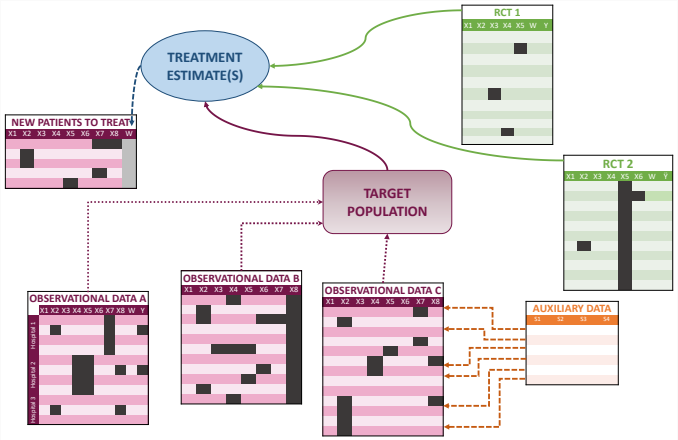
Personalized Care Through Causal Learning: From Data to Decisions

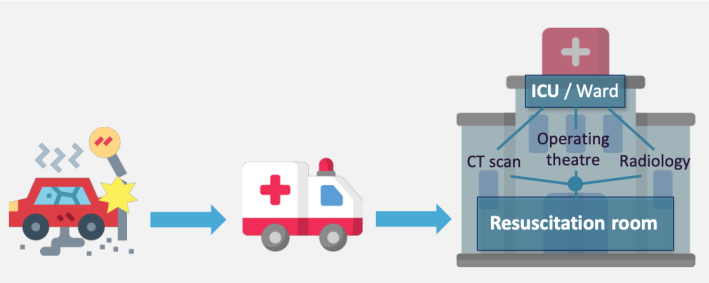
Julie Josse. Senior Researcher Inria 2020-; Prof. CMAP X 2016-2020

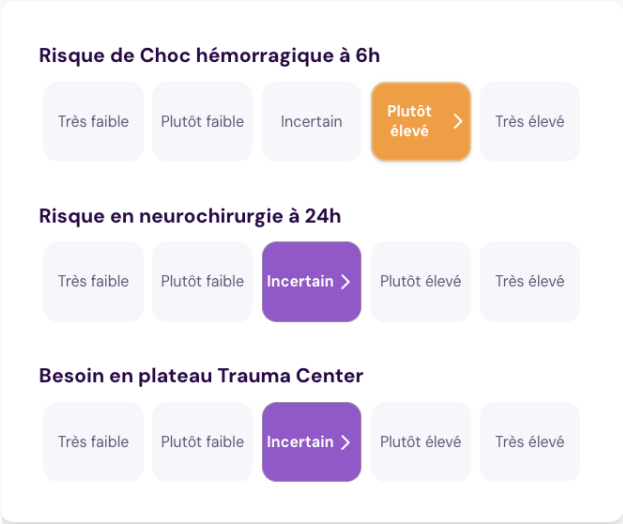
Lead Inria-Inserm PreMeDICaL team: personalized medicine by data integration & causal learning



(Online) Decision support tool with quantified uncertainty

Ex: Traumatrix project[1]: Reducing under and over triage for improved resource allocation in trauma care





Major trauma: brain injuries or hemorrhagic shock from car accidents, falls, stab wounds, etc.) requires specialized care/resources in "trauma centers"

Many patients are misdirected: human/ economical costs

Clinical trial launched in 2025: real-time implementation of Machine Learning models in ambulance dispatch via a mobile data collection application

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

Personalization of treatment recommendation

Ex: Estimating treatment e ect from the Traumabase data

- 40000 trauma patients
- 300 heterogeneous features from pre-hospital and in-hospital settings
- 40 trauma centers, 4000 new patients per year

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

) **Estimate causal effect** (with missing values[2]): Administration of the **treatment** *tranexamic acid (TXA)*, given within 3 hours of the accident, on the **outcome** (Y) *28 days in-hospital mortality* for trauma brain patients

[2] Mayer, I., Wager, S. & **J.J**. (2020). Doubly robust treatment effect estimation with incomplete confounders. *Annals Of Applied Statistics. (implemented in package `grf`).*

Causal inference: "what would happen if?"

Causal inference (simplest) question

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

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

Causal inference: "what would happen if?"

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

treatment

. (X ; W ; $Y(1); Y(0)$) 2 R^d $f0; 1g$ R R

covariates

potential outcomes

. Individual **causal effect** of the binary treatment: $_i = Y_i(1) \quad Y_i(0)$

Problem: $_i$ never observed (only one outcome is observed per 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	**?**

Causal inference: "what would happen if?"

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

. $(\underbrace{X}_{\text{covariates}}; \overbrace{W}^{\text{treatment}}; \underbrace{Y(1), Y(0)}_{\text{potential outcomes}}) \in \mathbb{R}^d \quad \{0,1\} \quad \mathbb{R} \quad \mathbb{R}$

. Individual **causal effect** of the binary treatment: $\tau_i = Y_i(1) - Y_i(0)$

Problem: τ_i never observed (only one outcome is observed per 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): $\tau = E[\tau_i] = E[Y_i(1) - Y_i(0)]$

ATE with **Risk Difference**: difference of the average outcome had everyone gotten treated and the average outcome had nobody gotten treatment

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

. **gold standard** (allocation)

. same covariate distributions in treated and control groups

) High **internal** validity

Observational data

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

- **gold standard** (allocation)
- same covariate distributions in treated and control groups
) High **internal** validity

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

Observational data

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

- **gold standard** (allocation)
- same covariate distributions in treated and control groups

) High **internal** validity

- expensive, long, ethical limitations
- small sample size: restrictive inclusion criteria

) No personalized medicine

- **trial sample different from the population eligible for treatment**

) Low **external** validity

Observational data

- low cost
- large amounts of data (registries, biobanks, EHR, claims)

) patient's heterogeneity

- **representative of the target populations**

) High **external** validity

Data sources & evidences to estimate the treatment effect

Randomized Controlled Trial (RCT)

- **gold standard** (allocation)
- same covariate distributions in treated and control groups
) High **internal** validity

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

Observational data

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

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

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 can use both to [3] ...

- ...validate observational methods, correct for confounding bias
- ...improve estimation of heterogeneous treatment e ects
- ...**generalize the treatment effect to a target population** (data fusion, transportability, recovery from selection bias)[4],[5]

[3] Colnet, et al. **J.J.** (2022). Causal inf. for combining RCT & obs. studies. *Statistical Science*.
[4] Elias Bareinboim & Judea Pearl. (2016). Causal inference & the data-fusion problem. *PNAS*.
[5] Dahabreh, Haneuse, Robins, Robertson, Buchanan, Stuart, Hernan. (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 can use both to [3] ...

- ...validate observational methods, correct for confounding bias
- ...improve estimation of heterogeneous treatment e ects
- ...**generalize the treatment effect to a target population** (data fusion, transportability, recovery from selection bias)[4],[5]

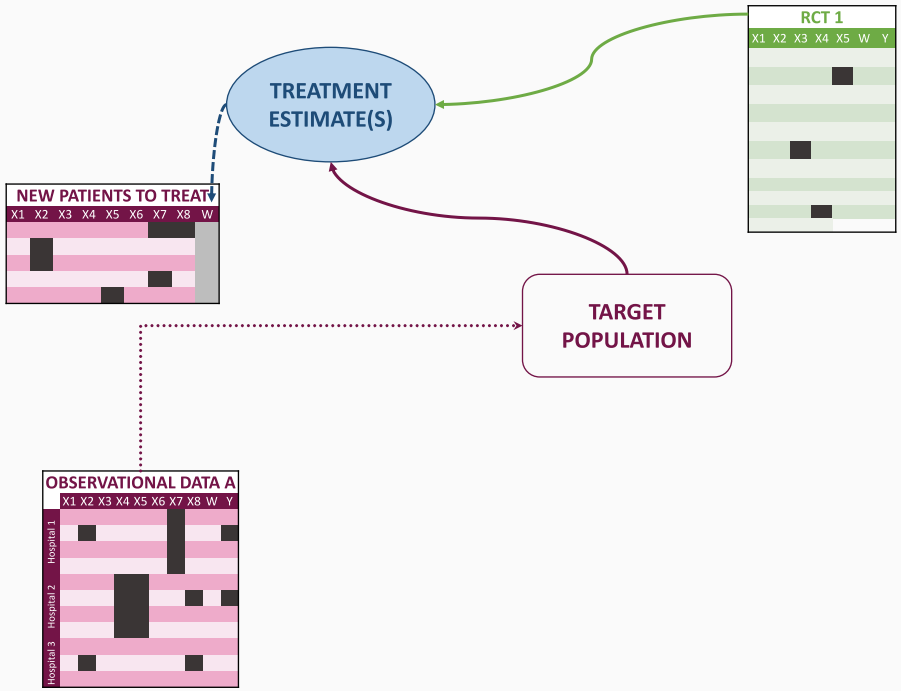
  The FDA has greenlighted the usage of the drug *Ibrance* to men with breast cancer, though clinical trials were performed only on women.

! Reduce drug approval times and costs

---

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

Predicting treatment effects from 1 trial to another population





Bénédicte Colnet (Corps des Mines, French social security's direction), Imke Mayer (Owkin)
Erwan Scornet (X - Sorbonne Université), Gaël Varoquaux (Inria)

Generalization task from one 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 one 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:

$p_R(x) \neq p_T(x)$) $\tau_R := E_R[Y(1) - Y(0)] \neq E_T[Y(1) - Y(0)] := \tau_T$

ATE in the RCT Target ATE

Assumptions for ATE identi ability in generalization

Overlap assumption[6]

$$8x \; 2 \; X ; p_R(x) > 0 \text{ and } \text{supp}(P_T(X)) \quad \text{supp}(P_R(X))$$

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

Transportability of the conditional average treatment e ect (CATE) [7]

$$\underbrace{E_R[Y(1) \; Y(0) \; j \; X]}_{R(X)} = \underbrace{E_T[Y(1) \; Y(0) \; j \; X]}_{T(X)}$$

Need to know which variables are shifted treatment e ect modi ers

The treatment e ect depends on covariates in the same way in the source (RCT) and target population

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

[7] Equivalent formulation with sampling mechanism S: $(Y(1) \quad Y(0)) \; ?? \; S \; j \; X$

Identi ability and estimation for generalization: weighting

Generalization of local e ects (i.e. conditional e ects/strata)

$$
\begin{aligned}
_T &= E_T[Y_i(1) \quad Y_i(0)] = E_T[E_T[Y_i(1) \quad Y_i(0)jX]] \\
&= E_T[\ _T(X)] = E_T[\ _R(X)] \quad \text{Transportability CATE} \\
&= E_R \ \frac{p_T(X)}{p_R(X)} \ _R(X)
\end{aligned}
$$

IPSW: inverse propensity sampling weighting

$$
\hat{}_{; n;m} = \frac{1}{n} \sum_{i2R} \frac{\hat{p}_T(X_i)}{\hat{p}_R(X_i)} Y_i \frac{W_i}{} \frac{1 \quad W_i}{1} ;
$$

$\hat{p}_{R;n}(x) := \frac{1}{n} \sum_{i2R} 1_{X_i = x}$; proba. of treatment assignment in trial

Re-weight, so that the trial follows the target sample's distribution

Idea of relying on an external representative sample to reweight is recent[8]

[8] Cole & Stuart. (2010). Generalizing from RCT to target pop. American J. of Epidemiology.

Identi ability and estimation for generalization: weighting

Generalization of local e ects (i.e. conditional e ects/strata)

$$
\begin{aligned}
_T &= E_T[Y_i(1) \quad Y_i(0)] = E_T[E_T[Y_i(1) \quad Y_i(0)jX]] \\
&= E_T[\ _T(X)] = E_T[\ _R(X)] \quad \text{Transportability CATE} \\
&= E_R \ \frac{p_T(X)}{p_R(X)} \ _R(X)
\end{aligned}
$$

IPSW: inverse propensity sampling weighting

$$
\hat{}_{; n;m} = \frac{1}{n} \sum_{i2R} \frac{\hat{p}_T(X_i)}{\hat{p}_R(X_i)} Y_i \frac{W_i}{} \frac{1 \quad W_i}{1} ;
$$

$\hat{p}_{R;n}(x) := \frac{1}{n} \sum_{i2R} 1_{X_i = x}$; proba. of treatment assignment in trial

Re-weight, so that the trial follows the target sample's distribution

Idea of relying on an external representative sample to reweight is recent[8]

[8] Cole & Stuart. (2010). Generalizing from RCT to target pop. American J. of Epidemiology.

Identi ability and estimation for generalization: weighting

Generalization of local e ects (i.e. conditional e ects/strata)

$$
\begin{aligned}
_T &= E_T[Y_i(1) \quad Y_i(0)] = E_T[E_T[Y_i(1) \quad Y_i(0)jX]] \\
&= E_T[\ _T(X)] = E_T[\ _R(X)] \quad \text{Transportability CATE} \\
&= E_R \ \frac{p_T(X)}{p_R(X)} \ _R(X)
\end{aligned}
$$

IPSW: inverse propensity sampling weighting

$$
\hat{}_{; n;m} = \frac{1}{n} \sum_{i2R} \frac{\hat{p}_T(X_i)}{\hat{p}_R(X_i)} Y_i \frac{W_i}{} \frac{1 \quad W_i}{1} ;
$$

$\hat{p}_{R;n}(x) := \frac{1}{n} \sum_{i2R} 1_{X_i = x}$; proba. of treatment assignment in trial

Re-weight, so that the trial follows the target sample's distribution

Idea of relying on an external representative sample to reweight is recent[8]

[8] Cole & Stuart. (2010). Generalizing from RCT to target pop. American J. of Epidemiology.

Identi ability and estimation for generalization: weighting

Generalization of local e ects (i.e. conditional e ects/strata)

$$\tau = E_T[Y_i(1) - Y_i(0)] = E_T[E_T[Y_i(1) - Y_i(0)|X]]$$
$$= E_T[\tau_T(X)] = E_T[\tau_R(X)] \quad \text{Transportability CATE}$$
$$= E_R\left[\frac{p_T(X)}{p_R(X)}\tau_R(X)\right]$$

IPSW: inverse propensity sampling weighting

$$\hat{\tau}_{IPSW;n;m} = \frac{1}{n}\sum_{i\in R}\frac{\hat{p}_T(X_i)}{\hat{p}_R(X_i)}Y_i\frac{W_i}{1}\frac{1-W_i}{1};$$

$\hat{p}_{R;n}(x) := \frac{1}{n}\sum_{i\in R} 1_{X_i=x}$; proba. of treatment assignment in trial

Re-weight, so that the trial follows the target sample's distribution

Idea of relying on an external representative sample to reweight is recent[8]

Open questions remain: Impact of the two data sources' sizes n & m?

[8] Cole & Stuart. (2010). Generalizing from RCT to target pop. American J. of Epidemiology.

Reweighting the RCT: nite sample & asymptotic analysis [9]

$$\hat{\tau}_{n;m} = \frac{1}{n}\sum_{i \in R} \frac{\hat{p}_T(X_i)}{\hat{p}_R(X_i)} Y_i \frac{W_i}{1} \quad \frac{1 - W_i}{1} \quad ;$$

Estimated with T

Estimated with R

Asymptotic properties

Letting $\lim_{n;m \to 1} m/n = \quad 2 [0; 1]$

$$\lim_{n;m \to 1} \min(n; m) \operatorname{Var}[\hat{\tau}_{n;m}] = \min(1 ; \quad) \frac{\operatorname{Var}[\quad(X)]}{} + V_{so}$$

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

- If target >> trial, $m/n \to 1$, (i.e., $= 1$): asymptotic variance = Semi-Oracle's one and depends on the ratio of probabilities.
- If trial >> target $m/n \to 0$, (i.e., $= 0$): asymptotic variance depends on var. of Conditional Average Treatment E ect Var[(X)].

[9] Colnet, J.J et al. 2022. Reweighting the RCT for generalization: nite sample analysis and variable selection. JRSSA.

Reweighting the RCT: nite sample & asymptotic analysis [9]

$$\hat{\tau}_{n;m} = \frac{1}{n} \sum_{i \in R} \frac{\overbrace{\hat{p}_T(X_i)}^{\text{Estimated with T}}}{\underbrace{\hat{p}_R(X_i)}_{\text{Estimated with R}}} Y_i \frac{W_i}{} \frac{1 \quad W_i}{1} ;$$

Impact for data collection

[9] Colnet, J.J et al. 2022. Reweighting the RCT for generalization: nite sample analysis and variable selection. JRSSA.

Generalization from Crash 3 trial [10] to the Traumabase

CRASH3

- Multi-centric RCT - 29 countries
- 9000 individuals - develp. countries
- Positive e ect for moderately injured patients

Traumabase

- Observational sample
- 8200 patients with brain trauma
- Deleterious/No evidence for an e ect of TXA

Comparison of trials, observational data, and generalization estimates

x-axis: Estimation of the Average Treatment E ect, Con dence intervals with bootstrap
y-axis: Estimation methods (estimation of nuisances: parametric: logistic regression - non parametric: forests)

[10](2019). E ects of tranexamic acid on death in patients with acute trauma. brain injury. Lancet.

Many medical and statistical challenges

1) Shifted e ect modi ers not available in Traumabase[11]. Missing covariates in one/both sets: sensitivity analysis

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

[11]Colnet, J.J, et al. 2022. Generalizing a causal e ect: sensitivity analysis.J. of Causal Inference

[12]Mayer, J.J. 2021. Generalizing e ects with incomplete covariates Biometrical Journal.

[13]Colnet, J.J et al. 2023. Reweighting the RCT for generalization: nite sample analysis and variable selection. JRSSC.

[14]Colnet, J.J et al. 2024. Risk-Ratio, Odds-ratio, wich causal measure is easier to generalize?

Many medical and statistical challenges

1) Shifted e ect modi ers not available in Traumabase[11]. Missing covariates in one/both sets: sensitivity analysis

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

2) Missing values: Missing values (NA) in both RCT and Obs data[12]

[11]Colnet, J.J, et al. 2022. Generalizing a causal e ect: sensitivity analysis.J. of Causal Inference

[12]Mayer, J.J. 2021. Generalizing e ects with incomplete covariates Biometrical Journal.

[13]Colnet, J.J et al. 2023. Reweighting the RCT for generalization: nite sample analysis and variable selection. JRSSC.

[14]Colnet, J.J et al. 2024. Risk-Ratio, Odds-ratio, wich causal measure is easier to generalize?

Many medical and statistical challenges

1) Shifted e ect modi ers not available in Traumabase[11]. Missing covariates in one/both sets: sensitivity analysis

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

2) Missing values: Missing values (NA) in both RCT and Obs data[12]

3) Which covariates should be include? Would adding prognostic variables reduce the variance as in the classical case?[13]

[11] Colnet, J.J, et al. 2022. Generalizing a causal e ect: sensitivity analysis.J. of Causal Inference

[12] Mayer, J.J. 2021. Generalizing e ects with incomplete covariates Biometrical Journal.

[13] Colnet, J.J et al. 2023. Reweighting the RCT for generalization: nite sample analysis and variable selection. JRSSC.

[14] Colnet, J.J et al. 2024. Risk-Ratio, Odds-ratio, wich causal measure is easier to generalize?

Many medical and statistical challenges

1) Shifted e ect modi ers not available in Traumabase[11]. Missing covariates in one/both sets: sensitivity analysis

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

2) Missing values: Missing values (NA) in both RCT and Obs data[12]

3) Which covariates should be include? Would adding prognostic variables reduce the variance as in the classical case?[13]

4) Clinicians are more interested in the risk ratio than the risk di erence[14]

[11] Colnet, J.J, et al. 2022. Generalizing a causal e ect: sensitivity analysis. J. of Causal Inference.
[12] Mayer, J.J. 2021. Generalizing e ects with incomplete covariates. Biometrical Journal.
[13] Colnet, J.J et al. 2023. Reweighting the RCT for generalization: nite sample analysis and variable selection. JRSSC.
[14] Colnet, J.J et al. 2024. Risk-Ratio, Odds-ratio, wich causal measure is easier to generalize?

Comparing two average situations

Binary outcome: $P[Y(w) = 1] = E[Y(w)]$ and $P[Y(w) = 0] = 1 \quad E[Y(w)]$.

Absolute measures

$$^{RD} := E[Y(1)] \quad E[Y(0)]; \qquad ^{NNT} := (\ ^{RD})\ ^{1}:$$

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

Relative measures

$$^{RR} := \frac{E[Y(1)]}{E[Y(0)]}; \qquad ^{SR} := \frac{P[Y(1) = 0]}{P[Y(0) = 0]} = \frac{1 \quad E[Y(1)]}{1 \quad E[Y(0)]};$$

$$^{OR} := \frac{P[Y(1) = 1]}{P[Y(1) = 0]} \quad \frac{P[Y(0) = 1]}{P[Y(0) = 0]} \qquad ^{1}$$

A null e ect now corresponds to a Risk Ratio of 1

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

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

Di erent treatment measures give di erent impressions

An example: Randomized Control Trial (RCT) from Cook and Sackett (1995)

Y = 1 stroke in 5 years andY = 0 no stroke

W antihyperintensive therapy

Feature X (blood pressure),X = 1 low baseline risk (15/1000 versus 2/10)

$$P[Y(0) = 1 \mid X = 0] \qquad P[Y(0) = 1 \mid X = 1]$$

	RD	RR	SR	NNT	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 0045 the probability to su er from a stroke

RR: the treated has 0.6 the risk of having a stroke comp. with the control

SR: increased chance of not having a stroke when treated (factor 1.05).

NNT: one has to treat 22 people to prevent one additional stroke

OR RR in a stratum where prevalence of the outcome is low

Di erent treatment measures give di erent impressions

An example: Randomized Control Trial (RCT) from Cook and Sackett (1995)

Y = 1 stroke in 5 years andY = 0 no stroke

W antihyperintensive therapy

Feature X (blood pressure),X = 1 low baseline risk (15/1000 versus 2/10)

$$P[Y(0) = 1 \mid X = 0] \qquad P[Y(0) = 1 \mid X = 1]$$

	RD	RR	SR	NNT	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 is heterogeneous with X while RR is homogeneous with X

Heterogeneity's property de ned w.r.t. (i) covariates & (ii) a measure

Impact of the baseline risk: with 3% baseline mortality reduced to 1% by treatment, RD shows a 0.02 drop, while RR shows controls have three times the risk: RD suggests a small e ect; RR highlights a larger one

Formalization of causal measures's properties: toward guidance

A desirable property: collapsibility

Collapsibility : Population's e ect is equal to a weighted sum of local e ects (conditional e ects)

Direct collapsibility - weights are equal to population's proportions

$$= E[\ (X)]$$

Risk Di erence is directly collapsible

	RD	RR	SR	NNT	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

$$\overset{RD}{R} = p_R(X = 1) \quad \overset{RD}{R}(X = 1) + p_R(X = 0) \quad \overset{RD}{R}(X = 0)$$

0:0452 = 0:47 0:006 0:53 0:08:

Useful for generalization! (replacing p_R by p_T)

Summary of causal measure properties

Direct collapsibility

$$E[\ (X)] =$$

Collapsibility: weights depend on the baseline distribution $Y(0)$

$E[w(X;P(X;Y(0)))\quad(X)] = \qquad$ with $w \quad 0;\ E[w(X;P(X;Y(0)))] = 1$

Logic respecting (Simpson paradox)

$$2\quad \min_{x}(\ (x)) ; \max_{x}(\ (x))\quad :$$

Ex. OR: Overall population, $_{OR}$ 0:26 $_{OR|F=1}$ 0:167 and $_{OR|F=0}$ 0:166

Measure	Dir. collapsible	Collapsible	Logic-respecting
Risk Di erence	Yes	Yes	Yes
Number Needed to Treat	No	No	Yes
Risk Ratio	No	Yes	Yes
Survival Ratio	No	Yes	Yes
Odds Ratio	No	No	No

Back to generalizability from one RCT to a Target pop.

Generalizing	Conditional Outcome	Local e ects/CATE
Assumption	$E_R[Y(w) \mid X] = E_T[Y(w) \mid X]$	$_R(X) = \ _T(X)$
Variables	All shifted prognostic covariates	All shifted e ect modi ers
Identi cation	$E_T[Y(w)] = E_T[E_R[Y(w) \mid X]]$	$E_R\left[\frac{p_T(X)}{p_R(X)} w_T(Y(0); X) \ _R(X)\right]$
Estimation	Regression (G-formula)	Weighting

Generalize local e ects only for collapsible measures, need info. on $Y^{(0)}$

Generalizing conditional outcome require stronger assumptions

Estimation challenges: using with parametric models (logistic reg.) or non-parametric ones (i.e. random forests). Ex for RR[15]: doubly robust estim. using in uence function theory, convergence rate, con dence intervals, etc.

[15] Boughdiri, J.J., Scornet. (2024). Estimating Risk Ratios in Causal Inference.

From one to multiple Randomized Control Trials (RCTs)

Meta-analysis (aggregating estimated e ects from multiple studies) is at the top of the pyramid of evidence based medicine.

Meta-analysis still faces signi cant challenges:

- Be careful with aggregation of causal measures
- Heterogeneity across studies: sample size, population, center e ects
- Di culty to share individual-level data : data silos & regulations

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Going beyond meta-analysis with federated causal inference[16]

Bridging causal inference and federated learning to improve treatment e ect estimation from decentralized data sources

[16] Rémi Khellaf, Aurelien Bellet, J.J. (2025). Multi-centric ATE estimation AISTAT .

Our setting: decentralized heterogeneous RCTs

We consider K decentralized and potentially heterogeneous RCTs (studies) from di erent sources and want to estimate the ATE given by

$$= E \; E(Y^{(1)} \quad Y^{(0)} \; j \; H)$$

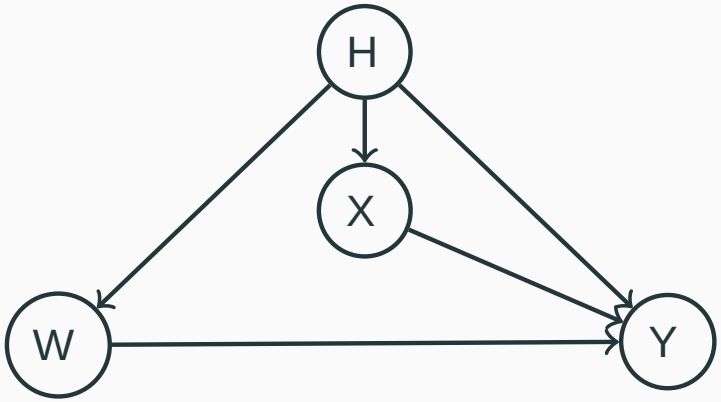
Source	Obs.	Covariates			Treat.	Outcome
H	i	X_1	X_2	X_3	W	Y
1	1	2.3	1.5	M	1	3.2
1	2	2.2	3.1	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	4.5	5.0	F	1	4.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	1	3.7	2.0	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	n_K	2.5	1.7	M	0	3.2

Our setting: decentralized heterogeneous RCTs

We consider K decentralized and potentially heterogeneous RCTs (studies) from different sources and want to estimate the ATE given by

$$= E \; E(Y^{(1)} \quad Y^{(0)} \; j \; H)$$

Source	Obs.	Covariates			Treat.	Outcome
H	i	X_1	X_2	X_3	W	Y
1	1	2.3	1.5	M	1	3.2
1	2	2.2	3.1	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	4.5	5.0	F	1	4.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	1	3.7	2.0	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	n_K	2.5	1.7	M	0	3.2



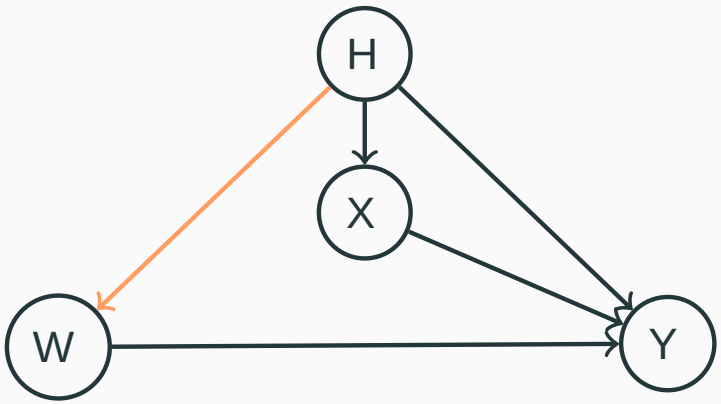
Our setting: decentralized heterogeneous RCTs

We consider K decentralized and potentially heterogeneous RCTs (studies) from different sources and want to estimate the ATE given by

$$= E \; E(Y^{(1)} \quad Y^{(0)} \; j \; H)$$

Source	Obs.	Covariates			Treat.	Outcome
H	i	X_1	X_2	X_3	W	Y
1	1	2.3	1.5	M	1	3.2
1	2	2.2	3.1	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	4.5	5.0	F	1	4.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	1	3.7	2.0	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	n_K	2.5	1.7	M	0	3.2

Heterogeneity in treatment allocation



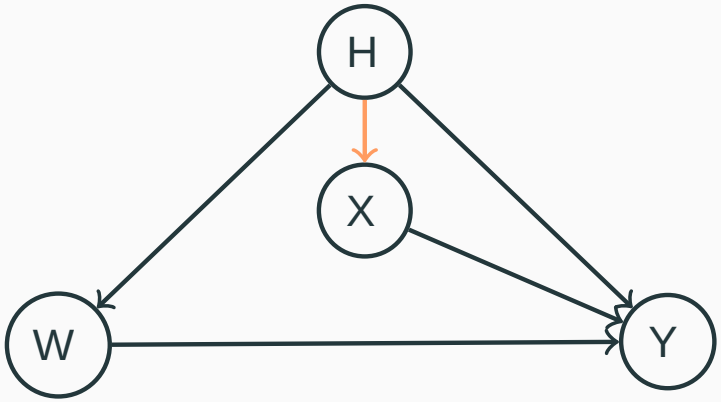
Our setting: decentralized heterogeneous RCTs

We consider K decentralized and potentially heterogeneous RCTs (studies) from different sources and want to estimate the ATE given by

$$= E\ E(Y^{(1)} \quad Y^{(0)} \mid H)$$

Source	Obs.	Covariates			Treat.	Outcome
H	i	X_1	X_2	X_3	W	Y
1	1	2.3	1.5	M	1	3.2
1	2	2.2	3.1	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	4.5	5.0	F	1	4.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	1	3.7	2.0	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	n_K	2.5	1.7	M	0	3.2

Heterogeneity in covariates distribution

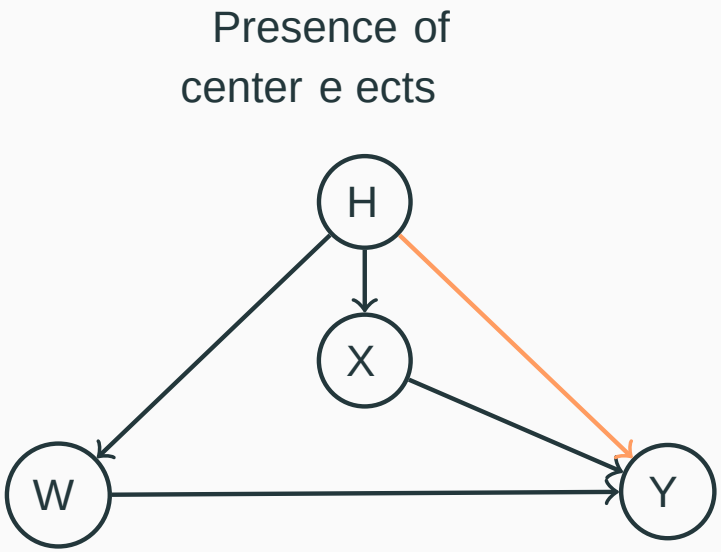


Our setting: decentralized heterogeneous RCTs

We consider K decentralized and potentially heterogeneous RCTs (studies) from different sources and want to estimate the ATE given by

$$= E\ E(Y^{(1)} \quad Y^{(0)} \mid H)$$

Source	Obs.	Covariates			Treat.	Outcome
H	i	X_1	X_2	X_3	W	Y
1	1	2.3	1.5	M	1	3.2
1	2	2.2	3.1	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	4.5	5.0	F	1	4.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	1	3.7	2.0	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	n_K	2.5	1.7	M	0	3.2

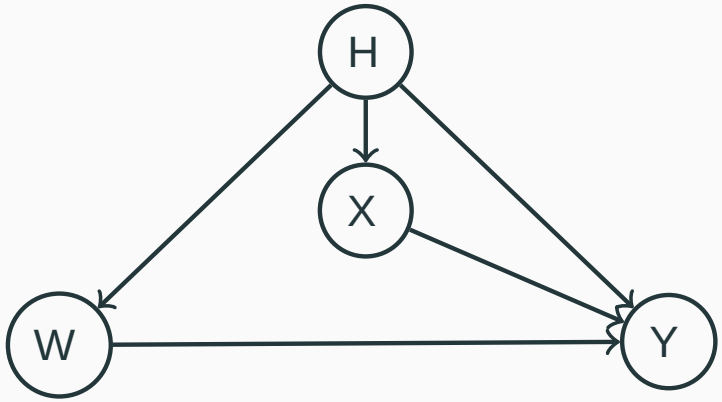


Our setting: decentralized heterogeneous RCTs

We consider K decentralized and potentially heterogeneous RCTs (studies) from different sources and want to estimate the ATE given by

$$= E \quad E(Y^{(1)} \quad Y^{(0)} \; j \; H)$$

Source	Obs.	Covariates			Treat.	Outcome
H	i	X_1	X_2	X_3	W	Y
1	1	2.3	1.5	M	1	3.2
1	2	2.2	3.1	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	1	4.5	5.0	F	1	4.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	1	3.7	2.0	F	0	2.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	n_K	2.5	1.7	M	0	3.2



How to estimate without pooling together individual-level data?

Three types of federated estimators

Ex: linear outcome model for all studies k: $Y_{k;i}^{(w)} = c^{(w)} + X_{k;i}^{(w)} + \varepsilon_{k;i}^{(w)}$

Baseline: estimator $\hat{\beta}_{pool} = \frac{1}{n} \sum_{i=1}^{n} X_i^0 (\hat{\beta}_{pool}^{(1)} - \hat{\beta}_{pool}^{(0)})$ on pooled data

$\hat{\beta}_{pool}^{(w)} = (\hat{c}_{pool}^{(w)}; \hat{\beta}_{pool}^{(w)}) = (X^{0(w)\top} X^{0(w)})^{-1} X^{0(w)\top} Y^{(w)}$ with $X^{0(w)} = [1; X^{(w)}]$

Aggregation w_k: sample size weights (SW) or inverse variance weights (IVW)

Three types of federated estimators

Ex: linear outcome model for all studies k: $Y_{k;i}^{(w)} = c^{(w)} + X_{k;i}^{(w)} + \varepsilon_{k;i}^{(w)}$

Baseline: estimator $\hat{\beta}_{pool} = \frac{1}{n} \sum_{i=1}^{n} X_i^0 (\hat{\beta}_{pool}^{(1)} - \hat{\beta}_{pool}^{(0)})$ on pooled data

$\hat{\beta}_{pool}^{(w)} = (\hat{c}_{pool}^{(w)}; \hat{\beta}_{pool}^{(w)}) = (X^{0(w)\top} X^{0(w)})^{-1} X^{0(w)\top} Y^{(w)}$ with $X^{0(w)} = [1; X^{(w)}]$

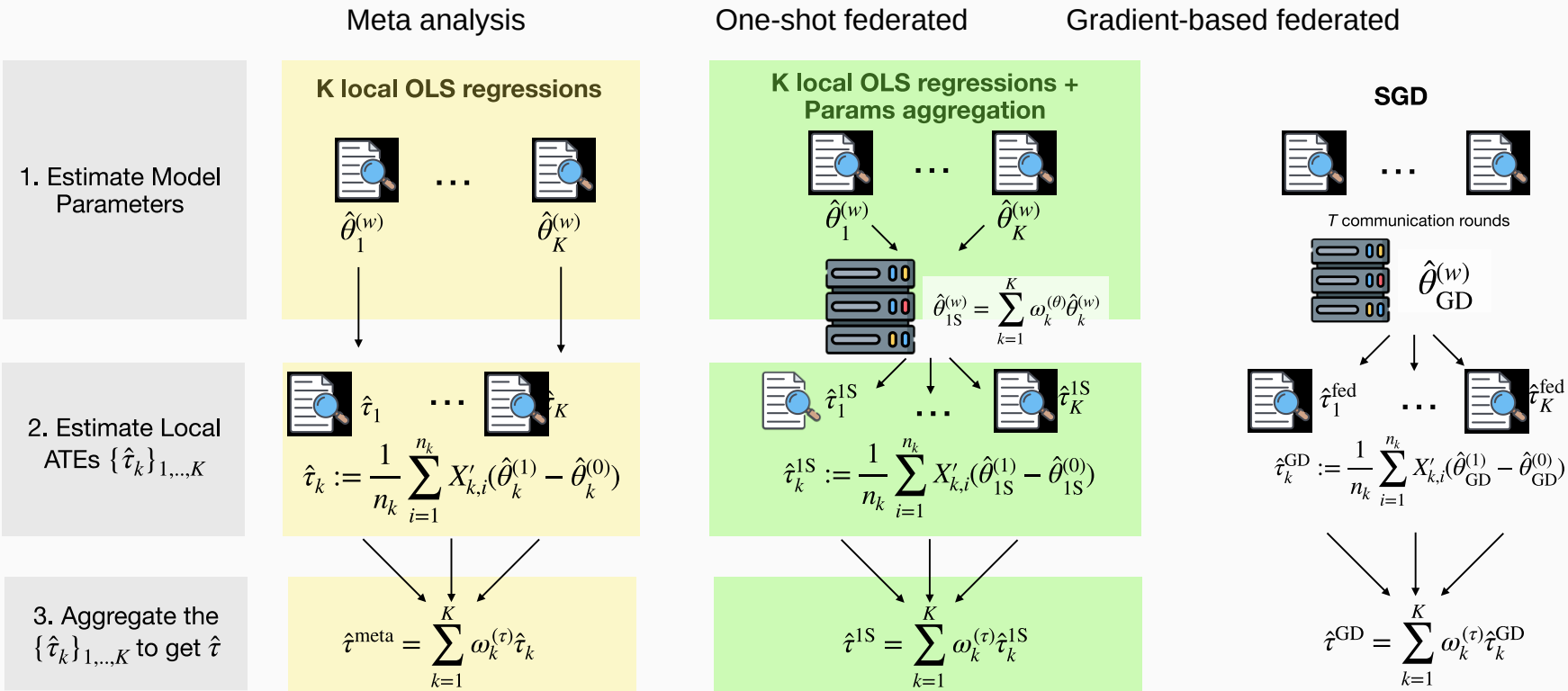
Aggregation w_k: sample size weights (SW) or inverse variance weights (IVW)

Three types of federated estimators

Ex: linear outcome model for all studies $8k$: $Y_{k;i}^{(w)} = c^{(w)} + X_{k;i}\ ^{(w)} + \text{"}_{k;i}^{(w)}$

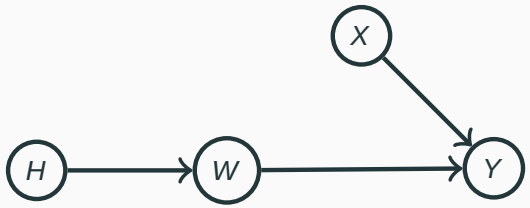
Baseline: estimator $\hat{}_{\text{pool}} = \frac{1}{n}\ ^{\text{P}}{}_{i=1}^{n} X_i^{0}(\hat{}_{\text{pool}}^{(1)}\ \ \hat{}_{\text{pool}}^{(0)})$ on pooled data

$\hat{}_{\text{pool}}^{(w)} = (\hat{c}_{\text{pool}}^{(w)}; \hat{}_{\text{pool}}^{(w)}) = \ X^{0(w)>} X^{0(w)}\ \ ^{1} X^{0(w)>} Y^{(w)}$ with $X^{0(w)} = [1; X^{(w)}]$



Aggregation w_k: sample size weights (SW) or inverse variance weights (IVW)

Statistical perf. & communication costs



Heterogeneity: Source membership H only affects treatment allocation:
$W_{k;i} \quad B(p_k)$

Unbiased estimators but different asymptotic variance & communication costs:

Estimator	V [1]	Com. rounds	Com. cost
$\hat{\tau}_{\text{Meta-SW}}$	$\frac{2}{n}\sum_{k=1}^{K}\frac{k}{p_k(1\quad p_k)}+\frac{1}{n}k\ {}^{(1)}\quad {}^{(0)}k^2_\Sigma$	1	$O(1)$
$\hat{\tau}_{\text{Meta-IVW}}$	$\Big(\sum_{k=1}^{K}\big(\ {}^{2}\frac{n\ k}{p_k(1\quad p_k)}+\frac{1}{n_k}k\ {}^{(1)}\quad {}^{(0)}k^2_\Sigma\big)\ {}^{1}\Big)\ {}^{1}$	1	$O(1)$
$\hat{\tau}_{\text{1S-SW}}$	V_{pool}	2	$O(d)$
$\hat{\tau}_{\text{1S-IVW}}$	V_{pool}	2	$O(d^2)$
$\hat{\tau}_{\text{GD}}$	V_{pool}	$T+1$	$O(Td)$
$\hat{\tau}_{\text{pool}}$	$V_{\text{pool}}=\frac{2}{n}\frac{1}{p(1\quad p)}+\frac{1}{n}\lVert\beta^{(1)}-\beta^{(0)}\rVert^2_\Sigma$	—	—

with $\quad_k = \mathrm{P}(H=k) = \mathrm{E}\ \frac{n_k}{n}$ and $p = \ {}^{\mathrm{P}}{}_{k=1}^{K}\ \frac{n_k}{n}p_k$

Federated Causal Inference/Generalization

Federated RCTs: Guidelines Meta & GD predilection regimes

. Small **sample size**: Gradient Descent: other need $n_k^{(w)}$ d for $k; w$

. **Heterogeneity**: Shift across sources (^$_{\text{meta}}$ $_{\text{IVW}}$ biased); di erent baseline outcomes (^$_{\text{meta}}$ handles center e ects, ^$_{\text{GD}}$ needs adjustment/prior knowledge on the model)

Multiple Randomized Control Trials, Multiple Observational data, Multiple Causal Measures

Real world data strenghten clinical evidences