Causal inference from heterogeneous data with missing data

Application to critical care management

Imke Mayer

École des Hautes Études en Sciences Sociales & Inria

Advisors: Julie Josse and Jean-Pierre Nadal

Founding: École doctorale de l'EHESS (ED 286)

Paris, September 20th 2021

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

Critical care management – Major trauma

- **Major trauma**: any injury that endangers a person's life or functional integrity
  - a major source of death and disability, 3rd cause of loss of disability adjusted life years (after cancer and CVD)
  - its socio-economic impact constitutes a public health challenge[1]
- **Critical care management**:
  - **multiple agents and sites**, different levels of care (scene of the accident, control center, ambulance, resuscitation room, . . .)

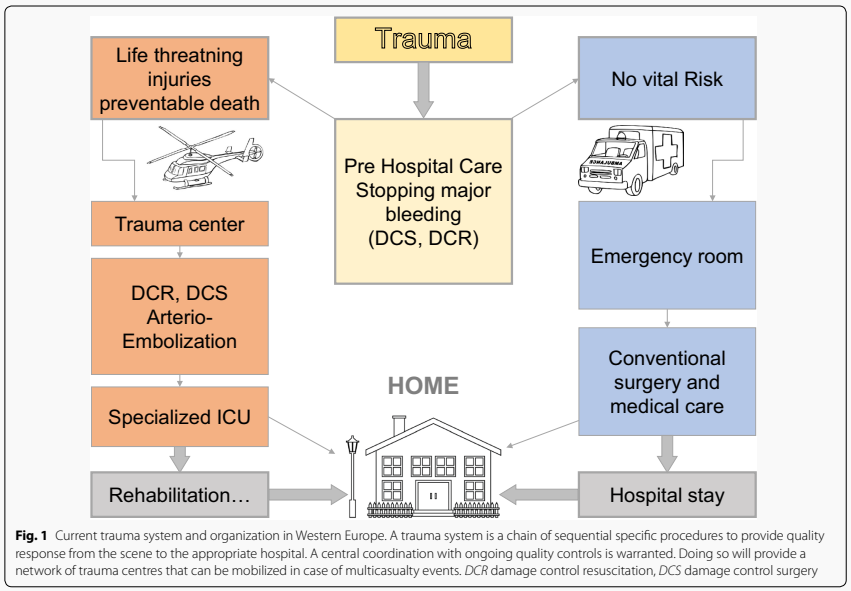


Fig. 1 Current trauma system and organization in Western Europe. A trauma system is a chain of sequential specific procedures to provide quality response from the scene to the appropriate hospital. A central coordination with ongoing quality controls is warranted. Doing so will provide a network of trauma centres that can be mobilized in case of multicasualty events. *DCR* damage control resuscitation, *DCS* damage control surgery

From Asehnoune et al. (2017)

[1] Hay et al., "Global, regional, and national disability-adjusted life-years (DALYs): a systematic analysis for the Global Burden of Disease Study 2016", 2017; Gauss et al., "Strategic proposal for a national trauma system in France", 2019.

Critical care management – Major trauma

- **Major trauma**: any injury that endangers a person's life or functional integrity
  - a major source of death and disability, 3rd cause of loss of disability adjusted life years (after cancer and CVD)
  - its socio-economic impact constitutes a public health challenge[1]
- **Critical care management**:
  - **multiple agents and sites**, different levels of care (scene of the accident, control center, ambulance, resuscitation room, ...)
  - quick decisions in complex context under **time and resource constraints**, with **high levels of uncertainty** and stress

[1] Hay et al., "Global, regional, and national disability-adjusted life-years (DALYs): a systematic analysis for the Global Burden of Disease Study 2016", 2017; Gauss et al., "Strategic proposal for a national trauma system in France", 2019.

Traumabase – A registry for major trauma patients in France

Id	Center	Accident	Age	Sex	Weight	Height	BMI	BP	SBP	SpO2	Lactates	Hb	Glasgow	Transfusion	...
1	Beaujon	Fall	54	m	85	NA	NA	180	110	97	NA	12.7	12	yes	
2	Lille	Other	33	m	80	1.8	24.69	130	62	100	4.8	11.1	15	no	
3	Pitie	Gun	26	m	NA	NA	NA	131	62	100	3.9	11.4	3	no	
4	Beaujon	AVP moto	63	m	80	1.8	24.69	145	89	100	1.66	13	15	yes	
6	Pitie	AVP bicycle	NA	m	75	NA	NA	104	86	100	NA	14.4	15	no	
6	Pitie	AVP pedestrian	30	w	NA	NA	NA	107	66	100	NA	14.3	15	yes	
7	HEGP	White weapon	16	NA	98	1.92	26.58	118	54	100	13	15.9	NA	yes	
9	Toulon	White weapon	20	m	NA	NA	NA	124	73	100	NA	13.7	15	no	
⋮							...								⋱

- ▷ 2012 – Motivation: gather information to learn from, improve decisions and assist patient care (initiated by Tobias Gauss & Sophie Hamada).
- ▷ Today:
  - ◇ >30,000 patients, 244 variables, 23 hospitals, >4,000 new patients/year
  - ◇ Complex and data-rich problem(s) → many different problems and solutions (different phases, different targets/indicators, etc.)

Traumatic brain injury in major trauma patients

Focus of this thesis

Patients with **traumatic brain injury** (TBI)
& treatment with **tranexamic acid** (TXA)

- TBI: any identified cerebral injury; $>$ 60M cases/year, first cause of death and disability among major trauma[2].
- 8,248 patients with TBI in our reference database.
- **Various treatments exist** for TBI (intracranial pressure control, maintenance of cerebral perfusion pressure and avoidance of secondary injuries, decompressive craniectomy).
- TXA: an **antifibrinolytic agent** (prevents plasmin from binding to fibrin).

[2] Rubiano et al., "Global neurotrauma research challenges and opportunities", 2015.

Traumatic brain injury in major trauma patients

Focus of this thesis

- Patients with **traumatic brain injury** (TBI)
  & treatment with **tranexamic acid** (TXA)

Challenges with the Traumabase® data

incomplete **heterogeneous** **observational**

Goal of my thesis

Address these challenges from a causal inference perspective.

- *Can we estimate the effect from TXA on TBI patients with evidence from the Traumabase®?*
- *How do the results compare to other findings on this question?*

Contributions of this thesis

I Consistently and efficiently estimate treatment effects with incomplete and heterogeneous attributes
- ◇ Impact of missingness on identifiability
- ◇ Doubly robust machine learning for informative missingness

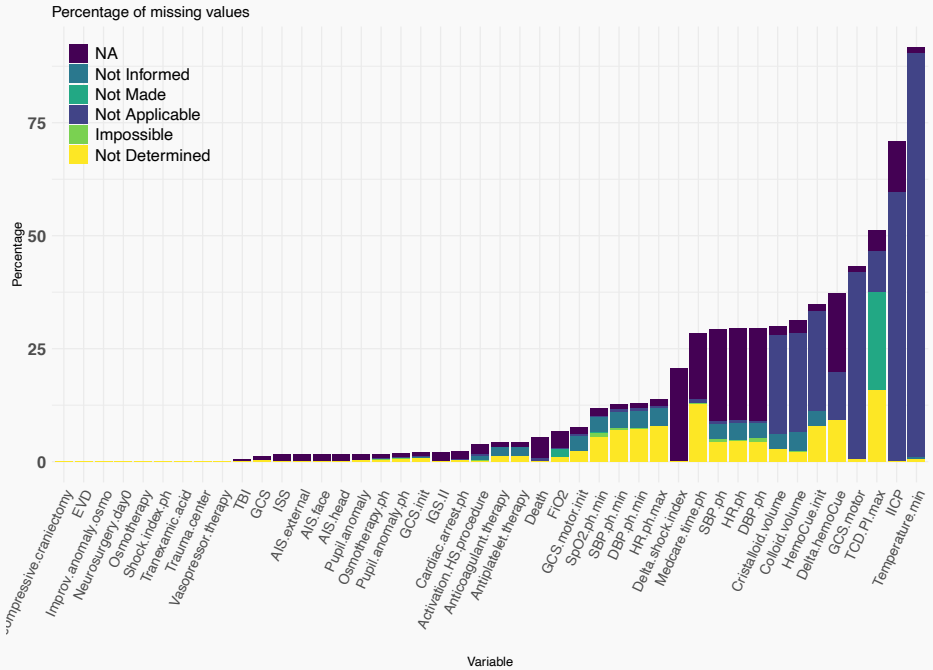
II Generalize treatment effects to different target populations
- ◇ Context and state of the art
- ◇ Multiple imputation strategies for incomplete multi-source attributes

III Provide ready-to-use and easily accessible tools for other applications
- ◇ R-miss-tastic – Platform for missing values problems and methods
- ◇ Traumabase® data analysis, integrative RCT and registry data analysis, AP–HP COVID-19 data analysis,

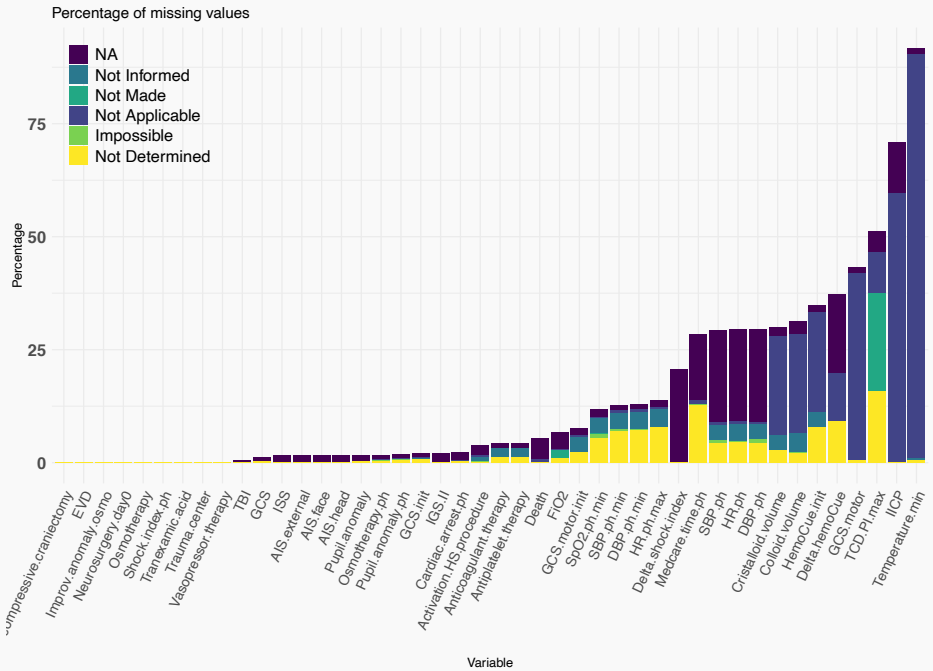
Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

The role of missing data



The role of missing data



- ▷ Standard (implicit) practice: *list-wise deletion* or *complete case analysis*
- ▷ But *"One of the ironies of Big Data is that missing data play an ever more significant role" (R. Samworth, 2019)*
  - ◇ **Loss of information**

    An $n \times p$ matrix, each entry is missing with probability 0.01.

    $p = 5 \implies \approx 95\%$ of rows kept; $p = 300 \implies \approx 5\%$ of rows kept.
  - ◇ **Bias in the analysis**

    Resulting sample is generally not representative of the target population.

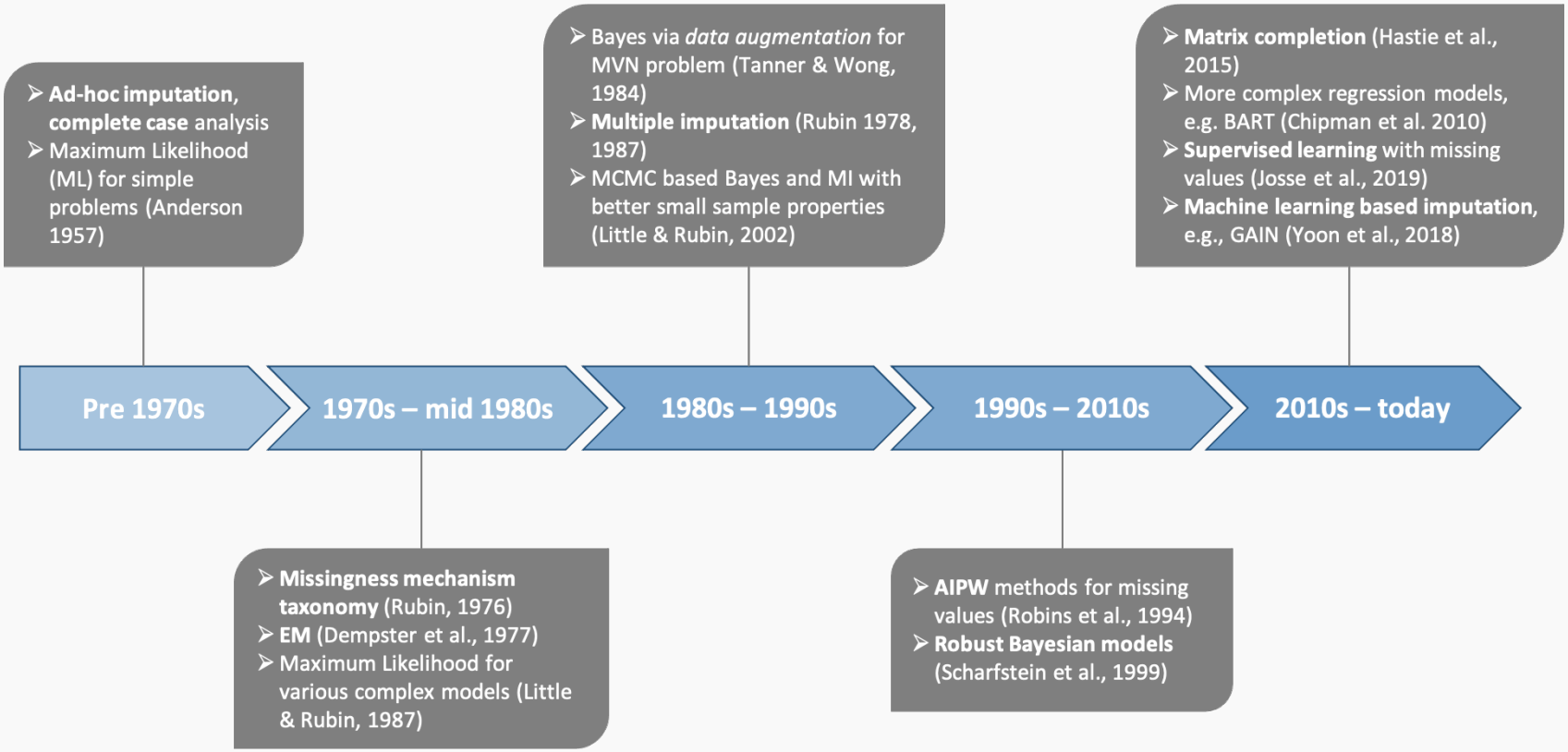
$\rightarrow$ *How to deal with missing values?*

How to deal with missing values?

There is **no single best solution** because it depends on

- ▷ the type of missing values
- ▷ the purpose of the statistical analysis.

A brief (and incomplete) history of missing values in statistical analyses[2]



[2] Based on a talk by R. Little (2020).

Rubin's missing values mechanisms taxonomy[3]

Idea: characterize the link between the (full) data and the missing values.

1. **Missing Completely At Random (MCAR)**

Probability to be missing depends neither on observed information nor on unobserved information.

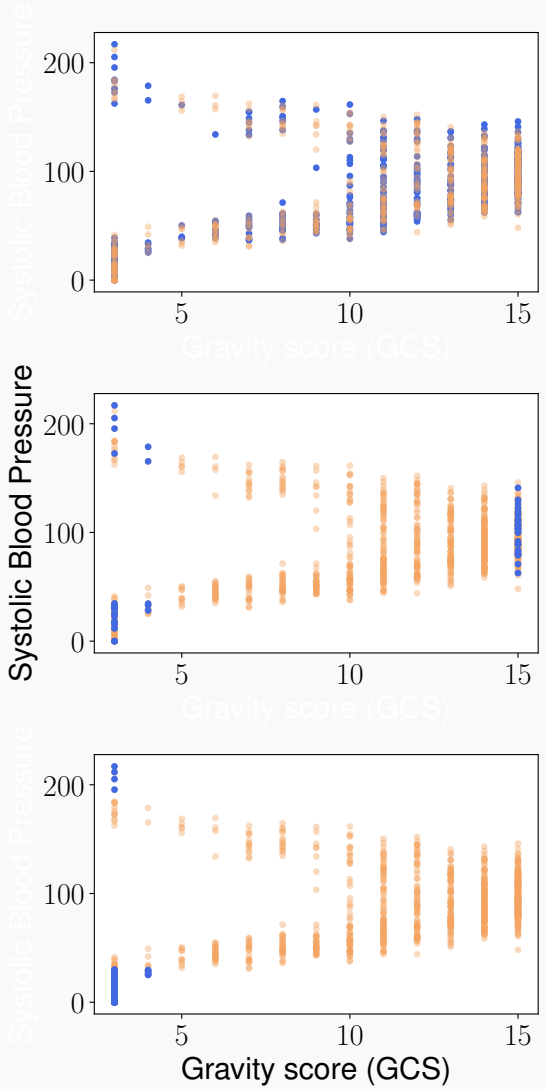
2. **Missing At Random (MAR)**

Probability to be missing depends on **observed** information.

3. **Missing Not At Random (MNAR)**

Probability to be missing depends on **unobserved** information.

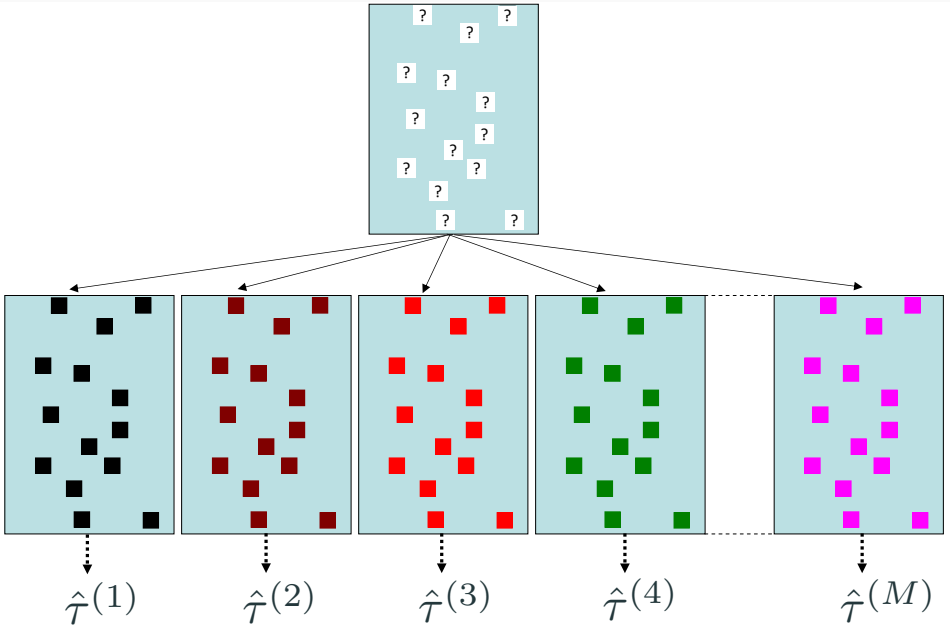
Missing values in y (Blood Pressure) – x (Gravity) always observed



[3] Rubin, "Inference and missing data", 1976.

Estimation and inference with missing values

One of the most popular methods: multiple imputation[4].



$$\Longrightarrow \quad \hat{\tau} = \frac{1}{M}\sum_{m=1}^{M}\hat{\tau}^{(m)} \quad \text{and} \quad \hat{V} = \frac{1}{M}\sum_{m=1}^{M}\widehat{Var}\left(\hat{\tau}^{(m)}\right) + \frac{1+\frac{1}{M}}{M-1}\sum_{m=1}^{M}(\hat{\tau}^{(m)} - \hat{\tau})^2$$

→ Rubin's rules for aggregation (estimator and its variance).

→ Variance estimation **reflects uncertainty due to the missing values**.

[4] Rubin, "Bayesian inference for causal effects: The role of randomization", 1978; Buuren, *Flexible Imputation of Missing Data. Second Edition*, 2018.

Supervised learning with missing values

Different from classical regression & inference tasks.

- Goal: **Predict an outcome** Y given $X^* \triangleq \begin{cases} X & \text{if } X \text{ is observed} \\ \texttt{NA} & \text{otherwise} \end{cases}$
- Data: train & test sets with missing values

[5] Josse et al., "On the consistency of supervised learning with missing values", 2019; Morvan et al., "What's a good imputation to predict with missing values?", 2021.

Supervised learning with missing values

Different from classical regression & inference tasks.

Goal: **Predict an outcome** Y given $X^* \triangleq \begin{cases} X & \text{if } X \text{ is observed} \\ \texttt{NA} & \text{otherwise} \end{cases}$

Data: train & test sets with missing values

Existing solutions[5]:

1. For almost all imputation functions, an impute-then-regress procedure with a powerful learner is Bayes optimal; e.g. mean imputation.
2. Possibility to skip the imputation and directly regress: random forest predictors with a different splitting criterion handling half-discrete X^*: *missing incorporated in attributes*.

[5] Josse et al., "On the consistency of supervised learning with missing values", 2019; Morvan et al., "What's a good imputation to predict with missing values?", 2021.

Missing incorporated in attributes[7]

Random trees with a different splitting criterion to account for NA

Method: Recursively, find which partition $\mathcal{P}$ minimizes

$$\mathbb{E}\left[(Y - \mathcal{P}(\mathrm{X}^*))^2\right],$$

where, for each feature j and each threshold θ, there are three possible partitions,

$$\begin{aligned} \{X_j^* \leq \theta \text{ or } X_j^* = \texttt{NA}\} \quad &\textbf{VS} \quad \{X_j^* > \theta\} \\ \{X_j^* \leq \theta\} \quad &\textbf{VS} \quad \{X_j^* > \theta \text{ or } X_j^* = \texttt{NA}\} \\ \{X_j^* \neq \texttt{NA}\} \quad &\textbf{VS} \quad \{X_j^* = \texttt{NA}\} \end{aligned}$$

$\rightarrow$ targets the Bayes estimate $\mathbb{E}[Y|\mathrm{X}^*]$

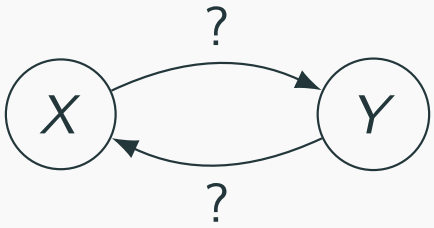
Implemented in the `grf` R package.[6]

[7] Twala, Jones, and Hand, "Good methods for coping with missing data in decision trees", 2008.
[6] Tibshirani et al., *grf: Generalized Random Forests*, 2020.

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

Causality in statistics



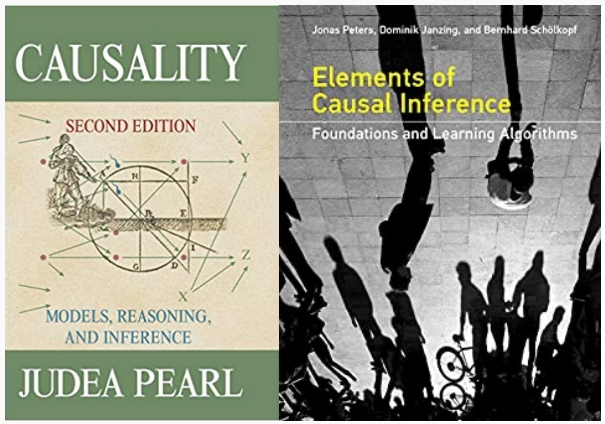
Correlation is not causation.

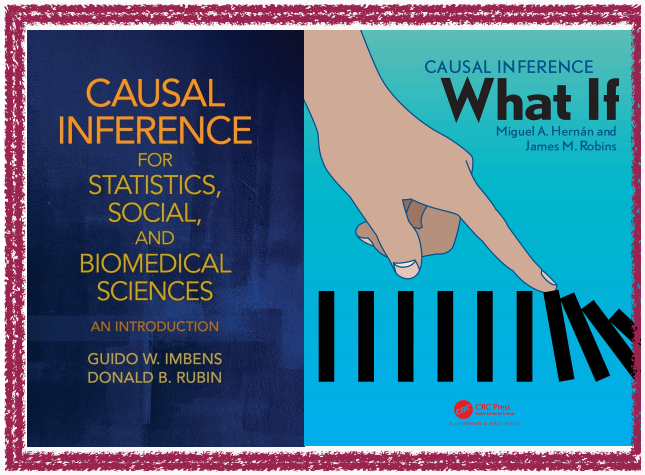
Overall mortality rate in TBI patients in the Traumabase®: 16%.

- ▷ Mortality rate among the TXA treated: 28%
- ▷ Mortality rate among the control: 13%

Is the treatment harmful?

- ▷ *'What causes what?' is not a question we can or aim to answer. But we can answer to 'what is the effect of a defined manipulation?'* (D. Rubin, OCIS 2021)





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

Causal effect

▷ n i.i.d. samples $\big(\underbrace{X_i}_{\text{context}}, \overbrace{W_i}^{\text{treatment}}, \underbrace{Y_i(1), Y_i(0)}_{\textit{potential outcomes}}\big) \in \mathcal{X} \times \{0, 1\} \times \mathcal{Y} \times \mathcal{Y}$

▷ Individual causal effect of the treatment:

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

$\rightarrow$ Missingness problem: Δ_i never observed (observe 1 outcome/unit)

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

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

Causal effect

▷ n i.i.d. samples $(\underbrace{X_i}_{\text{context}}, \overbrace{W_i}^{\text{treatment}}, \underbrace{Y_i(1), Y_i(0)}_{\textit{potential outcomes}}) \in \mathcal{X} \times \{0, 1\} \times \mathcal{Y} \times \mathcal{Y}$

▷ Individual causal effect of the treatment:

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

$\rightarrow$ Missingness problem: Δ_i never observed (observe 1 outcome/unit)

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

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

Identifiability of the ATE

- Straightforward in experimental data (randomized controlled trial, RCT) — by design ().
- Requires assumptions in case of non-randomized or observational data.
  Treatment assignment W depends on covariates X
  $\Rightarrow$ Treated and control groups **differ at baseline**.
  $\Rightarrow$ The data is **confounded**.

Assumptions for ATE identifiability in observational data

1. SUTVA

$$Y = WY(1) + (1 - W)Y(0)$$

2. Unconfoundedness - selection on observables

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \,|\, X_i$$

Treatment assignment W_i is random, conditionally on covariates X_i.

3. Overlap

Define propensity score as $e(x) \triangleq P(W_i = 1 \,|\, X_i = x), \quad \forall\, x \in \mathcal{X}.$

Assume

$$\exists \eta > 0,\ s.t.\ \eta < e(x) < 1 - \eta, \quad \forall\, x \in \mathcal{X}.$$

Estimation of the ATE

Different estimators have been proposed since the 1980's and can be summarized by 4 different classes:

1. Regression adjustment
2. Balance the differences between the two groups: inverse propensity weighting (IPW), matching
3. Extrapolate fitted models from one group to the other: g-formula
4. Combine the two: CBPS, AIPW, and other **doubly robust estimators**[8]

Except for approach 1, all methods consider τ as a (population) **causal parameter**, not as a model parameter to estimate directly.

[8] Robins, Rotnitzky, and Zhao, "Estimation of Regression Coefficients When Some Regressors are not Always Observed", 1994.

Doubly robust ATE estimation

Idea: combine different models to efficiently use the data and to protect against mis-specification.

Model the propensity score & the conditional outcomes

$$\text{nuisance parameters} \equiv \begin{cases} W \sim X, & e(x) \\ Y(w) \sim X, & \mu_{(w)}(x) \triangleq \mathbb{E}[Y_i(w) \,|\, X_i = x] \end{cases}$$

Augmented IPW

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

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

Doubly robust ATE estimation

Idea: combine different models to efficiently use the data and to protect against mis-specification.

Model the propensity score & the conditional outcomes

$$\text{nuisance parameters} \equiv \begin{cases} W \sim X, & e(x) \\ Y(w) \sim X, & \mu_{(w)}(x) \triangleq \mathbb{E}[Y_i(w) \,|\, X_i = x] \end{cases}$$

Augmented IPW

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

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

Recent result from 2018[9]: **Double Machine Learning**

Extends the previous to use **any (machine learning) procedure** such as **random forests**, deep nets, etc. to estimate $\hat{e}(x)$ and $\hat{\mu}_{(w)}(x)$ without harming the interpretability of the causal effect estimation.

[9] Chernozhukov et al., "Double/debiased machine learning for treatment and structural parameters", 2018.

Contributions of this thesis

I Consistently and efficiently estimate treatment effects with incomplete and heterogeneous attributes
- ◇ Impact of missingness on identifiability
- ◇ Doubly robust machine learning for informative missingness

II Generalize treatment effects to different target populations
- ◇ Context and state of the art
- ◇ Multiple imputation strategies for incomplete multi-source attributes

III Provide ready-to-use and easily accessible tools for other applications
- ◇ R-miss-tastic – Platform for missing values problems and methods
- ◇ Traumabase® data analysis, integrative RCT and registry data analysis, AP–HP COVID-19 data analysis,

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

Missing values in the covariates

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
NA	20	F	1	?	Alive
-6	45	NA	0	Dead	?
0	NA	M	1	?	Alive
NA	32	F	1	?	Dead
1	63	M	1	Dead	?
-2	NA	M	0	Alive	?

Three families of methods with different sets of assumptions

1. Unconfoundedness despite missingness
2. Full data unconfoundedness + classical missing values mechanisms
3. Latent unconfoundedness + classical missing values mechanisms

Joint work with E. Sverdrup, T. Gauss, J.-D. Moyer, S. Wager, J. Josse[10]

[10] Mayer et al., "Doubly robust treatment effect estimation with missing attributes", 2020.

Missing values in the covariates

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	Y(0)	Y(1)
NA	20	F	1	?	Alive
-6	45	NA	0	Dead	?
0	NA	M	1	?	Alive
NA	32	F	1	?	Dead
1	63	M	1	Dead	?
-2	NA	M	0	Alive	?

Three families of methods with different sets of assumptions

1. Unconfoundedness despite missingness
   → unconfoundedness holds conditionally on incomplete X
2. Full data unconfoundedness + classical missing values mechanisms
   → missing values are ignorable and don't affect causal identifiability
3. Latent unconfoundedness + classical missing values mechanisms
   → confounders are latent, we observe incomplete proxies

Joint work with E. Sverdrup, T. Gauss, J.-D. Moyer, S. Wager, J. Josse[10]

[10] Mayer et al., "Doubly robust treatment effect estimation with missing attributes", 2020.

1. Treatment is unconfounded given X^*

Notation:

- response pattern $R \in \{0,1\}^p$, $R_j \triangleq \mathbb{1}_{\{X_j \text{ is observed}\}}$,
- $\mathrm{X}^* \triangleq \mathrm{R} \odot \mathrm{X} + (1 - \mathrm{R}) \odot \mathrm{NA} \in \{\mathbb{R} \cup \mathrm{NA}\}^{\mathrm{p}}$

$$X^* \equiv \textbf{observed} \text{ covariates} + \textbf{response pattern}.$$

Unconfoundedness despite missingness (UDM)[11] ,[12]

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i \mid X^*$$

Note: **no** assumption on the missingness mechanism.

Doctors decide to treat a patient based on what they observe.
& We have access to the same information as the doctors.

Example

For patient 1, the doctor observes temperature, heart rate and blood pressure, and makes the decision based on this.
For patient 2, the doctor observes temperature and heart rate and cannot measure BP, and bases the treatment decision on these 3 elements of information.

[11] Mattei and Mealli, "Estimating and using propensity score in presence of missing background data: an application to assess the impact of childbearing on wellbeing", 2009.
[12] Blake et al., "Estimating treatment effects with partially observed covariates using outcome regression with missing indicators", 2020.

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

Under UDM: Tree-based estimation with missing values

Generalized nuisance parameters[13]

$$e^*(x^*) \triangleq P(W = 1 \,|\, X^* = x^*) \text{ and } \mu^*_{(w)}(x^*) \triangleq \mathbb{E}[Y(w) \,|\, X^* = x^*]$$

$\equiv$ 1 model / pattern: $\sum_{r \in \{0,1\}^d} \mathbb{E}\left[Z | X_{obs(r)}, R = r\right] \mathbb{1}_{R=r}$, $Z \in \{W, Y(0), Y(1)\}$.

AIPW with missing values

$$\widehat{\tau^*}_{AIPW} \triangleq \frac{1}{n} \sum_i \left(\widehat{\mu^*_{(1)}}(X_i) - \widehat{\mu^*_{(0)}}(X_i) + W_i \frac{Y_i - \widehat{\mu^*_{(1)}}(X_i)}{\widehat{e^*}(X_i)} - (1 - W_i) \frac{Y_i - \widehat{\mu^*_{(0)}}(X_i)}{1 - \widehat{e^*}(X_i)} \right)$$

Under mild assumptions on the nuisance parameter estimators[14], $\widehat{\tau^*}_{AIPW}$ is $\sqrt{n}$-consistent and asymptotically normal.

$\rightarrow$ Recall the supervised learning with missing values.[15]

- ▷ Mean imputation is consistent with a powerful learner.
- ▷ Alternative for tree-based predictors: Missing Incorporate in Attributes (MIA).

[15] Morvan et al., "What's a good imputation to predict with missing values?", 2021.
[14] Wager and Athey, "Estimation and inference of heterogeneous treatment effects using random forests", 2018
[13] Rosenbaum and Rubin, "Reducing bias in observational studies using subclassification on the propensity score", 1984

Methods to handle missing values for ATE estimation

	Covariates		Missingness		Unconfoundedness			Models for (W, Y)	
	multiva-riate normal	general	M(C)AR	MNAR	Case 1 UDM	Case 2 Classical	Case 3 Latent	logistic-linear	non-param.
(SA)EM	✓	✗	✓	✗	✓	✗	✗	✓	✗
MIA	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mult. Imputation	✓	✓	✓	✗	✗	✓	✗	✓	(✗)
MissDeepCausal	✓	✓	✓	✗	✗	✗	✓	✓	✓

✓ can be handled, ✗ not applicable in theory, (✗) no theoretical guarantees but heuristics

Methods to handle missing values for ATE estimation

	Covariates		Missingness		Unconfoundedness			Models for (W, Y)	
	multivariate normal	general	M(C)AR	MNAR	Case 1 UDM	Case 2 Classical	Case 3 Latent	logistic-linear	non-param.
(SA)EM	✓	✗	✓	✗	✓	✗	✗	✓	✗
MIA	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mult. Imputation	✓	✓	✓	✗	✗	✓	✗	✓	(✗)
MissDeepCausal	✓	✓	✓	✗	✗	✗	✓	✓	✓

✓ can be handled, ✗ not applicable in theory, (✗) no theoretical guarantees but heuristics

Apply under UDM assumption

No assumption on the missingness mechanism is made.

- ▷ MIA → seen today
- ▷ (SA)EM → maximum-likelihood approximation of observed likelihood using EM algorithm[16]. Contribution in Mayer et al., *AOAS* (2020)

[16] Jiang et al., "Logistic regression with missing covariates—Parameter estimation, model selection and prediction within a joint-modeling framework", 2020

Methods to handle missing values for ATE estimation

	Covariates		Missingness		Unconfoundedness			Models for (W, Y)	
	multivariate normal	general	M(C)AR	MNAR	Case 1 UDM	Case 2 Classical	Case 3 Latent	logistic-linear	non-param.
(SA)EM	✓	✗	✓	✗	✓	✗	✗	✓	✗
MIA	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mult. Imputation	✓	✓	✓	✗	✗	✓	✗	✓	(✗)
MissDeepCausal	✓	✓	✓	✗	✗	✗	✓	✓	✓

✓ can be handled, ✗ not applicable in theory, (✗) no theoretical guarantees but heuristics

Applies under full data unconfoundedness and MAR

Multiple imputation solutions.

- ▷ $\hat{\tau}^{MI}_{IPW}$ → existing works on consistency and applications[16]
- ▷ $\hat{\tau}^{MI}_{AIPW}$ → contribution in Mayer et al., *AOAS* (2020)

[16] Seaman and White, "Inverse probability weighting with missing predictors of treatment assignment or missingness", 2014; Mattei and Mealli, "Estimating and using propensity score in presence of missing background data: an application to assess the impact of childbearing on wellbeing", 2009

Methods to handle missing values for ATE estimation

	Covariates		Missingness		Unconfoundedness			Models for (W, Y)	
	multiva-riate normal	general	M(C)AR	MNAR	Case 1 UDM	Case 2 Classical	Case 3 Latent	logistic-linear	non-param.
(SA)EM	✓	✗	✓	✗	✓	✗	✗	✓	✗
MIA	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mult. Imputation	✓	✓	✓	✗	✗	✓	✗	✓	(✗)
MissDeepCausal	✓	✓	✓	✓	✗	✗	✓	✓	✓

✓ can be handled, ✗ not applicable in theory, (✗) no theoretical guarantees but heuristics

Applies under latent unconfoundedness

True confounders are **latent** variables, we observe **incomplete proxies**, e.g., IQ as a proxy for intelligence or temperature and CRP for an infection.

- ▷ Matrix factorization and regression adjustment → existing results on consistency and applications[16]
- ▷ MissDeepCausal → non-linear latency structure via variational auto-encoders (VAE), contribution in Mayer, Vert & Josse (2020)[17]

[17] Mayer et al., *MissDeepCausal: Causal Inference from Incomplete Data Using Deep Latent Variable Models*, 2020
[17] Kallus, Mao, and Udell, "Causal Inference with Noisy and Missing Covariates via Matrix Factorization", 2018

Methods to handle missing values for ATE estimation

	Covariates		Missingness		Unconfoundedness			Models for (W, Y)	
	multivariate normal	general	M(C)AR	MNAR	Case 1 UDM	Case 2 Classical	Case 3 Latent	logistic-linear	non-param.
(SA)EM	✓	✗	✓	✗	✓	✗	✗	✓	✗
MIA	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mult. Imputation	✓	✓	✓	✗	✗	✓	✗	✓	(✗)
MissDeepCausal	✓	✓	✓	✗	✗	✗	✓	✓	✓

✓ can be handled, ✗ not applicable in theory, (✗) no theoretical guarantees but heuristics

Performances

Our extensive simulation study corroborates that due to the different identifiability assumptions there is no overall best performing method, but the proposed methods perform well under the corresponding assumptions.

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

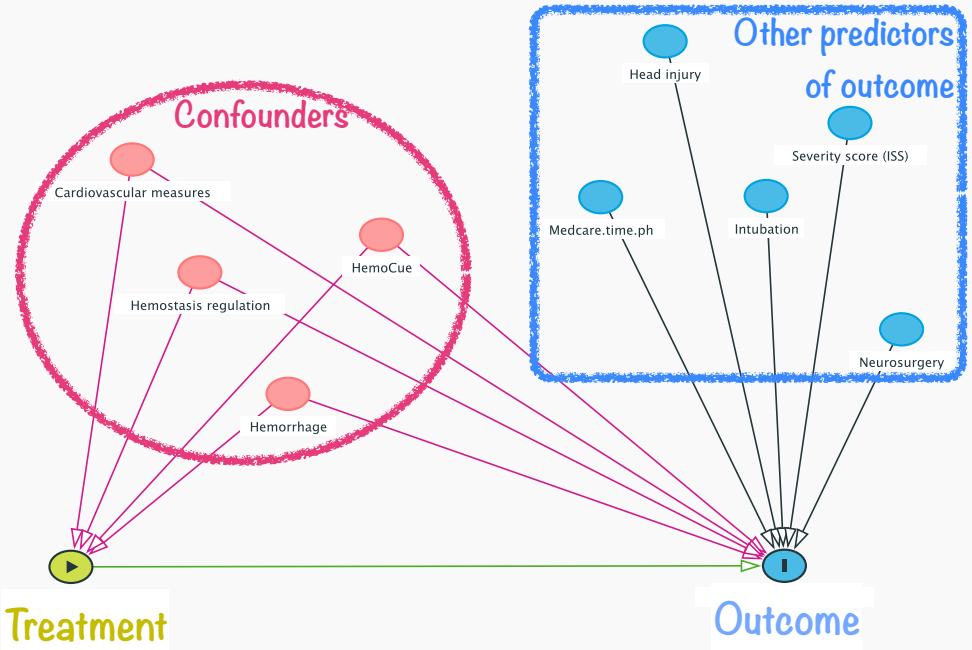
Recall our initial problem and question

- Question: *Is there a benefit from tranexamic acid (TXA) for traumatic brain injury (TBI) patients in terms of mortality reduction?*
- Data: Traumabase® registry with 8,248 TBI patients.

[17] Jones and Hunter, "Consensus methods for medical and health services research.", 1995
[16] Textor, Hardt, and Knüppel, "DAGitty: a graphical tool for analyzing causal diagrams", 2011

Recall our initial problem and question

- Question: *Is there a benefit from tranexamic acid (TXA) for traumatic brain injury (TBI) patients in terms of mortality reduction?*
- Data: Traumabase® registry with 8,248 TBI patients.



Simplified causal graph produced using DAGitty[16]

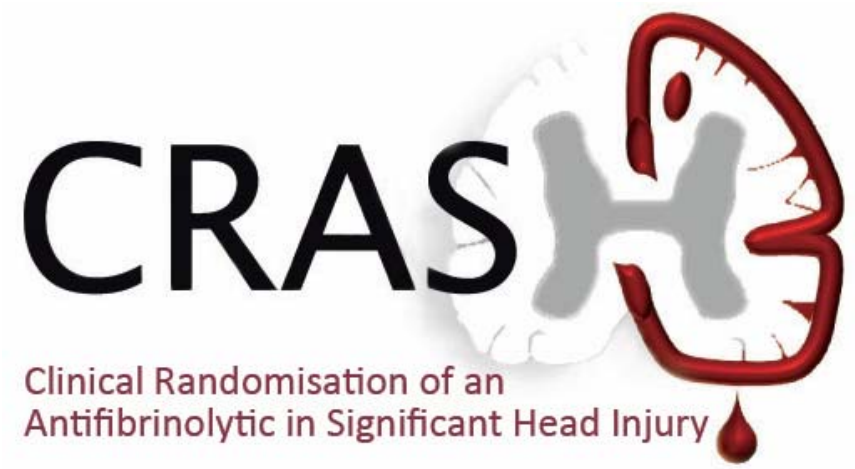
Step 1: Identify relevant covariates through a Delphi process[17]

- 18 confounders
- 22 predictors of Y only

[17] Jones and Hunter, "Consensus methods for medical and health services research.", 1995
[16] Textor, Hardt, and Knüppel, "DAGitty: a graphical tool for analyzing causal diagrams", 2011

Results

Emulation of a recent RCT[18]

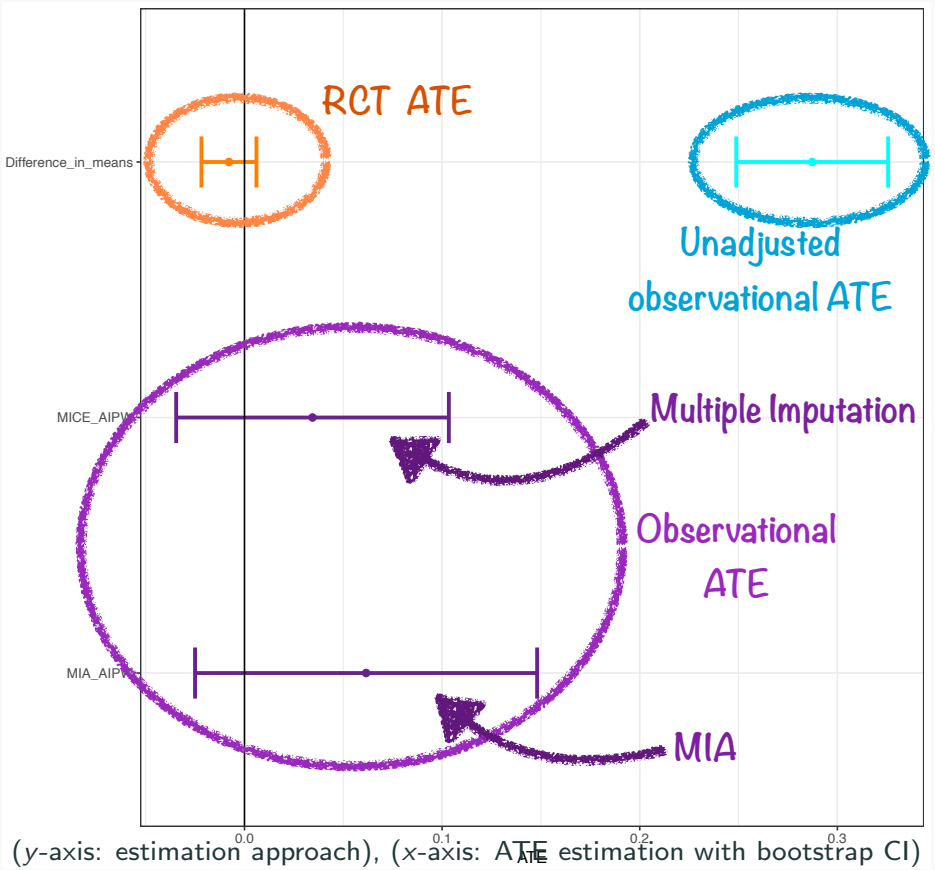


Tranexamic acid for the treatment of significant traumatic brain injury: an international randomised, double blind placebo controlled trial

[18] Cap, "CRASH-3: a win for patients with traumatic brain injury", 2019.

Results

ATE estimation for the effect of tranexamic acid on D-28 head-injury related mortality for TBI patients.



(*y*-axis: estimation approach), (*x*-axis: ATE estimation with bootstrap CI)

$\tau = 0$: "No average effect", $\tau < 0$: "TXA reduces mortality".

Contributions of this thesis

I Consistently and efficiently estimate treatment effects with incomplete and heterogeneous attributes
- ◇ Impact of missingness on identifiability
- ◇ Doubly robust machine learning for informative missingness

II Generalize treatment effects to different target populations
- ◇ Context and state of the art
- ◇ Multiple imputation strategies for incomplete multi-source attributes

III Provide ready-to-use and easily accessible tools for other applications
- ◇ R-miss-tastic – Platform for missing values problems and methods
- ◇ Traumabase® data analysis, integrative RCT and registry data analysis, AP–HP COVID-19 data analysis,

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

RCT - Ground truth?

"Gold standard" to assess the causal effect of an intervention or treatment on an outcome. τ is identifiable by design.
$\rightarrow$**The covariate distributions of treated and control groups are balanced** Control group looks like treatment group: difference in response is attributable to treatment.

Randomized Controlled Trial (RCT)

- Simple unbiased estimate of the ATE, but often on narrowly defined populations
- Examples:
  - Evidence-based medicine,
  - Economic experiments,
  - A/B testing.
- High **internal** validity

Observational data

- Large amounts of data reflecting day-to-day practice, but with **confounding**
- Examples:
  - Electronic Health Records (EHR),
  - Public policy evaluations,
  - Social sciences usage.
- High **external** validity

Leverage both RCT and observational data

RCT

– Narrowly defined population

\+ High **internal** validity

Observational data

– **Confounding**

\+ High **external** validity

We could use both to . . .

▷ . . . validate observational methods. → Contribution in Mayer et al. (2021)[18]

▷ . . . correct confounding bias, ground the observational data.

▷ . . . improve estimation of heterogeneous treatment effects and long-term effects.

▷ . . . **generalize the ATE to a (broader) target population.**

[18] Mayer et al., "Machine Learning Augmented Causal Inference To Estimate The Treatment Effect of Tranexamic Acid In Traumatic Brain Injury", 2021

[19] Dagan et al., "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting", 2021.

Leverage both RCT and observational data

RCT

– Narrowly defined population
\+ High **internal** validity

Observational data

– **Confounding**
\+ High **external** validity

We could use both to . . .

▷ . . . validate observational methods. Contribution in Mayer et al. (2021)[18]

▷ . . . correct confounding bias, ground the observational data.

▷ . . . improve estimation of heterogeneous treatment effects.

▷ . . . **generalize the ATE to a (broader) target population.**

Differences between findings from RCT on Pfizer COVID-19 vaccine efficacy and emulated trial from large obs. data on vaccine effectiveness[19].

→ Reduce the time and cost to approve a drug for patients who could benefit from it.

[18] Mayer et al., "Machine Learning Augmented Causal Inference To Estimate The Treatment Effect of Tranexamic Acid In Traumatic Brain Injury", 2021

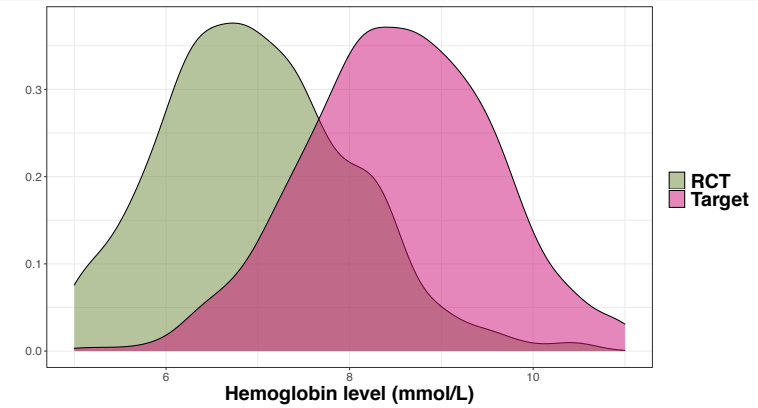
[19] Dagan et al., "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting", 2021.

Notations for joint analysis

Joint work with Bénédicte Colnet, Julie Josse, Gaël Varoquaux, Jean-Philippe Vert, Shu Yang, and others.

- We introduce S an indicator of eligibility for the trial & willingness to participate
- The distribution of covariates X is not the same in the target population and in the RCT,

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



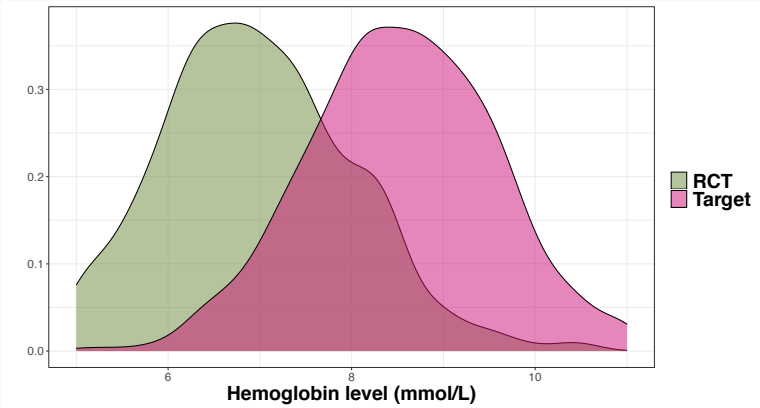
Notations for joint analysis

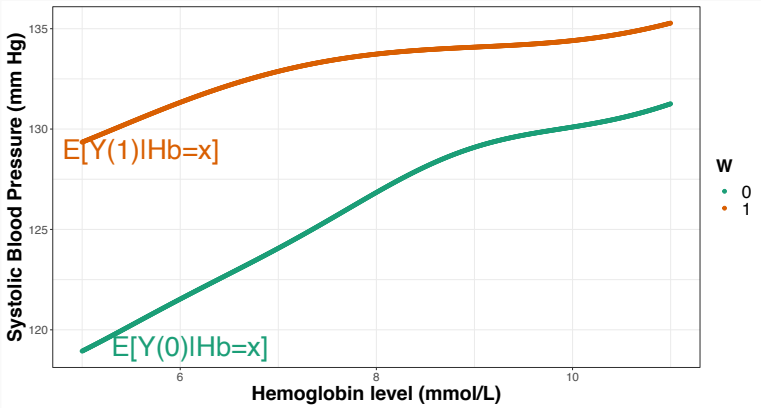
Joint work with Bénédicte Colnet, Julie Josse, Gaël Varoquaux, Jean-Philippe Vert, Shu Yang, and others.

- We introduce S an indicator of eligibility for the trial & willingness to participate
- The distribution of covariates X is not the same in the target population and in the RCT,

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

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





Intuition of the generalization task

	Set	Covariates			Treatment	Outcome under W
		X_1	X_2	X_3	W	Y
1	$\mathcal{R}$	1.1	20	5.4	1	24.1
...	$\mathcal{R}$		...		...	...
$n-1$	$\mathcal{R}$	-6	45	8.3	0	26.3
n	$\mathcal{R}$	0	15	6.2	1	23.5
$n+1$	$\mathcal{O}$	-2	52	7.1	NA	NA
$n+2$	$\mathcal{O}$	-1	35	2.4	NA	NA
...	$\mathcal{O}$		...		NA	NA
$n+m$	$\mathcal{O}$	-2	22	3.4	NA	NA

Available data with observed treatment and outcome only in the RCT.

Idea: Use a sample of the target population to generalize τ.[20]

[20] Other terms are *data fusion*, *transportability*, *covariate shift*.

Intuition of the generalization task

	Set	Covariates			Treatment	Outcome under W
		X_1	X_2	X_3	W	Y
1	$\mathcal{R}$	1.1	20	5.4	1	24.1
...	$\mathcal{R}$		...		...	...
$n-1$	$\mathcal{R}$	-6	45	8.3	0	26.3
n	$\mathcal{R}$	0	15	6.2	1	23.5
$n+1$	$\mathcal{O}$	-2	52	7.1	NA	NA
$n+2$	$\mathcal{O}$	-1	35	2.4	NA	NA
...	$\mathcal{O}$		...		NA	NA
$n+m$	$\mathcal{O}$	-2	22	3.4	NA	NA

Available data with observed treatment and outcome only in the RCT.

Idea: Use a sample of the target population to generalize τ.[20]

Typical estimators of τ rely on different formulae and are obtained by:

- ▷ **weighting the RCT sample** so that it fits the target population distribution (**IPSW**)
- ▷ modeling the conditional outcomes and **extrapolate** to the target population sample (**G-formula**)
- ▷ combining the previous two ideas (**doubly robust approaches**: **AIPSW**, **C**alibration **W**eighting)

[20] Other terms are *data fusion*, *transportability*, *covariate shift*.

Assumptions for generalized ATE identifiability

Ignorability assumption on trial participation (S-ignorability)

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

X contains all covariates that are treatment effect modifiers and with a distributional shift.

Positivity of trial participation

$$\text{Selection score: } \pi_S(x) \triangleq P(S_i = 1 \mid X_i = x) \quad \forall\, x \in \mathcal{X}.$$

Assume $\exists\, c > 0$, such that $\forall x \in \mathcal{X}$, $\pi_S(x) \geq c > 0$.

Each individual from the target population had a non-zero probability to be eligible for the trial.

State of the art for generalization

Review of the state of the art

The state of the art has been reviewed from a theoretical, practical and empirical perspective in Colnet, Mayer, et al. (under review at *Statistical Science*, 2020).[21]

[23] Dong et al., "Integrative analysis of randomized clinical trials with real world evidence studies", 2020.
[22] Dahabreh et al., "Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals", 2019.
[21] Colnet et al., "Causal inference methods for combining randomized trials and observational studies: a review", 2020

State of the art for generalization

Review of the state of the art

The state of the art has been reviewed from a theoretical, practical and empirical perspective in Colnet, Mayer, et al. (under review at *Statistical Science*, 2020).[21]

Augmented inverse probability of sampling weighting (AIPSW)[22]

$$\widehat{\tau}_{\mathrm{AIPSW},n,m} \triangleq \frac{1}{n}\sum_{i=n}^{n} \frac{1}{\widehat{\alpha}_{n,m}(X_i)}\left[\frac{W_i\left\{Y_i-\widehat{\mu}_{1,1,n}(X_i)\right\}}{e_1(X_i)} - \frac{(1-W_i)\left\{Y_i-\widehat{\mu}_{0,1,n}(X_i)\right\}}{1-e_1(X_i)}\right]$$
$$+\frac{1}{m}\sum_{i=n+1}^{n+m}\left\{\widehat{\mu}_{1,1,n}(X_i)-\widehat{\mu}_{0,1,n}(X_i)\right\}.$$

where $\alpha(x)$ is the conditional odds of RCT selection.

$\widehat{\tau}_{\mathrm{AIPSW},n,m}$ is a doubly robust estimator of τ.

▸ Details on this estimator

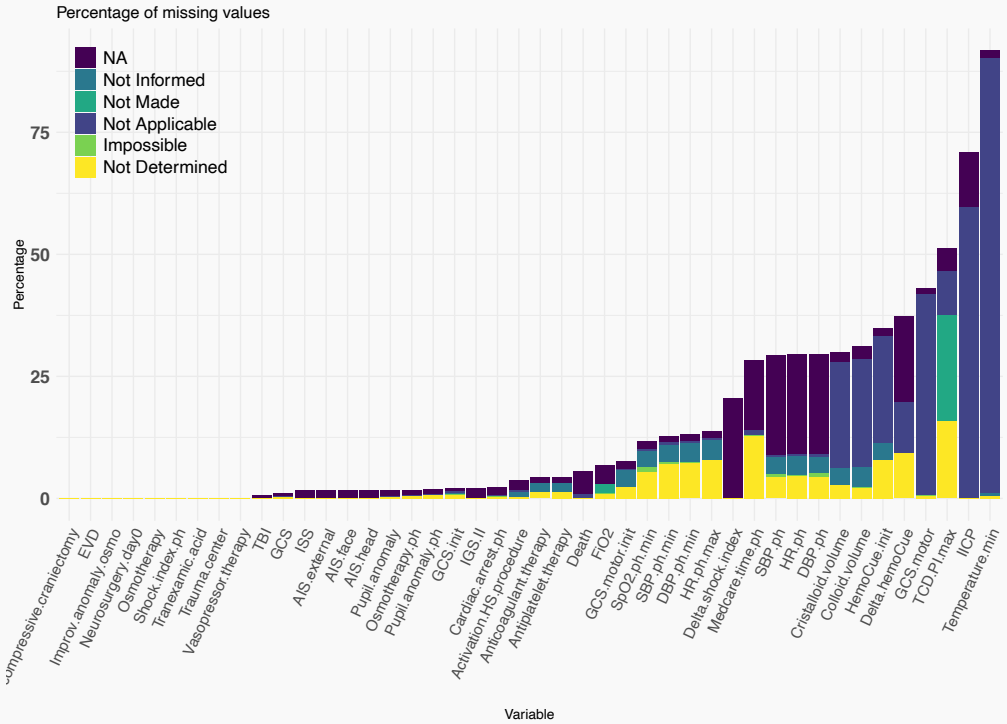
Alternative doubly robust estimator: (Augmented) Calibration Weighting[23]

[23] Dong et al., "Integrative analysis of randomized clinical trials with real world evidence studies", 2020.
[22] Dahabreh et al., "Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals", 2019.
[21] Colnet et al., "Causal inference methods for combining randomized trials and observational studies: a review", 2020

Outline

1. Introduction
   - Critical care management & Traumabase
   - Missing data
   - Causal inference
2. Treatment effect estimation with incomplete attributes
   - Identifiability with incomplete attributes
   - Doubly robust treatment effect estimation with incomplete attributes
   - Data analysis on the Traumabase® registry
3. Generalizing treatment effects
   - Context and state of the art
   - Generalizing with incomplete source and target samples
4. Conclusion

Recall again the missing data challenge of the Traumabase®



	Set	Covariates X_1^*	X_2^*	X_3^*	Treatment W	Outcome Y
1	$\mathcal{R}$	**NA**	20	5.4	1	24.1
...	$\mathcal{R}$		...		...	...
$n-1$	$\mathcal{R}$	-6	**NA**	8.3	0	26.3
n	$\mathcal{R}$	0	15	6.2	1	23.5
$n+1$	$\mathcal{O}$	**NA**	**NA**	7.1	NA	NA
$n+2$	$\mathcal{O}$	-1	35	**NA**	NA	NA
...	$\mathcal{O}$		...		NA	NA
$n+m$	$\mathcal{O}$	-2	**NA**	3.4	NA	NA

$\rightarrow$ *How do these missing values impact identifiability and estimation of the generalized ATE?*

Impact of missing values on identifiability

Identifiability in the complete data case in a nutshell:

Everyone has a non-zero chance to be eligible and that conditionally on attributes, the treatment effect is stable across populations.

Two solutions for identifiability with missing values

1. Conditionally independent selection (CIS)

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

→ eligibility and selection depend on the missingness pattern

e.g., trial with a list of 10 eligibility criteria and only 5 out of these need to be satisfied. For ind. 1, criteria C_1, C_9, C_7, C_2, C_3 are observed and he is included before recording $C_4, C_5, C_6, C_8, C_{10}$.

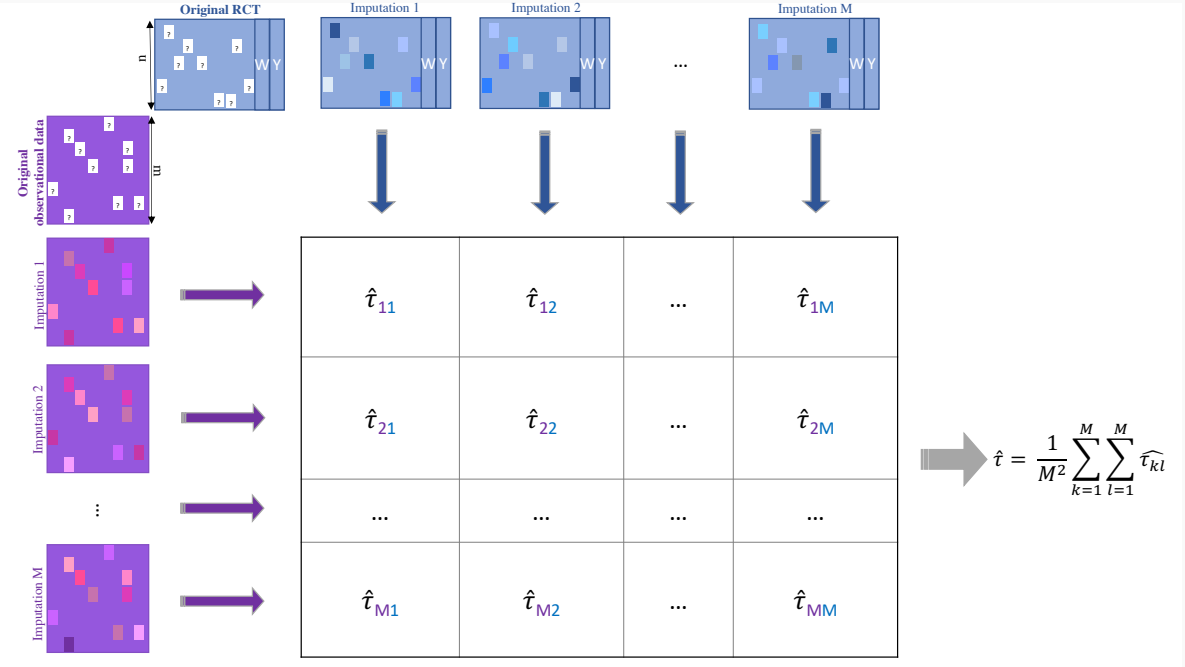
2. Full data S-ignorability + classical missingness assumptions (MCAR, MAR)

→ missing values don't alter selection or outcome models

Estimation under Assumption 2 – Multiple imputation

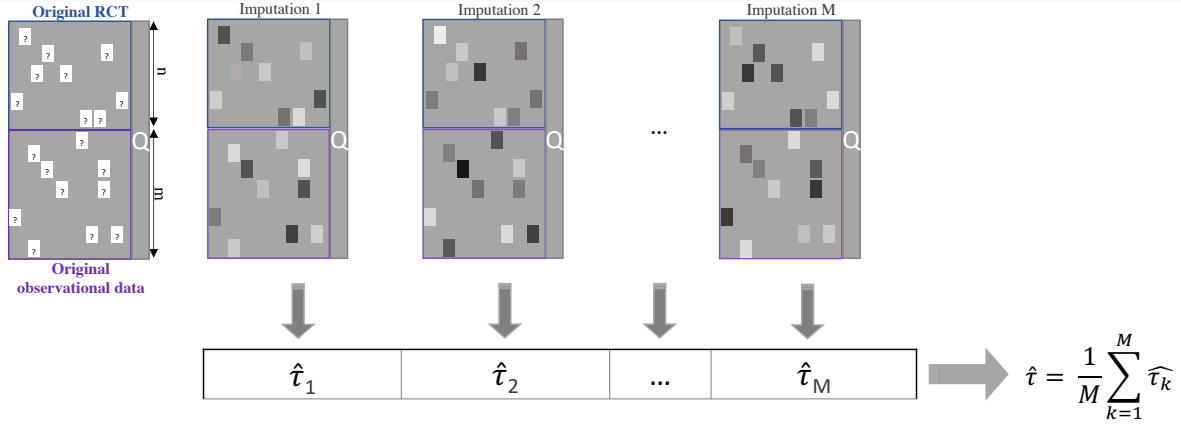
In case of integrative analysis, less straightforward.

We explore several strategies with different imputation models for the multi-source case:



Estimation under Assumption 2 – Multiple imputation

In case of integrative analysis, less straightforward.
We explore several strategies with different imputation models for the multi-source case:



$$Q_i \triangleq \mathcal{R}\mathbb{1}_{\{i \in RCT\}} + \mathcal{O}\mathbb{1}_{\{i \in Obs.\}}$$

Best performance in simulation study[24]: **joint fixed effect multiple imputation** (joint dataset, with source indicator Q).

[24] In terms of bias of the ATE estimator. Different scenarios varying $S-$ignorability, missingness mechanism and proportion, absolute and relative sample sizes.

Simulation study – Bias of $\hat{\tau}$

Under full data $S-$ignorability and *MCAR* & *MAR* mechanisms.

$n_{RCT} = n = 1000$, $n_{Obs} = m = 10 \times n$.

			Covariates				Treatment	Outcome
	S	Source	X_1	X_2	X_3	X_4	W	Y
1	1	RCT	NA	20	F	5	1	-166
.	1	RCT		...			:	:
.	1	RCT	-6	45	F	6	0	111
.	1	RCT	0	15	M	NA	1	-48
$n+1$	0	Obs.	-2	52	M	18		
.	0	Obs		...				
.	1	Obs	-1	NA	NA	1		
.								
$n+m$	0	Obs	-2	NA	M	32		

- Sample of size $50n$, $X_i \sim \mathcal{N}((1,1,1,1), \mathbb{I}_4)$.
- Generate S: $\text{logit}\{\pi_S(X)\} = \eta_0 - 0.5X_1 - 0.3X_2 - 0.5X_3 - 0.4X_4$, (where η_0 such that $\mathbb{E}[\pi_S(X)] = 1/50$). Keep $S = 1$ observations as RCT.
- Generate W: $W_i \sim \mathcal{B}(0.5)$.
- Generate $Y(w)$:

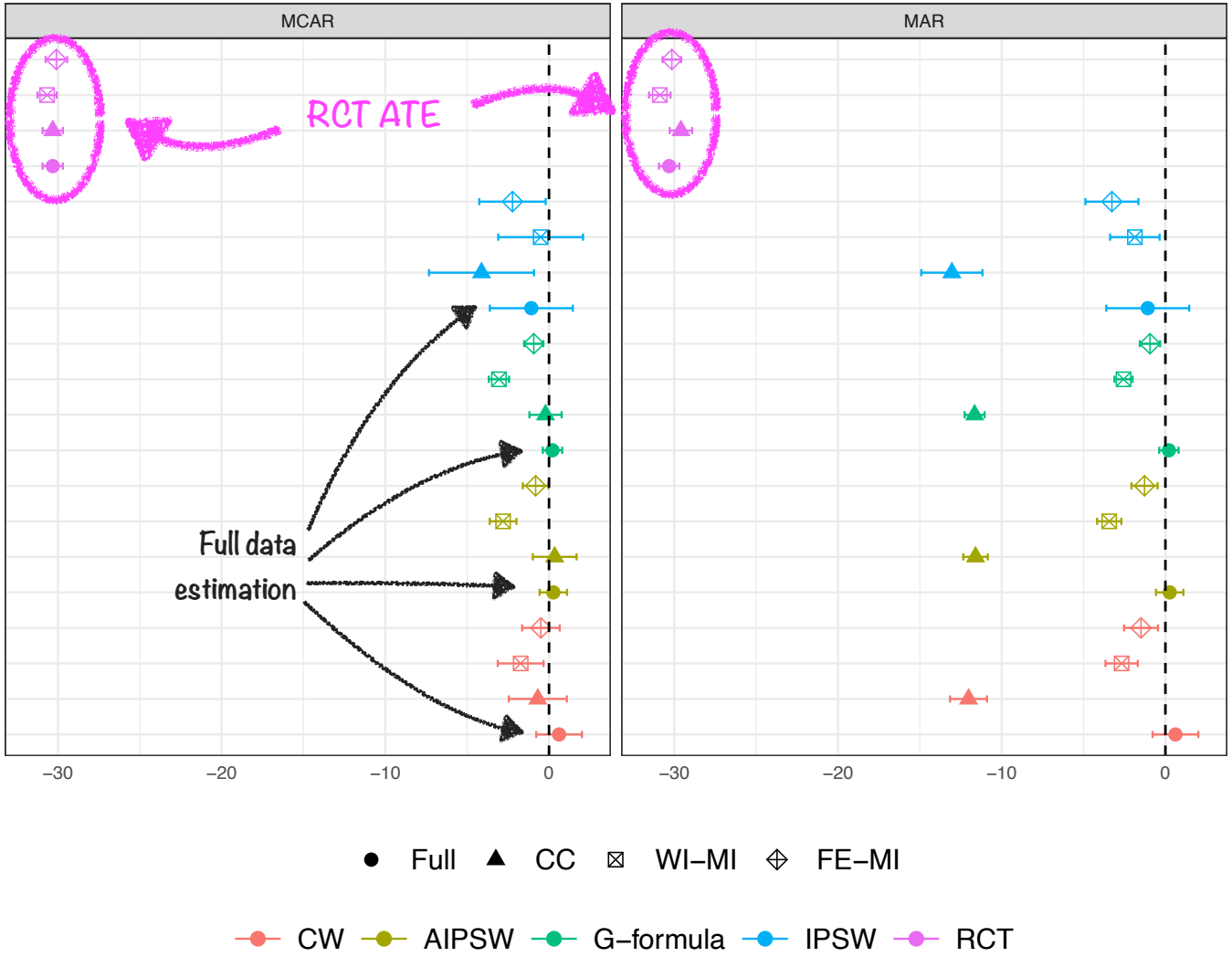
  $Y(w) = -100 + 27.4wX_1 + X_2 + 13.7X_3 + 13.7 + X_4 + \epsilon$ with $\epsilon \sim \mathcal{N}(0,1)$
- Sample of size m, $X_i \sim \mathcal{N}((1,1,1,1), \mathbb{I}_4)$ as observational data.
- Generate R for RCT and observational data under MCAR or MAR:

  $\text{logit}(P(R_{i\cdot} = r|X_i)) = \beta_0 + X_{i,obs(r)}\beta$, where β is chosen such that we have 30% of missing values in X.

Simulation study – Bias of $\hat{\tau}$

Under full data $S-$ignorability and *MCAR* & *MAR* mechanisms.

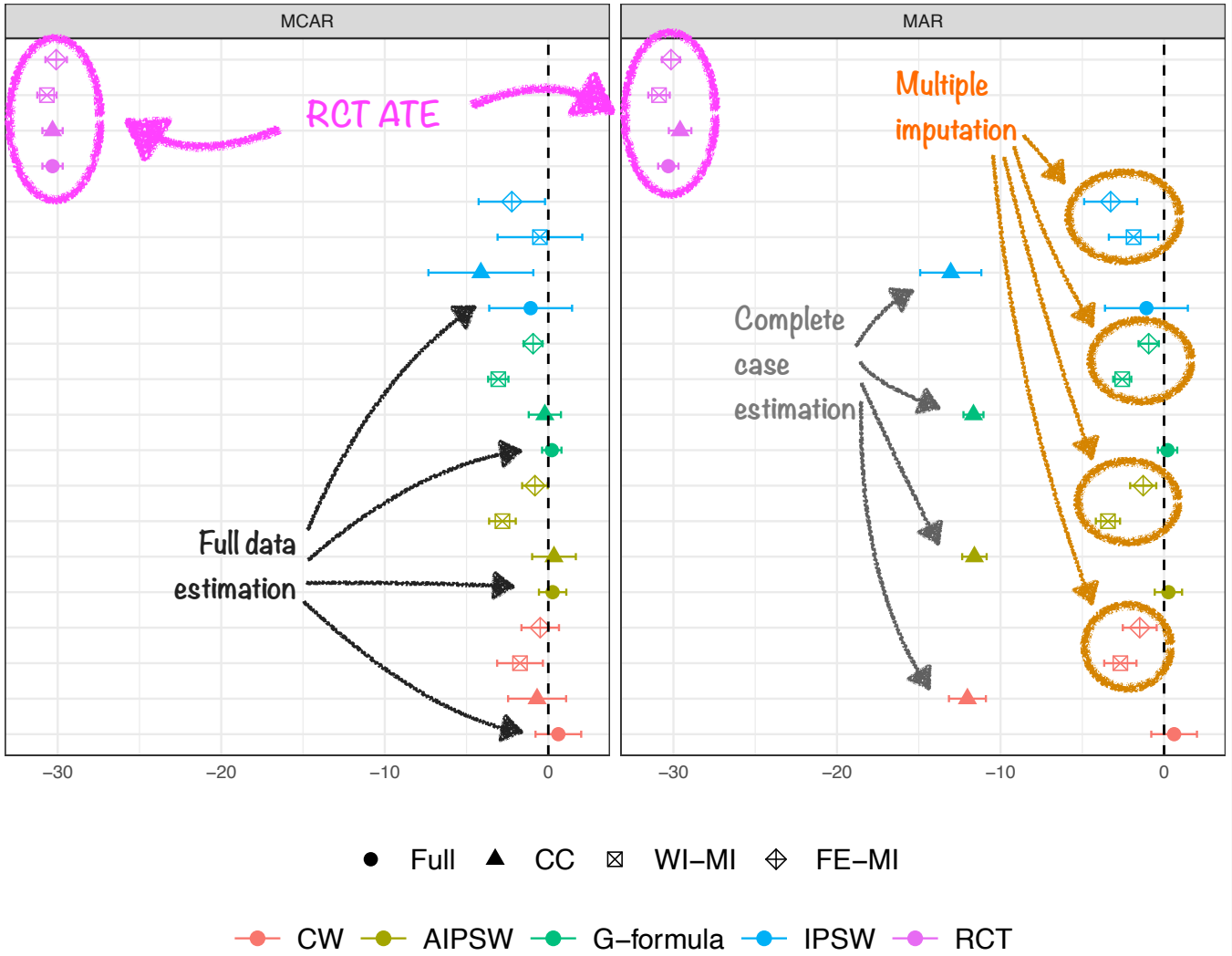
$n_{RCT} = n = 1000$, $n_{Obs} = m = 10 \times n$.



Simulation study – Bias of $\hat{\tau}$

Under full data $S-$ignorability and *MCAR* & *MAR* mechanisms.

$n_{RCT} = n = 1000$, $n_{Obs} = m = 10 \times n$.



Back to our Traumabase® and medical question...

Is there an effect of tranexamic acid (TXA) on mortality among patients with severe brain injuries (TBI)?

Randomized Controlled Trial
CRASH-2

- ▷ 40 different countries
- ▷ 3727 patients

Concludes on **beneficial effect of TXA for TBI with severe extracranial hemorrhage**.

Target population
Traumabase®

- ▷ 23 French Trauma centers
- ▷ 8270 patients

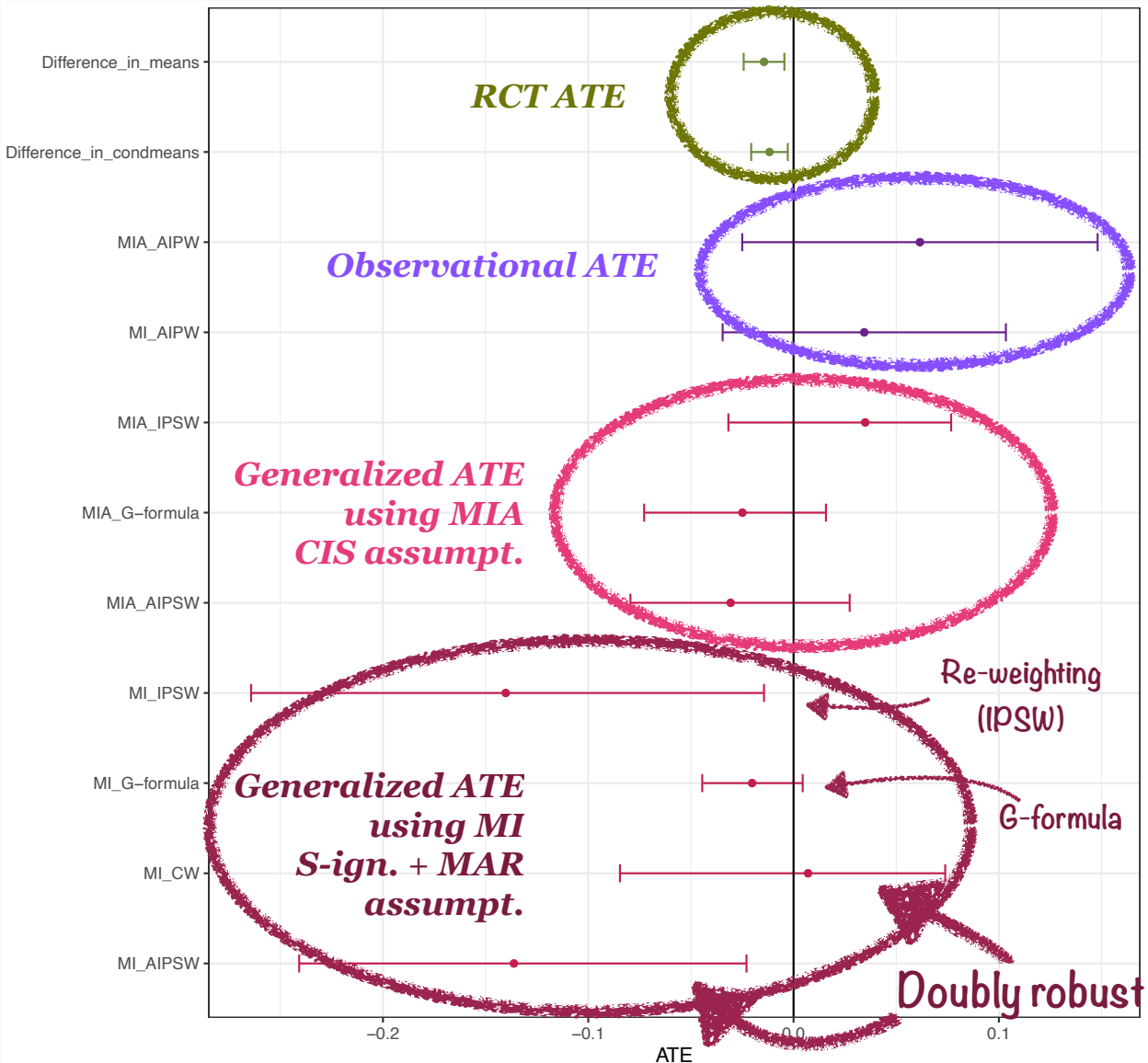
Concludes on **no significant effect of TXA for TBI**.

→ *Generalize the ATE from CRASH-2*[25] *to the Traumabase patients.*

[25] Shakur-Still et al., "Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): A randomised, placebo-controlled trial", 2009.

Back to our Traumabase® and medical question...

Is there an effect of tranexamic acid (TXA) on mortality among patients with severe brain injuries (TBI)?



General conclusion

Goal

Proposal of theoretical and methodological elements to reduce the gap between classical statistical analysis frameworks and real world data and application of the proposed solutions into practice.

General conclusion

Goal

Proposal of theoretical and methodological elements to reduce the gap between classical statistical analysis frameworks and real world data and application of the proposed solutions into practice.

Contributions

Study of the impact of missing values in causal analyses.

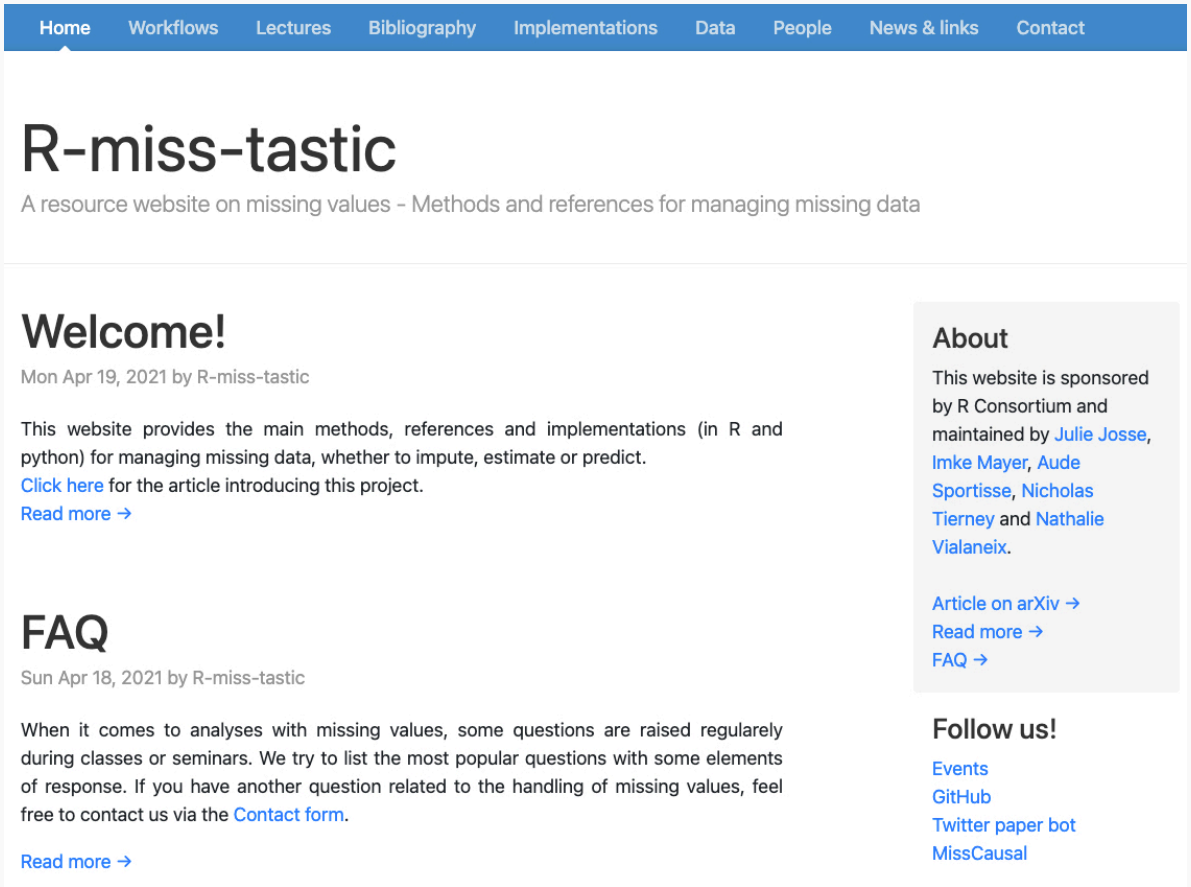
- ▷ Classification into classical and novel modeling of missing values in causal identifiability.
- ▷ Estimation on incomplete and heterogeneous observational data.
- ▷ Generalization from experimental data to target populations described by observational data.
- ▷ Implementation and application on critical care management data.

R-miss-tastic – A unified platform on missing values methods

Despite the large range of standard references for missing values problems, it is not (always) obvious where to go with a specific problem at hand.

A broad and accessible overview is given with the R-miss-tastic platform[25].

Joint work with Aude Sportisse, Nathalie Vialaneix, Julie Josse, Nick Tierney and many other contributors.



[25] Mayer et al., "R-miss-tastic: a unified platform for missing values methods and workflows", 2019.

Publications

Published articles and articles under review

Presented today

- Doubly robust treatment effect estimation with missing attributes, I. Mayer, E. Sverdrup, J.-D. Moyer, T. Gauss, S. Wager, J. Josse, *Annals of Applied Statistics*, 2020.
- Causal inference methods for combining experimental and observational studies: a review, B. Colnet, I. Mayer, G. Chen, A. Dieng, R. Li, G. Varoquaux, J.-P. Vert, J. Josse, S. Yang, *under review at Statistical Science*.
- Generalizing treatment effects with incomplete covariates, I. Mayer, J. Josse, Traumabase Group, *under review at Biometrical Journal*.
- R-miss-tastic: a unified platform for missing values methods and workflows, I. Mayer, A. Sportisse, J. Josse, N. Tierney, N. Vialaneix, *under review at R-Journal*.
- Machine Learning augmented causal inference to estimate the treatment effect of Tranexamic Acid in Traumatic Brain Injury, I. Mayer, J.-D. Moyer, J.-P. Nadal, J. Josse, T. Gauss, and others, *under review at BMC Research Methodology*.

Ongoing works and technical reports

- MissDeepCausal: causal inference from incomplete data using deep latent variable models, with J.-P. Vert, J. Josse.
- CRAN Task View on Causal Inference, with P. Zhao, J. Josse.
- Survival causal inference, with P. Roussel, J. Josse, B. Sebastien.
- HCQ with or without azithromycin and in-hospital mortality or discharge in patients hospitalized for COVID-19 infection: a cohort study of 4,642 in-patients in France, with E. Sbidian, E.

Perspectives

From a methodological point of view

- ▷ Propose sensitivity analyses to assess the different identifiability assumptions with missing values and quantify the bias of different estimators.
- ▷ Generalizing ATE with different missingness mechanisms in the RCT and the observational data.
- ▷ Extend the generalization results to target populations defined by combinations of populations represented by different observational cohorts.

From an applied/medical point of view

- ▷ Study treatment effect heterogeneity in TBI patients and compare with known patho-physiological heterogeneities.
- ▷ Provide easy-to-use tools (such as R package) to allow for direct deployment by practitioners.

$\rightarrow$ Towards translational (personalized) medicine.

Acknowledgements

Julie Josse

Jean-Pierre Nadal

Stefan Wager

Jean-Philippe Vert

Shu Yang

Tobias Gauss

Jean-Denis Moyer

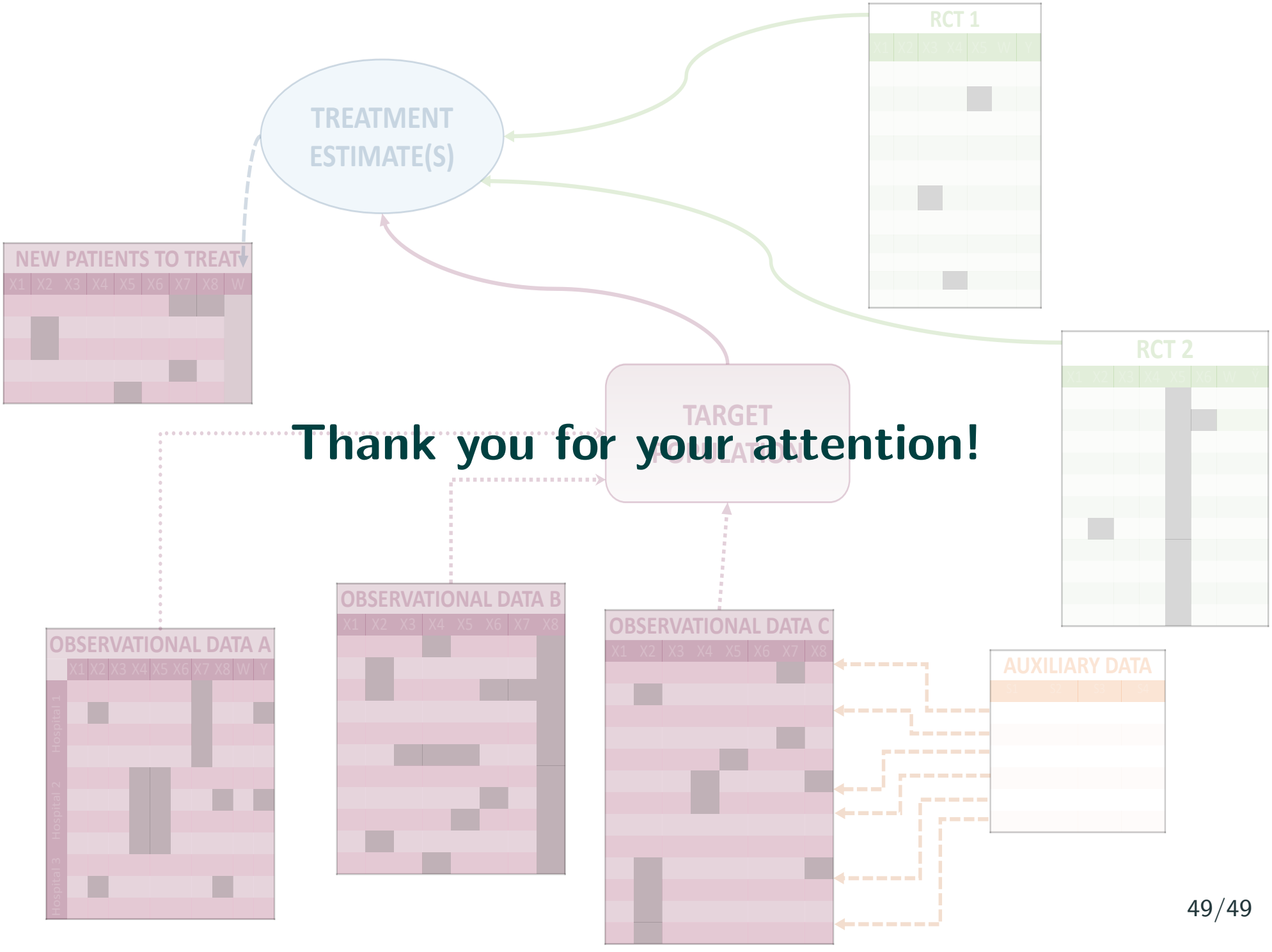
Nathalie Vialaneix

Bénédicte Colnet

Aude Sportisse

Aliénor Dreyfus

And many others!



Thank you for your attention!

References i

- Blake, Helen A. et al. "Estimating treatment effects with partially observed covariates using outcome regression with missing indicators". In: *Biometrical Journal* 62.2 (2020), pp. 428–443.
- Breiman, Leo et al. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984. ISBN: 9780412048418.
- Buuren, Stef van. *Flexible Imputation of Missing Data. Second Edition*. Boca Raton, FL: Chapman and Hall/CRC, 2018. URL: https://stefvanbuuren.name/fimd/.
- Cap, Andrew P. "CRASH-3: a win for patients with traumatic brain injury". In: *The Lancet* 394.10210 (2019), pp. 1687 –1688. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(19)32312-8. URL: http://www.sciencedirect.com/science/article/pii/S0140673619323128.

References ii

- Chen, Tianqi and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM. 2016, pp. 785–794.
- Chernozhukov, Victor et al. "Double/debiased machine learning for treatment and structural parameters". In: *The Econometrics Journal* 21.1 (2018), pp. C1–C68.
- Colnet, Bénédicte et al. "Causal inference methods for combining randomized trials and observational studies: a review". In: *arXiv preprint arXiv:2011.08047* (2020).
- Dagan, Noa et al. "BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting". In: *New England Journal of Medicine* 384.15 (2021), pp. 1412–1423.

References iii

- Dahabreh, Issa J. et al. “Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals”. In: *Biometrics* 75 (2019), pp. 685–694.
- Dong, Lin et al. “Integrative analysis of randomized clinical trials with real world evidence studies”. In: *arXiv preprint arXiv:2003.01242* (2020).
- Gauss, Tobias et al. “Strategic proposal for a national trauma system in France”. In: *Anaesthesia Critical Care & Pain Medicine* 38.2 (2019), pp. 121–130.
- Hay, Simon I. et al. “Global, regional, and national disability-adjusted life-years (DALYs): a systematic analysis for the Global Burden of Disease Study 2016”. In: *The Lancet* 390.10100 (2017), pp. 1260–1344.

References iv

- Hill, Jennifer L. "Bayesian nonparametric modeling for causal inference". In: *Journal of Computational and Graphical Statistics* 20.1 (2011), pp. 217–240.
- Jiang, Wei et al. "Logistic regression with missing covariates—Parameter estimation, model selection and prediction within a joint-modeling framework". In: *Computational Statistics & Data Analysis* 145 (2020), p. 106907.
- Jones, Jeremy and Duncan Hunter. "Consensus methods for medical and health services research.". In: *BMJ: British Medical Journal* 311.7001 (1995), p. 376.
- Josse, Julie et al. "On the consistency of supervised learning with missing values". In: *arXiv preprint* (2019). arXiv: `1902.06931 [stat.ML]`.
- Kallus, Nathan, Xiaojie Mao, and Madeleine Udell. "Causal Inference with Noisy and Missing Covariates via Matrix Factorization". In: *Advances in Neural Information Processing Systems*. 2018, pp. 6921–6932.

References v

- Louizos, Christos et al. "Causal effect inference with deep latent-variable models". In: *Advances in Neural Information Processing Systems*. 2017, pp. 6446–6456.
- Mattei, Alessandra and Fabrizia Mealli. "Estimating and using propensity score in presence of missing background data: an application to assess the impact of childbearing on wellbeing". In: *Statistical Methods and Applications* 18.2 (2009), pp. 257–273. DOI: 10.1007/s10260-007-0086-0.
- Mattei, Pierre-Alexandre and Jes Frellsen. "MIWAE: Deep Generative Modelling and Imputation of Incomplete Data Sets". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. Long Beach, California, USA: PMLR, 2019, pp. 4413–4423.

References vi

- Mayer, Imke et al. "Doubly robust treatment effect estimation with missing attributes". In: *Ann. Appl. Statist.* 14.3 (2020), pp. 1409–1431. ISSN: 1932-6157. DOI: 10.1214/20-AOAS1356.
- Mayer, Imke et al. "Machine Learning Augmented Causal Inference To Estimate The Treatment Effect of Tranexamic Acid In Traumatic Brain Injury". In: (2021).
- Mayer, Imke et al. *MissDeepCausal: Causal Inference from Incomplete Data Using Deep Latent Variable Models*. Tech. rep. 2020. arXiv: 2002.10837 [stat.ME].
- Mayer, Imke et al. "R-miss-tastic: a unified platform for missing values methods and workflows". In: *arXiv preprint arXiv:1908.04822* (2019).
- Morvan, Marine Le et al. "What's a good imputation to predict with missing values?" In: *arXiv preprint arXiv:2106.00311* (2021).
- Quinlan, JR. "PROBABILISTIC DECISION TREES". In: *Machine Learning: An Artificial Intelligence Approach* 3 (2014), p. 140.

References vii

- Robins, James M., Andrea Rotnitzky, and Lue P. Zhao. "Estimation of Regression Coefficients When Some Regressors are not Always Observed". In: *Journal of the American Statistical Association* 89.427 (1994), pp. 846–866. DOI: 10.1080/01621459.1994.10476818.
- Rosenbaum, Paul R. and Donald B. Rubin. "Reducing bias in observational studies using subclassification on the propensity score". In: *Journal of the American Statistical Association* 79.387 (1984), pp. 516–524. DOI: 10.2307/2288398.
- Rubiano, Andrés M et al. "Global neurotrauma research challenges and opportunities". In: *Nature* 527.7578 (2015), S193–S197.
- Rubin, Donald B. "Bayesian inference for causal effects: The role of randomization". In: *The Annals of statistics* (1978), pp. 34–58. DOI: 10.1214/aos/1176344064.
- – . "Inference and missing data". In: *Biometrika* 63.3 (1976), pp. 581–592.

References viii

- Seaman, Shaun R. and Ian R. White. "Inverse probability weighting with missing predictors of treatment assignment or missingness". In: *Communications in Statistics-Theory and Methods* 43.16 (2014), pp. 3499–3515. DOI: 10.1080/03610926.2012.700371.
- Shakur-Still, Haleema et al. "Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): A randomised, placebo-controlled trial". In: *Lancet* 376 (Nov. 2009), pp. 23–32.
- Textor, Johannes, Juliane Hardt, and Sven Knüppel. "DAGitty: a graphical tool for analyzing causal diagrams". In: *Epidemiology* 22.5 (2011), p. 745.
- Tibshirani, Julie et al. *grf: Generalized Random Forests*. R package version 1.1.0. 2020. URL: https://github.com/grf-labs/grf.


Twala, BETH, MC Jones, and David J Hand. "Good methods for coping with missing data in decision trees". In: *Pattern Recognition Letters* 29.7 (2008), pp. 950–956.


Wager, Stefan and Susan Athey. "Estimation and inference of heterogeneous treatment effects using random forests". In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1228–1242.