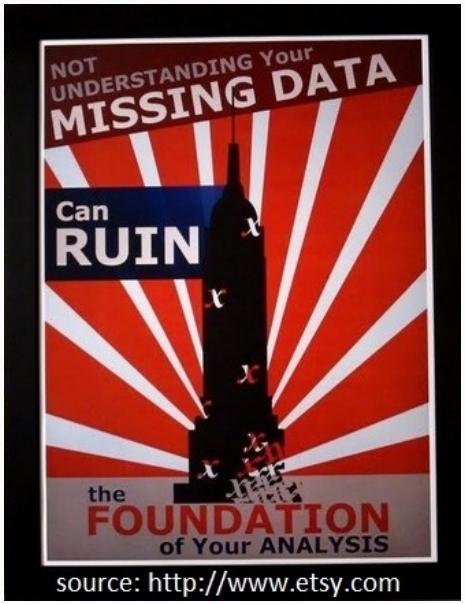
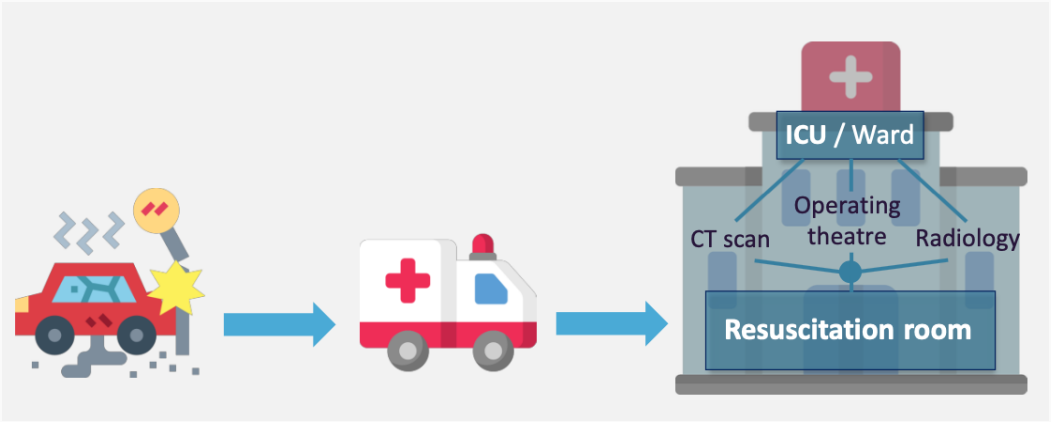
Generalization of trial's findings toward a target population: choice of a causal measure





Julie Josse. Senior Researcher Inria, Montpellier, France

Lead Inria-Inserm PreMeDICaL team: precision medicine by data integration & causal learning

Presentation Julie Josse: Stat and ML for bio-sciences

Academic background:

- Engineer & Assistant Prof. **Agronomy Univ.** Rennes (2007 - 2015)
- Visiting researcher + teaching **Stanford Univ.** (18 months)
- Professor at **Ecole Polytechnique** (IP Paris) (2016 - 2020)
- Visiting Researcher at **Google Brain** Paris (2019 - 2020)
- Senior Researcher at **Inria** Montpellier (Sept. 2020 -)
  Head of Premedical Inria-Inserm (National health research center) team

Research topics:

- Dimensionality reduction to visualize high dimensional heterogeneous data
- Missing values: supervised learning, inference, matrix completion, MNAR
- Causal inference: estimating treatment effect, combining RCT & observational data, personalized recommendation
- Medical collaborations: Critical care, Inst. Gustave Roussy, etc.

Implementations: R community

- book R for Stat., Elected member **R foundation**, Founding of Rforwards to **increase minority participation**, packages: FactoMineR (4 500 download/day, $>$ 5 million in total, $>$7000 citations), Rmisstastic, taskviews

PreMeDICaL research axes

Personalized medicine by integration of different data sources

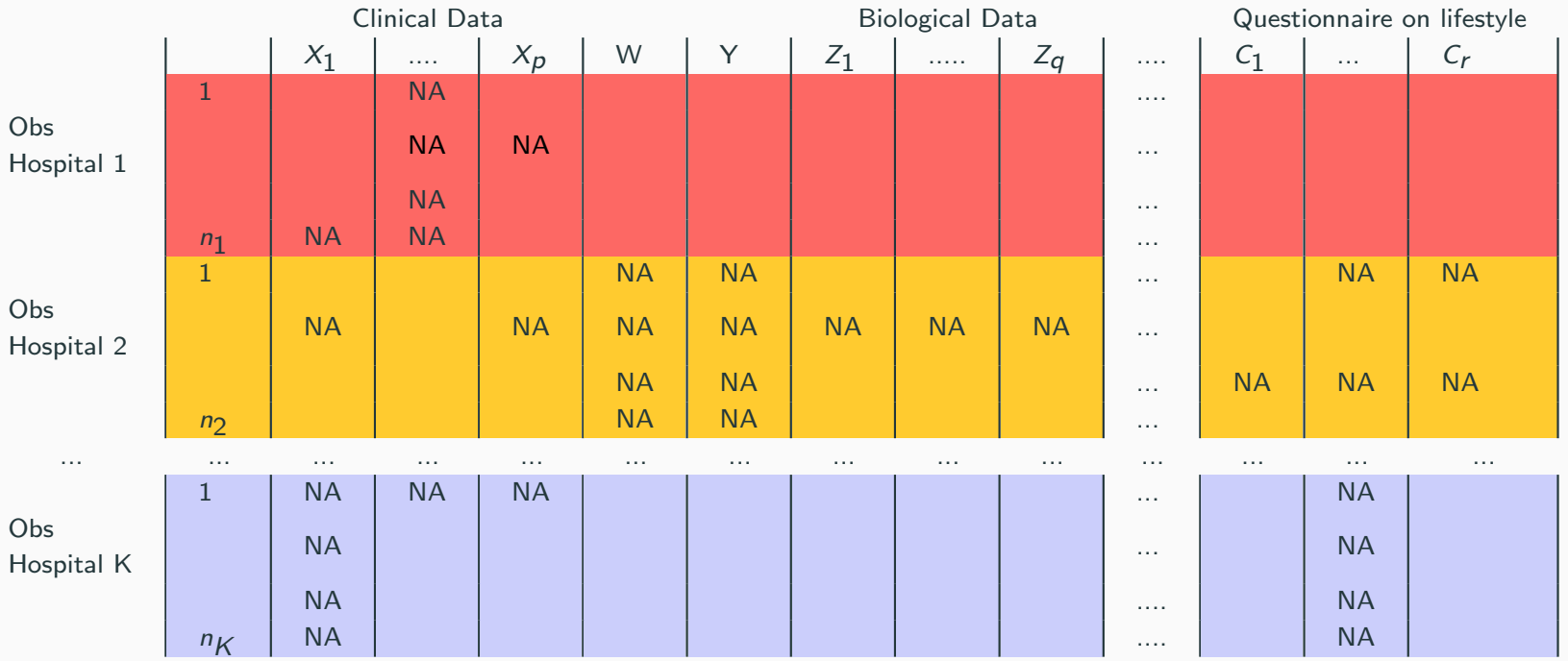
- ▷ relationship between data sources (relevance of each source?)
- ▷ solutions to handle complexe structure of missing values
- ▷ confidence in ML/AI algo. (uncertainty quantification - conformal)
- ▷ **federated learning**, privacy

⇒ Translate research **into clinically actionable** solutions

⇒ Push methodological innovation up to patients, clinicians, regulators

⇒ Leverage ML, data, clinical expertise & existing recommendations

Data integration comes with methodological challenges



State-of-the-art ML/causal methods struggle with high dimensional multi-sources data with distributional shifts & **missing data** (systematic/sporadic)

PreMeDICaL research axes

Personalized medicine by integration of different data sources

- ▷ relationship between data sources (relevance of each source?)
- ▷ solutions to handle complexe structure of missing values
- ▷ confidence in machine learning: **uncertainty quantification**
- ▷ **federated learning**, privacy

$\Rightarrow$ Translate research **into clinically actionable** solutions

$\Rightarrow$ Push methodological innovation up to patients, clinicians, regulators

$\Rightarrow$ Leverage ML, data, clinical expertise & existing recommendations

PreMeDICaL research axes

Personalized medicine by integration of different data sources

- ▷ relationship between data sources (relevance of each source?)
- ▷ solutions to handle complexe structure of missing values
- ▷ confidence in machine learning: **uncertainty quantification**
- ▷ **federated learning**, privacy

Personalized medicine by optimal prescription of treatment

- ▷ causal inference techniques for (dynamic) policy learning
  ⇒ who to treat and when
- ▷ leverage both randomized control trials (RCTs) and observational data
  ⇒ launch a drug without running RCTs
  ⇒ rethink evidence needed to bring treatments to the market faster

⇒ Translate research **into clinically actionable** solutions
⇒ Push methodological innovation up to patients, clinicians, regulators
⇒ Leverage ML, data, clinical expertise & existing recommendations

Traumabase: an observational French registry on trauma[1]

- ▷ 40000 patients
- ▷ 250 continuous and categorical variables
- ▷ 40 trauma centers
- ▷ 4000 new patients/ year

Center	Accident	Age	Sex	Weight	Lactactes	BP	TXA.	Y
Beaujon	fall	54	m	85	NM	180	treated	0
Pitie	gun	26	m	NR	NA	131	untreated	1
Beaujon	moto	63	m	80	3.9	145	treated	1
Pitie	moto	30	w	NR	Imp	107	untreated	0
HEGP	knife	16	m	98	2.5	118	treated	1
⋮								⋱

[1] www.traumabase.eu - https://www.traumatrix.fr/

Traumabase: an observational French registry on trauma[1]

- ▷ 40000 patients
- ▷ 250 continuous and categorical variables
- ▷ 40 trauma centers
- ▷ 4000 new patients/ year

Center	Accident	Age	Sex	Weight	Lactactes	BP	TXA.	Y
Beaujon	fall	54	m	85	NM	180	treated	0
Pitie	gun	26	m	NR	NA	131	untreated	1
Beaujon	moto	63	m	80	3.9	145	treated	1
Pitie	moto	30	w	NR	Imp	107	untreated	0
HEGP	knife	16	m	98	2.5	118	treated	1
⋮								⋱

⇒ **Explain and Predict** hemorrhagic shock given pre-hospital features.
Ex: logistic regression/ random forests with covariates with missing values

Clinical Trial: real-time testing of models in the ambulance via a mobile data collection application (ShockMatrix Google play)

[1] www.traumabase.eu - https://www.traumatrix.fr/

Traumabase: an observational French registry on trauma[1]

- ▷ 40000 patients
- ▷ 250 continuous and categorical variables
- ▷ 40 trauma centers
- ▷ 4000 new patients/ year

Center	Accident	Age	Sex	Weight	Lactactes	BP	TXA.	Y
Beaujon	fall	54	m	85	NM	180	treated	0
Pitie	gun	26	m	NR	NA	131	untreated	1
Beaujon	moto	63	m	80	3.9	145	treated	1
Pitie	moto	30	w	NR	Imp	107	untreated	0
HEGP	knife	16	m	98	2.5	118	treated	1
⋮								⋱

$\Rightarrow$ **Estimate causal effect**: Administration of the **treatment** *tranexamic acid (TXA)*, given within 3 hours of the accident, on the **outcome** *28 days intra hospital mortality* for trauma brain patients

TXA decreases mortality for extra-cranial bleeding. Effect for intra-cranial bleeding? (detected by CT scan). TXA is one of the first treatments given

Traumabase: an observational French registry on trauma[1]

- ▷ 40000 patients
- ▷ 250 continuous and categorical variables
- ▷ 40 trauma centers
- ▷ 4000 new patients/ year

Center	Accident	Age	Sex	Weight	Lactactes	BP	TXA.	Y
Beaujon	fall	54	m	85	NM	180	treated	0
Pitie	gun	26	m	NR	NA	131	untreated	1
Beaujon	moto	63	m	80	3.9	145	treated	1
Pitie	moto	30	w	NR	Imp	107	untreated	0
HEGP	knife	16	m	98	2.5	118	treated	1
⋮								⋱

$\Rightarrow$ **Estimate causal effect**: Administration of the **treatment** *tranexamic acid (TXA)*, given within 3 hours of the accident, on the **outcome** *28 days intra hospital mortality* for trauma brain patients

Causal inference with covariates with missing values (implemented in the R package grf). 2020. Mayer, I., Wager, S. & **J.J**. Doubly robust treatment effect estimation with incomplete confounders. *Annals Of Applied Statistics.*

Data sources and evidence at hand for the effect of TXA [3]

CRASH3

- Multi-centric RCT - 29 countries
- 9000 individuals - develp. countries
- No evidence for a TXA effect (positive effect for moderate injured patients)

Traumabase

- Observational sample
- 8200 patients with brain trauma
- Slightly negative effect of TXA/ no evidence

[2] Lodi, Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

[3] (2019). Effects of tranexamic acid on death in patients with acute trauma. brain injury. *Lancet.*

Data sources and evidence at hand for the effect of TXA [3]

CRASH3

- Multi-centric RCT - 29 countries
- 9000 individuals - develp. countries
- No evidence for a TXA effect (positive effect for moderate injured patients)

Traumabase

- Observational sample
- 8200 patients with brain trauma
- Slightly negative effect of TXA/ no evidence

Is this expected? Is there a paradox?

[2] Lodi, Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

[3] (2019). Effects of tranexamic acid on death in patients with acute trauma. brain injury. *Lancet.*

Data sources and evidence at hand for the effect of TXA [3]

CRASH3

- Multi-centric RCT - 29 countries
- 9000 individuals - develp. countries
- No evidence for a TXA effect (positive effect for moderate injured patients)

Traumabase

- Observational sample
- 8200 patients with brain trauma
- Slightly negative effect of TXA/ no evidence

Is this expected? Is there a paradox?

- Treatment and outcome are not exactly the same[2]
- Traumabase may suffer from unobserved confounding - missing values

[2] Lodi, Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

[3] (2019). Effects of tranexamic acid on death in patients with acute trauma. brain injury. *Lancet.*

Data sources and evidence at hand for the effect of TXA [3]

CRASH3

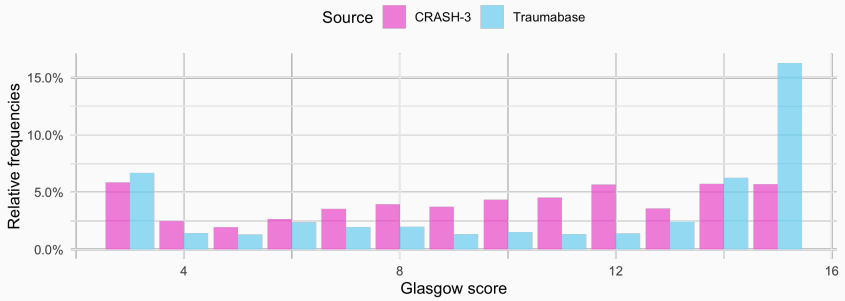
- Multi-centric RCT - 29 countries
- 9000 individuals - develp. countries
- No evidence for a TXA effect (positive effect for moderate injured patients)

Traumabase

- Observational sample
- 8200 patients with brain trauma
- Slightly negative effect of TXA/ no evidence

Is this expected? Is there a paradox?

- Treatment and outcome are not exactly the same[2]
- Traumabase may suffer from unobserved confounding - missing values
- Populations are different



[2] Lodi, Hernán et al. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol*.

[3] (2019). Effects of tranexamic acid on death in patients with acute trauma. brain injury. *Lancet.*

Leverage both RCT and observational data

RCT

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

(big) Observational data

− Confounding
+ Representative of the target population

We could use both to [4] ...

- ▷ ...validate observational methods, correct confounding bias
- ▷ ...improve estimation of heterogeneous treatment effects
- ▷ ...**generalize the treatment effect to a target population** (data fusion, transportability, recovery from selection biais)[5],[6]

[4] Colnet, **J.J.** (2022). Causal inference for combining RCT & obs. studies. *Statistical Science*.
[5] Elias Bareinboim & Judea Pearl. (2016). Causal inference & the data-fusion problem. *PNAS*.
[6] Dahabreh, Haneuse, Robins, Robertson, Buchanan, Stuart, Hernán. (2021). Study Designs for Extending Causal Inferences From a RCT to a Target Population *American J. of Epidemiology*.

Leverage both RCT and observational data

RCT

\+ No confounding

– Trial sample different from the population eligible for treatment

(big) Observational data

– Confounding

\+ Representative of the target population

We could use both to [4] . . .

▷ . . . validate observational methods, correct confounding bias

▷ . . . improve estimation of heterogeneous treatment effects

▷ . . . **generalize the treatment effect to a target population** (data fusion, transportability, recovery from selection biais)[5],[6]

The FDA has greenlighted the usage of the drug *Ibrance* to men with breast cancer, though clinical trials were performed only on women.

→ Reduce drug approval times and costs for patients who could benefit

---

[4] Colnet, **J.J.** (2022). Causal inference for combining RCT & obs. studies. *Statistical Science*.

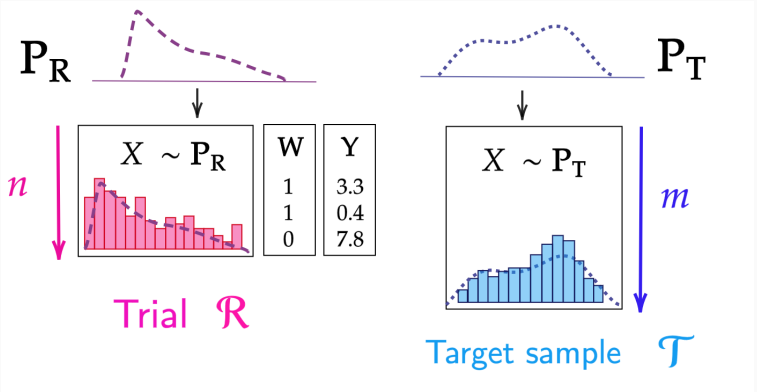
[5] Elias Bareinboim & Judea Pearl. (2016). Causal inference & the data-fusion problem. *PNAS*.

[6] Dahabreh, Haneuse, Robins, Robertson, Buchanan, Stuart, Hernán. (2021). Study Designs for Extending Causal Inferences From a RCT to a Target Population *American J. of Epidemiology*.

Generalization task from a RCT to a target population

Two data sources:

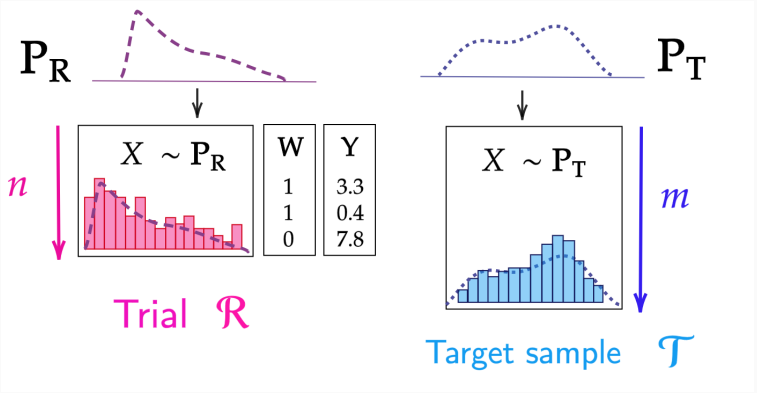
- A trial of size n with $p_R(x)$ the probability of observing individual with $X = x$,
- A sample of the target population of interest – for e.g. a national cohort (resp. m and $p_T(x)$).



Generalization task from a RCT to a target population

Two data sources:

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



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

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

Assumptions for ATE identifiability in generalization

Transportability (Ignorability on trial participation)[7]

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

Corresponds to shifted prognostic variables

[7] Equivalent formulation with sampling mechanism S ($S = 1$ trial eligibility & willingness to participate) in non-nested design, $\{Y(1), Y(0)\} \perp\!\!\!\perp S \mid X$

[8] If this is too strong, we could generalize on a different target population: the target population for which eligibility criteria of the trial are ensured

Assumptions for ATE identifiability in generalization

Transportability (Ignorability on trial participation)[7]

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

Corresponds to shifted prognostic variables

Overlap assumption[8]

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

The observational covariates support is included in the RCT's support. Every individual in the target population could have been selected into the trial

[7] Equivalent formulation with sampling mechanism S ($S = 1$ trial eligibility & willingness to participate) in non-nested design, $\{Y(1), Y(0)\} \perp\!\!\!\perp S \mid X$

[8] If this is too strong, we could generalize on a different target population: the target population for which eligibility criteria of the trial are ensured

Generalization of conditional outcome: identifiability

	Set	S	X_1	X_2	X_3	W	$Y(0)$	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

	Set	S	X_1	X_2	X_3	W	Y
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
$n+1$	$\mathcal{O}$	NA	-2	52	7.1	NA	NA
$n+2$	$\mathcal{O}$	NA	-1	35	2.4	NA	NA
...	$\mathcal{O}$	NA		...		NA	NA
$n+m$	$\mathcal{O}$	NA	-2	22	3.4	NA	NA

Data with observed treatment W and outcome Y only in the RCT.

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

$$
\begin{aligned}
\mathbb{E}_{\mathsf{T}}[Y(1)] &= \mathbb{E}_{\mathsf{T}}\left[\mathbb{E}_{\mathsf{T}}[Y(1) \mid X]\right] && \text{Law of total expectation} \\
&= \mathbb{E}_{\mathsf{T}}\left[\mathbb{E}_{\mathsf{R}}[Y(1) \mid X]\right] && \text{Ignorability} \\
&= \mathbb{E}_{\mathsf{T}}\left[\mathbb{E}_{\mathsf{R}}[Y(1) \mid X = x, W = 1]\right] && \text{Random treatment} \\
&= \mathbb{E}_{\mathsf{T}}\underbrace{\left[\mathbb{E}_{\mathsf{R}}[Y \mid X = x, W = 1]\right]}_{\mu_1(x)} && \text{Consistency } Y = Y(1)W + (1-W)Y(0)
\end{aligned}
$$

Generalization of conditional outcome: identifiability

	Set	S	X_1	X_2	X_3	W	$Y(0)$	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

	Set	S	X_1	X_2	X_3	W	Y
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
$n+1$	$\mathcal{O}$	NA	-2	52	7.1	NA	NA
$n+2$	$\mathcal{O}$	NA	-1	35	2.4	NA	NA
...	$\mathcal{O}$	NA		...		NA	NA
$n+m$	$\mathcal{O}$	NA	-2	22	3.4	NA	NA

Data with observed treatment W and outcome Y only in the RCT.

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

$$
\begin{aligned}
\mathbb{E}_{\mathsf{T}}[Y(1)] &= \mathbb{E}_{\mathsf{T}}\left[\mathbb{E}_{\mathsf{T}}[Y(1) \mid X]\right] \text{ Law of total expectation} \\
&= \mathbb{E}_{\mathsf{T}}\left[\mathbb{E}_{\mathsf{R}}[Y(1) \mid X]\right] \text{ Ignorability} \\
&= \mathbb{E}_{\mathsf{T}}\left[\mathbb{E}_{\mathsf{R}}[Y(1) \mid X = x, W = 1]\right] \text{ Random treatment} \\
&= \mathbb{E}_{\mathsf{T}}\underbrace{\left[\mathbb{E}_{\mathsf{R}}[Y \mid X = x, W = 1]\right]}_{\mu_1(x)} \text{ Consistency } Y = Y(1)W + (1-W)Y(0)
\end{aligned}
$$

Regression adjustment - plug-in gformula

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

Plug-in gformula: difference between conditional mean

Plug-in gformula

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

$$\mu_w(x) = \mathbb{E}_R[Y \mid X = x, W = w]$$

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

- Fit two models of the outcome (Y) on covariates (X)

among trial participants ($\mathcal{R}$) for treated and for control to get $\widehat{\mu}_{1,n}$ & $\widehat{\mu}_{0,n}$

Plug-in gformula: difference between conditional mean

Plug-in gformula

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

$$\mu_w(x) = \mathbb{E}_R[Y \mid X = x, W = w]$$

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

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

Assumptions for identifiability with fewer covariates

Transportability (Ignorability on trial participation)[9]

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

Corresponds to shifted prognostic variables

[9] Equivalent formulation with sampling mechanism S ($S = 1$ trial eligibility & willingness to participate) in non-nested design, $(Y(1), Y(0)) \perp\!\!\!\perp S \mid X$

[10] Equivalent formulation with sampling mechanism S: $(Y(1) - Y(0)) \perp\!\!\!\perp S \mid X$

Assumptions for identifiability with fewer covariates

Transportability (Ignorability on trial participation)[9]

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

Corresponds to shifted prognostic variables

Transportability of the CATE[10]

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

Corresponds to shifted treatment effect modifiers

[9] Equivalent formulation with sampling mechanism S ($S = 1$ trial eligibility & willingness to participate) in non-nested design, $(Y(1), Y(0)) \perp\!\!\!\perp S \mid X$

[10] Equivalent formulation with sampling mechanism S: $(Y(1) - Y(0)) \perp\!\!\!\perp S \mid X$

Identifiability and estimation for generalization: weighting

Generalization of local effects[11]

$$\begin{aligned}\tau_T &= \mathbb{E}_T[Y_i(1) - Y_i(0)] = \mathbb{E}_T[\mathbb{E}_T[Y_i(1) - Y_i(0)|X]] \\ &= \mathbb{E}_T\left[\tau_T(X)\right] = \mathbb{E}_T\left[\tau_R(X)\right] \text{ Transportability CATE} \\ &= \mathbb{E}_R\left[\frac{p_T(X)}{p_R(X)}\tau_R(X)\right]\end{aligned}$$

IPSW: inverse propensity sampling weighting[12]

$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n}\sum_{i\in\mathcal{R}} \frac{\hat{p}_T(X_i)}{\hat{p}_R(X_i)} \left(\frac{WY}{e_R(x)} - \frac{(1-W)Y}{1-e_R(x)}\right),$$

$e_R(x) = P(W = 1 \mid X = x)(= \pi = 0.5.)$

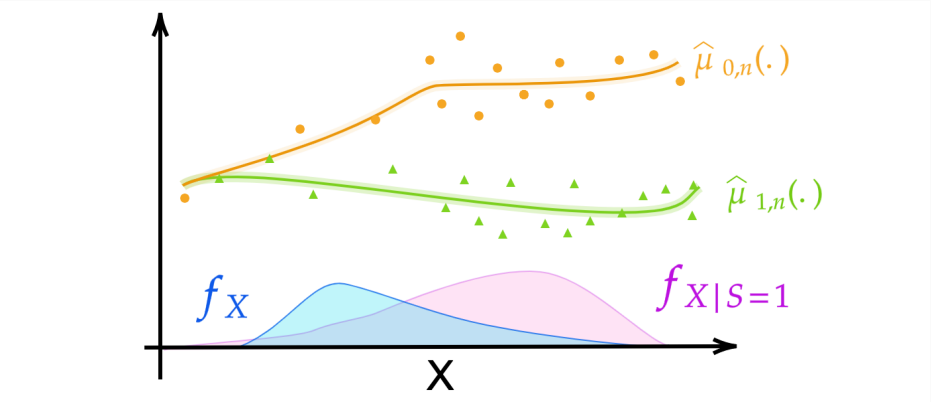
Re-weight, so that the trial follows the target sample's distribution: if proba to be in trial when old is small, then up-weight old in trial

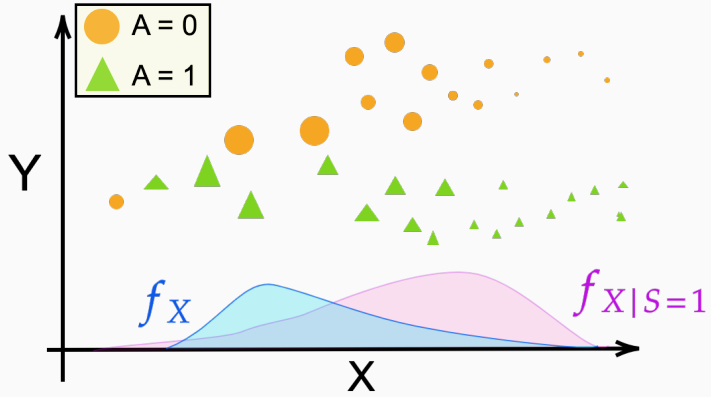
- ▷ Re-weighting can be found in the 2000's (*standardization, Rothman & Greenland)*
- ▷ But the idea of relying on an external representative sample is recent

[11] When the measure is collapsible

[12] Cole & Stuart. (2010). Generalizing evidence from RCT to target pop.. *American J. of*

Generalization estimators: illustrative schematics





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

Theorem - consistency[13]

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

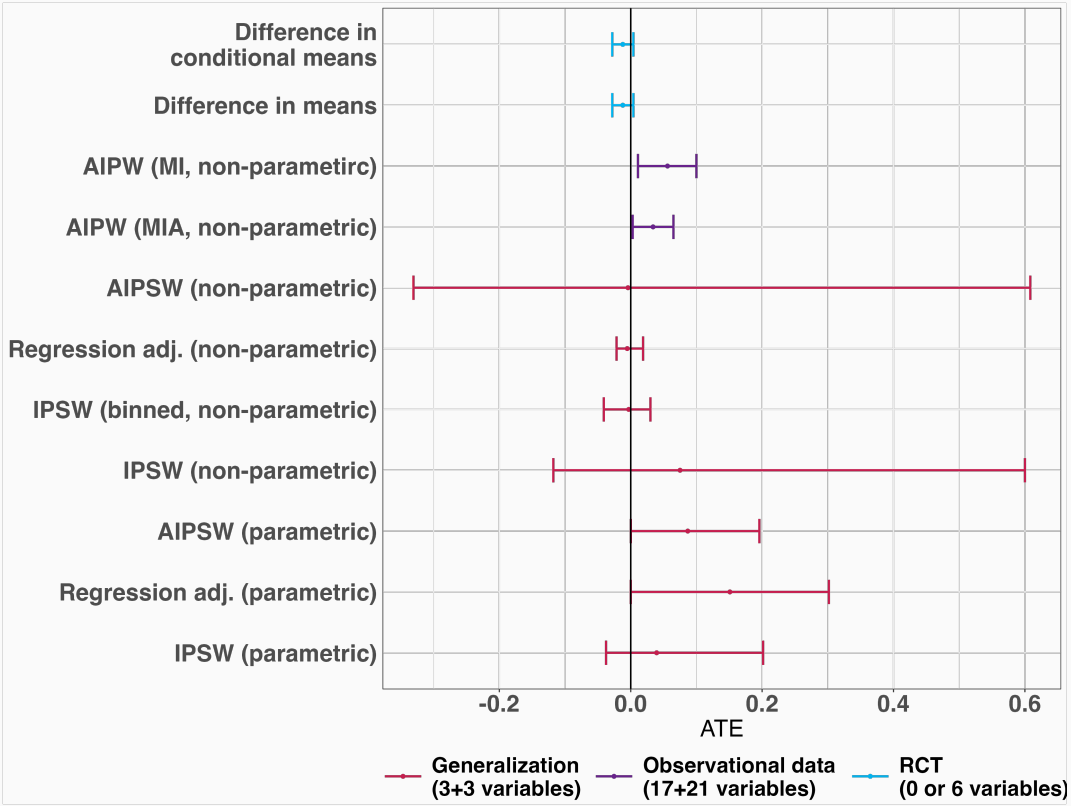
$$\hat{\tau}_{\text{IPSW},n,m} \xrightarrow[n,m\to\infty]{L^1} \tau_{\text{T}}$$

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

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

Generalization from Crash 3 trial to the Traumabase

Comparison of trials, observational data, and generalization estimates



Confidence intervals obtained with stratified bootstrap

x-axis: Estimat. of the ATE ($\times 100$), bootstrap CI

y-axis: Methods - parametric: logistic regression or non parametric forests for nuisances

Missing values handled with multiple imputation MI or missing incorportaed in attributes MIA for forests

Many medical and statistical challenges

• 1) Shifted effect modifiers not available in Traumabase[14]. Missing covariates in one/both sets: **sensitivity analysis**

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

Difficult to give sensitivity parameters, semi-parametric model (linear CATE), shift on means, sensitivity plot for one missing variable, etc.

[14] Colnet, **J.J**, et al. 2022. Generalizing a causal effect: sensitivity analysis. *J. of Causal Inference*.
[15] Mayer, **J.J.** 2021. Generalizing effects with incomplete covariates *Biometrical Journal.*
[16] Colnet, **J.J** et al. 2022. Reweighting the RCT for generalization: finite sample analysis and variable selection. *In revision JRSSC.*

Many medical and statistical challenges

• 1) Shifted effect modifiers not available in Traumabase[14]. Missing covariates in one/both sets: **sensitivity analysis**

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

Difficult to give sensitivity parameters, semi-parametric model (linear CATE), shift on means, sensitivity plot for one missing variable, etc.

[14] Colnet, **J.J**, et al. 2022. Generalizing a causal effect: sensitivity analysis. *J. of Causal Inference*.
[15] Mayer, **J.J.** 2021. Generalizing effects with incomplete covariates *Biometrical Journal.*
[16] Colnet, **J.J** et al. 2022. Reweighting the RCT for generalization: finite sample analysis and variable selection. *In revision JRSSC.*

Many medical and statistical challenges

• 1) Shifted effect modifiers not available in Traumabase[14]. Missing covariates in one/both sets: **sensitivity analysis**

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

Difficult to give sensitivity parameters, semi-parametric model (linear CATE), shift on means, sensitivity plot for one missing variable, etc.

• 2) **Missing values:** Missing values in both RCT and Obs data[15]

[14] Colnet, **J.J**, et al. 2022. Generalizing a causal effect: sensitivity analysis. *J. of Causal Inference*.
[15] Mayer, **J.J.** 2021. Generalizing effects with incomplete covariates *Biometrical Journal.*
[16] Colnet, **J.J** et al. 2022. Reweighting the RCT for generalization: finite sample analysis and variable selection. *In revision JRSSC.*

Many medical and statistical challenges

• 1) Shifted effect modifiers not available in Traumabase[14]. Missing covariates in one/both sets: **sensitivity analysis**

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

Difficult to give sensitivity parameters, semi-parametric model (linear CATE), shift on means, sensitivity plot for one missing variable, etc.

• 2) **Missing values:** Missing values in both RCT and Obs data[15]

• 3) Effect of **finite sample**? Which covariate to include? Would adding prognostic variables reduce the variance as in the classical case?[16]

[14] Colnet, **J.J**, et al. 2022. Generalizing a causal effect: sensitivity analysis. *J. of Causal Inference*.
[15] Mayer, **J.J.** 2021. Generalizing effects with incomplete covariates *Biometrical Journal.*
[16] Colnet, **J.J** et al. 2022. Reweighting the RCT for generalization: finite sample analysis and variable selection. *In revision JRSSC.*

Reweighting the RCT: finite sample & asymptotic analysis[17]

$$\hat{\tau}_{\pi,n,m} = \frac{1}{n}\sum_{i\in\mathcal{R}} \frac{\overbrace{\hat{p}_{\mathrm{T}}(X_i)}^{\text{Estimated with } \mathcal{T}}}{\underbrace{\hat{p}_{\mathrm{R}}(X_i)}_{\text{Estimated with } \mathcal{R}}} Y_i \left(\frac{W_i}{\pi} - \frac{1-W_i}{1-\pi}\right),$$

Asymptotic properties - completely estimated

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

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

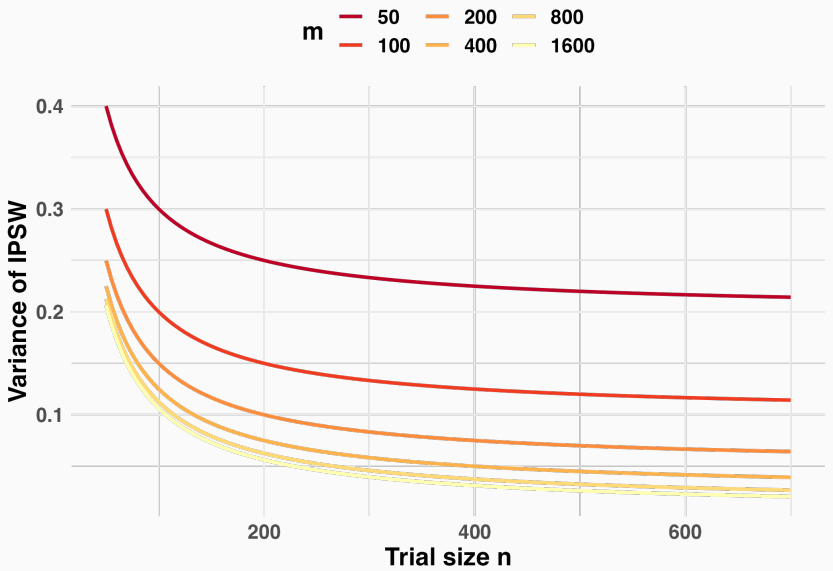
Variance depends on the size of the two data sets, n and m

- If target >> trial, $m/n \to \infty$, (i.e., $\lambda = \infty$): asymptotic variance = Semi-Oracle's one and depends on the ratio of proba.
- If trial >> target $m/n \to 0$, (i.e., $\lambda = 0$): asymptotic variance depends on $\mathrm{Var}\left[\tau(X)\right]$.

[17] Colnet, **J.J** et al. 2022. Reweighting the RCT for generalization: finite sample analysis and variable selection. *In revision JRSSC.*

Reweighting the RCT: finite sample & asymptotic analysis[17]

$$\hat{\tau}_{\pi,n,m} = \frac{1}{n} \sum_{i \in \mathcal{R}} \frac{\overbrace{\hat{p}_{\mathrm{T}}(X_i)}^{\text{Estimated with } \mathcal{T}}}{\underbrace{\hat{p}_{\mathrm{R}}(X_i)}_{\text{Estimated with } \mathcal{R}}} Y_i \left(\frac{W_i}{\pi} - \frac{1 - W_i}{1 - \pi} \right),$$

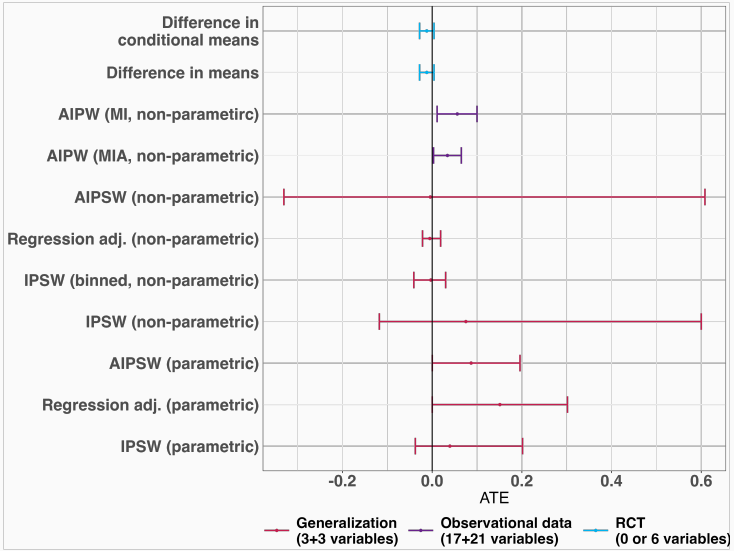


- Adding shifted and independent covariates: loss in variance
- Adding non-shifted treatment effect modifiers: gain in variance

[17] Colnet, **J.J** et al. 2022. Reweighting the RCT for generalization: finite sample analysis and variable selection. *In revision JRSSC.*

Generalization from Crash 3 trial to the Traumabase

Comparison of trials, observational data, and generalization estimates



Confidence intervals obtained with stratified bootstrap

- 1) Shifted effect modifiers not available in Traumabase[18]
- 2) **Missing values** in both RCT and Obs data[19]
- 3) Effect of **finite sample**? Which covariate to include? Would adding prognostic variables reduce the variance as in the classical case?[20]
- 4) Clinicians rather interested in the ratio than the difference

---

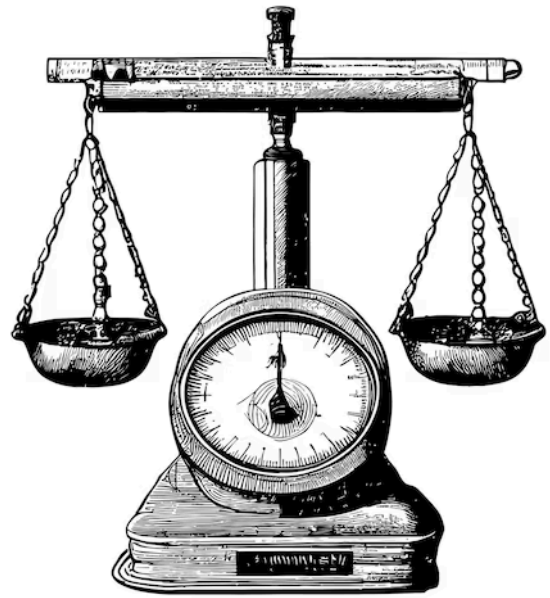
[18] Colnet, **J.J.** 2022. Generalizing a causal effect: **sensitivity analysis**, *Journal of Causal Inference*.

[19] Mayer, **J.J.** 2021. Generalizing effects with incomplete covariates *Biometrical Journal.*

[20] Colnet, **J.J.** 2022. Reweighting the RCT for generalization: finite sample analysis & variable selection. *Submitted.*

Risk ratio, odds ratio, risk difference

Which causal measure is easier to generalize?



Comparing two average situations

Binary outcome: $\mathbb{P}[Y(w) = 1] = \mathbb{E}[Y(w)]$ and $\mathbb{P}[Y(w) = 0] = 1 - \mathbb{E}[Y(w)]$.

Absolute measures

$$\tau^{\text{RD}} := \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)], \qquad \tau^{\text{NNT}} := {\tau_{\text{RD}}}^{-1}.$$

• Number Needed to Treat (NNT): how many individuals should be treated to observe one individual answering positively to treatment.

Relative measures

$$\tau^{\text{RR}} := \frac{\mathbb{E}[Y(1)]}{\mathbb{E}[Y(0)]}, \quad \tau^{\text{SR}} := \frac{\mathbb{P}[Y(1) = 0]}{\mathbb{P}[Y(0) = 0]} = \frac{1 - \mathbb{E}[Y(1)]}{1 - \mathbb{E}[Y(0)]},$$

$$\tau^{\text{OR}} := \frac{\mathbb{P}[Y(1) = 1]}{\mathbb{P}[Y(1) = 0]} \left(\frac{\mathbb{P}[Y(0) = 1]}{\mathbb{P}[Y(0) = 0]}\right)^{-1}$$

• A null effect now corresponds to a Risk Ratio of 1

• Survival Ratio (SR) corresponds to the RR with swapped labels Y

• RR is not symmetric to the choice of outcome 0 and 1 –e.g. counting the living or the dead while Odds Ratio (OR) is

Different treatment measures give different impressions

RCT from Cook and Sackett (1995)

- $Y = 1$ stroke in 5 years and $Y = 0$ no stroke
- $X = 1$ low baseline risk $\mathbb{P}[Y(0) = 1 \mid X = 0] \geq \mathbb{P}[Y(0) = 1 \mid X = 1]$

	τ_{RD}	τ_{RR}	τ_{SR}	τ_{NNT}	τ_{OR}
All (P_R)	-0.0452	**0.6**	1.05	22	0.57
X = 1	-0.006	**0.6**	1.01	167	0.6
X = 0	-0.08	**0.6**	1.1	13	0.545

- RD: treatment reduces by 0.045 the probability to suffer from a stroke
- RR: the treated has 0.6 $\times$ the risk of having a stroke comp. with the control
- SR: there is an increased chance of not having a stroke when treated compared to the control by a factor 1.05.
- NNT: one has to treat 22 people to prevent one additional stroke
- OR $\approx$ RR in a stratum where prevalence of the outcome is low

- RD is heterogeneous with X while RR is homogeneous with X

$\Rightarrow$ Heterogeneity's property defined w.r.t. *(i)* covariates & *(ii)* a measure

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.

Simpson paradox:

(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

Marginal effect larger than subgroups' effects

Unfortunately, not all measures are collapsible

Writting marginal effect as a weighed sum of conditional effects

Direct collapsibility

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

$\Rightarrow$ RD directly collapsible:

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

Ex: $0.0452 = -0.47 \times 0.006 - 0.53 \times 0.08$

Weights are equal to the population's proportions

Useful for generalization! (replacing p_{R} by p_{T})

Writting marginal effect as a weighed sum of conditional effects

Direct collapsibility

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

$\Rightarrow$ RD directly collapsible:

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

Ex: $0.0452 = -0.47 \times 0.006 - 0.53 \times 0.08$

Weights are equal to the population's proportions

Useful for generalization! (replacing p_{R} by p_{T})

Collapsibility (require knowing $Y(0)$)

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

$\Rightarrow$ RR collapsible:

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

Measures' properties

Direct collapsibility

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

Collapsibility (require knowing $Y(0)$)

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

Logic respecting

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

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

Non parametric generative models

Continuous outcome

Assuming that $\mathbb{E}\big[|Y(1)|\,\big|X\big] < \infty$ and $\mathbb{E}\big[|Y(0)|\,\big|X\big] < \infty$

$$\begin{aligned} Y(0) &= f(0,X) + \varepsilon_0, \quad \text{with} \quad f(0,X) = \mathbb{E}\left[Y(0) \mid X\right] \\ Y(1) &= f(1,X) + \varepsilon_1, \quad \text{with} \quad f(1,X) = \mathbb{E}\left[Y(1) \mid X\right] \end{aligned}$$

$$Y(w) = \underbrace{f(0,X)}_{:=b(X)} + w \underbrace{\left(f(1,X) - f(0,X)\right)}_{:=m(X)} + \underbrace{w\varepsilon_1 + (1-w)\varepsilon_0}_{:=\varepsilon_w}.$$

Additive decomposition

$$Y(w) = b(X) + w\, m(X) + \varepsilon_w,$$

Baseline $b(X) := \mathbb{E}[Y(0) \mid X]$

Modification $m(X) := \mathbb{E}[Y(1) - Y(0) \mid X]$

$\mathbb{E}\left[\varepsilon_w \mid X\right] = 0$

Spirit of Robinson's decomposition (1998), also in Nie et al. (2020)

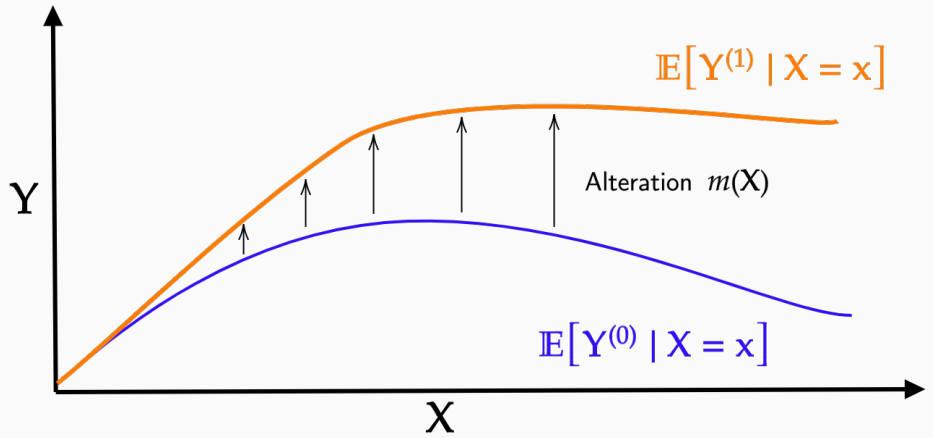
Linking models and causal measures

Nonparametric generative model, continuous outcome

$$Y(w) = b(X) + w\, m(X) + \varepsilon_w,$$

Baseline $b(X) := \mathbb{E}[Y(0) \mid X]$

Modification $m(X) := \mathbb{E}[Y(1) - Y(0) \mid X]$



$$\tau^{\text{RD}} = \mathbb{E}\left[m(X)\right], \qquad \tau^{\text{RR}} = 1 + \frac{\mathbb{E}\left[m(X)\right]}{\mathbb{E}\left[b(X)\right]}$$

- With RR, baseline $b(X)$ is entangled with effect $m(X)$
- RD independent of baseline

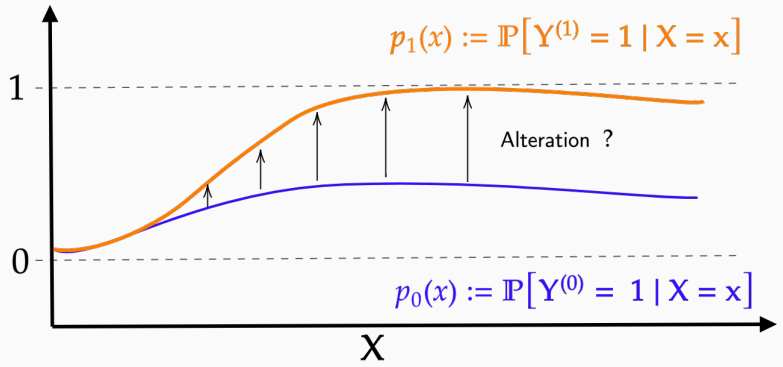
Linking models and causal measures

Ex. of the Russian Roulette: Harmful "homogeneous" treatment effect

Nonparametric generative model, binary outcome

$$\mathbb{P}[Y(w) = 1 \mid X = x] = b(x) + w \underbrace{(1 - b(x))}_{\text{Entanglement}} \frac{1}{6}.$$

Baseline $b(X) := \mathbb{P}[Y(0) = 1 \mid X = x]$



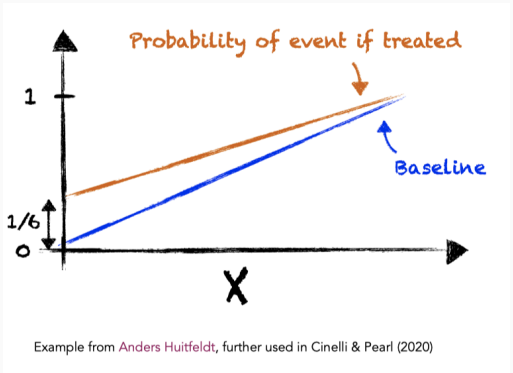
Linking models and causal measures

Ex. of the Russian Roulette: Harmful "homogeneous" treatment effect

Nonparametric generative model, binary outcome

$$\mathbb{P}\left[Y(w)=1 \mid X=x\right]=b(x)+w\underbrace{(1-b(x))}_{\text{Entanglement}}\frac{1}{6}.$$

Baseline $b(X):=\mathbb{P}\left[Y(0)=1 \mid X=x\right]$



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

$$\tau^{\text{RD}}=\frac{1}{6}\left(1-\mathbb{E}\left[b(x)\right]\right), \qquad \lim_{\mathbb{E}[b(x)]\to 1}\tau_{\text{RD}}=0. \qquad \tau_{\text{SR}}=1-\frac{1}{6}$$

- With RD, baseline $b(X)$ is entangled with effect $1/6$
- SR independent of baseline

Linking models and causal measures

Heterogeneous treatment effect

Nonparametric generative model, binary outcome

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

$$\mathbb{P}\left[Y(w) = 1 \mid X = x\right] = b(x) \underbrace{+ w\ (1 - b\,(x))\, m_b\,(x)}_{\nearrow} \underbrace{- w\, b\,(x)\, m_g\,(x)}_{\searrow},$$

A beneficial effect ($m_b(x) = 0$) is visible on a high baseline ($b(x) \approx 1$).
A deleterious effect ($m_g(x) = 0$) is visible on low baseline ($1 - b(x) \approx 1$)

Linking models and causal measures

Heterogeneous treatment effect

Nonparametric generative model, binary outcome

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

$$\mathbb{P}\left[Y(w) = 1 \mid X = x\right] = b(x) \underbrace{+ w\ (1 - b\,(x))\, m_b\,(x)}_{\nearrow} \underbrace{- w\, b\,(x)\, m_g\,(x)}_{\searrow},$$

A beneficial effect ($m_b(x) = 0$) is visible on a high baseline ($b(x) \approx 1$).
A deleterious effect ($m_g(x) = 0$) is visible on low baseline ($1 - b(x) \approx 1$)

Disantanglement of treatment effect and baseline

Assuming that the treatment is beneficial (i.e. $\forall x, m_b(x) = 0$),

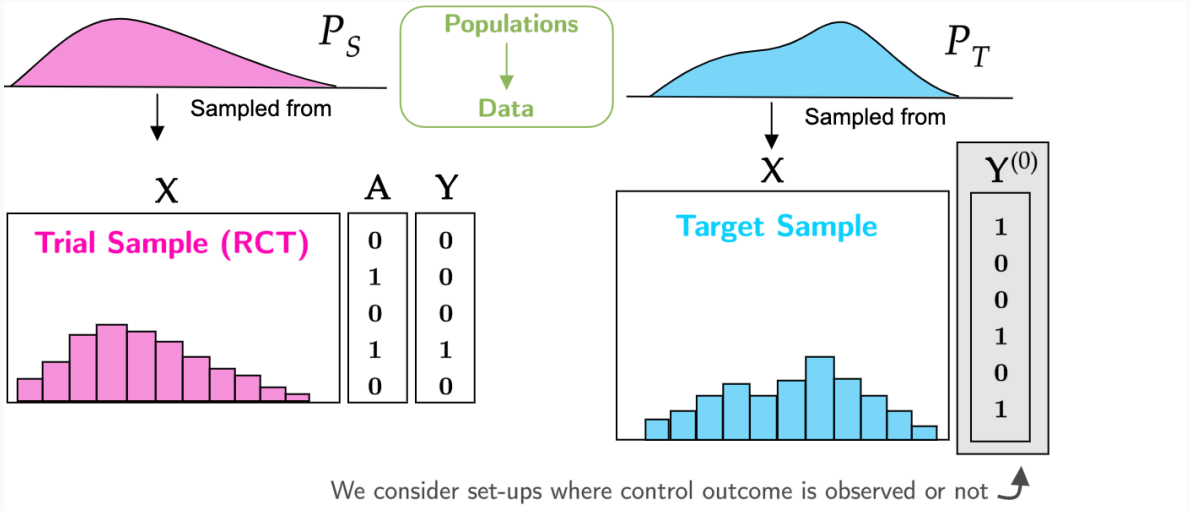
$$\tau_{\text{RR}}(x) = 1 - m_g(x).$$

Computing group effects on X affecting $b(.)$: constant effect

Harmful effect: conditional SR (like the Russian Roulette)
Beneficial effect: conditional RR

Back to generalizability

Aim: transport trial finding's to a target population, using the trial data and a sample of the target population

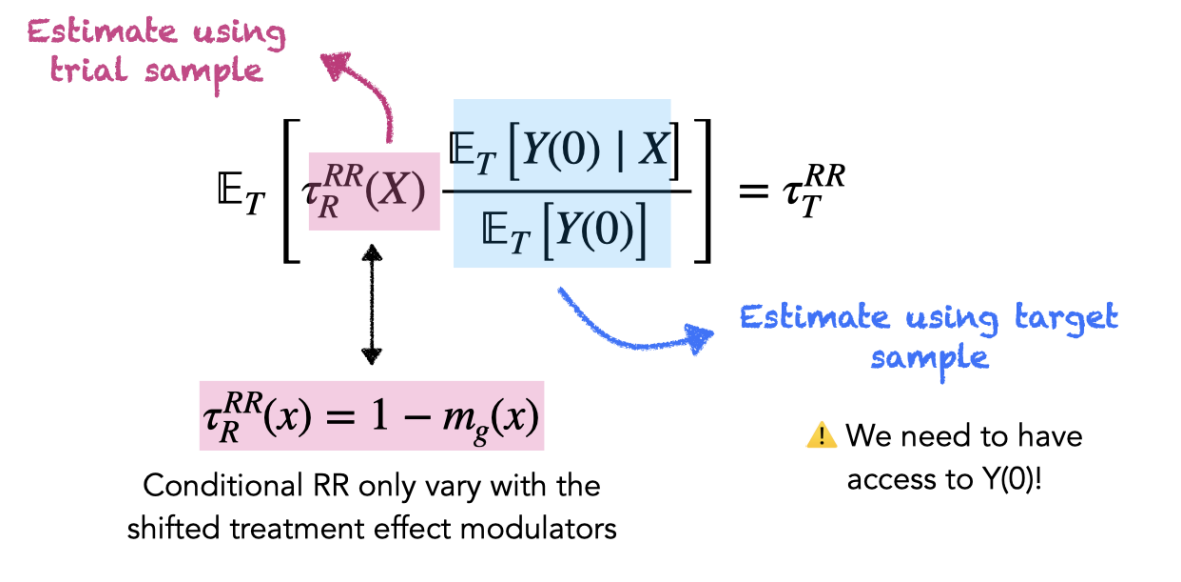


Back to generalizability

Generalizing	Conditional Outcome	Local effects
Assumption	$\mathbb{E}_{R}[Y(w) \mid X] = \mathbb{E}_{T}[Y(w) \mid X = x]$	$\tau_R(X) = \tau_T(X)$
Unformal	All shifted prognostic covariates	All shifted effect modifiers
Identification	$\mathbb{E}_{T}\left[Y(w)\right] = \mathbb{E}_{T}\left[\mathbb{E}_{R}\left[Y(w) \mid X\right]\right]$	$\mathbb{E}_{R}\left[\frac{p_T(X)}{p_R(X)}\, g_T(Y(0), X)\, \tau_R(X)\right]$

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

Generalize local effect: Y binary and a beneficial effect



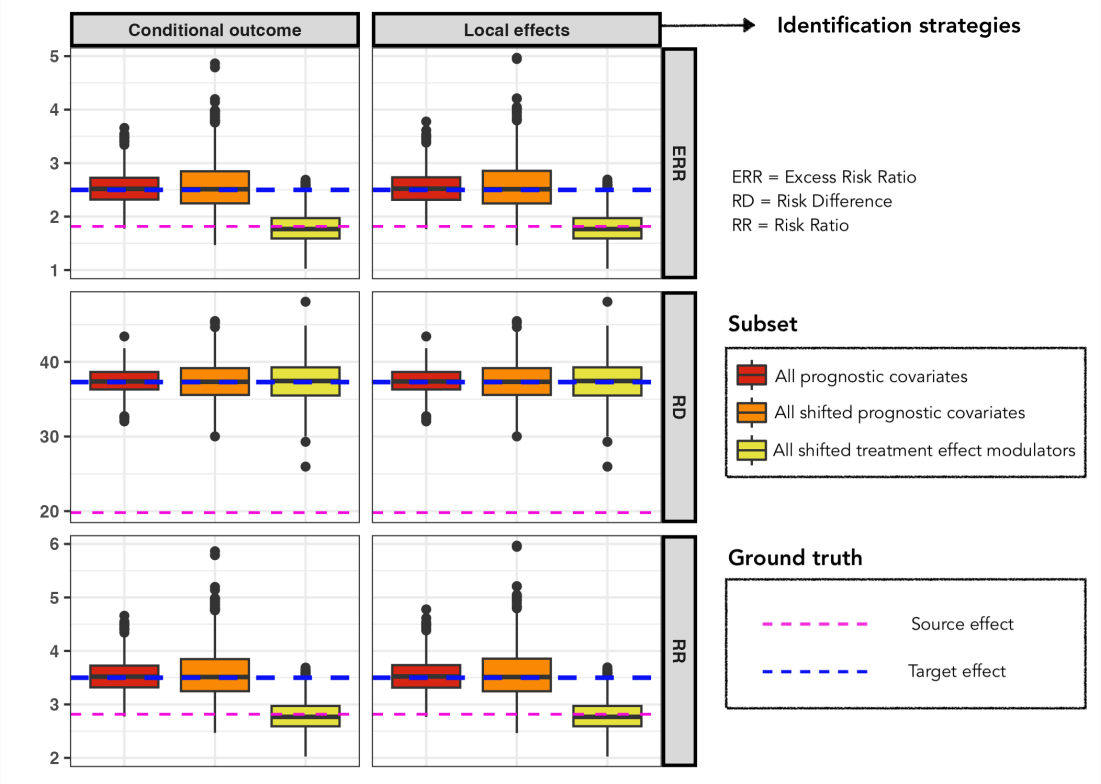
Different covariates sets are required to retrieve the target population effect depending on

- *(i)* the causal measure
- *(ii)* the nature of the outcome (continuous, binary),
- *(iii)* the method to generalize (conditional outcome or local effect).

Simulations: continuous outcome

$$Y = b\left(X_1, X_2, X_3, X_4, X_5, X_6\right) + W\, m\left(X_1, X_2, X_5\right) + \varepsilon.$$

$X_1, \ldots X_4$ shifted. Linear setting

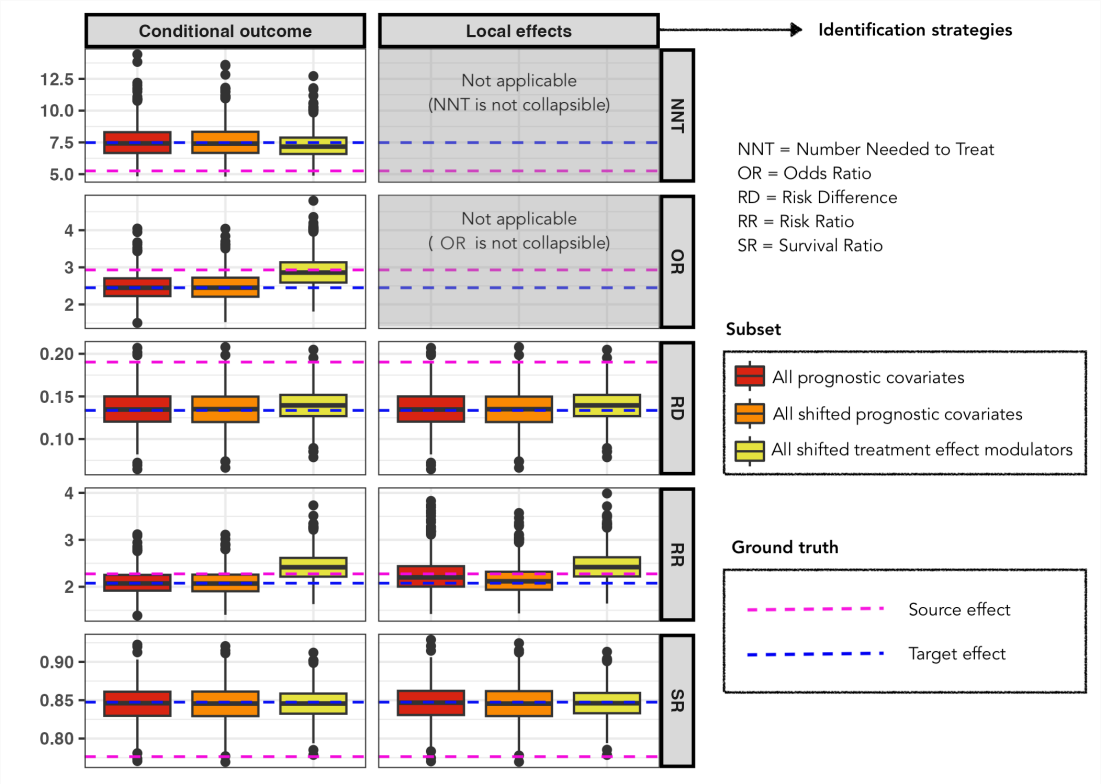


- RD can be generalized with only X_1, X_2
- Generalizing conditional outcome & local effects similar
- Adding prognostic variables improve precisions

Simulations: binary outcome

$$\mathbb{P}\left[Y(w) = 1 \mid X = x\right] = b(X_1, X_2, X_3) + a\ (1 - b\left(X_1, X_2, X_3\right))\ m_b(X_2, X_3),$$

$X_1 = \texttt{lifestyle}$, $X_2 = \texttt{stress}$, $X_3 = \texttt{gender}$; gender not shifted



- SR can generalize with Stress only when generalizing local effects

Conclusion: many medical and statistical challenges

Some measures are easier to generalize (i.e. needs less covariates to adjust on): less sensitive to a population's shift

Summary

- ▷ Collapsibility is an important property
- ▷ Generalize by conditional outcomes requires shifted prognostic var.
- ▷ Generalize by local effects requires shifted treatment effect modifiers
- ▷ Continuous outcome: RD depends on the modification
- ▷ Binary outcome: Conditional SR for harmful, RR for beneficial effect

[21] https://cran.r-project.org/web/views/CausalInference.html

Conclusion: many medical and statistical challenges

Some measures are easier to generalize (i.e. needs less covariates to adjust on): less sensitive to a population's shift

Summary

- Collapsibility is an important property
- Generalize by conditional outcomes requires shifted prognostic var.
- Generalize by local effects requires shifted treatment effect modifiers
- Continuous outcome: RD depends on the modification
- Binary outcome: Conditional SR for harmful, RR for beneficial effect

On going

- Estimation strategies to estimate/generalize the RR
- Insights from/Impact for Meta Analyses
- Usefulness of the model - both beneficial and harmful...
- Methods used are methods implemented. Taskview on causal inference [21]
- Federated causal inference

[21] https://cran.r-project.org/web/views/CausalInference.html

Conclusion: many medical and statistical challenges

Some measures are easier to generalize (i.e. needs less covariates to adjust on): less sensitive to a population's shift

Summary

- Collapsibility is an important property
- Generalize by conditional outcomes requires shifted prognostic var.
- Generalize by local effects requires shifted treatment effect modifiers
- Continuous outcome: RD depends on the modification
- Binary outcome: Conditional SR for harmful, RR for beneficial effect

On going

- Estimation strategies to estimate/generalize the RR
- Insights from/Impact for Meta Analyses
- Usefulness of the model - both beneficial and harmful...
- Methods used are methods implemented. Taskview on causal inference [21]
- Federated causal inference

- Logistic (non parametric) regression model - Constant conditional OR

[21] https://cran.r-project.org/web/views/CausalInference.html

Collaborators

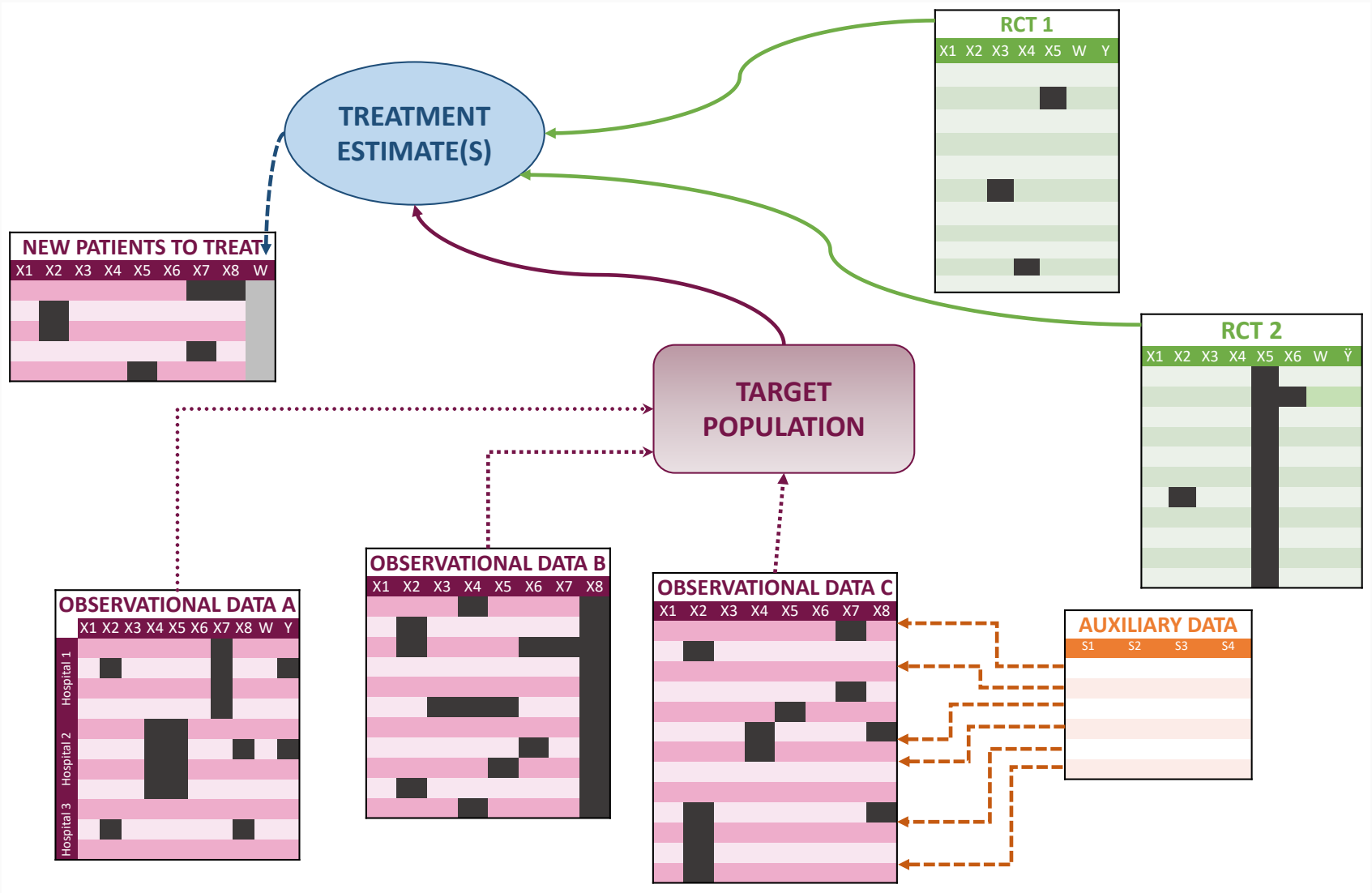

Bénédicte Colnet (DSS), Erwan Scornet (LPSM),



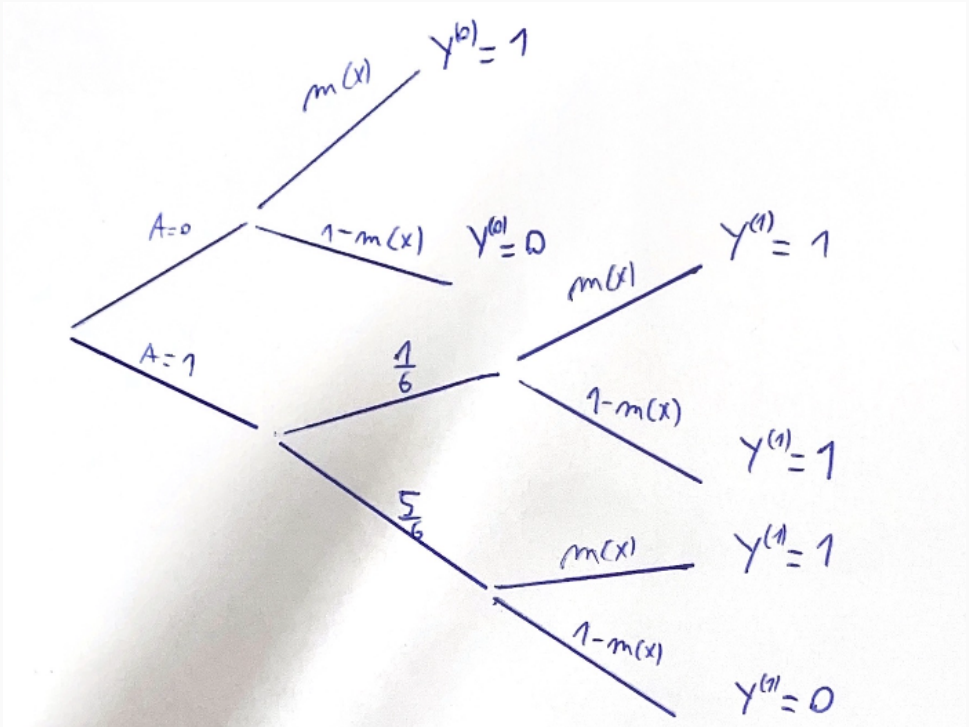
Imke Mayer (Owkin), Tobias Gauss (Traumabase), G. Varoquaux (Inria)

• Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize? *Submitted*
• Mayer, et al. Doubly robust treatment effect estimation with missing attributes. *Annals of Applied Statistics*, 14(3), 2020.
• Colnet, Mayer et al. Causal inference methods for combining randomized trials and observational studies: a review. *Statitsical Science*. 2023.
• Mayer, Josse. Generalizing treatment effects with incomplete covariates: identifying assumptions and multiple imputation algorithms. *Biometrical Journal*. 2022.
• Colnet et al. Causal effect on a target population: a sensitivity analysis to handle missing covariates. *Journal of Causal Inference*. 2022.
• Colnet et al. Reweighting the RCT for generalization: finite sample analysis and variable selection. *Submitted*. 2022.

Real-world data can help strengthen clinical evidence



black correspond to sporadically & systematic missing covariates



$$\mathbb{P}(Y^{(0)}=1 \mid X=x) = m(x)$$

$$\mathbb{P}(Y^{(1)}=1 \mid X=x) = \frac{m(x)}{6} + \frac{1-m(x)}{6} + \frac{5}{6}m(x)$$

$$= m(x) + \frac{1-m(x)}{6}$$

$$\Rightarrow \mathbb{P}(Y^{(a)}=1 \mid X=x) = m(x) + \frac{a(1-m(x))}{6}$$

Resources

R-miss-tastic https://rmisstastic.netlify.com/R-miss-tastic

J., I. Mayer, N. Tierney & N. Vialaneix

Project funded by the R consortium (Infrastructure Steering Committee)[22]

Aim: a reference platform on the theme of missing data management

- ▷ list existing packages
- ▷ available literature
- ▷ tutorials
- ▷ analysis workflows on data
- ▷ main actors

⇒ Federate the community

⇒ Contribute!

[22] https://www.r-consortium.org/projects/call-for-proposals

Appendix

Semi-synthetic simulation

- All the results are illustrated on semi-synthetic simulations;
- Build from two large clinical data bases, reflecting a real-world situation
  - CRASH3 $\sim$ 9 000 individuals.
  - Traumabase $\sim$ 30 000 individuals.
- The outcome is the only synthetic part,

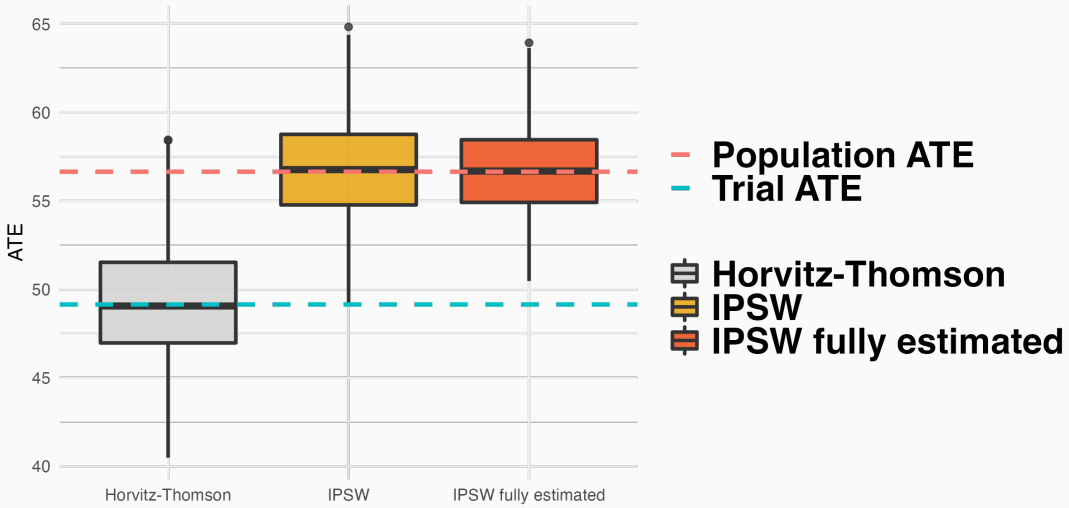
$$Y := f(\texttt{GCS}, \texttt{Gender}) + A\,\tau(\texttt{TTT}, \texttt{Blood Pressure}) + \epsilon_{\texttt{TTT}},$$

More precisely,

$$\begin{aligned} Y = 10 - \texttt{Glasgow} + (\texttt{if Girl:} - 5 \texttt{ else:} 0) \\ + A\left(15(6 - \texttt{TTT}) + 3 * (\texttt{Systolic.blood.pressure} - 1)^2\right) + \varepsilon_{\texttt{TTT}}, \end{aligned}$$

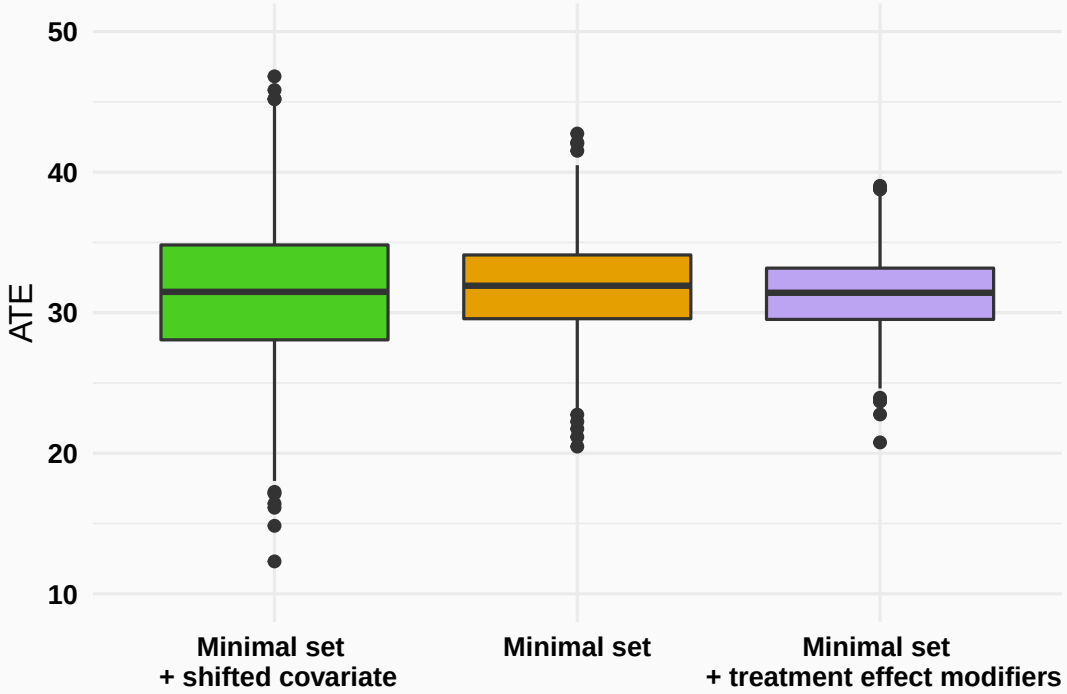
where $\varepsilon_{\texttt{TTT}}$ is a random Gaussian noise with a standard deviation depending on the value of the covariate TTT.

Results from the semi-synthetic simulation



- ▷ Reduced variance for IPSW fully estimated (π is also estimated).
- ▷ The re-weighted trial has not necessarily a larger variance.

Results from the semi-synthetic simulation



- ▷ Reduced variance for IPSW fully estimated (π is also estimated).
- ▷ The re-weighted trial has not necessarily a larger variance.

Papier:

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins *

Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu). Initially submitted December 9, 2014; accepted for publication September 8, 2015.

Protocol component	CRASH2	CRASH3	Traumabase
Eligibility criteria	Adult trauma patients with, or at risk of, significant bleeding who were within 8 hours of injury.	Adults with TBI who were within 3 h of injury, had a Glasgow Coma Scale (GCS) score of 12 or lower or any intracranial bleeding on CT scan, and no major extracranial bleeding	Patients of age ≥ 16 years. Until 2019, the treatment TXA was only recommended for patients with hemorrhage (external or internal) and not specifically for TBI (which is not necessarily caused by intracranial hemorrhage)
Treatment strategies	Tranexamic acid (loading dose 1 g over 10 minutes then infusion of 1 g over 8 hours)	TXA (loading dose 1 g over 10 min then infusion of 1 g over 8 h) in less than 3 hours after injury	TXA is given before/at entry in the hospital in case of extracranial bleeding, time and dose are not registered
Assignment procedures	Participants are randomized (1:1)	Participants are randomized (1:1)	Observational study
Follow-up period	28 days after accident	28 days after accident	28 days after accident
Outcome	Death in hospital within 4 weeks of injury	Head injury-related death in hospital within 28 days of injury	Head injury-related death within 28 days of injury
Causal contrasts of interest	Intention-to-treat effect	Intention-to-treat effect	Intention-to-treat effect

TBI is a complex condition caused by severe trauma (e.g. in an accident) and it can be assumed that the severity of TBI is determined/fixed early after the accident i.e., the severity of TBI is “hard coded” once TBI occurs and the following treatments won’t affect the severity, they only affect the chances of survival or other functional outcomes conditionally on the TBI severity.

In theory TBI severity could be assessed quickly after its occurrence (at the scene of the accident). But it is assessed and treated only at the hospital, once the patient is stable enough (i.e. without ongoing strong hemorrhage) to be placed in a CT scan.

One of the limit when combining data or comparing evidence is that TBI is not assessed in the same way in the different data set

Causal inference[23]

Causal inference (simplest) question

Assume a policy/intervention/**treatment** W causes an **outcome** Y

Aim: **estimate the effect** as acurately as possible (bias & variance)

[23] Taskview to organize all packages on causal inference.

Causal inference[23]

Potential Outcome framework (Neyman, 1923; Rubin, 1974)

- n iid sample ($\underbrace{X_i}_{\text{covariates}}$, $\overbrace{W_i}^{\text{treatment}}$, $\underbrace{Y_i(1), Y_i(0)}_{\textit{potential outcomes}}$) $\in \mathbb{R}^d \times \{0, 1\} \times \mathbb{R} \times \mathbb{R}$
- Individual **causal effect** of the binary treatment: $\Delta_i = Y_i(1) - Y_i(0)$

Problem: Δ_i never observed (only observe one outcome/indiv)

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
1.1	20	F	1	**?**	200
-6	45	F	0	10	**?**
0	15	M	1	**?**	150
	...		...	...	...
-2	52	M	0	100	**?**

[23] Taskview to organize all packages on causal inference.

Causal inference[23]

Potential Outcome framework (Neyman, 1923; Rubin, 1974)

- n iid sample ($\underbrace{X_i}_{\text{covariates}}$, $\overbrace{W_i}^{\text{treatment}}$, $\underbrace{Y_i(1), Y_i(0)}_{\text{potential outcomes}}$) $\in \mathbb{R}^d \times \{0, 1\} \times \mathbb{R} \times \mathbb{R}$
- Individual **causal effect** of the binary treatment: $\Delta_i = Y_i(1) - Y_i(0)$

Problem: Δ_i never observed (only observe one outcome/indiv)

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
1.1	20	F	1	**?**	200
-6	45	F	0	10	**?**
0	15	M	1	**?**	150
	...		...	...	...
-2	52	M	0	100	**?**

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

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

[23] Taskview to organize all packages on causal inference.

Randomized Controlled Trial

Identifiability assumptions

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

Flip a coin to assign the treatment

$$\begin{aligned} \text{We can check that } \tau &= \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\ &= \mathbb{E}[Y_i(1)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0] \\ &= \mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] \end{aligned}$$

$\Rightarrow$ Although Δ_i never observe, τ is identifiable and can be estimated

Difference-in-means estimator

$$\hat{\tau}_{DM} = \frac{1}{n_1} \sum_{W_i=1} Y_i - \frac{1}{n_0} \sum_{W_i=0} Y_i, \text{ where } n_w = \sum_{i=1}^{n} \mathbb{1}_{W_i=w}$$

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

Randomized Controlled Trial

Identifiability assumptions

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

Flip a coin to assign the treatment

$$\begin{aligned} \text{We can check that } \tau &= \mathbb{E}[\Delta_i] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)] \\ &= \mathbb{E}[Y_i(1)|W_i = 1] - \mathbb{E}[Y_i(0)|W_i = 0] \\ &= \mathbb{E}[Y_i|W_i = 1] - \mathbb{E}[Y_i|W_i = 0] \end{aligned}$$

$\Rightarrow$ Although Δ_i never observe, τ is identifiable and can be estimated

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
1.1	20	F	1	?	200
-6	45	F	0	10	?
0	15	M	1	?	150
	...		...	...	...
-2	52	M	0	100	?

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

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

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

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

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

- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria
  $\Rightarrow$ No personalized medicine
- **trial sample different from the population eligible for treatment**
  $\Rightarrow$ Low **external** validity

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

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

- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria
  $\Rightarrow$ No personalized medicine
- **trial sample different from the population eligible for treatment**
  $\Rightarrow$ Low **external** validity

Observational data

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

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

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

- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria
  $\Rightarrow$ No personalized medicine
- **trial sample different from the population eligible for treatment**
  $\Rightarrow$ Low **external** validity

Observational data

- "big data": low quality
- lack of a controlled design opens the door to **confounding bias**
  $\Rightarrow$ Low **internal** validity

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