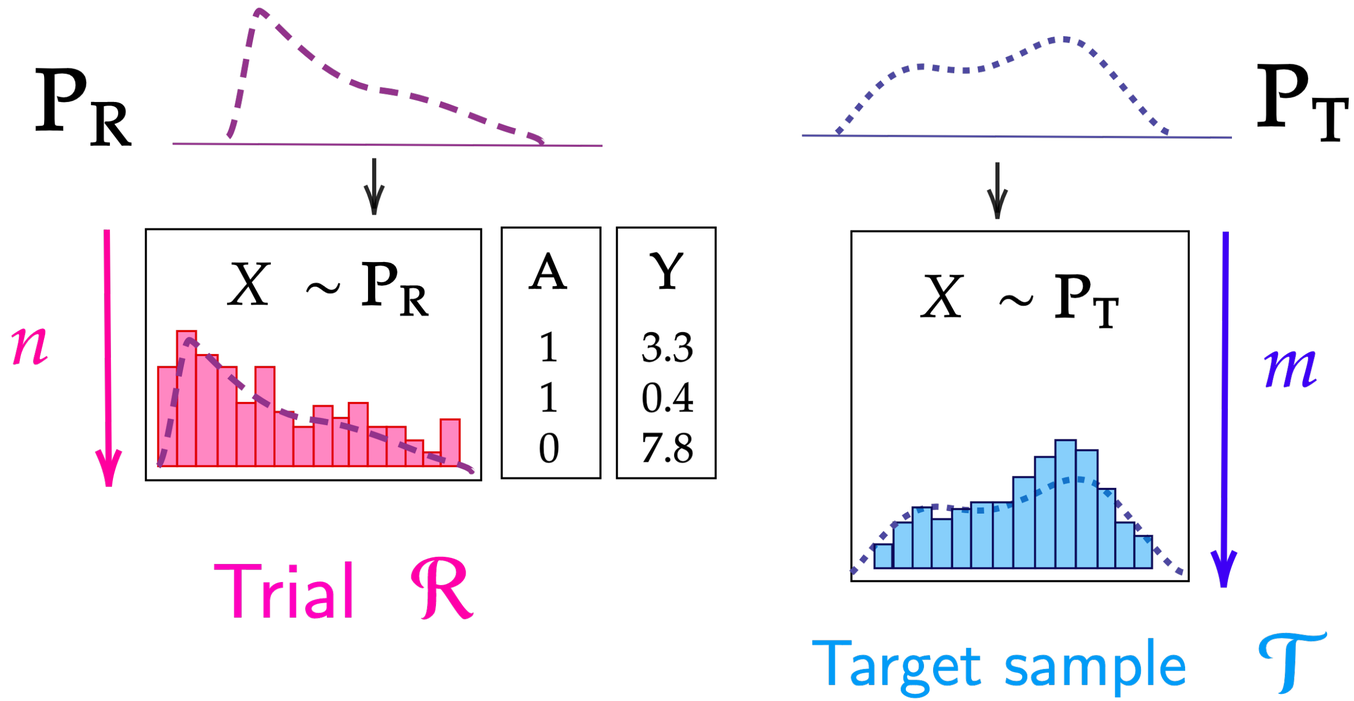
The potential outcomes framework for generalization

Denoting,

- **A** the binary treatment
- **X** the covariates
- **Y** the observed outcome

Two samples,

- A **trial** of size **n** sampled from a population $p_R(X)$,
- A data set of size **m** sampled from $p_T(X)$ the **target** population of interest.



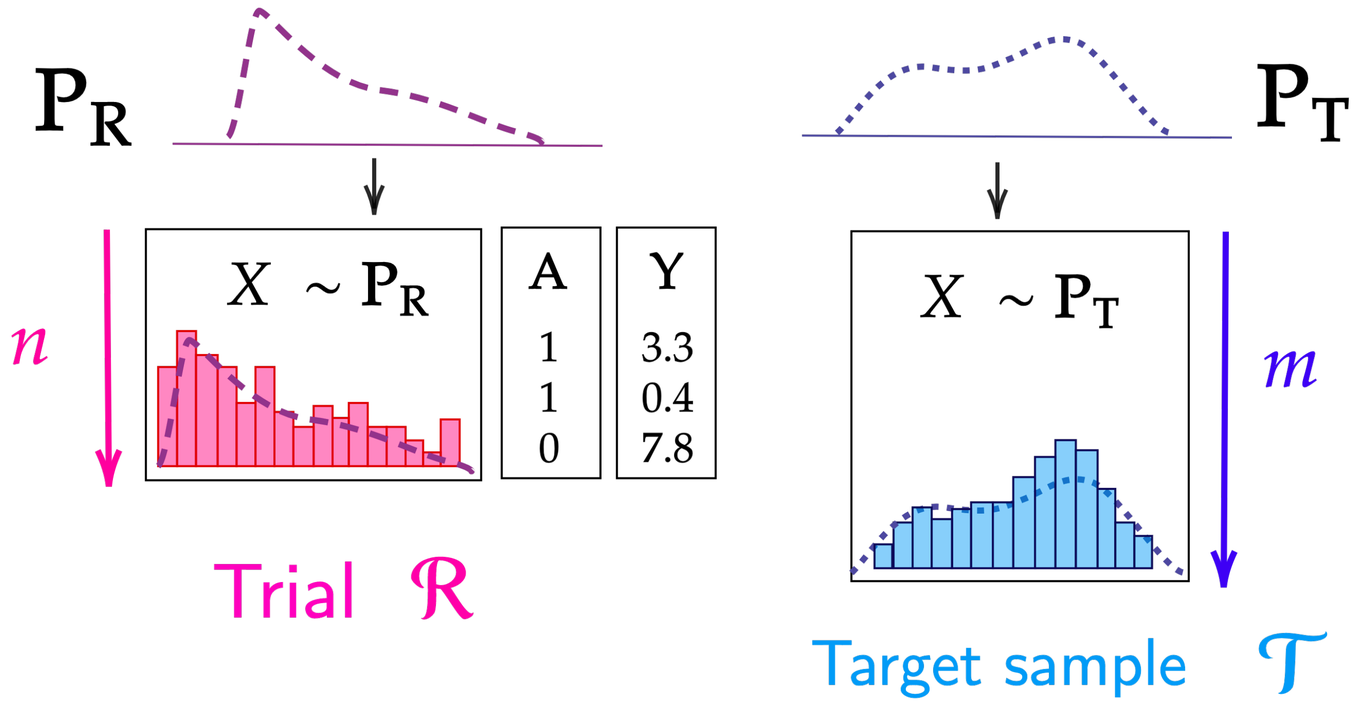
The potential outcomes framework for generalization

Denoting,

- **A** the binary treatment
- **X** the covariates
- **Y** the observed outcome

Two samples,

- A **trial** of size **n** sampled from a population $p_R(X)$,
- A data set of size **m** sampled from $p_T(X)$ the **target** population of interest.



$$p_R(x) \neq p_T(x) \Rightarrow \underbrace{\tau_R := \mathbb{E}_R[Y(1) - Y(0)]}_{\text{ATE in the RCT}} \neq \underbrace{\mathbb{E}_T[Y(1) - Y(0)] := \tau}_{\text{Target ATE}}$$

Generalization's *causal* assumptions

Transportability assumption

$$\forall x \in \mathbb{X}, \quad \mathbb{P}_R(Y^{(1)} - Y^{(0)} \mid X = x) = \mathbb{P}_T(Y^{(1)} - Y^{(0)} \mid X = x)$$

— Needed covariates are shifted treatment effect modifiers.

Spirit of ignobility assumption for a single observational data set

Generalization's *causal* assumptions

Transportability assumption

$$\forall x \in \mathbb{X}, \quad \mathbb{P}_R(Y^{(1)} - Y^{(0)} \mid X = x) = \mathbb{P}_T(Y^{(1)} - Y^{(0)} \mid X = x)$$

— Needed covariates are shifted treatment effect modifiers.

Several versions in practice

e.g. of a lighter version

$$\forall x \in \mathbb{X}, \quad \mathbb{E}_R\left[Y^{(1)} - Y^{(0)} \mid X = x\right] = \mathbb{E}_T\left[Y^{(1)} - Y^{(0)} \mid X = x\right]$$

Dahabreh et al. 2020

Most common notation where S denotes the sample's indicator

$$Y^{(1)} - Y^{(0} \perp\!\!\!\perp S \mid X$$

Nguyen et al. 2017

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

Stuart et al. 2011

Stronger assumption

Generalization's *causal* assumptions

Transportability assumption

$$\forall x \in \mathbb{X}, \quad \mathbb{P}_R(Y^{(1)} - Y^{(0)} \mid X = x) = \mathbb{P}_T(Y^{(1)} - Y^{(0)} \mid X = x)$$

— Needed covariates are shifted treatment effect modifiers.

Positivity assumption

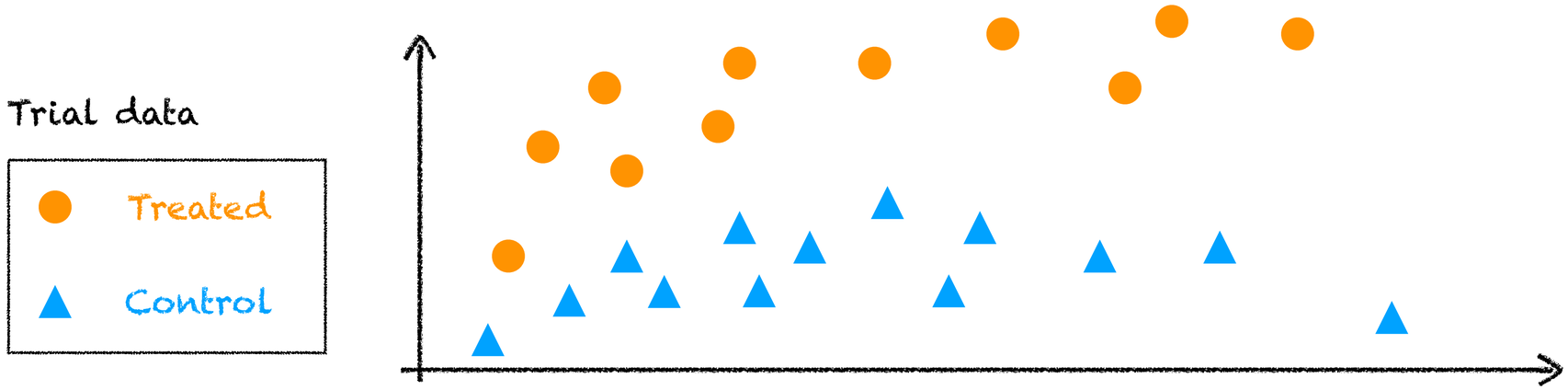
$$\text{supp}(P_T(X)) \subset \text{supp}(P_R(X))$$

— Each individuals in the target population has to be represented in the trial.

Also found as $P(S=1|X) > 1$

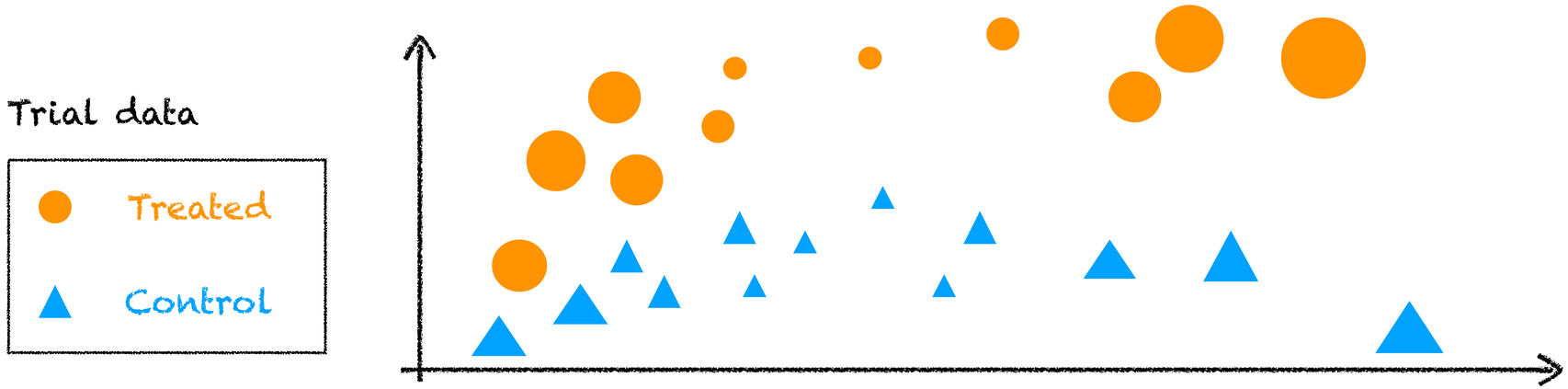
2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*



2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*



Definition

Spirit of IPW

$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n} \sum_{i \in Trial} \hat{w}_{n,m}(X_i) \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1-A_i)}{1-\pi} \right)$$

Weights

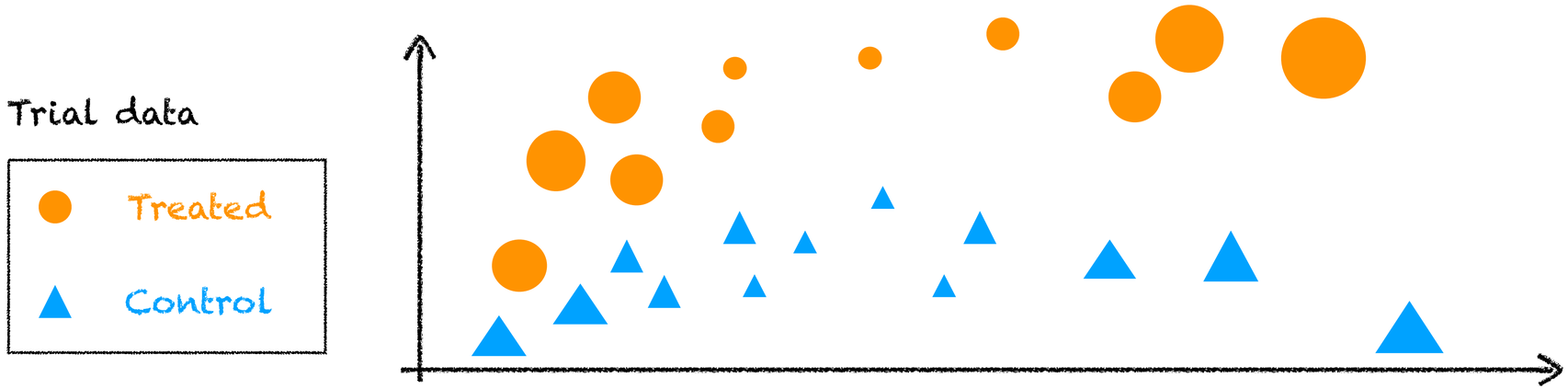
Trial only

$\pi := P_{RCT}(A=1)$

Typically $\pi = 0.5$

2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*



Definition

Spirit of IPW

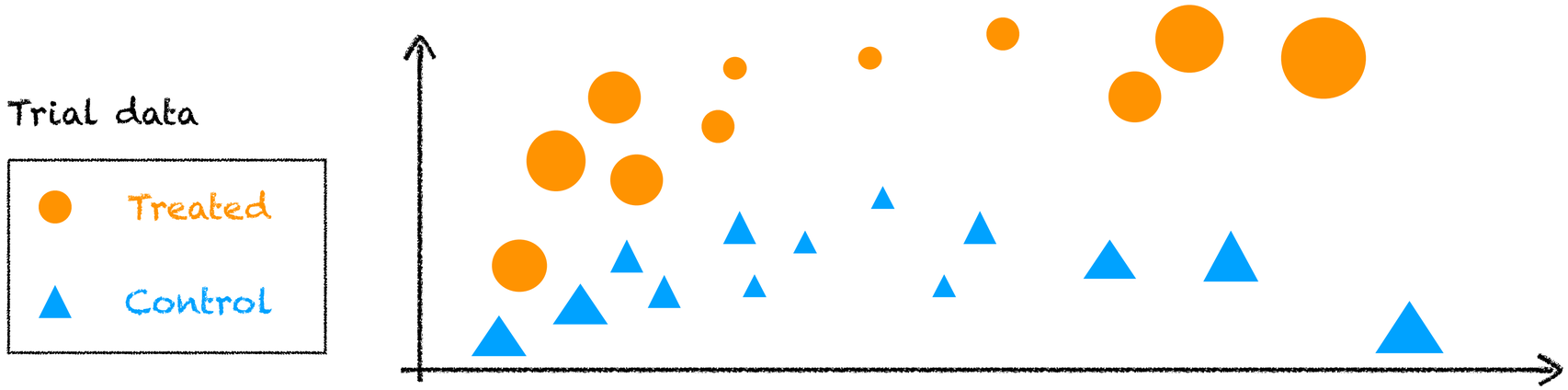
$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n} \sum_{i \in Trial} \frac{\hat{p}_{T,m}(X_i)}{\hat{p}_{R,n}(X_i)} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1-A_i)}{1-\pi} \right)$$

$\pi = P_{RCT}(A=1)$

Typically $\pi = 0.5$

2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*



Consistency **Assuming that Y is square integrable, and that**

$$(H1) \qquad \sup_{x \in \mathcal{X}} | \hat{w}_{n,m}(x) - \frac{p_T}{p_R}(x) | = \epsilon_{n,m} \xrightarrow[n,m \to \infty]{a.s.} 0$$

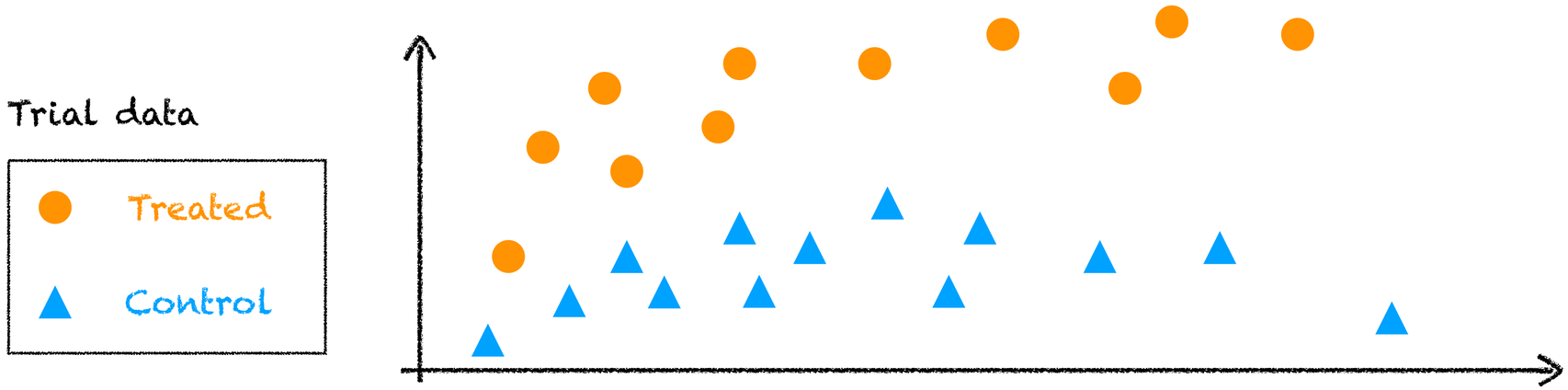
$$(H2) \qquad \mathbb{E}[\varepsilon_{n,m}^2] \xrightarrow[n,m \to \infty]{a.s.} 0$$

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

See Colnet et al. 2021, published in *Journal of Causal Inference*

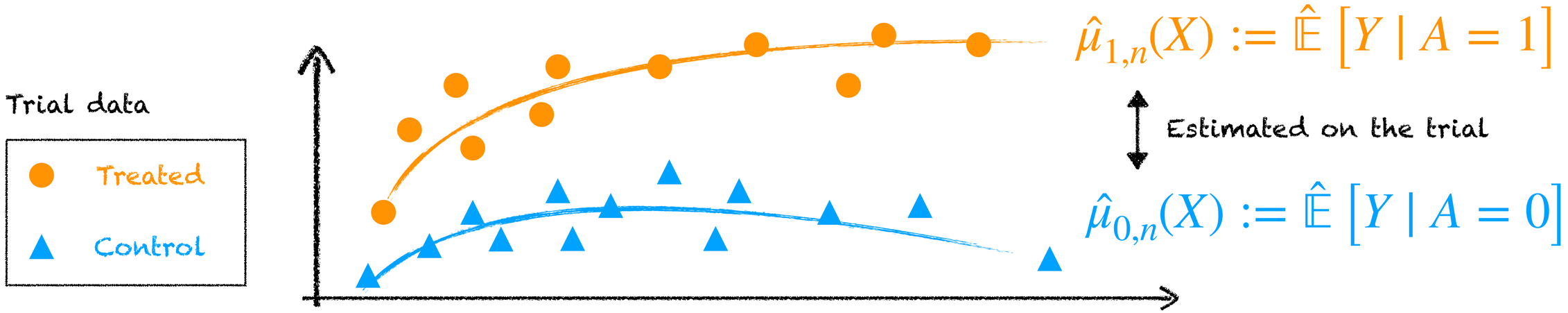
2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*
2. **Model the response** on the trial and impute the target sample — *plug-in G-formula*



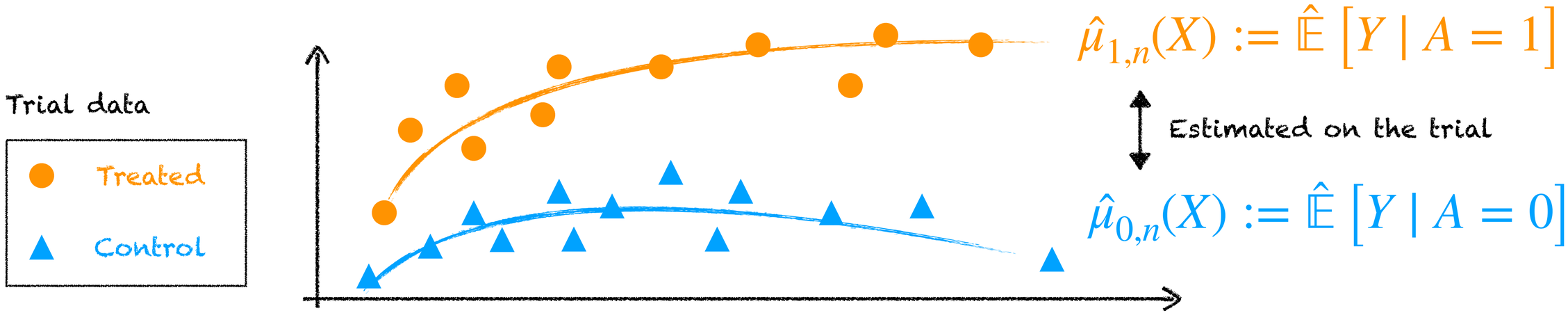
2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*
2. **Model the response** on the trial and impute the target sample — *plug-in G-formula*



2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*
2. **Model the response** on the trial and impute the target sample — *plug-in G-formula*



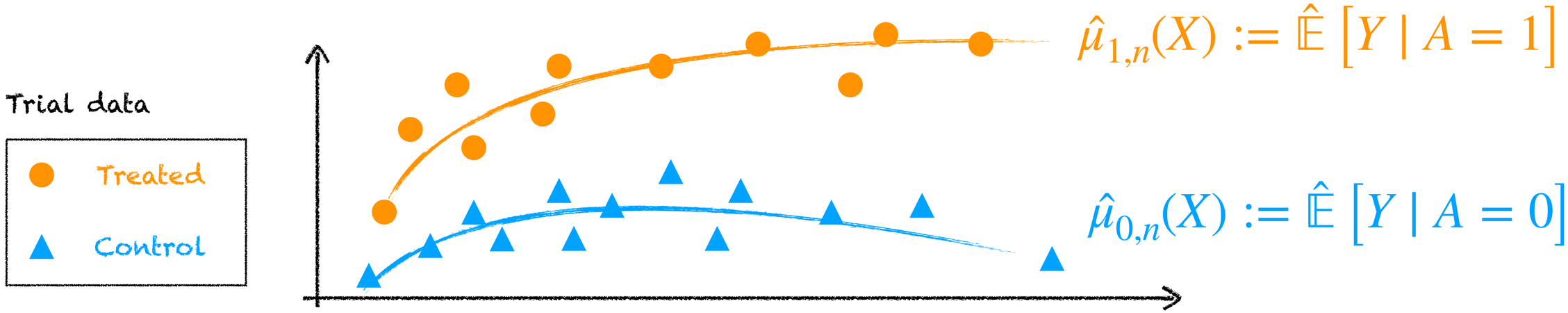
Marginalised on the target sample

Definition

$$\hat{\tau}_{G,n,m} := \frac{1}{m} \sum_{i \in Target} \hat{\mu}_{1,n}(X_i) - \hat{\mu}_{0,n}(X_i)$$

2 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*
2. **Model the response** on the trial and impute the target sample — *plug-in G-formula*



Consistency

$$(H1) \qquad \mathbb{E}\left[\,|\hat{\mu}_{a,n}(X) - \mu_a(X)|\;|\;T\right] \xrightarrow[n\to\infty]{p} 0$$

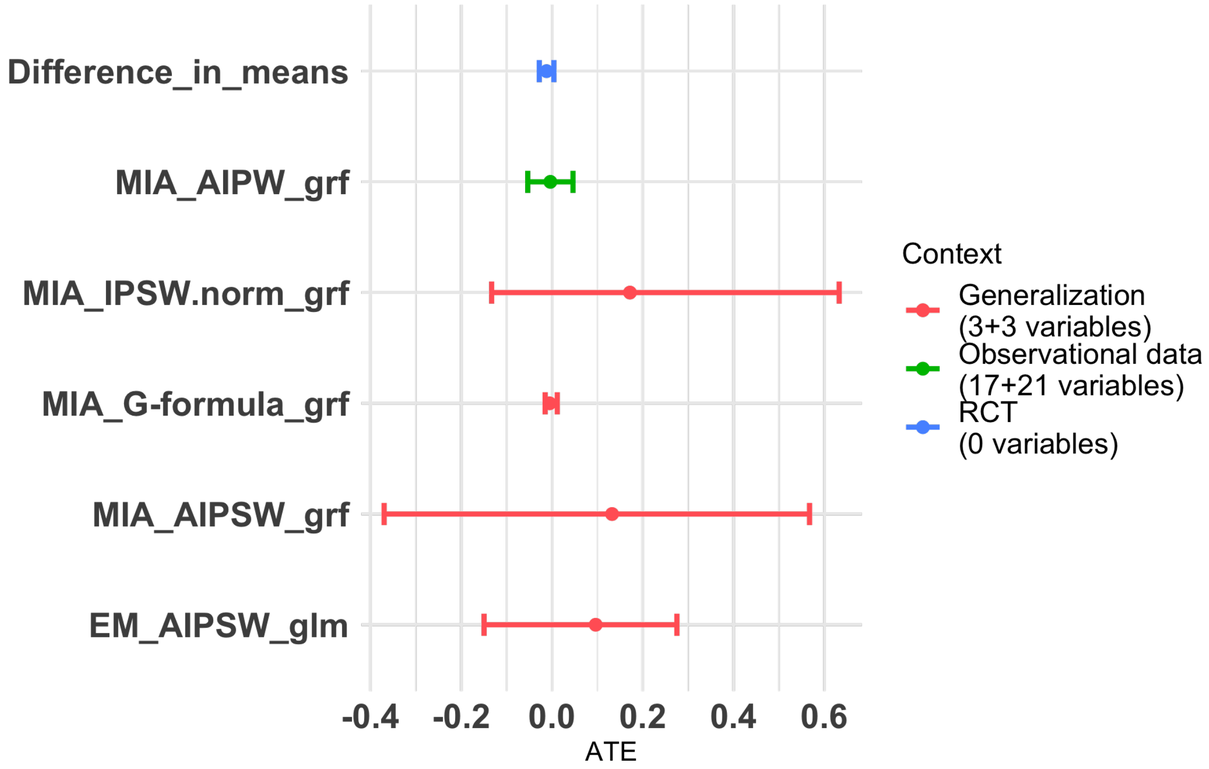
$$(H2) \quad \exists C_1, N_1 \quad \forall n \geq N_1, \quad \mathbb{E}[\hat{\mu}^2_{a,n}(X) \mid \mathscr{D}_n] \leqslant C_1$$

then $\hat{\tau}_{G,n,m} \xrightarrow[n,m\to\infty]{L^1} \tau_T$

See Colnet et al. 2021, published in *Journal of Causal Inference*

Application on the CRASH-3 & Traumabase example

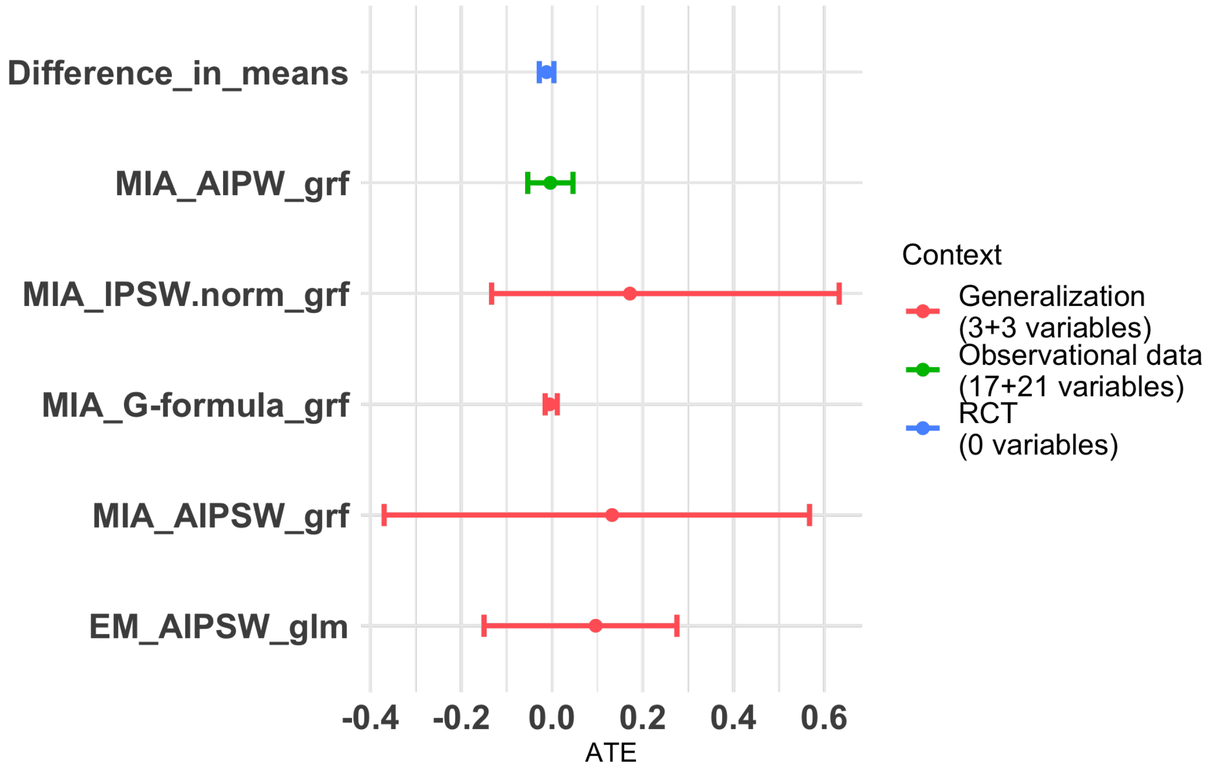
Widely varying results!



Extract of the applied results published in Statistical sciences.

Application on the CRASH-3 & Traumabase example

Widely varying results!



Extract of the applied results published in Statistical sciences.

List of open questions

- Effect of finite sample?
- Which covariate to include? -— would adding prognostic variables reduce the variance as in the classical case?
- Clinicians collaborators where rather interested in the ratio, rather than the difference

Contributions

1. A review of methods to combine experimental and observational data

   — *Causal inference methods for combining randomized trials and observational studies: a review*, co-authored with Imke Mayer, Statistical Science, 2022

2. Consistency proofs and sensitivity analysis for generalisation

   — *Causal effect on a target population: A sensitivity analysis to handle missing covariates, Journal of Causal Inference, 2022*

3. Properties of IPWS and discussion on covariates selection

   — *Reweighting the RCT for generalization: finite sample error and variable selection*, in revision in JRRS-A

4. Extension of generalization to other causal measures than the difference

   — *Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize?*, submitted to *Stat. In Med.*

Contributions

1. A review of methods to combine experimental and observational data

— *Causal inference methods for combining randomized trials and observational studies: a review*, co-authored with Imke Mayer, Statistical Science, 2022

2. Consistency proofs and sensitivity analysis for generalisation

— *Causal effect on a target population: A sensitivity analysis to handle missing covariates, Journal of Causal Inference, 2022*

3. Properties of IPWS and discussion on covariates selection

— *Reweighting the RCT for generalization: finite sample error and variable selection*, in revision in JRRS-A

4. Extension of generalization to other causal measures than the difference

— *Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize?*, submitted to *Stat. In Med.*

Recalling what is done on a classical clinical randomized trial

Horvitz-Thomson estimator

$$\hat{\tau}_{HT,n} = \frac{1}{n} \sum_{i \in Trial} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1 - A_i)}{1 - \pi} \right)$$

Probability to receive treatment, usually 0.5

Recalling what is done on a classical clinical randomized trial

Horvitz-Thomson estimator

$$\hat{\tau}_{HT,n} = \frac{1}{n} \sum_{i \in Trial} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1 - A_i)}{1 - \pi} \right)$$

Probability to receive treatment, usually 0.5

Properties

$$\mathbb{E}\left[\hat{\tau}_{HT,n}\right] = \tau_R$$

Unbiased

$$n \operatorname{Var}\left[\hat{\tau}_{HT,n}\right] = \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{\pi} + \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{1 - \pi} - \tau^2 := V_{HT}$$

Finite sample variance

Enriching the trial data with the target sample data

Depends on n and m !

$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n} \sum_{i \in Trial} \frac{\hat{p}_{T,m}(X_i)}{\hat{p}_{R,n}(X_i)} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1 - A_i)}{1 - \pi} \right)$$

Same as single RCT

Wished properties?

$$\mathbb{E}\left[\hat{\tau}_{IPSW,n}\right] = \tau_T$$

Unbiased

$$n \operatorname{Var}\left[\hat{\tau}_{IPSW,n,m}\right] = ?$$

Theoretical guarantees of IPSW with oracle weights

$$\hat{\tau}^{*}_{\pi,T,R,n} = \frac{1}{n} \sum_{i \in \mathscr{R}} \frac{p_T(X_i)}{p_R(X_i)} Y_i \left(\frac{A_i}{\pi} - \frac{1 - A_i}{1 - \pi} \right)$$

Theoretical guarantees of IPSW with oracle weights

$$\hat{\tau}^*_{\pi,T,R,n} = \frac{1}{n} \sum_{i \in \mathscr{R}} \frac{p_T(X_i)}{p_R(X_i)} Y_i \left(\frac{A_i}{\pi} - \frac{1 - A_i}{1 - \pi} \right)$$

Finite-sample properties — Oracle weights

$$\mathbb{E}\left[\hat{\tau}^*_{\pi,T,R,n}\right] = \tau_T \qquad \text{Var}\left[\hat{\tau}^*_{\pi,T,R,n}\right] = \frac{V_o}{n}$$

where $$V_o = \text{Var}_R\left[\frac{p_T(X)}{p_R(X)}\tau(X)\right] + \mathbb{E}_R\left[\left(\frac{p_T(X)}{p_R(X)}\right)^2 V_{HT}(X)\right]$$

How do we estimate weights in practice?

Assumption: assume **X** is composed of categorical covariates — e.g. smoking status, gender, …

$$\hat{\tau}^{*}_{\pi,T,n} = \frac{1}{n}\sum_{i\in\mathscr{R}} \frac{p_T(X_i)}{\hat{p}_{R,n}(X_i)} Y_i\left(\frac{A_i}{\pi} - \frac{1-A_i}{1-\pi}\right) \quad \textbf{where} \quad \hat{p}_{R,n}(x) := \frac{1}{n}\sum_{i\in\mathbb{R}} 1_{X_i=x}$$

How do we estimate weights in practice?

Assumption: assume **X** is composed of categorical covariates — e.g. smoking status, gender, …

$$\hat{\tau}^*_{\pi,T,n} = \frac{1}{n}\sum_{i\in\mathcal{R}} \frac{p_T(X_i)}{\hat{p}_{R,n}(X_i)} Y_i\left(\frac{A_i}{\pi} - \frac{1-A_i}{1-\pi}\right) \quad \textbf{where} \quad \hat{p}_{R,n}(x) := \frac{1}{n}\sum_{i\in\mathbb{R}} 1_{X_i=x}$$

Finite-sample properties — Semi oracle weights

$$\mathbb{E}\left[\hat{\tau}^*_{\pi,T,n}\right] - \tau = -\sum_{x\in\mathbb{X}} p_T(x)\left(1-p_R(x)\right)^n \tau(x)$$

$$\mathrm{Var}\left[\hat{\tau}^*_{\pi,T,n}\right] \leq \frac{2V_{so}}{n+1} + \left(1 - \min_{x\in\mathbb{X}} p_R(x)\right)^n \mathbb{E}_T\left[\tau(X)^2\right]$$

where $$V_{so} := \mathbb{E}_R\left[\left(\frac{p_T(X)}{p_R(X)}\right)^2 V_{HT}(X)\right] = V_o - \mathrm{Var}_R\left[\frac{p_T(X)}{p_R(X)}\tau(X)\right]$$

- Positive but exponentially small bias compared to the oracle estimate due to undercoverage of some categories in the trial

- Smaller asymptotic variance than the oracle estimate[4]

(4) Robins et al. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*.

Theoretical guarantees of IPSW with completely estimated weights

$$\hat{\tau}_{\pi,n,m} = \frac{1}{n} \sum_{i \in \mathcal{R}} \frac{\hat{p}_{T,m}(X_i)}{\hat{p}_{R,n}(X_i)} Y_i \left(\frac{A_i}{\pi} - \frac{1 - A_i}{1 - \pi} \right)$$

Finite-sample properties — Fully estimated weights

$$\mathbb{E}\left[\hat{\tau}^*_{\pi,T,n}\right] - \tau = -\sum_{x \in \mathbb{X}} p_T(x)\left(1 - p_R(x)\right)^n \tau(x)$$

$$\begin{aligned}\mathrm{Var}\left[\hat{\tau}_{\pi,n,m}\right] \leq \frac{2V_{so}}{n+1} + \frac{\mathrm{Var}_T\left[\tau(X)\right]}{m} \\ + \frac{2}{m\,(n+1)} \mathbb{E}_R\left[\frac{p_T(X)(1 - p_T(X))}{p_R(X)^2} V_{HT}(X)\right] \\ + \left(1 - \min_x p_R(x)\right)^{n/2} \mathbb{E}_T\left[\tau(X)^2\right] \left(1 + \frac{4}{m}\right)\end{aligned}$$

- Same bias as the semi oracle: bias can only be explained by a limited RCT

- Two sample size: RCT (n) and observational study (m)
  - Additional term decreasing as 1/m compared to the semi oracle estimate
  - Consistent if both n and m → ∞. In this case, the first two terms dominate.

IPSW Large sample properties

Large sample properties — Fully estimated weights

Letting $\lim_{n,m\to\infty} m/n = \lambda \in [0,\infty]$,

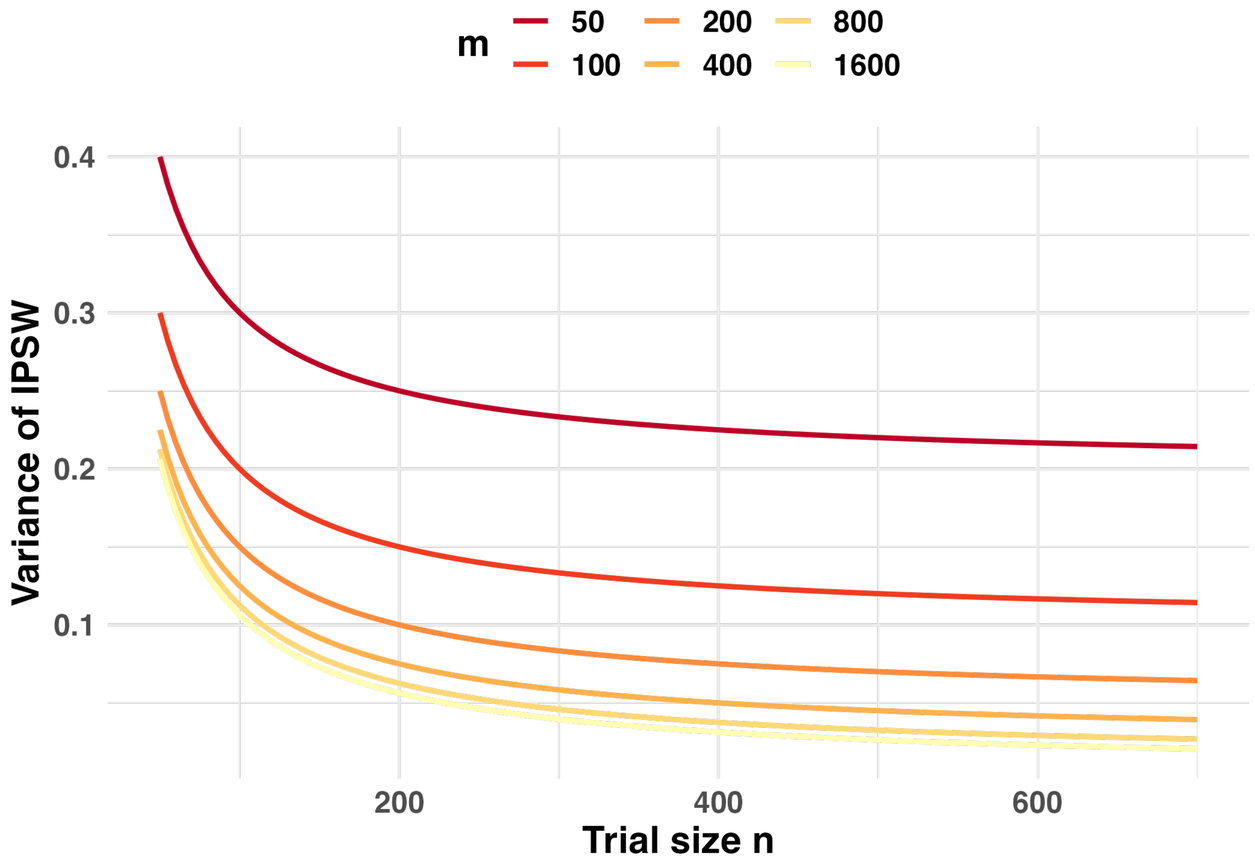
$$\lim_{n,m\to\infty} \min(n,m)\mathrm{Var}\left[\hat{\tau}_{\pi,n,m}\right] = \min(1,\lambda)\left(\frac{\mathrm{Var}\left[\tau(X)\right]}{\lambda} + V_{so}\right)$$

Two data samples sizes dictating two asymptotic variance

If target >> trial (i.e. $\lambda = \infty$), asymptotic variance = semi-oracle's one and depends on the ratio of probabilities

If target << trial (i.e. $\lambda = 0$), asymptotic variance = conditional treatment effect variance

IPSW Large sample properties - Illustration



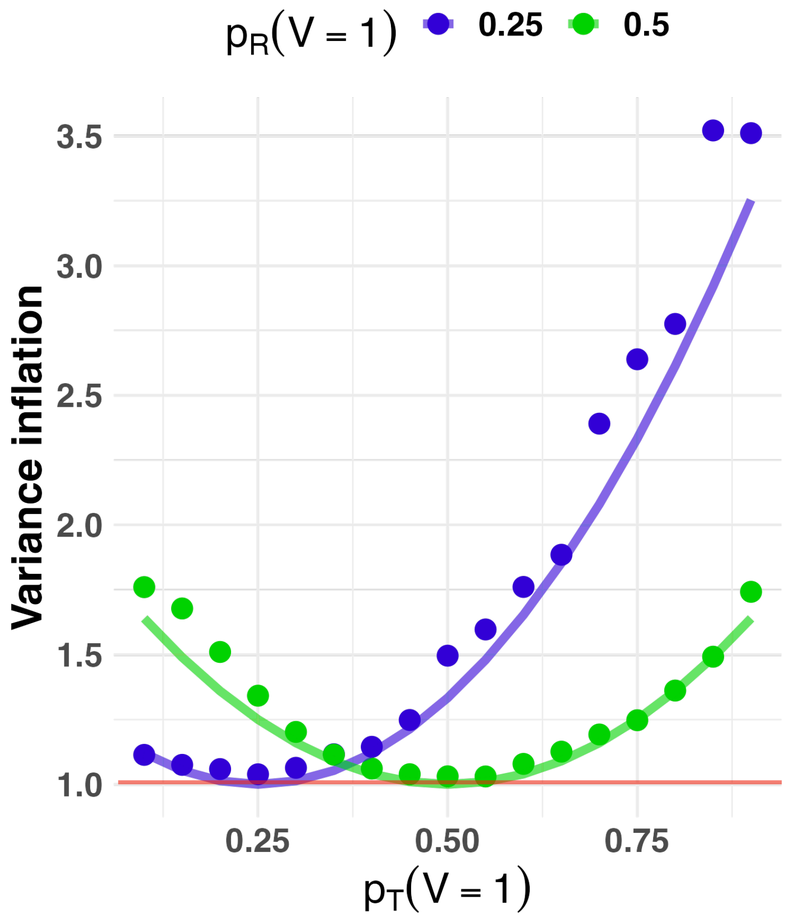
Practical recommandation

e.g. When n = 200 and m = 50, it is better to double the size of the observational data than that of the RCT.

Impact of additional covariates: for the worse?

- Covariates needed to generalize are,
  - Treatment effect modifier
    A covariate along which the treatment effect is modulated
  - Shifted
    Not the same proportion in each population

- In practice, one may be tempted to add many covariates
  - It does prevent to miss important ones
  - But what happen if gender is added but is only shifted?



Dots are simulations, plain lines are the theory introduced on next slide

Impact of adding a shifted covariate which is not treatment effect modifier

Non treatment effect modifier

V does not modulate treatment effect, that is

$$\forall v \in \mathbb{V}, \forall s \in \{T, R\}, \quad \mathbb{P}_s(Y^{(1)} - Y^{(0)} \mid X = x, V = v) = \mathbb{P}_s(Y^{(1)} - Y^{(0)} \mid X = x)$$

Impact of adding a shifted covariate which is not treatment effect modifier

Non treatment effect modifier

V does not modulate treatment effect modifier, that is

$$\forall v \in \mathbb{V}, \forall s \in \{T, R\}, \quad \mathbb{P}_s(Y^{(1)} - Y^{(0)} \mid X = x, V = v) = \mathbb{P}_s(Y^{(1)} - Y^{(0)} \mid X = x)$$

Shifted covariate which is not a treatment effect modifier

Consider the semi-oracle IPSW estimator and a set of additional shifted covariates V, independent of X, which are not treatment effect modifier, then

$$\lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}^*_{T,n,m}(X, V)\right] = \left(\sum_{v\in\mathscr{V}} \frac{p_T(v)^2}{p_R(v)}\right) \lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}^*_{T,n,m}(X)\right]$$

Including non-necessary covariates can seriously damage precision

Impact of adding a non-shifted covariate which is a treatment effect modifier

Non-shifted covariate

V is not shifted, that is

$$\forall v \in \mathbb{V}, p_T(v) = p_R(v)\,.$$

Non-shifted covariate which is a treatment effect modifier

Consider the semi-oracle IPSW estimator and a set of additional non-shifted treatment effect modifier set V, independent of X. Then,

$$\lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}^*_{T,n}(X,V)\right] = \lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}^*_{T,n}(X)\right] - \mathbb{E}_R\left[\frac{p_T(X)}{p_R(X)} \operatorname{Var}\left[\tau(X,V) \mid X\right]\right]$$

Including non-necessary covariates can improve precision

Semi-synthetic simulation

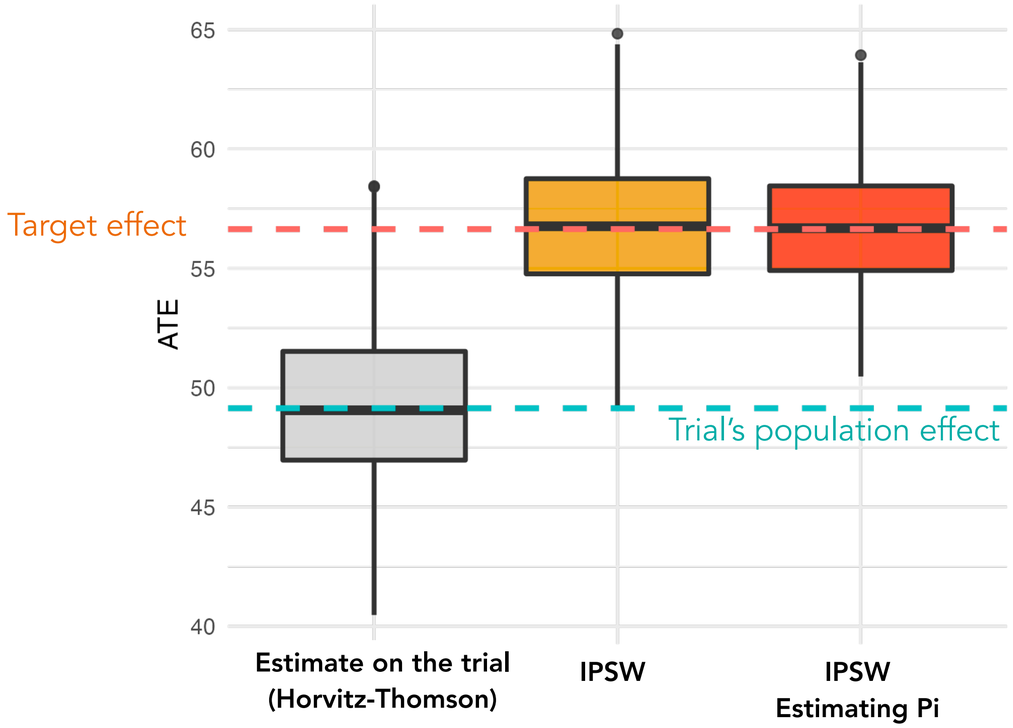
We illustrate the results on semi-synthetic simulations

- Simulations are built from CRASH-3 (~ 9,000 individuals) and Traumabase (~30,000 individuals);
- Doing so, this reflects a real-world shift;
- Covariates are : Glasgow score, gender, time-to-treatment (TTT), blood pressure;
- Time to treatment is simulated as not present in the Traumabase;
- As all covariates are shifted (even a little), a non-shifted treatment effect modifier Z is created
- The outcome is synthetic.

$$Y = 10 - \textbf{Glasgow} + \left(\textbf{if Girl:} - 5 \textbf{ else:} 0\right) + A\left(15(6 - \textbf{TTT}) + 3 * (\textbf{Blood.pressure} - 1)^2 + 50Z\right) + \varepsilon_{TTT}$$

Random gaussian noise whose variance depends on the value of TTT

Results from the semi-synthetic simulations (1)



This simulation does not include Z as the focus is not on adding non-useful covariates

1. Re-weighting allows to recover the target effect

2. Two additional theoretical results not detailed above

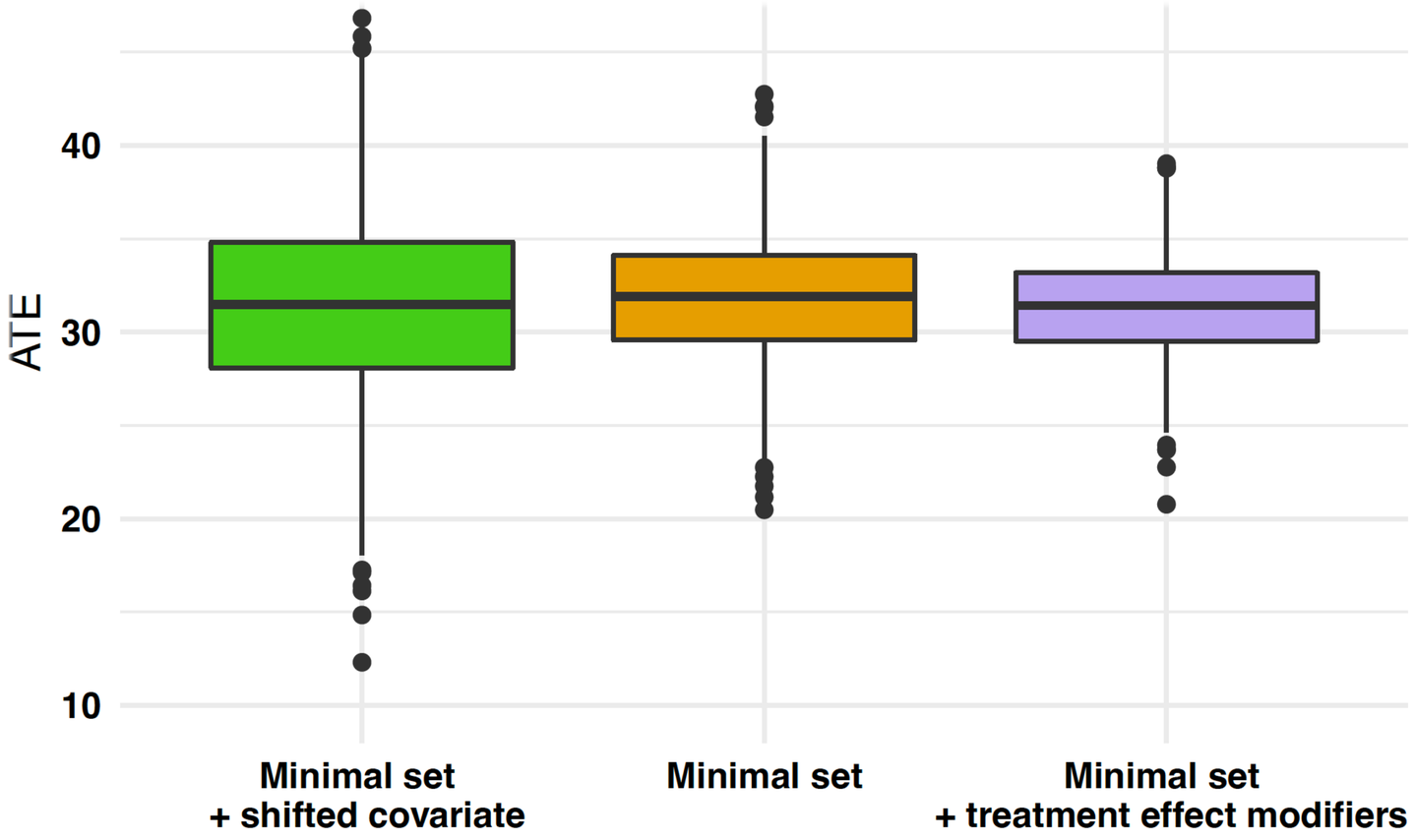
- Reducing variance when estimating the probability to be treated in the trial Pi,
- Re-weighted trial has not necessarily a larger variance.

Results from the semi-synthetic simulations (2)

Effect of non-necessary covariates on the variance

IPSW with n = 3000 and m = 10000 and 1,000 repetitions

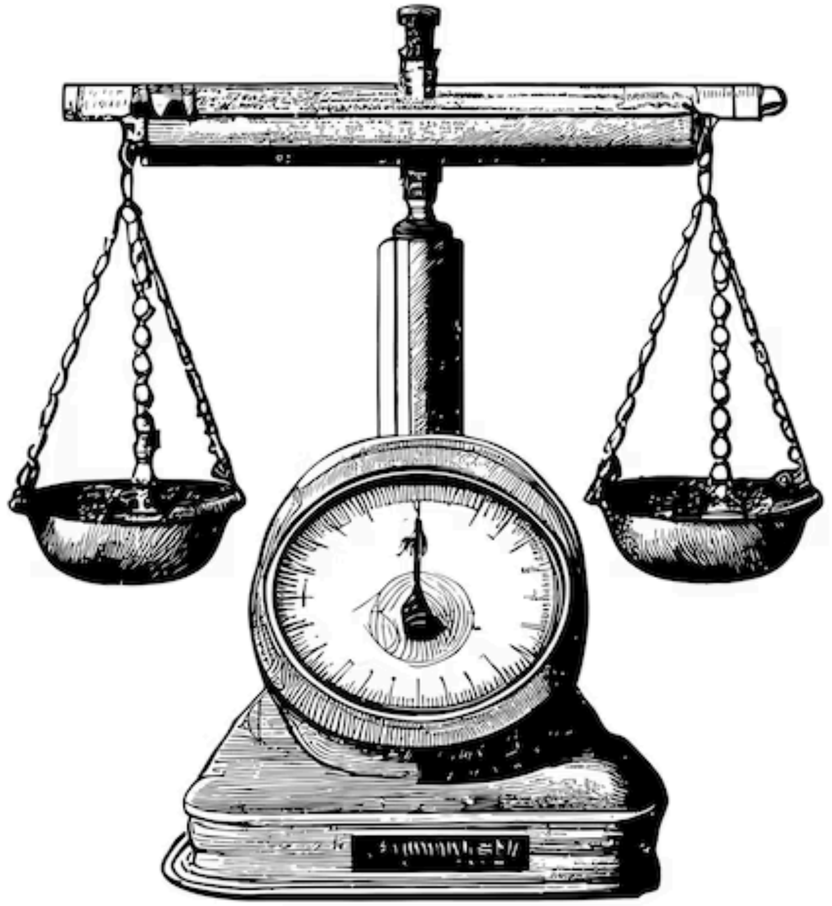
- The addition of the covariate **GCS** increases the variance,
- while the addition of a **non-shifted treatment effect modifier** leads to an improvement in variance.



This simulation includes Z as the focus is on adding non-useful covariates

Risk ratio, odds ratio, risk difference

Which causal measure is easier to generalize?



Contributions

1. A review of methods to combine experimental and observational data

   — *Causal inference methods for combining randomized trials and observational studies: a review*, co-authored with Imke Mayer, Statistical Science, 2022

2. Consistency proofs and sensitivity analysis for generalisation

   — *Causal effect on a target population: A sensitivity analysis to handle missing covariates, Journal of Causal Inference, 2022*

3. Properties of IPWS and discussion on covariates selection

   — *Reweighting the RCT for generalization: finite sample error and variable selection*, in revision in JRRS-A

4. Extension of generalization to other causal measures than the difference

   — *Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize?*, submitted to *Stat. In Med.*

Illustrative example

RCT from Cook and Sackett (1995)

- **Y** the observed binary outcome
- **A** binary treatment assignment
- **X** baseline covariates

Stroke after 5 years

11.1% Control — vs — **6.7 % Treated**

Usually referring to an *effect*, is related to how one *contrasts* those two

e.g. Ratio = 6.7/11.1 = 0.6 **or** Diff = - 0.04

Illustrative example

Potential outcomes framework → $\mathbb{E}\left[Y^{(0)}\right]$, $\mathbb{E}\left[Y^{(1)}\right]$

RCT from Cook and Sackett (1995)

- **Y** the observed binary outcome
- **A** binary treatment assignment
- **X** baseline covariates

Stroke after 5 years

11.1% Control — vs — **6.7 % Treated**

Count the stroke

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

Count the non-stroke

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

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

Risk Difference

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

Number Needed to Treat

Odds Ratio

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

— A variety of causal measures exist

Note that for binary Y,
E[Y(a)] = P(Y(a)=1)

Computing all the measures on the illustrative clinical example

X = 1 <-> low baseline risk

	τ_{RD}	τ_{RR}	τ_{SR}	τ_{NNT}	τ_{OR}
All (P_{S})	−0.0452	**0.6**	1.05	22	0.57
$\mathbf{X = 1}$	−0.006	**0.6**	1.01	167	0.6
$\mathbf{X = 0}$	−0.08	**0.6**	1.1	13	0.545

Marginal effects τ

Conditional effects $\tau(x)$

Computed from Cook & Sackett (1995)

👩‍⚕️ *``Treated group has 0.6 times the risk of having a stroke outcome when compared with the placebo." **or**` `The Number Needed to Treat is 22."*

— leads to different impressions and heterogeneity patterns

The age-old question of how to report effects

Source: Wikipedia

`` We wish to decide whether we shall count the failures or the successes and whether we shall make relative or absolute comparisons"

— Mindel C. Sheps, New England Journal of Medicine, in 1958

The choice of the measure is still actively discussed

e.g. Spiegelman and VanderWeele, 2017; Baker and Jackson, 2018; Feng et al., 2019; Doi et al., 2022; Xiao et al., 2021, 2022; Huitfeldt et al., 2021; Lapointe-Shaw et al., 2022; Liu et al., 2022 …

— CONSORT guidelines recommend to report all of them

A desirable property: collapsibility

i.e. population's effect is equal to a weighted sum of local effects

Discussed in Greenland, 1987; Hernàn et al. 2011; Huitfeldt et al., 2019; Daniel et al., 2020; Didelez and Stensrud, 2022 and many others.

A desirable property: collapsibility

i.e. population's effect is equal to a weighted sum of local effects

 Discussed in Greenland, 1987; Hernàn et al. 2011; Huitfeldt et al., 2019; Daniel et al., 2020; Didelez and Stensrud, 2022 and many others.

A very famous example: the Simpson paradox

Marginal effect bigger than subgroups' effects

(a) Overall population, $\tau_{\text{OR}} \approx 0.26$

	Y=0	Y=1
A=1	1005	95
A=0	1074	26

(b) $\tau_{\text{OR}|F=1} \approx 0.167$ and $\tau_{\text{OR}|F=0} \approx 0.166$

F= 1	Y=0	Y=1
A=1	40	60
A=0	80	20

F=0	Y=0	Y=1
A=1	965	35
A=0	994	6

Toy example inspired from Greenland (1987).

— Unfortunately, not all measures are collapsible

Collapsibility and formalism

- Different definitions of collapsibility in the literature

Collapsibility and formalism

- Different definitions of collapsibility in the literature

- We propose three definitions encompassing previous works

  1. Direct collapsibility $\quad \mathbb{E}\left[\tau(X)\right] = \tau$

Collapsibility and formalism

- Different definitions of collapsibility in the literature

- We propose three definitions encompassing previous works

  1. Direct collapsibility $\mathbb{E}\left[\tau(X)\right] = \tau$

  2. Collapsibility $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\,\tau(X)\right] = \tau,$ **with** $w \geq 0,$ **and** $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\right] = 1$

e.g RR is collapsible, with $\mathbb{E}\left[\tau_{RR}(X)\dfrac{\mathbb{E}\left[Y^{(0)} \mid X\right]}{\mathbb{E}\left[Y^{(0)}\right]}\right] = \tau_{RR}$

Collapsibility and formalism

- Different definitions of collapsibility in the literature

- We propose three definitions encompassing previous works

  1. Direct collapsibility $\mathbb{E}\left[\tau(X)\right] = \tau$

  2. Collapsibility $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\,\tau(X)\right] = \tau,$ **with** $w \geq 0,$ **and** $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\right] = 1$

  3. Logic-respecting $\tau \in \left[\min_x(\tau(x)), \max_x(\tau(x))\right]$

Collapsibility and formalism

- Different definitions of collapsibility in the literature

- We propose three definitions encompassing previous works

1. Direct collapsibility $\mathbb{E}\left[\tau(X)\right] = \tau$

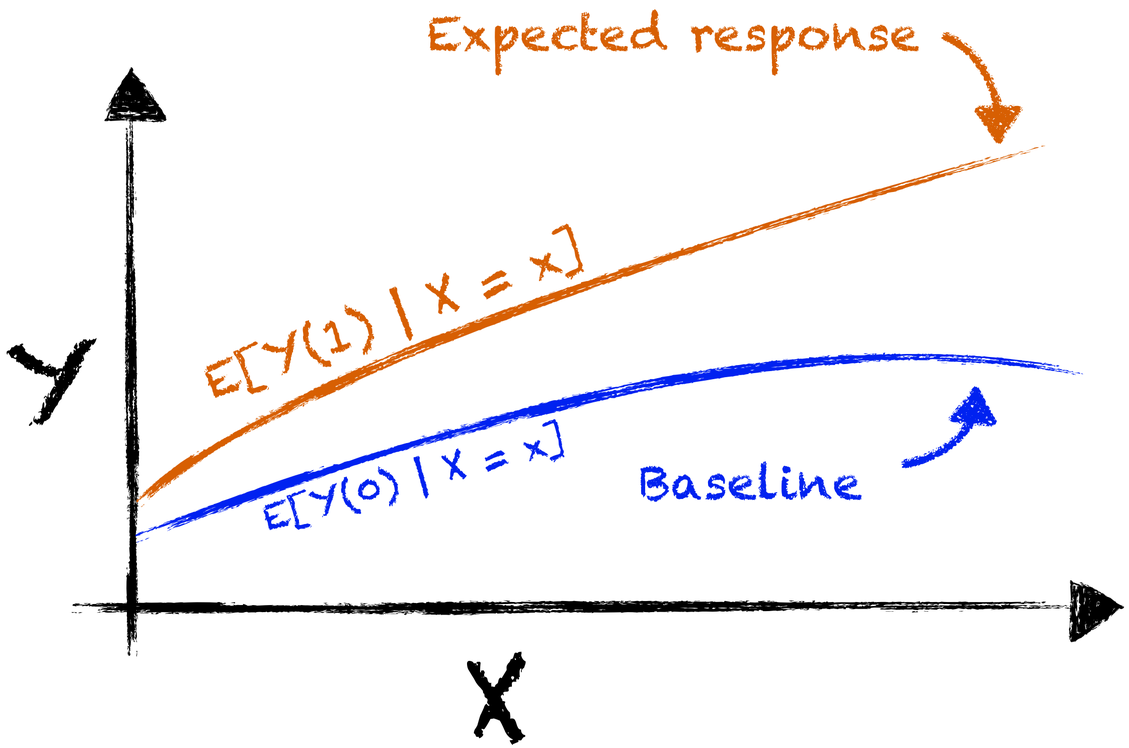
2. Collapsibility $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\,\tau(X)\right] = \tau,$ **with** $w \geq 0,$ **and** $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\right] = 1$

3. 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

Through the lens of non parametric generative models

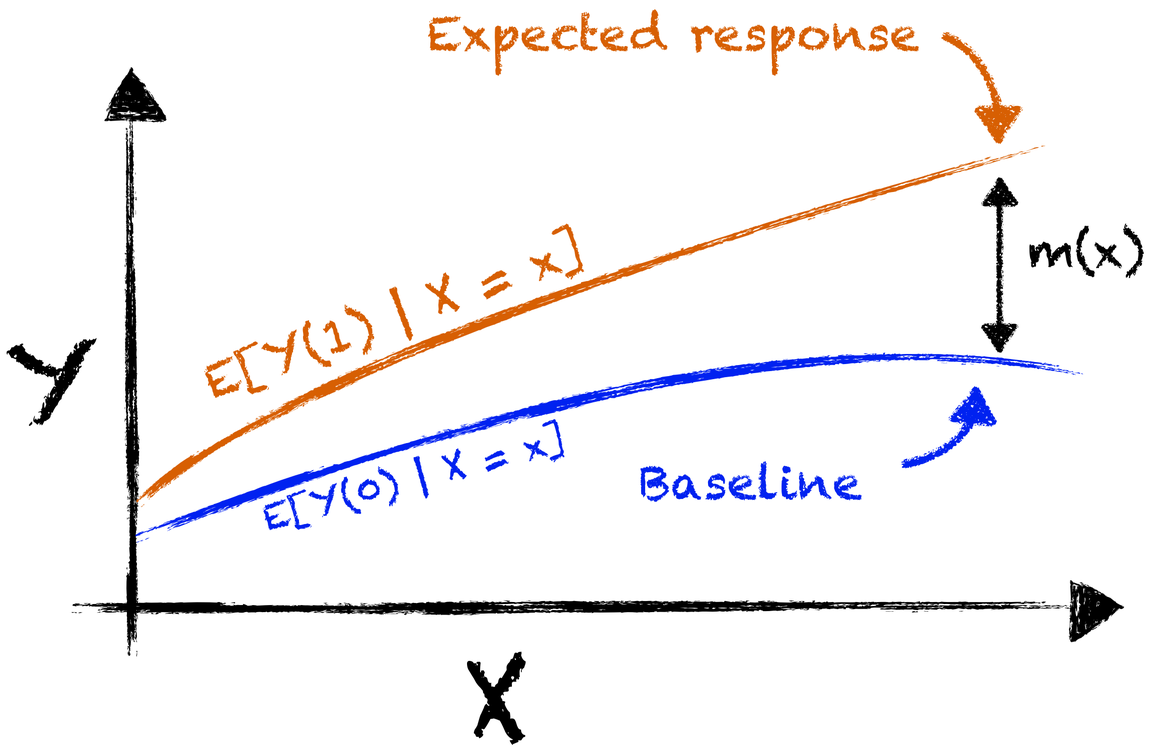
For Y <u>continuous</u>,



(*) This only assumes that conditional expected responses are defined for every x

Through the lens of non parametric generative models

For Y <u>continuous</u>,



(*) This only assumes that conditional expected responses are defined for every x

Lemma*

There exist two functions b(.) and m(.) such that,

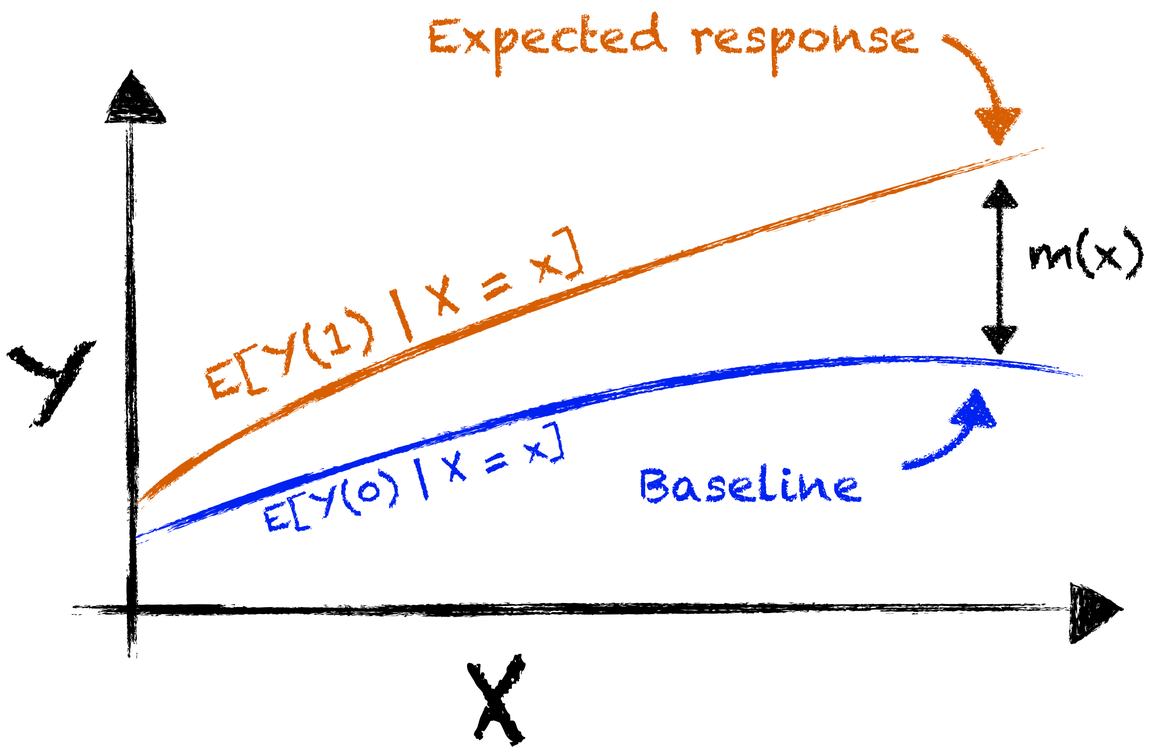
$$\mathbb{E}\left[Y^{(a)} \mid X\right] = b(X) + a\, m(X)$$

Additivity

Spirit of Robinson's decomposition (1988), further developed in Nie et al. 2020

Through the lens of non parametric generative models

For Y <u>continuous</u>,



(*) This only assumes that conditional expected responses are defined for every x

Lemma*

There exist two functions b(.) and m(.) such that,

$$\mathbb{E}\left[Y^{(a)} \mid X\right] = b(X) + a\, m(X)$$

Additivity

Spirit of Robinson's decomposition (1988), further developed in Nie et al. 2020

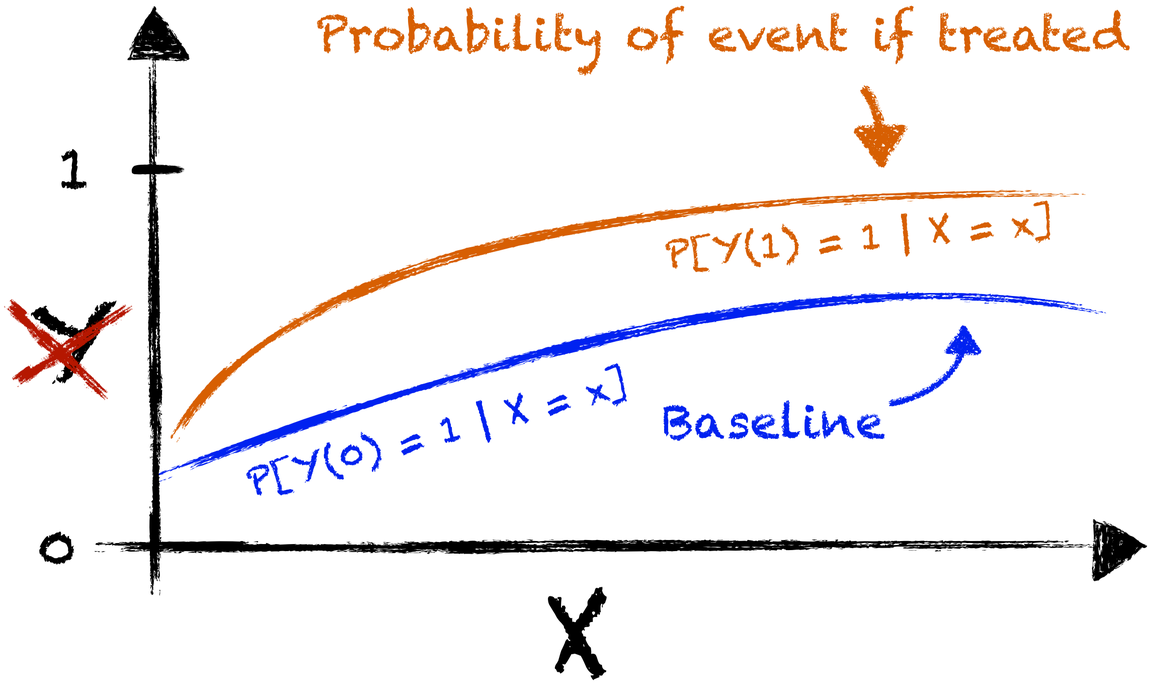
Linking generative functions with measures

$$\tau_{RR}(x) = 1 + m(x)/b(x)$$ Entanglement

$$\tau_{RD}(x) = m(x)$$ No entanglement

Through the lens of non parametric generative models

For Y <u>binary</u>,



Lemma

There exist two functions b(.) and m(.) such that,

$$\mathbb{E}\left[Y^{(a)} \mid X\right] = b(X) + a\,m(X)$$

Additivity

Adapted Lemma

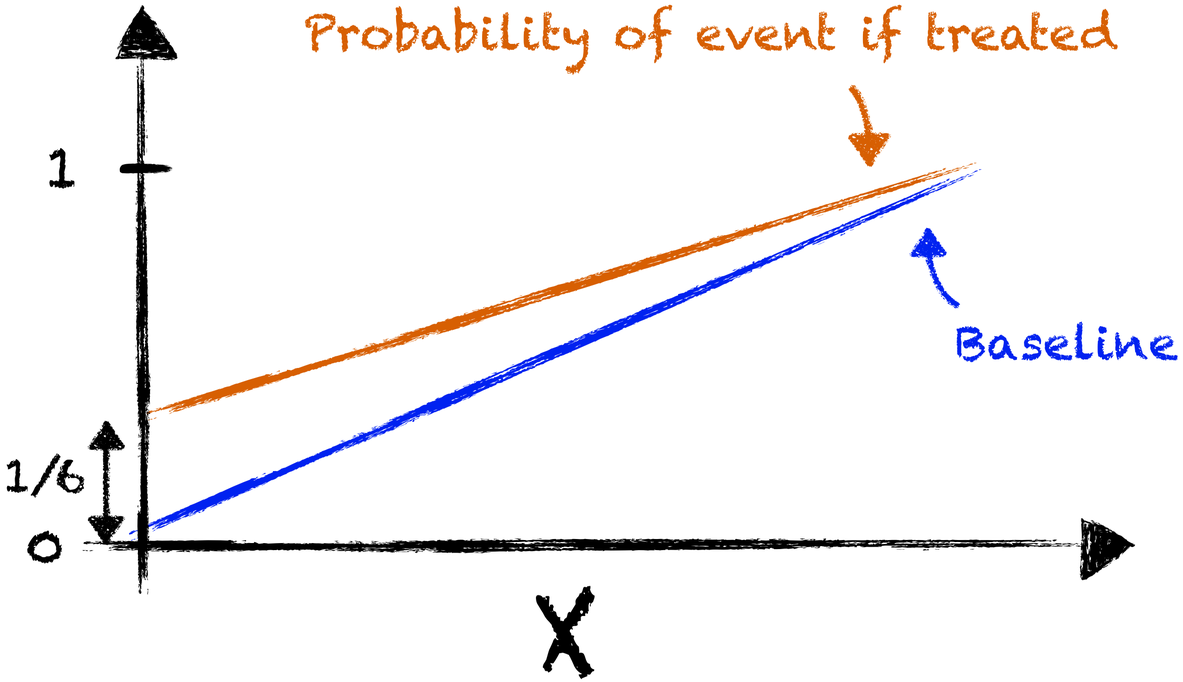
There exist two functions b(.) and m(.) such that,

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

The example of the Russian roulette

Harmful

For Y <u>binary</u>,

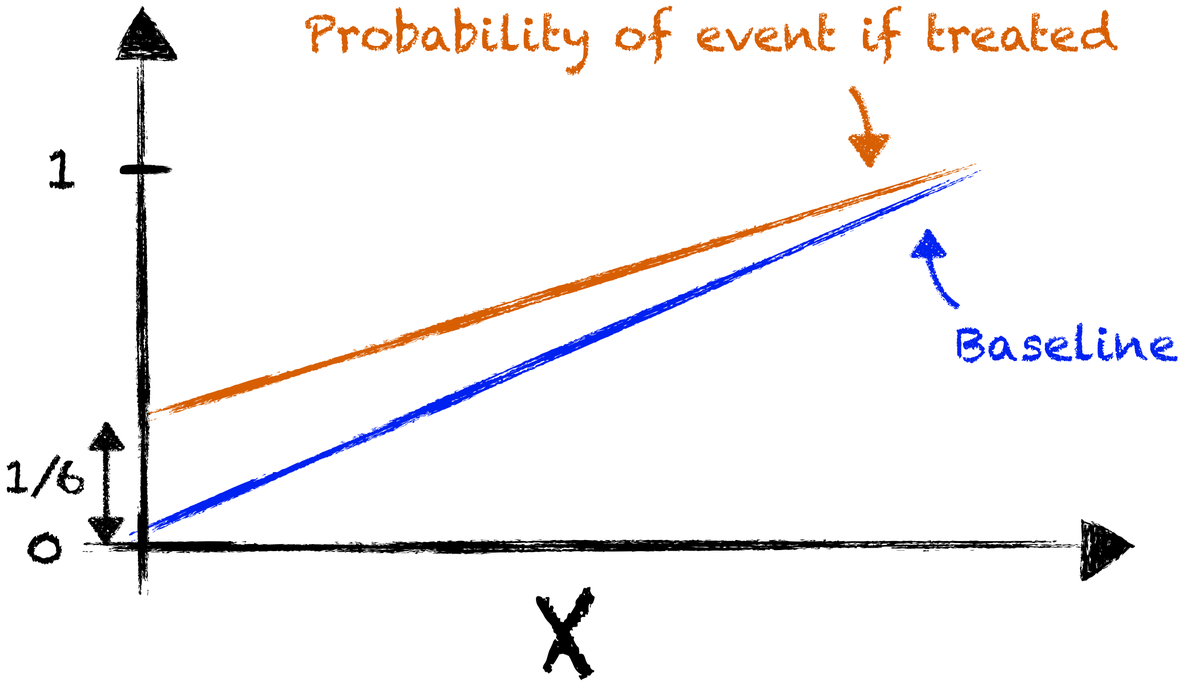


Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

The example of the Russian roulette

Harmful

For Y <u>binary</u>,



Lemma

There exist two functions b(.) and m(.) such that,

$$\mathbb{P}\left[Y^{(a)} = 1 \mid X\right] = b(X) + a\left(1 - b\left(X\right)\right)\frac{1}{6}$$

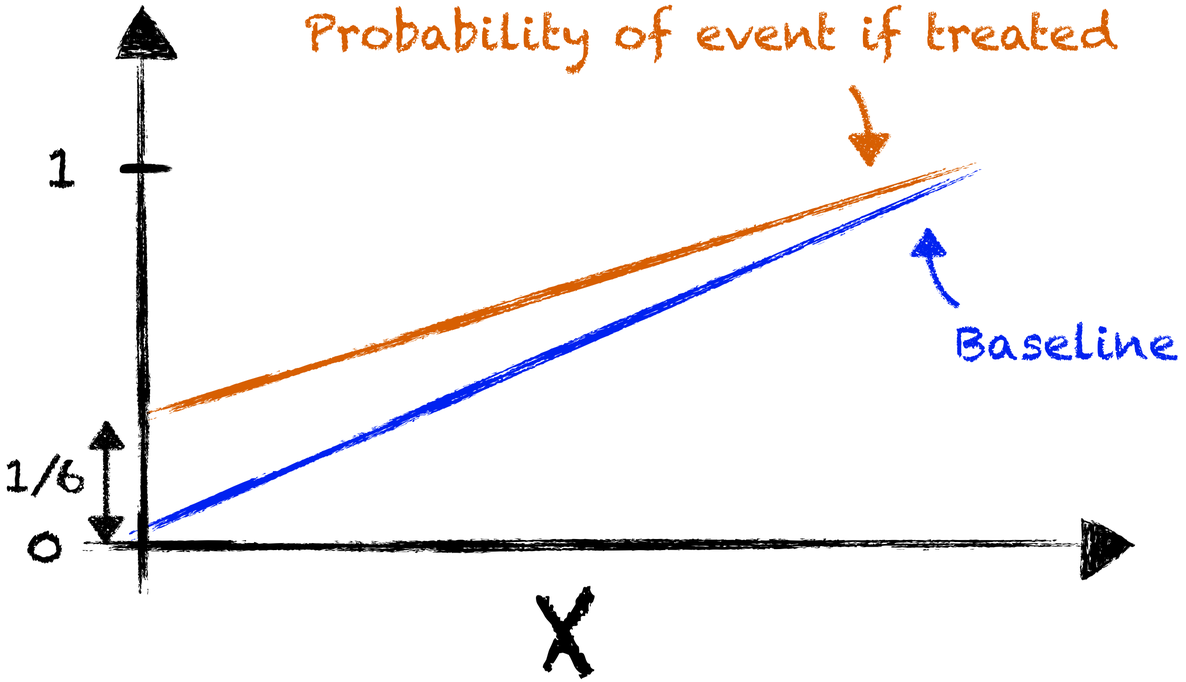
Simple additivity is not possible anymore

Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

The example of the Russian roulette

Harmful

For Y <u>binary</u>,



Lemma

There exist two functions b(.) and m(.) such that,

$$\mathbb{P}\left[Y^{(a)} = 1 \mid X\right] = b(X) + a\left(1 - b\left(X\right)\right)\frac{1}{6}$$

Simple additivity is not possible anymore

Linking generative functions with measures

$$\tau_{RD}(x) = (1 - b(x))\frac{1}{6}$$ Entanglement

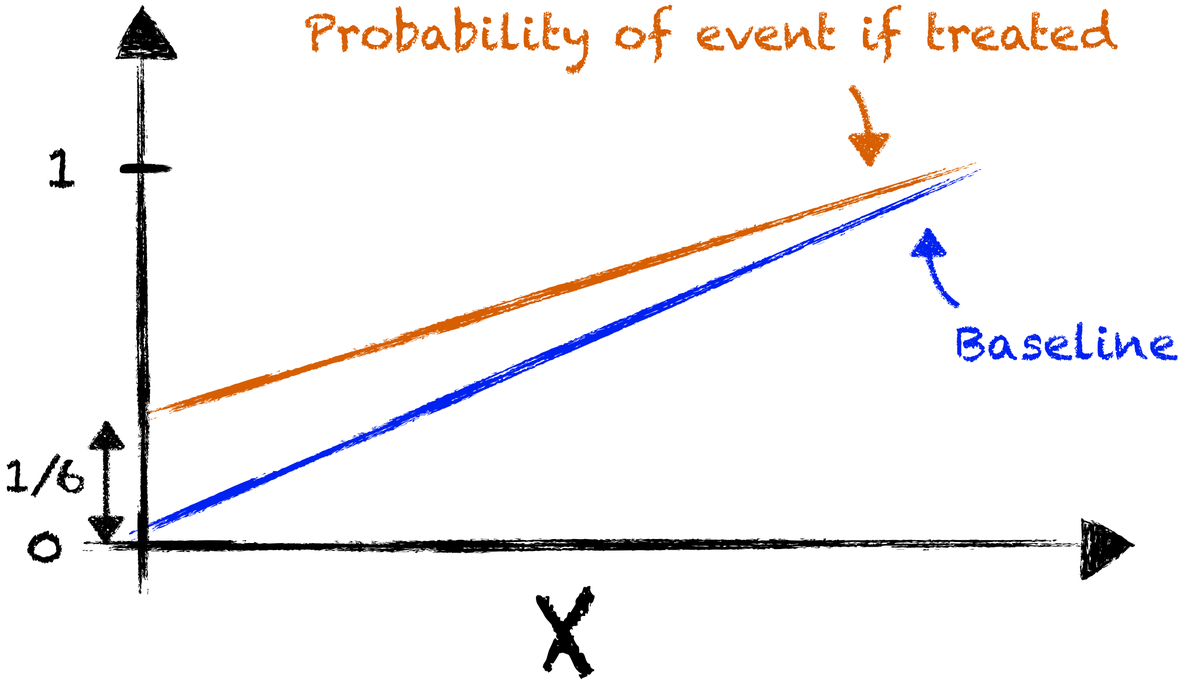
$$\tau_{SR}(x) = 1 - \frac{1}{6}$$ No entanglement

Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

The example of the Russian roulette

Harmful

For Y binary,



Lemma

There exist two functions b(.) and m(.) such that,

$$\mathbb{P}\left[Y^{(a)} = 1 \mid X\right] = b(X) + a\left(1 - b\left(X\right)\right) m(X)$$

Simple additivity is not possible anymore

Linking generative functions with measures

$$\tau_{RD}(x) = (1 - b(x))m(x)$$ Entanglement

$$\tau_{SR}(x) = 1 - m(x)$$ No entanglement

Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

Extension to all effect types (harmful and beneficial)

Considering a binary outcome, assume that

$$\forall x \in \mathbb{X}, \forall a \in \{0,1\}, \quad 0 < p_a(x) < 1, \quad \textbf{where } p_a(x) := \mathbb{P}\left[Y^{(a)} = 1 \mid X = x\right]$$

Assumptions

Introducing,

$$m_g(x) := \mathbb{P}\left[Y^{(1)} = 0 \mid Y^{(0)} = 1, X = x\right] \quad \textbf{and} \quad m_b(x) := \mathbb{P}\left[Y^{(1)} = 1 \mid Y^{(0)} = 0, X = x\right],$$

Extension to all effect types (harmful and beneficial)

Considering a binary outcome, assume that

$$\forall x \in \mathbb{X}, \forall a \in \{0,1\}, \quad 0 < p_a(x) < 1, \quad \textbf{where } p_a(x) := \mathbb{P}\left[Y^{(a)} = 1 \mid X = x\right]$$

Assumptions

Introducing,

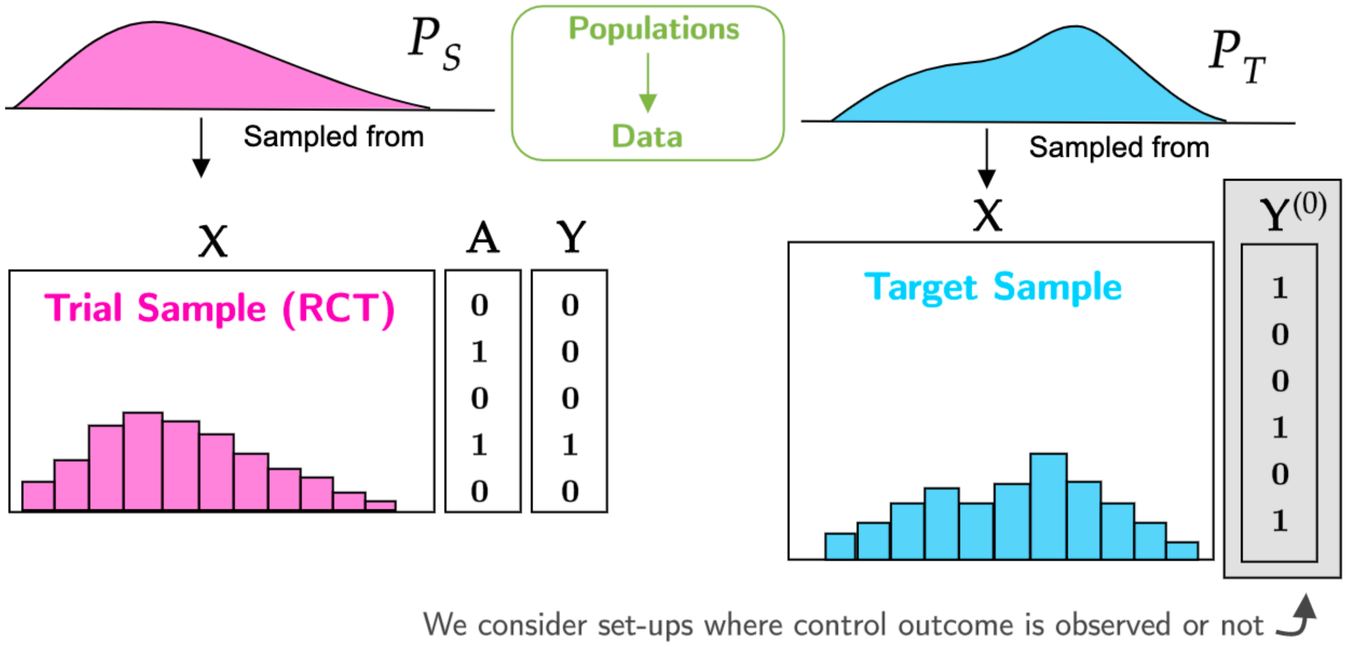
$$m_g(x) := \mathbb{P}\left[Y^{(1)} = 0 \mid Y^{(0)} = 1, X = x\right] \quad \textbf{and} \quad m_b(x) := \mathbb{P}\left[Y^{(1)} = 1 \mid Y^{(0)} = 0, X = x\right],$$

allows to have,

$$\mathbb{P}\left[Y^{(a)} = 1 \mid X = x\right] = b(x) + a\left(\underbrace{\left(1 - b\left(x\right)\right) m_b\left(x\right)}_{\text{More events}} - \underbrace{b\left(x\right) m_g\left(x\right)}_{\text{Less events}}\right), \quad \textbf{where } b(x) := p_0(x)\,.$$

Back to generalizability

Remember: we want to transport <u>trial findings</u> to a target population, using the trial data and a sample of the target population



Two methods, two assumptions

S is the indicator of population's membership

Generalizing	Conditional potential outcomes	Local effects
Assumptions for RD	$\{Y^{(0)}, Y^{(1)}\} \perp\!\!\!\perp S \mid X$	$Y^{(1)} - Y^{(0)} \perp\!\!\!\perp S \mid X$
Unformal	All shifted prognostic covariates	All shifted treatment effect modifiers
Identification		Less covariates if homogeneity

Two methods, two assumptions

S is the indicator of population's membership

Generalizing	Conditional potential outcomes	Local effects
Assumptions for RD	$\{Y^{(0)}, Y^{(1)}\} \perp\!\!\!\perp S \mid X$	$Y^{(1)} - Y^{(0)} \perp\!\!\!\perp S \mid X$
Unformal	All shifted prognostic covariates	All shifted treatment effect modifiers Less covariates if homogeneity
Identification	$\mathbb{E}^T\left[Y^{(a)}\right] = \mathbb{E}^T\left[\mathbb{E}^R\left[Y^{(a)} \mid X\right]\right]$	$\tau^T = \mathbb{E}\left[w(X, Y^{(0)})\tau^R(X)\right]$ Possible only if collapsible!

— Depending on the assumptions, either conditional outcome or local treatment effect can be generalised

Generalizing local effect, the example of a binary Y and a beneficial effect

i.e. reducing number of events

Estimate using trial sample

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

Estimate using target sample

$$\tau_{RR}(x) = 1 - m_g(x)$$

Conditional RR only vary with the shifted treatment effect modulators

⚠️ We need to have access to Y(0)!

A toy simulation

Introducing heterogeneities in the Russian roulette

- Probability to die varies
  - Stressed people can die from a heart attack
  - Executioner more merciful when facing women

A toy simulation

Introducing heterogeneities in the Russian roulette

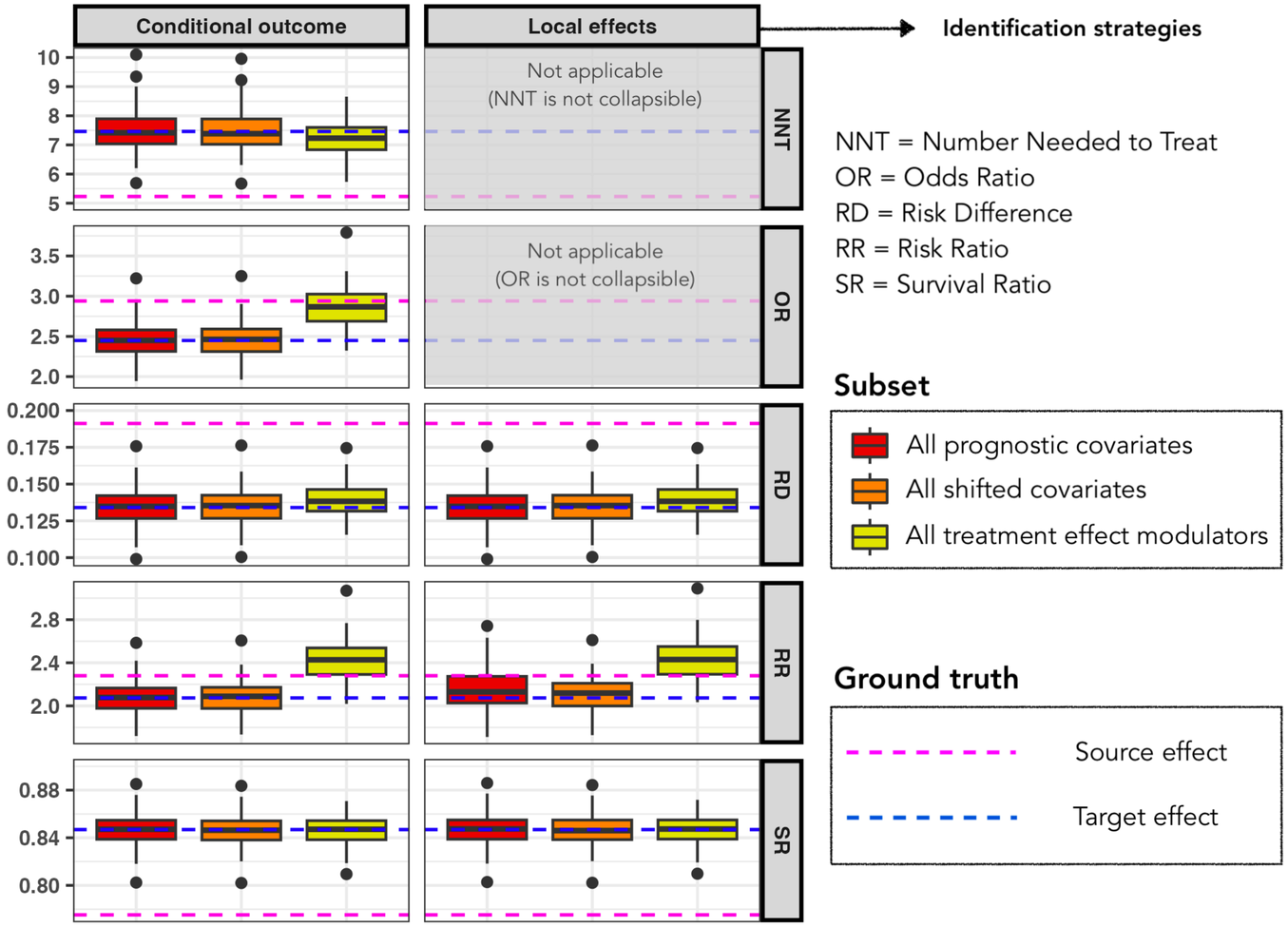
- Probability to die varies
  - Stressed people can die from a heart attack
  - Executioner more merciful when facing women

$$P[Y = 1 \mid X] = b(X_{1->3}) + (1 - b(X_{1->3}))\, m(X_{2->3})$$

X1 : lifestyle general level

X2 : stress

X3 : gender (not shifted)



— Local SR can be generalised using only stress. All others measures requires lifestyle and stress.

Contributions

All started from motivating example from critical care and two data samples with CRASH-3 & Traumabase

This leaded us to tackle a the broader scope : trial's findings generalisation.

We realised from application that many challenges remain: missing covariates, covariate selection, consistency, impact of the causal measures, etc.

Our contribution is to provide theoretical and methodological results to strengthen the practice:

- Consistency proofs
- Sensitivity analysis
- Finite and large sample results of IPSW
- Characterisation of the impact of adding non necessary covariates on precision
- Impact of the causal measure on transported treatment effect identification

Future work

This work opens new research questions

- Extension of the theoretical finite and/or large sample results for
  - G-formula and AIPSW, and not only IPSW,
  - In a context where covariates are not categorical,
  - When the ratio is targeted
    - using local effect or
    - conditional outcomes re-weighting.

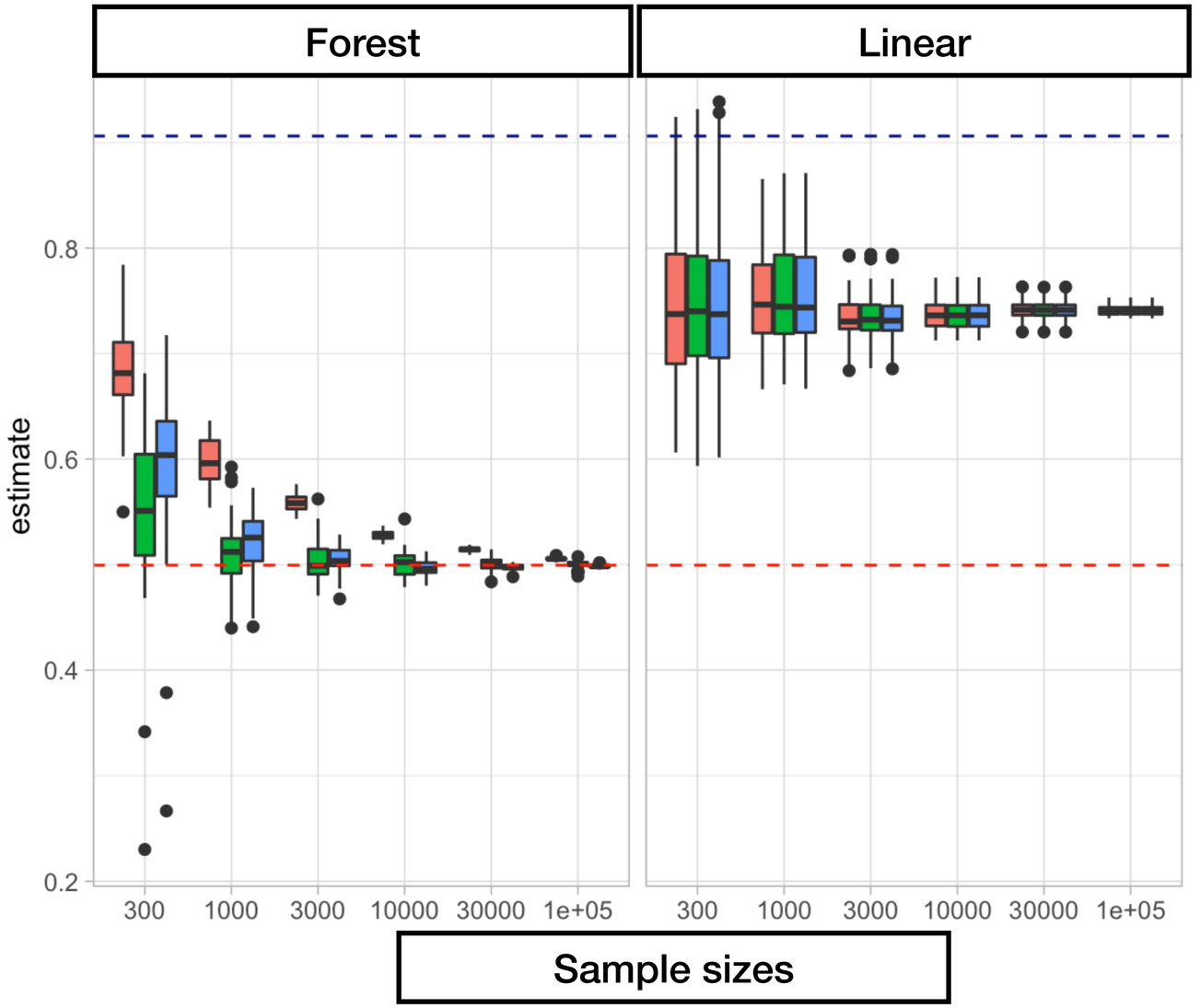
- Confront model with empirical data
  - Is the assumption of a completely beneficial or harmful effect valid in practice?
  - Using meta-analysis or different trials, investigate which causal measure is more or less dependent on the baseline level.

Why focusing on finite sample results? (1)

(1) Usual sample sizes in medicine remains small

(2) Results from simulations warned me and raised my interest

- Simulation set up from Nie and Wager
- Estimation with AIPW using either forest or linear models for nuisance parameters estimation

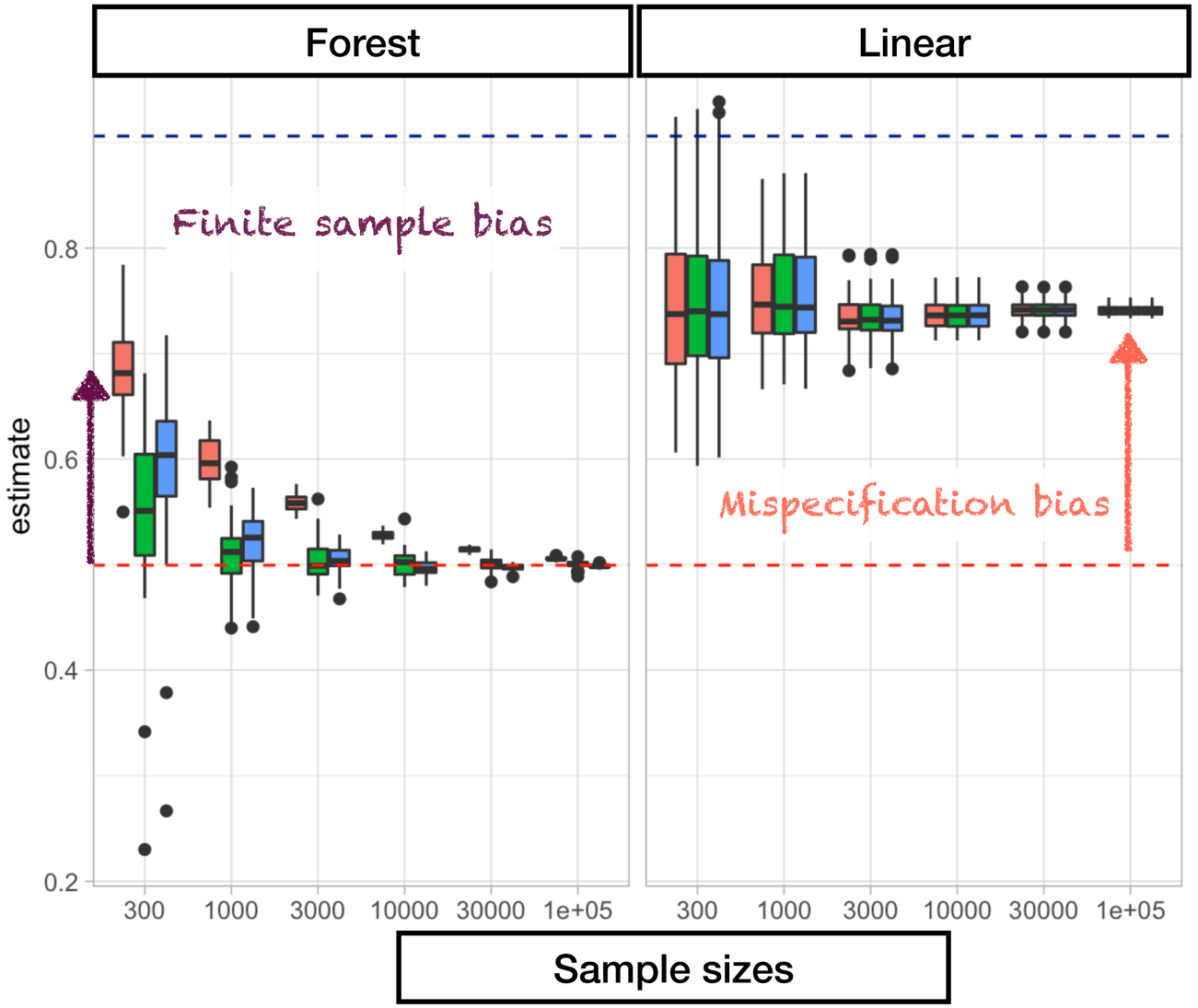


Why focusing on finite sample results? (1)

(1) Usual sample sizes in medicine remains small

(2) Results from simulations warned me and raised my interest

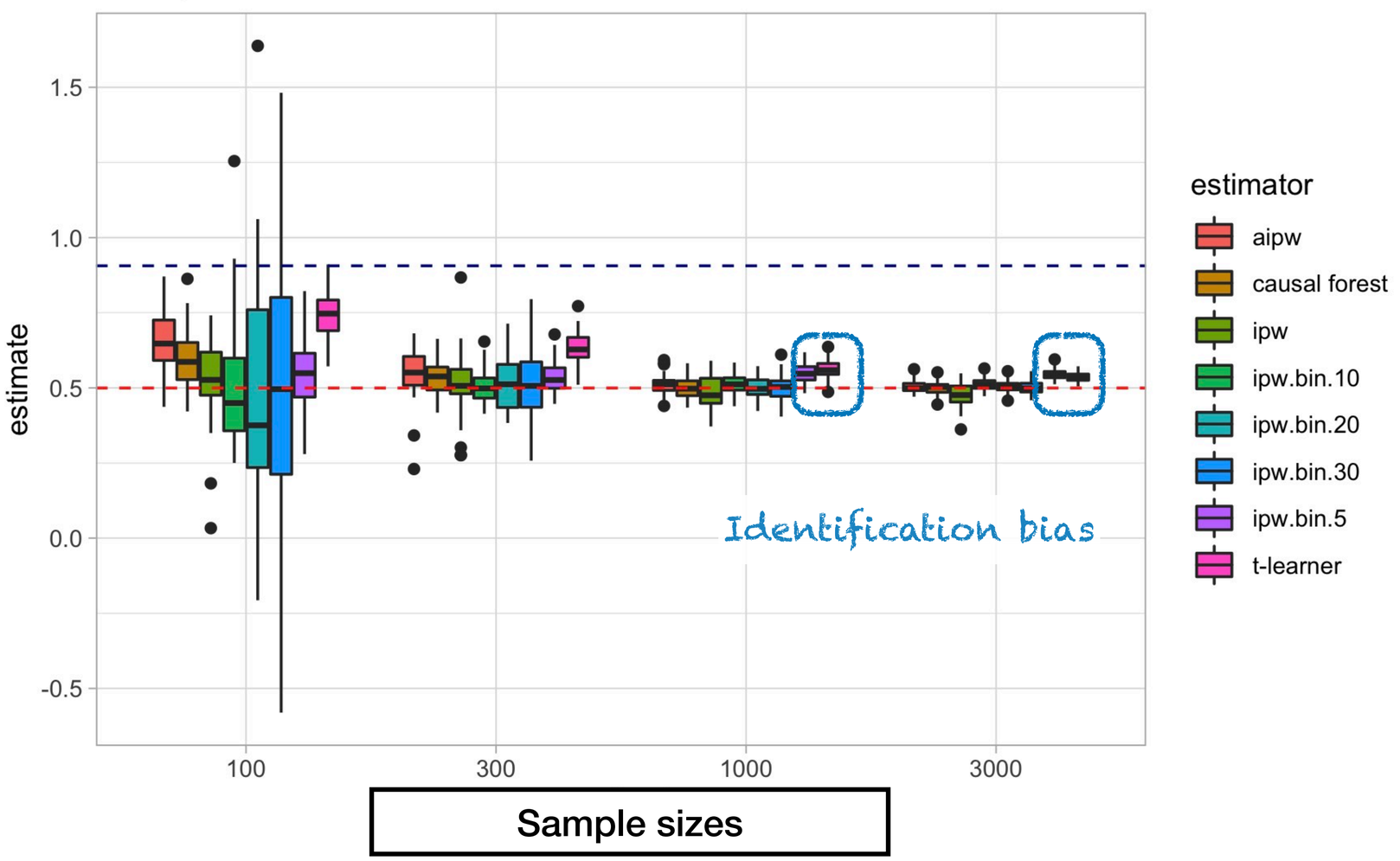
- Flexible estimation of the nuisance parameters guarantees large sample consistency…
- But at the cost of a finite sample bias!



Why focusing on finite sample results? (2)

- Flexible estimation of the nuisance parameters guarantees large sample consistency…
- But at the cost of a finite sample bias!

- Using a naive IPW with bins ensures a better finite sample risk than AIPW, at the cost of an identification bias that does not disappear with a bigger sample size.



Logistic regression and Russian roulette

Lemma 10 (Logit generative model for a binary outcome)**.** *Considering a binary outcome Y, assume that*

$$\forall x \in \mathbb{X},\, \forall a \in \{0,1\}, \quad 0 < p_a(x) < 1, \quad \textit{where } p_a(x) = \mathbb{P}(Y^{(a)} = 1 \mid X = x).$$

Then, there exist two functions $b, m : \mathcal{X} \to \mathbb{R}$ such that

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

Logistic regression and Russian roulette

Denoting $b_1(X)$ and $m_1(X)$ the functions for the intrication model, and $b_2(X)$ and $m_2(X)$ for the logistic model, one has:

$$b_2(X) = \ln\left(\frac{b_1(X)}{1-b_1(X)}\right)$$

and

$$m_2(X) = \ln\left(\frac{(m_1(X)+b_1(X))(1-b_1(X)))}{1-(m_1(X)+b_1(X))(1-b_1(X)))}\right) - \ln\left(\frac{b_1(X)}{1-b_1(X)}\right)$$

Taking the case of the Russian Roulette, one has

$$b_1(X) := p_0(X), \quad m_1(X) = \frac{1}{6}$$

so that

$$b_2(X) := \ln\left(\frac{X}{1-X}\right)$$

and

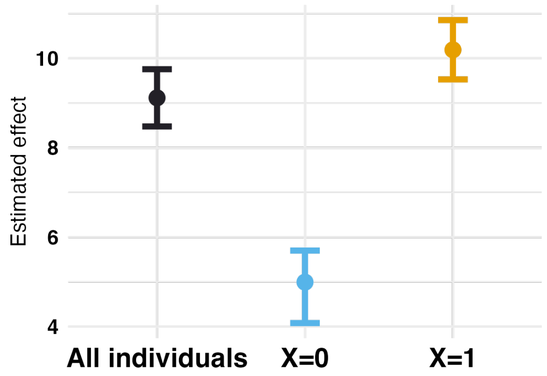
$$m_2(X) := \ln\left(\frac{\left(\frac{1}{6}+p_0(X)\right)}{1-\left(\frac{1}{6}+p_0(X)\right)(1-p_0(X)))}\right) - \ln\left(\frac{p_0(X)}{1-p_0(X)}\right).$$

Illustration on a toy simulation

Continuous outcome and binary baseline covariates X

	Target ($\mathcal{P}_{\mathbf{T}}$)	**Trial** ($\mathcal{P}_{\mathbf{R}}$)
$X = 1$	30%	75%
$X = 0$	70%	25%

Population's shift



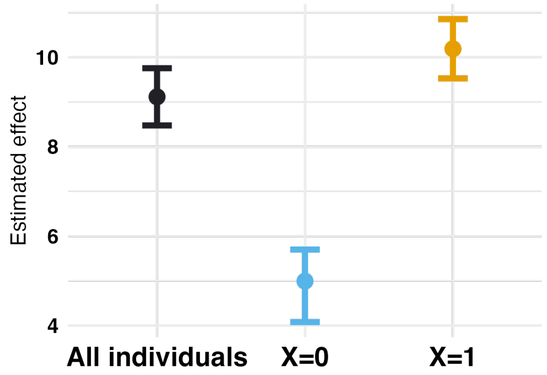
Hypothetical trial's results

Illustration on a toy simulation

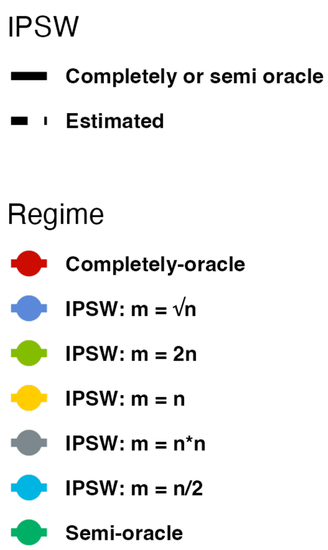
Continuous outcome and binary baseline covariates X

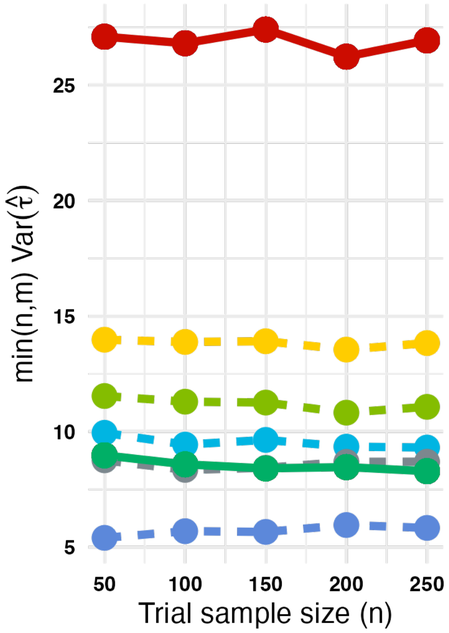
	Target ($\mathcal{P}_T$)	Trial ($\mathcal{P}_R$)
$X = 1$	30%	75%
$X = 0$	70%	25%

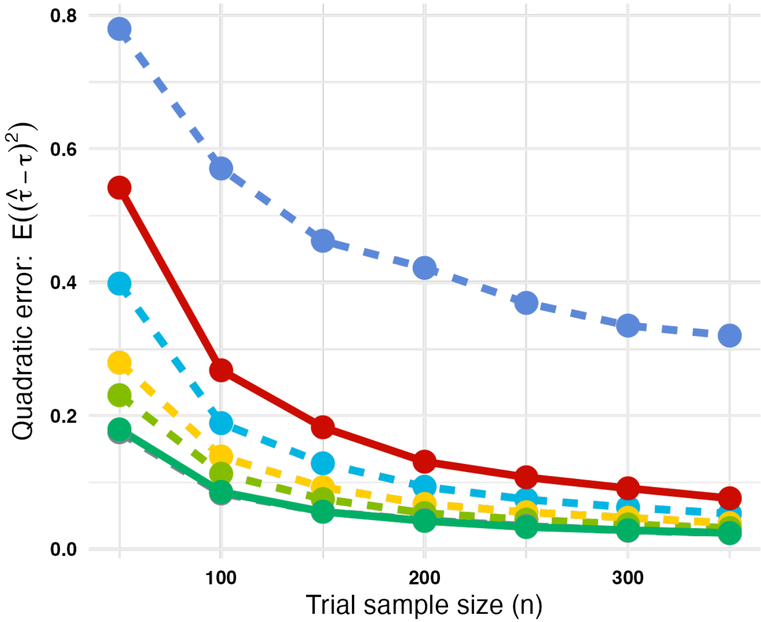
Population's shift



Hypothetical trial's results





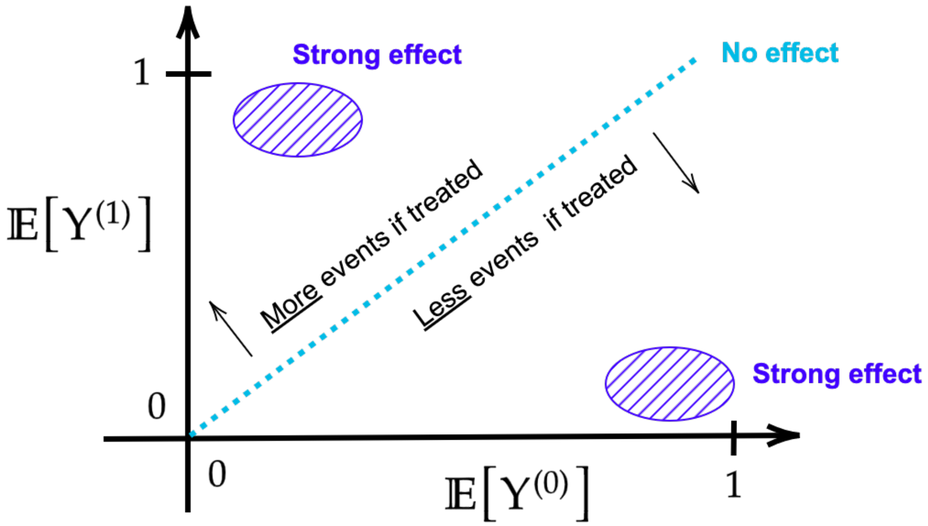


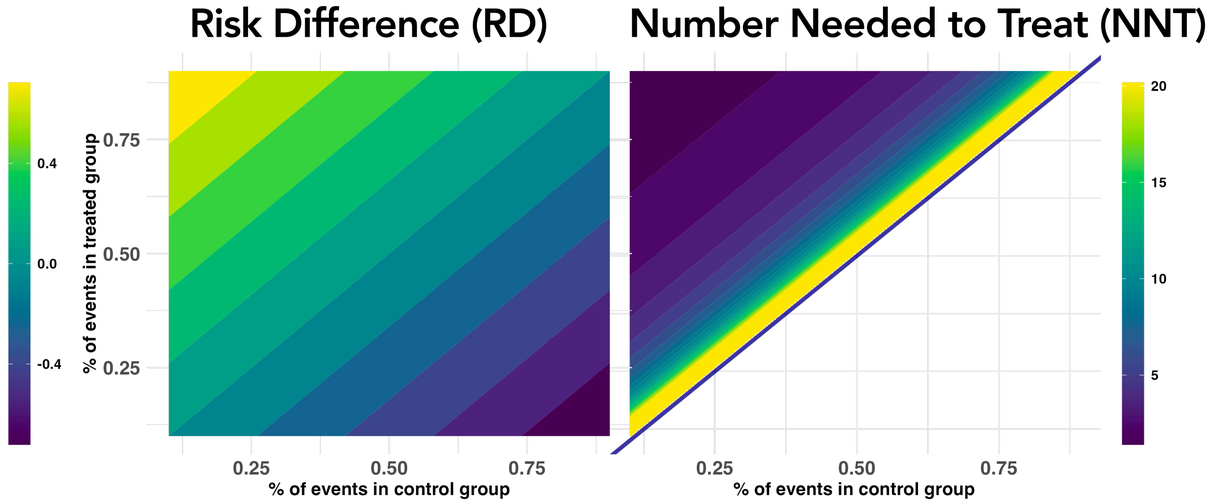
Empirical variance for different sizes **n** and **m** (6,000 repetitions for each dots) and different regimes

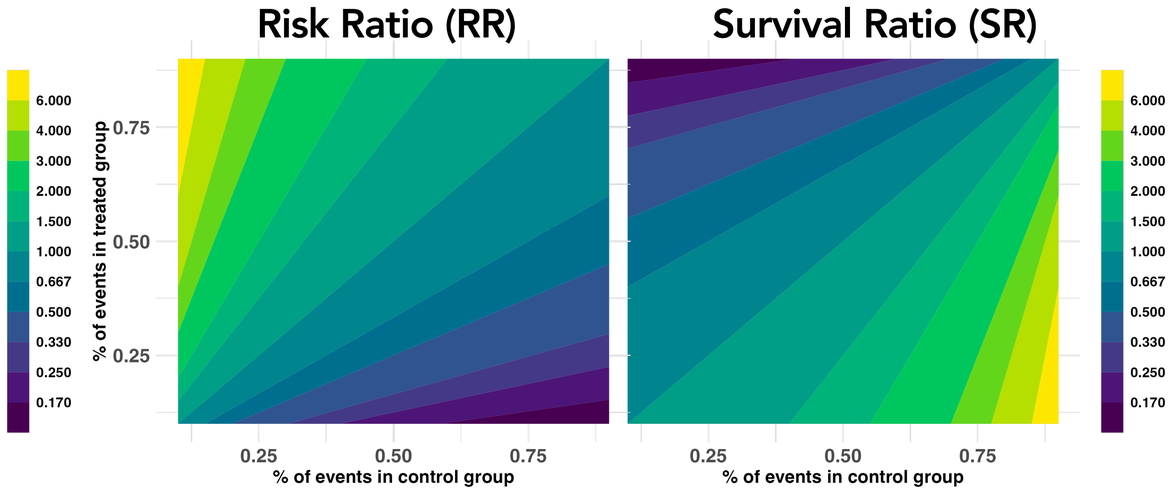
- Convergence speeds depend on the regime — i.e relative sizes of n and m,
- Completely oracle IPSW has a bigger variance than the semi-oracle IPSW.

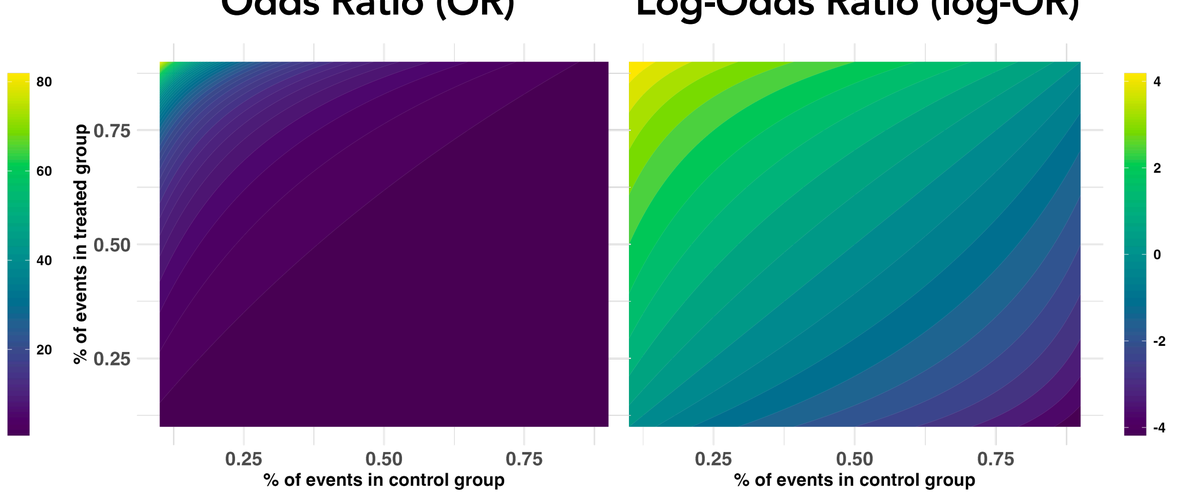
Ranges of effects

How to read plots







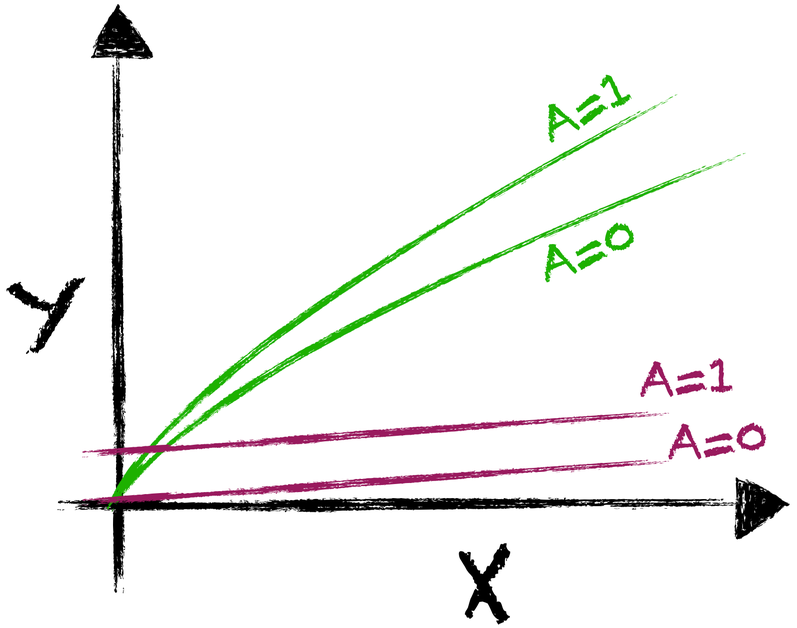


Common properties discussed

How the effect changes on sub-groups

Homogeneity $\forall x_1, x_2 \in \mathbb{X}, \quad \tau(x_1) = \tau(x_2) = \tau$

Heterogeneity $\exists x_1, x_2 \in \mathbb{X}, \quad \tau(x_1) \neq \tau(x_2)$



⚠️ *No non-zero effect can be homogeneous on all metrics*

How the effect changes with labelling

e.g. Odds Ratio is symmetric, while Risk Ratio is not

2+1 main approaches to generalize

1. **Re-weight** the trial individuals — *Inverse Propensity Sampling Weighting*
2. **Model the response** on the trial and impute the target sample — *plug-in G-formula*
3. **Combine the two** into a doubly robust approach — A(ugmented) IPSW

Consistency (Informal)

Considering that estimated surface responses are obtained following a cross-fitting estimation, then if IPSW **or** G-formula assumptions are ensured, then

$$\hat{\tau}_{AIPSW,n,m} \xrightarrow[n,m\to\infty]{L^1} \tau_T$$

See Colnet et al. 2021, published in *Journal of Causal Inference*