

Treatment effect estimation with missing attributes

Effect of tranexamic acid on patients with traumatic brain injury

Imke Mayer

EHESS; École Polytechnique

Séminaire Palaisien, Gif-sur-Yvette, 7 jan 2020

Introduction

Collaborators

Julie Josse (X, Inria), Jean-Pierre Nadal (ENS-EHESS), Stefan Wager (Stanford), Jean-Philippe Vert (Google Brain)

Traumabase (APHP): Tobias Gauss, Jean-Denis Moyer, Sophie Hamada
Capgemini Invent



INSTITUT
POLYTECHNIQUE
DE PARIS

L'ECOLE
DES HAUTES
ETUDES EN
SCIENCES SOCIALES

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Traumabase and TrauMatrix

- 20,000 patients
- 250 continuous and categorical variables: **heterogeneous**
- 16 hospitals: **multilevel data**
- 4,000 new patients/ year

Center	Accident	Age	Sex	Weight	Lactates	BP	shock	...
Beaujon	fall	54	m	85	NM	180	yes	
Pitie	gun	26	m	NR	NA	131	no	
Beaujon	moto	63	m	80	3.9	145	yes	
Pitie	moto	30	w	NR	Imp	107	no	
HEGP	knife	16	m	98	2.5	118	no	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

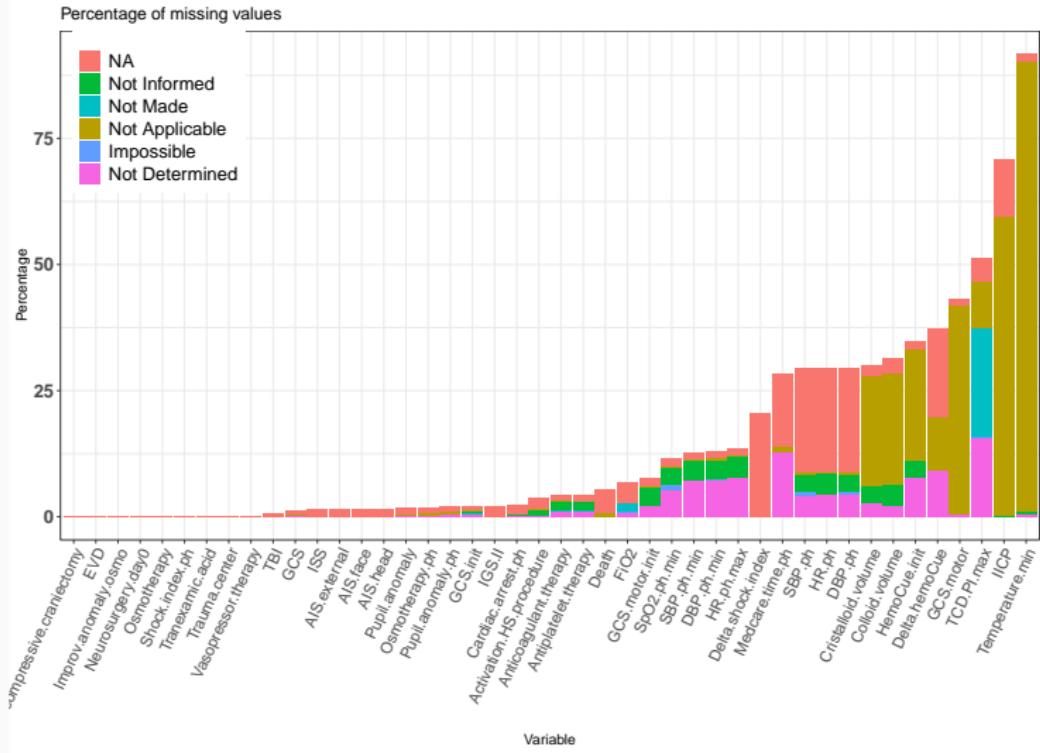
Traumabase and TrauMatrix

- 20,000 patients
- 250 continuous and categorical variables: **heterogeneous**
- 16 hospitals: **multilevel data**
- 4,000 new patients/ year

Center	Accident	Age	Sex	Weight	Lactates	BP	shock	...
Beaujon	fall	54	m	85	NM	180	yes	
Pitie	gun	26	m	NR	NA	131	no	
Beaujon	moto	63	m	80	3.9	145	yes	
Pitie	moto	30	w	NR	Imp	107	no	
HEGP	knife	16	m	98	2.5	118	no	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

⇒ **Estimate causal effect:** Administration of the **treatment** "tranexamic acid" (within 3 hours after the accident) on the **outcome** mortality for traumatic brain injury patients

Missing values



Causal inference: classical framework

Potential outcome framework (Neyman, 1923, Rubin, 1974)

Causal effect

Binary treatment $w \in \{0, 1\}$ on i -th individual with potential outcomes $Y_i(1)$ and $Y_i(0)$. Individual causal effect of the treatment:

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

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

Covariates			Treatment	Outcome(s)	
X_1	X_2	X_3	W	$Y(0)$	$Y(1)$
1.1	20	F	1	NA	T
-6	45	F	0	F	NA
0	15	M	1	NA	F

-2	52	M	0	T	NA

Potential outcome framework (Neyman, 1923, Rubin, 1974)

Causal effect

Binary treatment $w \in \{0, 1\}$ on i -th individual with potential outcomes $Y_i(1)$ and $Y_i(0)$. Individual causal effect of the treatment:

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

- Problem: Δ_i never observed (only observe one outcome/indiv). Causal inference as a missing value pb?
 - **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.
- ⇒ First solution: estimate τ with randomized controlled trials (RCT).

Observational data

Non random assignment → Confounding

Mortality rate 20% - among treated 46% - among not treated 18%: treatment kills?

	survived	deceased	Pr(survived treatment)	Pr(deceased treatment)
TA not administered	6,238 (76%)	1,327 (16%)	0.82	0.18
TA administered	367 (4%)	316 (4%)	0.54	0.46

Table 1: Occurrence and frequency table for traumatic brain injury patients (total number: 8,248).

Unconfoundedness and the propensity score

Assumptions

- n iid samples $(X_i, Y_i(0), Y_i(1), W_i)$,
- Treatment assignment is random conditionally on X_i :
 $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i | X_i \equiv \text{unconfoundedness assumption.}$

Propensity score and overlap assumption

$$e(x) \triangleq \mathbb{P}(W_i = 1 | X_i = x) \quad \forall x \in \mathcal{X}.$$

We will assume overlap, i.e. $0 < e(x) < 1 \quad \forall x \in \mathcal{X}.$

Key property

e is a **balancing score**, i.e. under unconfoundedness, it satisfies

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

Propensity based estimators

Inverse Propensity Weighted estimator

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

- ⇒ Balance the differences between the two groups
- ⇒ Consistent estimator of τ as long as $\hat{e}(\cdot)$ is consistent.

Propensity based estimators

Inverse Propensity Weighted estimator

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

- ⇒ Balance the differences between the two groups
- ⇒ Consistent estimator of τ as long as $\hat{e}(\cdot)$ is consistent.

Augmented IPW: a doubly robust estimator

Define $\mu_{(w)}(x) := \mathbb{E}[Y_i(w) | X_i = x]$.

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

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

Semiparametric efficiency for ATE estimation, Double ML

Efficient score estimator

Given unconfoundedness ($\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i | X_i$) but no further parametric assumptions on $\mu_{(w)}(x)$ and $e(x)$, the asymptotic variance,

$$V^* := \text{Var}(\tau(X)) + \mathbb{E} \left[\frac{\sigma^2(X)}{e(X)(1 - e(X))} \right],$$

is optimal and any estimator τ^* that attains it is asymptotically equivalent to $\hat{\tau}_{AIPW^*}$.

V^* is the semiparametric efficient variance for ATE estimation.

Extension: Double Machine Learning (Chernozhukov et al., 2018)

Possibility 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.

Many different proposals of such doubly robust estimators (CBPS, R-learner, causal forest, etc.). R packages `grf` (Athey et al., 2019) and `rlearner` (Nie and Wager, 2017).

Causal inference: with missing attributes?

Unconfoundedness with missing attributes?

Without any changes to the previous framework, the only straightforward – but generally biased – solution is complete-case analysis.

Covariates			Treatment	Outcome
X_1	X_2	X_3	W	Y
NA	20	F	1	T
-6	45	NA	0	F
0	NA	M	1	F
NA	32	F	1	T
1	63	M	1	F
-2	NA	M	0	T

Unconfoundedness with missing attributes?

Without any changes to the previous framework, the only straightforward
– but generally biased – solution is complete-case analysis.

Covariates			Treatment	Outcome
X_1	X_2	X_3	W	Y
NA	20	F	1	→ T
6	45	NA	0	→ F
0	NA	M	1	→ F
NA	32	F	1	→ T
1	63	M	1	F
2	NA	M	0	→ T

Unconfoundedness with missing attributes?

Without any changes to the previous framework, the only straightforward – but generally biased – solution is complete-case analysis.

→ Often not a good idea! What are the alternatives?

Three families of methods

- Unconfoundedness despite missingness
- Latent unconfoundedness
- Classical missing values mechanisms

1. Unconfoundedness despite missingness

Unconfoundedness despite missingness

Adapt the initial assumptions s.t. treatment assignment is unconfounded given only the **observed** covariates and the **response pattern**.

1. Unconfoundedness despite missingness

Notations

- response pattern $R \in \{NA, 1\}^P$, $R_j \triangleq \mathbb{1}_{\{X_j \text{ is observed}\}} + NA \mathbb{1}_{\{X_j \text{ is missing}\}}$,
- $\mathbf{X}^* = \mathbf{R} \odot \mathbf{X} \in \{\mathbb{R} \cup NA\}^P$

Unconfoundedness despite missingness

Treatment is unconfounded given X^* :

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

or alternatively:

$$\left\{ \begin{array}{ll} \{Y_i(1), Y_i(0)\} \perp\!\!\!\perp W_i | X_i, R_i, \\ \text{CIT: } \quad W_i \perp\!\!\!\perp X_i | X_i^*, R_i \\ \text{or} \\ \text{CIO: } \quad Y_i(w) \perp\!\!\!\perp X_i | X_i^*, R_i \quad \text{for } w \in \{0, 1\} \end{array} \right. \quad (2)$$

1. Unconfoundedness despite missingness

Unconfoundedness despite missingness

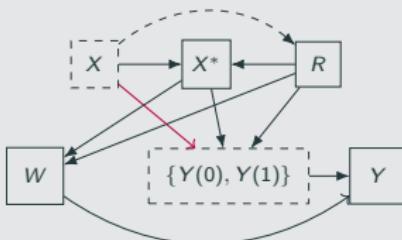
Treatment is unconfounded given X^* :

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

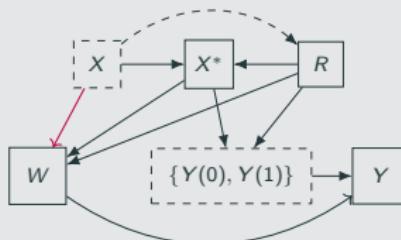
or alternatively:

$$\left\{ \begin{array}{ll} \text{CIT:} & W_i \perp\!\!\!\perp X_i | X_i^*, R_i \\ \text{or} & \\ \text{CIO:} & Y_i(w) \perp\!\!\!\perp X_i | X_i^*, R_i \quad \text{for } w \in \{0, 1\} \end{array} \right. \quad (2)$$

(a) CIT



(b) CIO



Under 1: Generalized propensity score and random forests

Generalized propensity score (Rosenbaum and Rubin, 1984)

$$e^*(X^*) = \mathbb{P}(W = 1 | X^*)$$

Under 1: Generalized propensity score and random forests

Generalized propensity score (Rosenbaum and Rubin, 1984)

$$e^*(X^*) = \mathbb{P}(W = 1 | X^*)$$

Random forest models

They allow incorporating missing values directly since they allow semi-discrete variables (e.g. $X^* \in (\mathbb{R} \times \text{NA})^P$).

- With mean imputation or specific representation/encoding of missing values (*MIA*, [more details](#):14), splits are possible either on observed variables or on response pattern ([Josse et al., 2019](#)).
- This procedure targets the Bayes estimate:

$$\mathbb{E}[W|X^*] = \sum_{r \in \{0,1\}^P} \mathbb{E}[W|X^*, R = r] \mathbb{1}_{R=r}.$$

- Only valid for **predictive** tasks!
- Handles **general missingness mechanisms**.

Missing incorporated in attribute (Twala et al., 2008)

CART with missing values

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

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

with, for each feature j and each threshold z , there are three possible partitions,

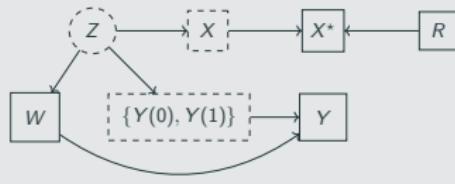
$$\begin{array}{lll} \{X_j^* \leq z \text{ or } X_j^* = \text{NA}\} & \text{VS} & \{X_j^* > z\} \\ \{X_j^* \leq z\} & \text{VS} & \{X_j^* > z \text{ or } X_j^* = \text{NA}\} \\ \{X_j^* \neq \text{NA}\} & \text{VS} & \{X_j^* = \text{NA}\} \end{array}$$

→ targets $\mathbb{E}[W|\mathbf{X}^*]$
(≡ generalized propensity score)

2. Latent unconfoundedness

Latent confounding assumption

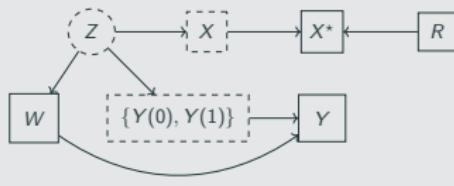
The covariates are **noisy incomplete proxies** of the true **latent confounders** (Kallus et al., 2018; Louizos et al., 2017).



2. Latent unconfoundedness

Latent confounding assumption

The covariates are **noisy incomplete proxies** of the true **latent confounders** (Kallus et al., 2018; Louizos et al., 2017).



MissDeepCausal (Josse et al., 2020, MDC)

Assume MAR mechanism, an unbiased estimator $\hat{f}(Z)$ of $\mathbb{E}[Y(1) - Y(0)|Z]$, and access to the distribution $P(Z|X^*)$. Then we can consistently estimate the $\tau = \mathbb{E}[\tau(Z)]$.

(Since $\mathbb{E}[Y(1) - Y(0)|X^*] = \mathbb{E}[\mathbb{E}[Y(1) - Y(0)|Z|X^*]]$)

Under 2: MissDeepCausal

Assume an unbiased estimator $\hat{f}(Z)$ of $\mathbb{E}[Y(1) - Y(0)|Z]$ and access to the distribution $P(Z|X^*)$

Latent variables estimation as a pre-processing step

- Heuristic nonlinear extension of Kallus et al. (2018)
- Regression model:
$$Y = \tau W + Z\beta + \varepsilon,$$

$$\varepsilon \sim \mathcal{N}(0, \sigma^2 I).$$

MDC-process

1. Estimate latent confounders with
$$\hat{Z}(x^*) = \mathbb{E}[Z|X^* = x^*].$$
2. Plug these $\hat{Z}(x^*)$ into regression model or define
$$\hat{\tau}_{process} = \mathbb{E}[f(\mathbb{E}[Z|X^*])].$$

Multiple imputation strategy

- Monte-Carlo approximation using posterior distribution $P(Z|X^*)$.

MDC-MI

1. Sample $(Z^{(j)})_{1 \leq j \leq B}$ from $\hat{P}(Z|X^*)$.
2. For each sample j , compute estimate $\hat{\tau}^{(j)} = f(Z^{(j)})$.
3. Aggregate into final estimate:
$$\hat{\tau}_{MI} = \frac{1}{B} \sum_{j=1}^B \hat{\tau}^{(j)} \approx \mathbb{E}[\mathbb{E}[f(Z|X^*)]].$$

MDC-process

1. Estimate latent confounders with $\hat{Z}(x^*) = \mathbb{E}[Z|X^* = x^*]$.
2. Plug these $\hat{Z}(x^*)$ into regression model or define
 $\hat{\tau}_{process} = \mathbb{E}[f(\mathbb{E}[Z|X^*])]$.

MDC-MI

1. Sample $(Z^{(j)})_{1 \leq j \leq B}$ from $\hat{P}(Z|X^*)$.
2. For each sample j , compute estimate $\hat{\tau}^{(j)} = f(Z^{(j)})$.
3. Aggregate into final estimate:
$$\hat{\tau}_{MI} = \frac{1}{B} \sum_{j=1}^B \hat{\tau}^{(j)} \approx \mathbb{E}[\mathbb{E}[f(Z|X^*)]]$$
.

Estimation of and sampling from $P(Z|X^*)$:

- 1) Use **missing data importance weight autoencoder** (Mattei and Frellsen, 2019, MIWAE): imputation by a constant maximizes the ELBO.
- 2) Approximate with **self-normalized importance sampling** on variational distribution $Q(Z|X^*)$:

$$\mathbb{E}[s(Z)|X^*] \approx \sum_{l=1}^L w_l s(Z^{(l)}),$$

where $w_l = \frac{r_l}{r_1 + \dots + r_L}$

and $r_l = \frac{p(X^*|Z^{(l)})p(Z^{(l)})}{q(Z^{(l)}|X^*)}$ for any measurable function s .

Methods for 1, 2, 3

	Confounding & Covariates		Missingness		Unconfoundedness			Models for (W, Y)	
	multiva- riate normal	general	M(C)AR	general	1	2	3	logistic- linear	non- param.
(SA)EM	✓	✗	✓	✗	✓	✗	✗	✓	✗
MIA.GRF	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mean.GRF	✓	✓	✓	✓	✓	✗	✗	✓	✓
Mult. Imp.	✓	✓	✓	✗	✓	✗	✓	✓	(✗)
Matrix Fact.	✓	✗	✓	✗	✗	✓	✗	✓	(✗)
MDC	✓	✓	✓	✗	✗	✓	✗	✓	(✗)

Methods and their assumptions on the underlying data generating process. (✓ indicates cases that can be handled by a method, whereas ✗ marks cases where a method is not applicable in theory; (✗) indicates cases without theoretical guarantees but with heuristic solutions.)

Simulations: importance of unconfoundedness assumption and choice of estimator

Setup

- Different data generating models (linear, nonlinear, latent, etc.)
- Different missingness mechanisms

Results

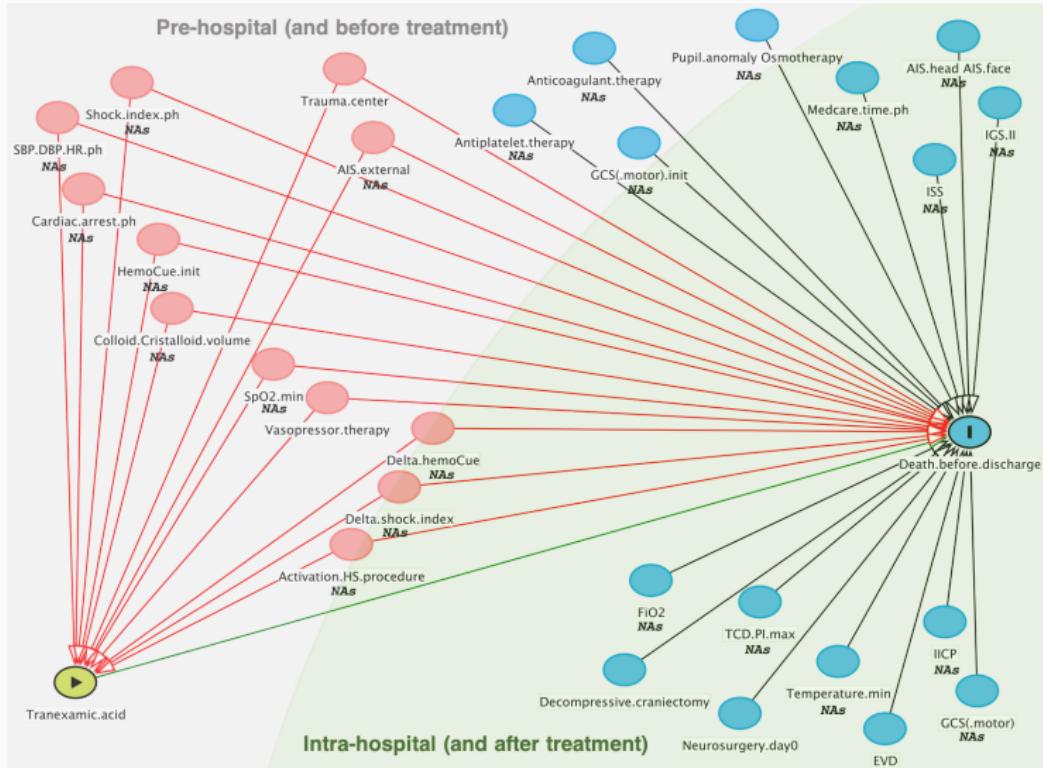
- AIPW estimators outperform their IPW counterparts.
- w/o appropriate unconfoundedness assumption, methods are biased.
- $\hat{\tau}_{mia}$ unbiased for all missingness mechanisms, especially for MNAR.
- Multiple imputation (mice) only requires standard unconfoundedness, but needs MAR.
- MDC-MI and MDC-process unbiased in case of linear and nonlinear latent confounding.

+ test on IHDP dataset (Hill, 2011)

MDC performs close to state-of-the-art on this benchmark dataset.

Application: Traumabase

Control for confounding

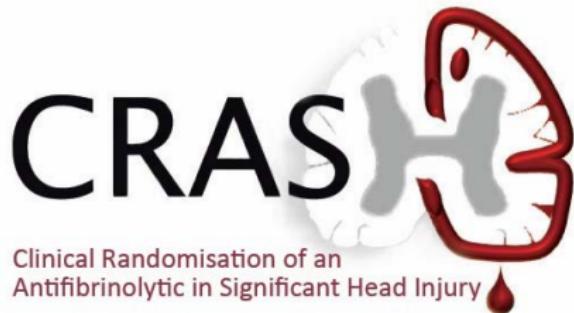


Graph produced using DAGitty ([Textor et al. \(2011\)](#))

Plausibility of underlying assumptions with Traumabase data

- Overlap: cannot be tested but high level of uncertainty at diagnosing severe (internal bleeding) makes it likely
- Unconfoundedness despite missingness: seems most plausible (physicians decide based on what they observe+record)
- Many variables have missing non at random data.

Tranexamic Acid & CRASH



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



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

The CRASH-3 trial collaborators*



Summary

Background Tranexamic acid reduces surgical bleeding and decreases mortality in patients with traumatic extracranial bleeding. Intracranial bleeding is common after traumatic brain injury (TBI) and can cause brain herniation and death. We aimed to assess the effects of tranexamic acid in patients with TBI.

Methods This randomised, placebo-controlled trial was done in 175 hospitals in 29 countries. 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 were eligible. The time window for eligibility was originally 8 h but in 2016 the protocol was changed to limit recruitment to patients within 3 h of injury. This change was made blind to the trial data, in response to external evidence suggesting that delayed treatment is unlikely to be effective. We randomly assigned (1:1) patients to receive tranexamic acid (loading dose 1 g over 10 min then infusion of 1 g over 8 h) or matching placebo. Patients were assigned by selecting a numbered treatment pack from a box containing eight packs that were identical apart from the pack number. Patients, caregivers, and those assessing outcomes were masked to allocation. The primary outcome was head injury-related death in hospital within 28 days of injury in patients treated within 3 h of injury. We prespecified a sensitivity analysis that excluded patients with a GCS score of 3 and those with bilateral unreactive pupils at baseline. All analyses were done by intention to treat. This trial was registered with ISRCTN (ISRCTN1508812), ClinicalTrials.gov (NCT01402882), EudraCT (2011-003669-14), and the Pan African Clinical Trial Registry (PACTR20121000441277).

Results Between July 20, 2012, and Jan 31, 2019, we randomly allocated 12 737 patients with TBI to receive tranexamic acid (6406 [50·3%] or placebo [6331 [49·7%], of whom 9202 (72·2%) patients were treated within 3 h of injury. Among patients treated within 3 h of injury, the risk of head injury-related death was 18·5% in the tranexamic acid group versus 19·8% in the placebo group (855 vs 892 events; risk ratio [RR] 0·94 [95% CI 0·86–1·02]). In the prespecified sensitivity analysis that excluded patients with a GCS score of 3 or bilateral unreactive pupils at baseline, the risk of head injury-related death was 12·5% in the tranexamic acid group versus 14·0% in the placebo group (485 vs 525 events; RR 0·89 [95% CI 0·80–1·00]). The risk of head injury-related death reduced with tranexamic acid in patients with mild-to-moderate head injury (RR 0·78 [95% CI 0·64–0·95]) but not in patients with severe head injury (0·99 [95% CI 0·91–1·07]; p value for heterogeneity 0·030). Early treatment was more effective than was later treatment in patients with mild and moderate head injury ($p=0\cdot005$) but time to treatment

Published Online
October 14, 2019
[https://doi.org/10.1016/S0140-6736\(19\)32233-0](https://doi.org/10.1016/S0140-6736(19)32233-0)

See Online/Comment
[https://doi.org/10.1016/S0140-6736\(19\)32312-8](https://doi.org/10.1016/S0140-6736(19)32312-8)

*Members listed at end of paper
For the Arabic translation of the abstract see Online for appendix 1
For the Chinese translation of the abstract see Online for appendix 2

For the French translation of the abstract see Online for appendix 3
For the Hindi translation of the abstract see Online for appendix 4

For the Japanese translation of the abstract see Online for appendix 5
For the Spanish translation of the abstract see Online for appendix 6

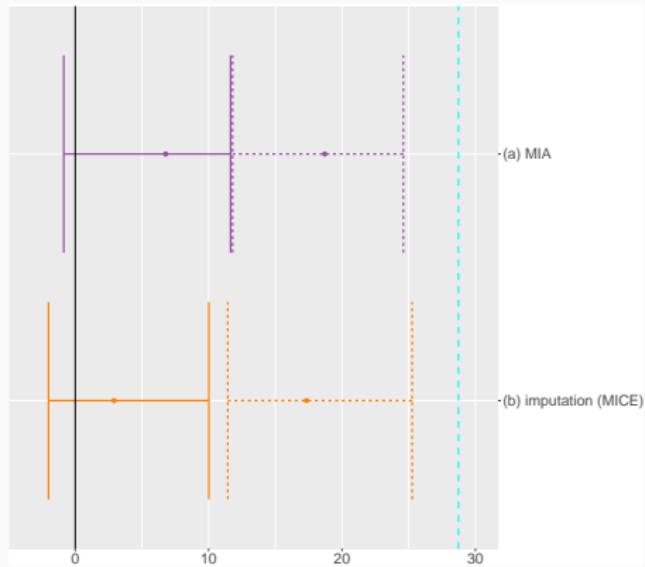
For the Urdu translation of the abstract see Online for appendix 7

Correspondence to:
Clinical Trials Unit, London

Results

40 covariates, 18 confounders. 8,248 patients.

ATE estimations ($\times 100$) for the effect of tranexamic acid on in-ICU mortality for TBI patients.



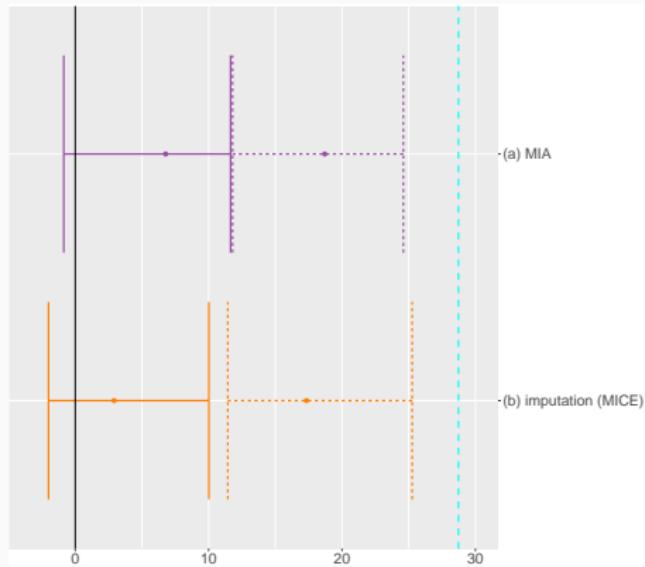
(y-axis: estimation approach, solid: DR, dotted: IPW, turquoise: without adjustment), (x-axis: ATE estimation with bootstrap CI)

We compute the mortality rate in the treated group and the mortality rate in the control group (after covariate balancing). The obtained value corresponds to the **difference in percentage points between mortality rates in treatment and control**.

Results

40 covariates, 18 confounders. 8,248 patients.

ATE estimations ($\times 100$) for the effect of tranexamic acid on in-ICU mortality for TBI patients.



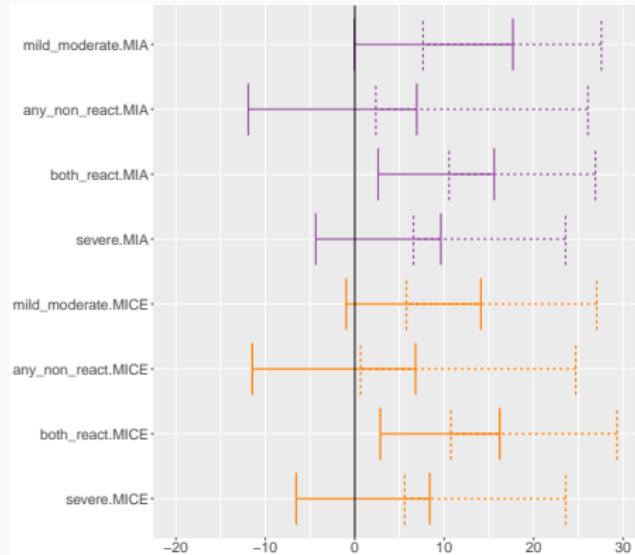
(y-axis: estimation approach, solid: DR, dotted: IPW, turquoise: without adjustment), (x-axis: ATE estimation with bootstrap CI)

Comparison with CRASH-3 study (Cap, 2019): same conclusion of “no average treatment effect”.

Results

40 covariates, 18 confounders. 8,248 patients.

ATE estimations on stratified population for the effect of tranexamic acid on in-ICU mortality for TBI patients.



(y-axis: estimation approach, solid: **DR**, dotted: **IPW**), (x-axis: ATE estimation with approx. asymptotic CI)

Comparison with CRASH-3 study ([Cap, 2019](#)): different conclusions on strata.

Conclusion

Conclusion and perspectives

Take-away messages

- **Missing attributes** alter causal analyses.
- Additional assumptions on appropriate **unconfoundedness**.
- Two new proposals to **handle missing values in causal inference**.
- Prefer **AIPW** to IPW estimators, in theory and in practice.
- Applicable on real data.

Conclusion and perspectives

Take-away messages

- **Missing attributes** alter causal analyses.
- Additional assumptions on appropriate **unconfoundedness**.
- Two new proposals to **handle missing values in causal inference**.
- Prefer **AIPW** to IPW estimators, in theory and in practice.
- Applicable on real data.

Further details in original paper (submitted) and working paper

M., Wager, S., Gauss, T., Moyer, J.-D. & Josse, J. (2019). Doubly robust treatment effect estimation with missing attributes. arXiv:1910.10624.

Josse, J., M. & Vert, J.-P. (2020). MissDeepCausal: causal inference from incomplete data using deep latent variable models.

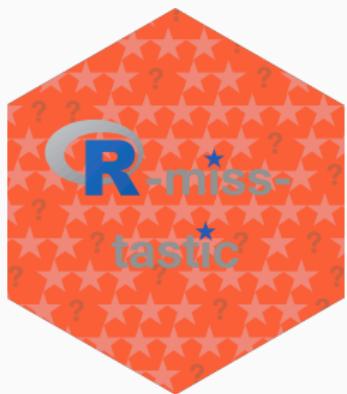
Ongoing work

- Coupling of observational data and RCT data (Yang and Ding, 2019).
- Matching under unconfoundedness despite missingness.
- Heterogeneous treatment effects (Athey and Imbens, 2015) and optimal policy learning (e.g., Imai and Ratkovic, 2013).

Missing value website

"One of the ironies of Big Data is that missing data play an ever more significant role" (R. Samworth, 2019)

More information and details on missing values: **R-miss-tastic** platform.



→ Theoretical and practical

tutorials, popular datasets, bibliography, workflows (in R and in python), active contributors/researchers in the community, etc.

rmisstastic.netlify.com

M., Josse, J., Tierney, N., & Vialaneix, N. (2019). R-miss-tastic: a unified platform for missing values methods and workflows. arXiv preprint arXiv:1908.04822.

MERCI

References

References i

- Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.
- Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.
- Cap, A. P. (2019). Crash-3: a win for patients with traumatic brain injury. *The Lancet*.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794. ACM.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.
- Hill, J. (2004). Reducing bias in treatment effect estimation in observational studies suffering from missing data. Technical report, Institute for Social and Economic Research and Policy, Columbia University.
- Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.
- Josse, J., Mayer, I., and Vert, J.-P. (2020). Missdeepcausal: causal inference from incomplete data using deep latent variable models.

References ii

- Josse, J., Prost, N., Scornet, E., and Varoquaux, G. (2019). On the consistency of supervised learning with missing values. *arXiv preprint*.
- Kallus, N., Mao, X., and Udell, M. (2018). Causal inference with noisy and missing covariates via matrix factorization. In *Advances in Neural Information Processing Systems*, pages 6921–6932.
- Leyrat, C., Seaman, S. R., White, I. R., Douglas, I., Smeeth, L., Kim, J., Resche-Rigon, M., Carpenter, J. R., and Williamson, E. J. (2019). Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Statistical methods in medical research*, 28(1):3–19.
- Louizos, C., Shalit, U., Mooij, J. M., Sontag, D., Zemel, R., and Welling, M. (2017). Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems*, pages 6446–6456.
- Mattei, P.-A. and Frellsen, J. (2019). MIWAE: Deep generative modelling and imputation of incomplete data sets. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4413–4423, Long Beach, California, USA. PMLR.
- Nie, X. and Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects. *arXiv preprint arXiv:1712.04912*.
- Quinlan, J. (2014). Probabilistic decision trees. *Machine Learning: An Artificial Intelligence Approach*, 3:140.

References iii

- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.
- Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.
- Seaman, S. and White, I. (2014). Inverse probability weighting with missing predictors of treatment assignment or missingness. *Communications in Statistics-Theory and Methods*, 43(16):3499–3515.
- Textor, J., Hardt, J., and Knüppel, S. (2011). Dagitty: a graphical tool for analyzing causal diagrams. *Epidemiology*, 22(5):745.
- Twala, B., Jones, M., and Hand, D. J. (2008). Good methods for coping with missing data in decision trees. *Pattern Recognition Letters*, 29(7):950–956.