Teresa Alves de Sousa – Rayan Charrier – Sofiane Hadji – Randy Kotti – Ilyas Malik

traumabase.eu

Effect of a treatment using causal inference

DECEMBER, TUESDAY 11TH

Introduction

Traumabase

- Data : from 7000 patients, 250 variables to 3000 patients, 35 variables
- Goal : perform causal inference to assess the effect of a treatment on the mortality of patients with head trauma

Introduction

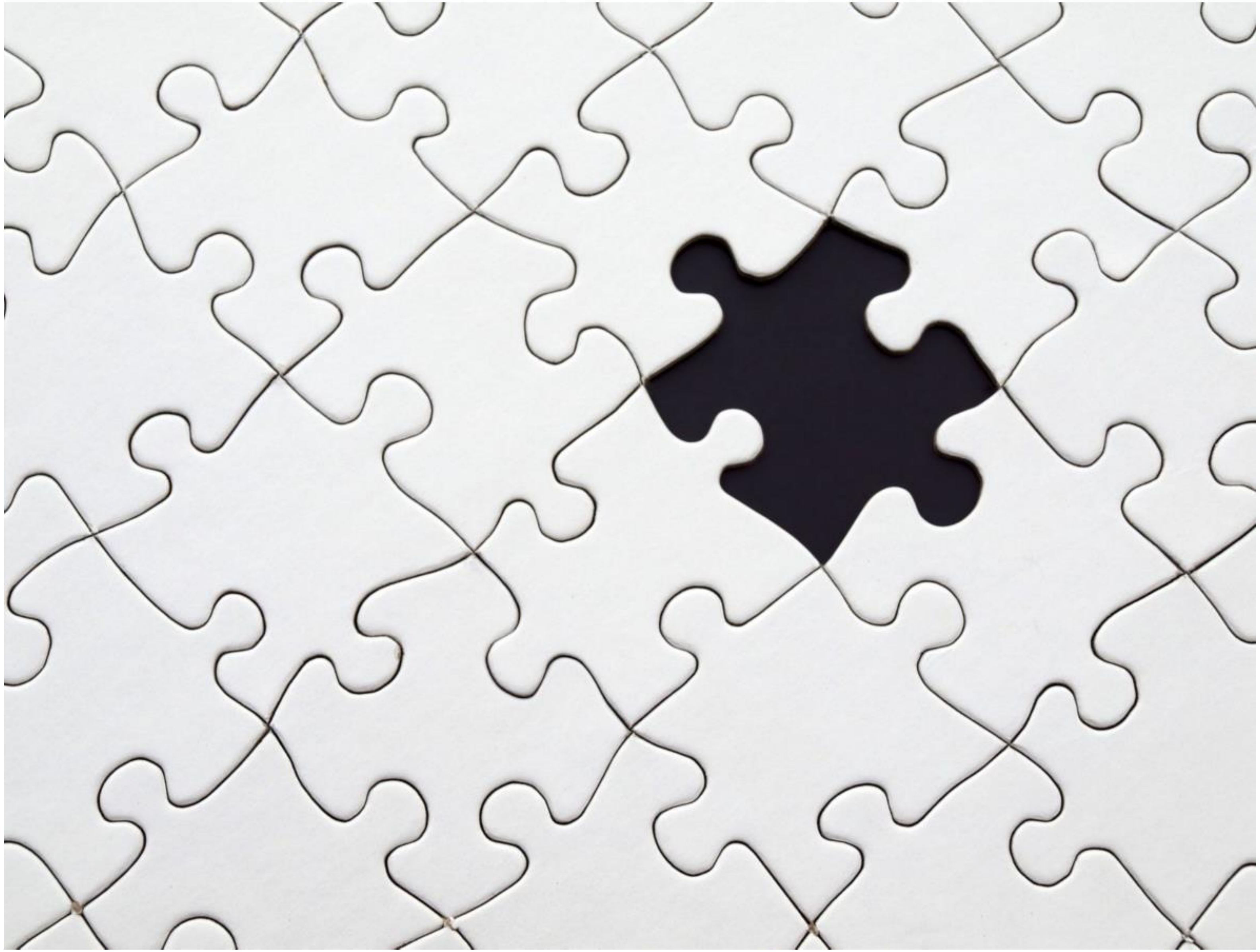
- Patients : Trauma.cranien == 1 || AIS.tete ≥ 2
- Treatment : Acide Tranexamique
- Outcome : DC.Trauma / Glasgow.sortie

SUMMARY

I. Preprocessing
   A. Missing values analysis
   B. Imputation
   C. Restriction

II. Descriptive analysis
   A. Principal component analysis
   B. Hierarchical clustering
   C. Visualization of main variables

III. Causal inference
   A. Matching
   B. Inverse Propensity Weighting
   C. Discussion

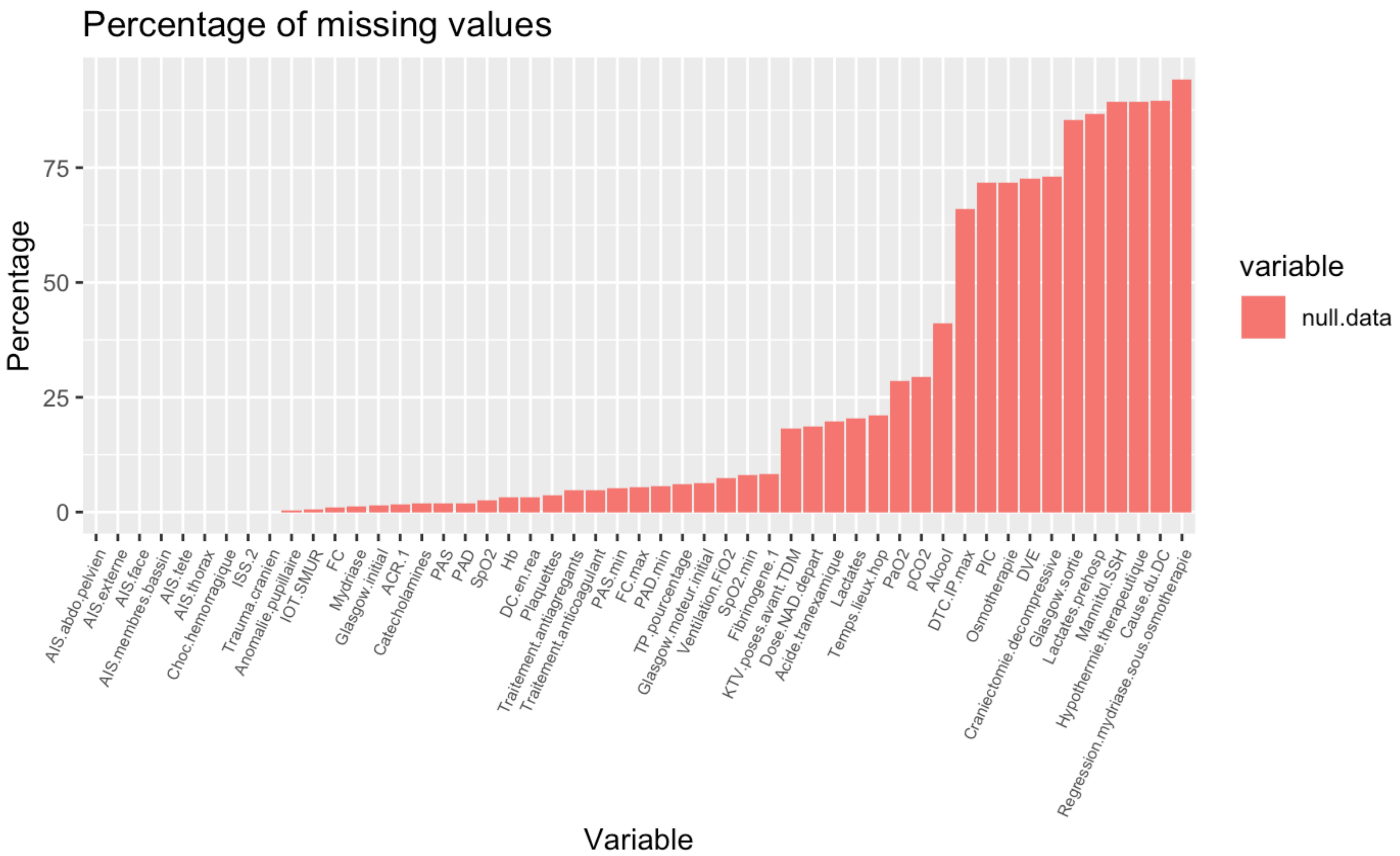


I. Preprocessing

I. Preprocessing

A. Missing values analysis

- From 244 variables, we keep 54 based on the doctors' input.
- Many of these null entries are not really missing – distinguishing them implies understanding the medical guidelines.

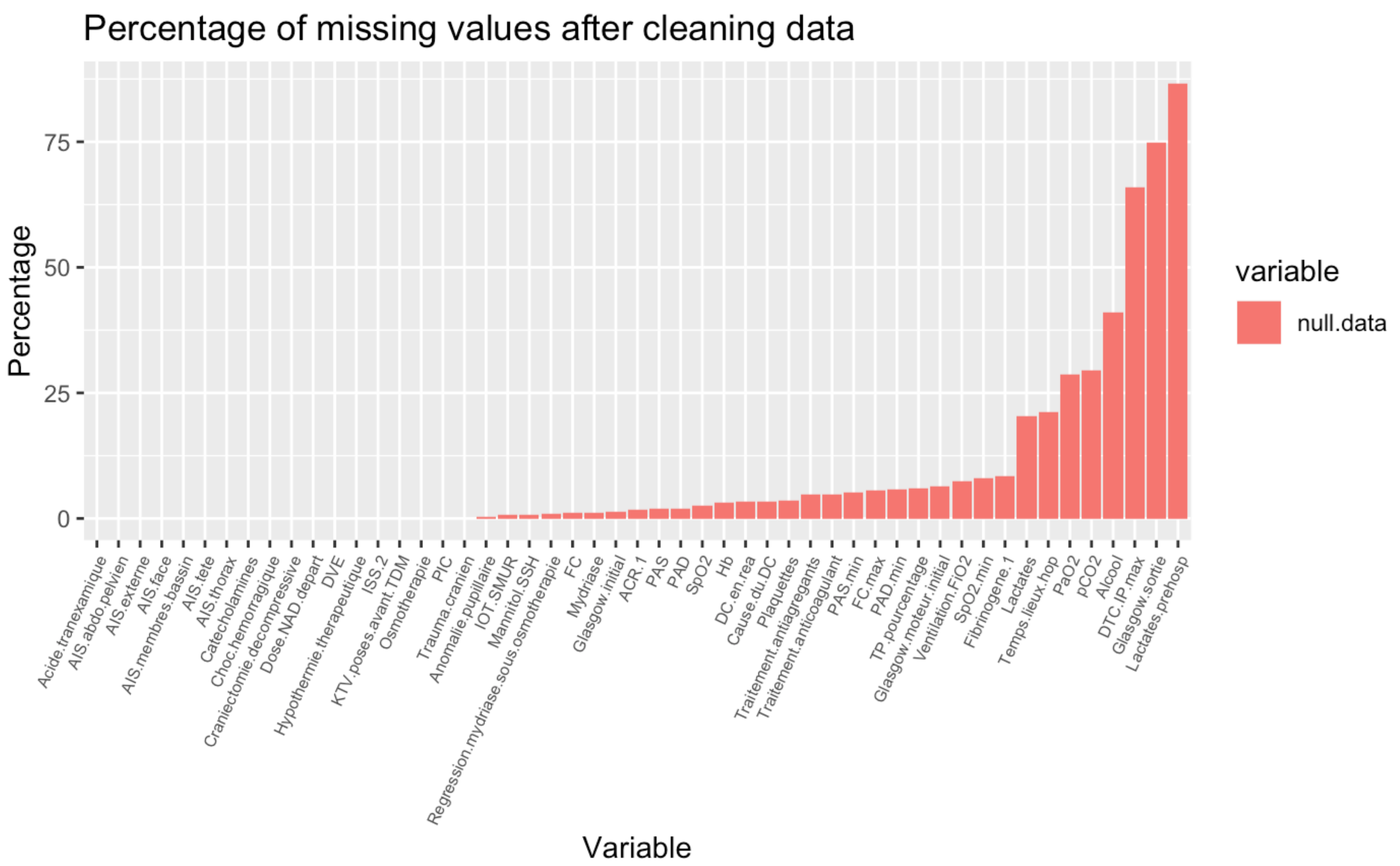


I. Preprocessing

A. Missing values analysis – not really missing entries

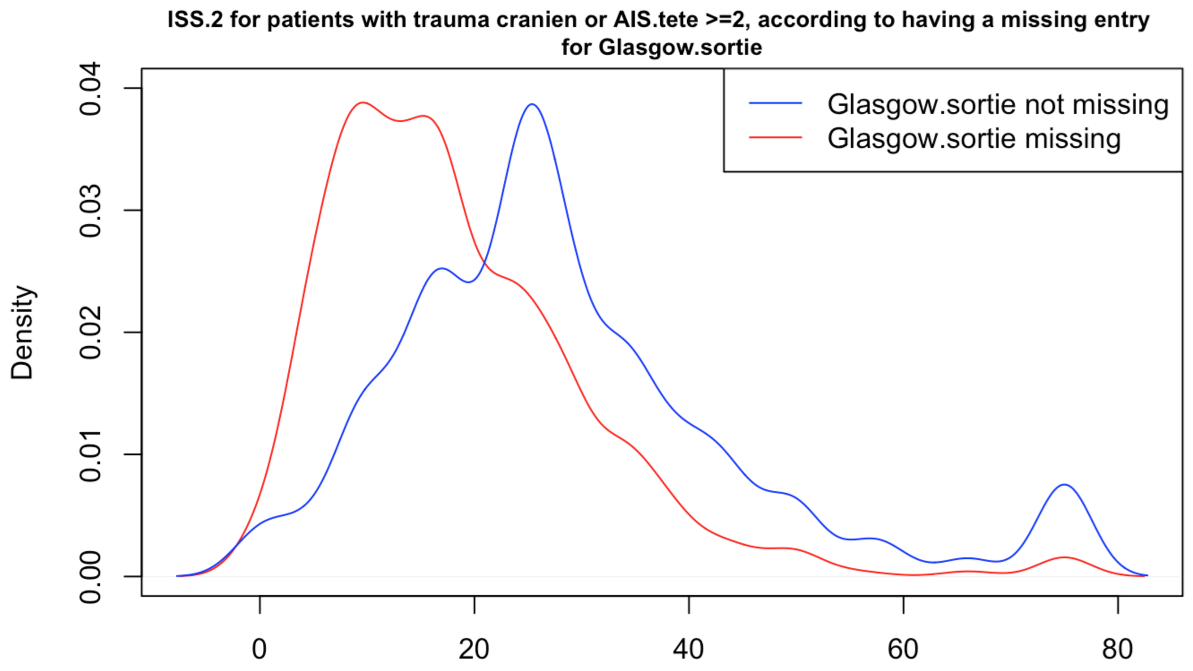
- We delete the only variable that has more than 75% of missing data: Lactates.prehosp

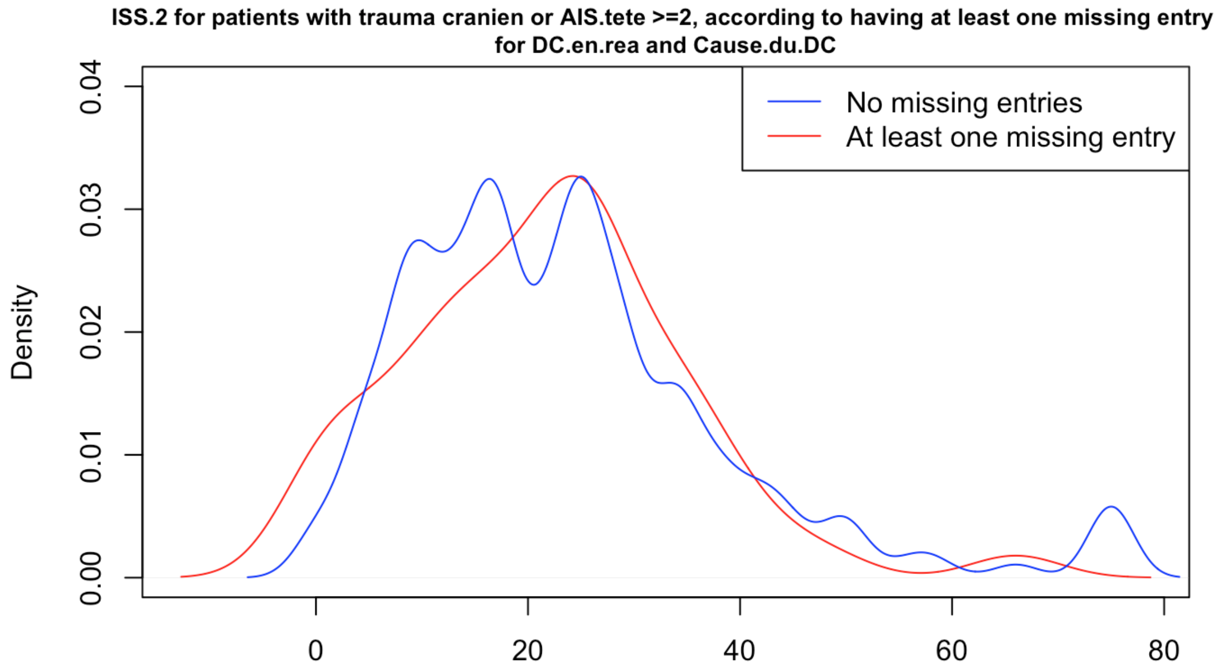
- Only 3 remaining variables that have more than 26% of missing values: Alcool, DTC.IP.max and Glasgow.sortie



I. Preprocessing

A. Missing values analysis – missing completely at random?

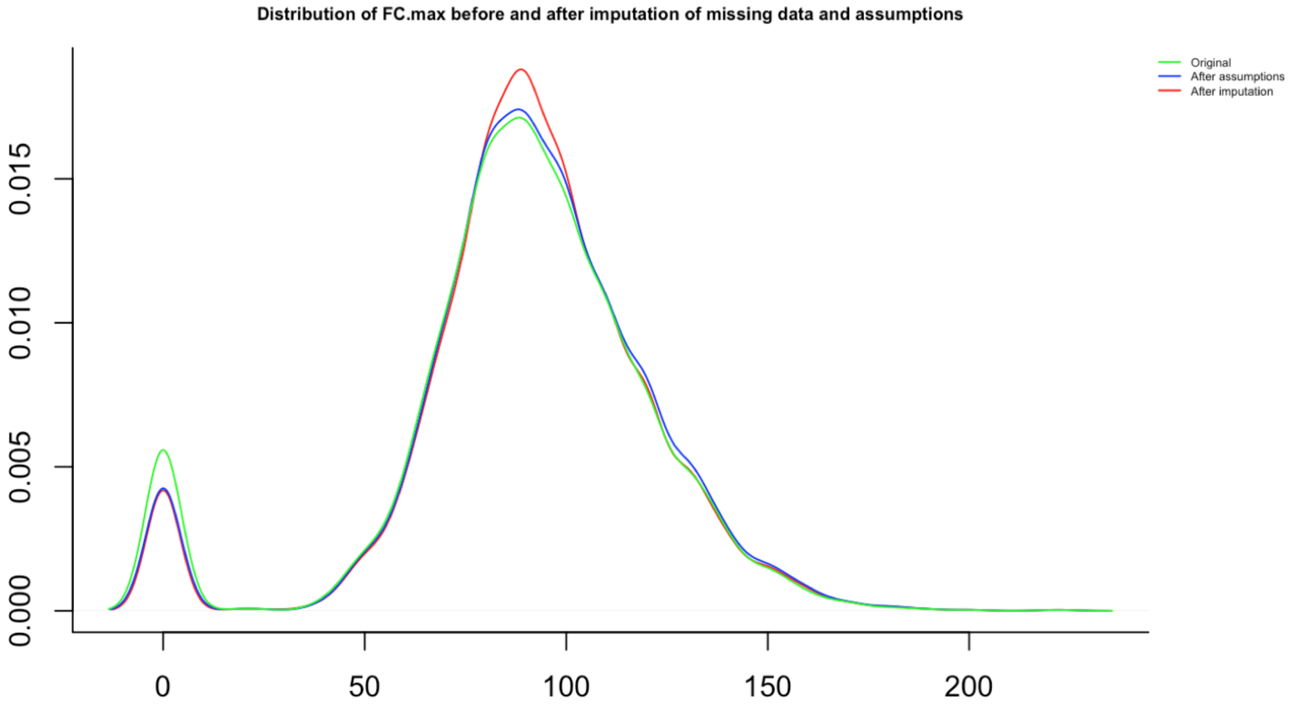




I. Preprocessing

B. Imputation

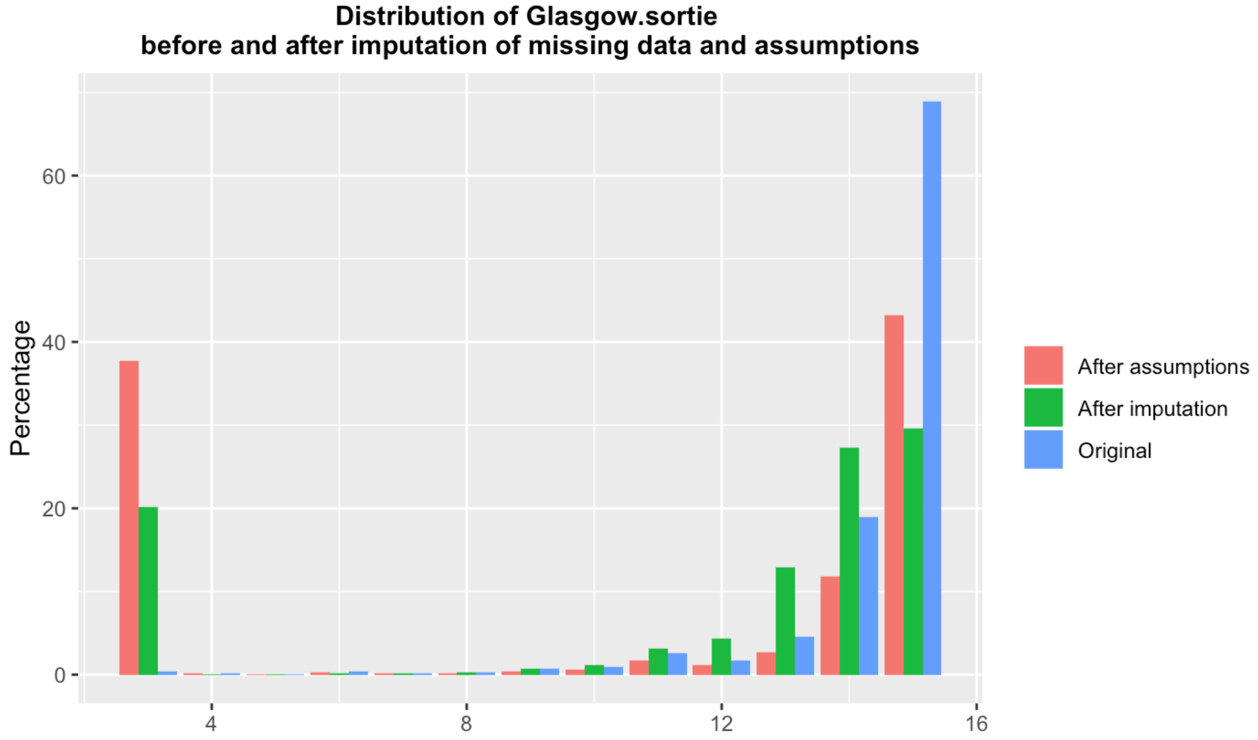
- Main method: iterative Factorial Analysis for Mixed Data model, package missMDA
- Most distributions unchanged - MCAR

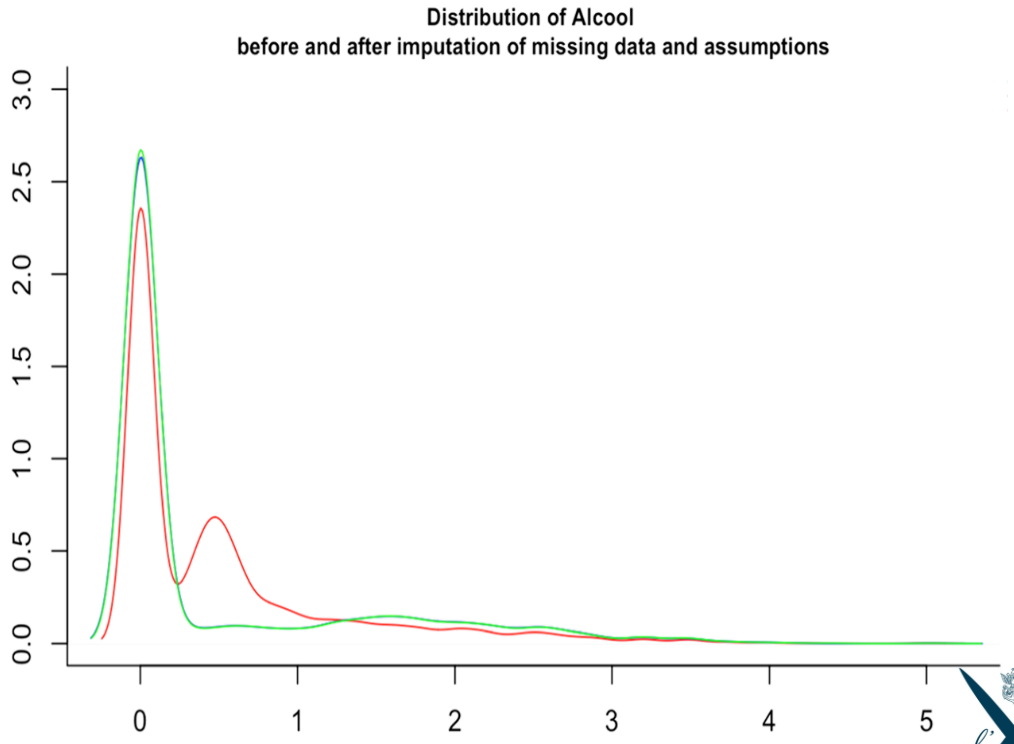


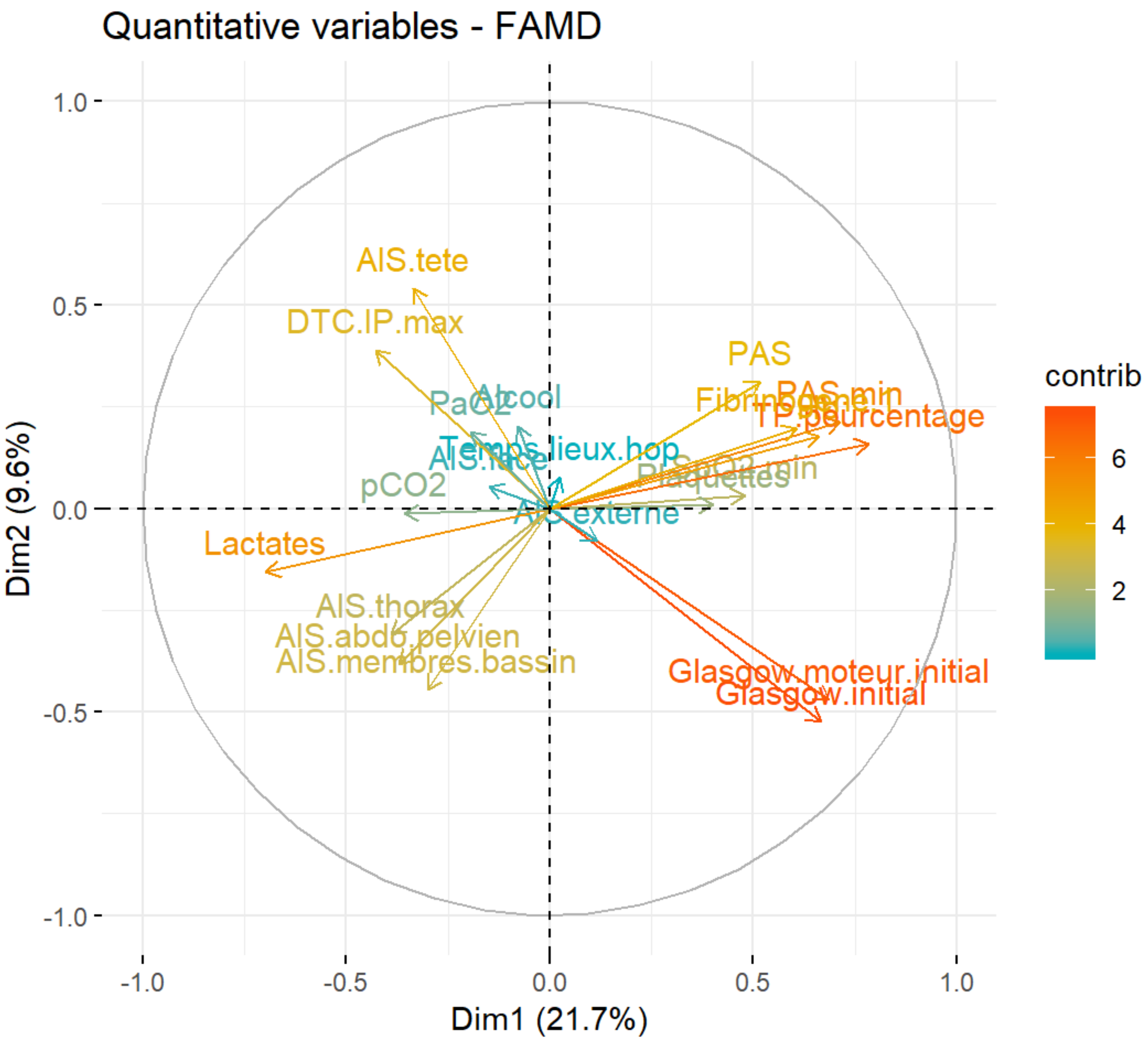
I. Preprocessing

B. Imputation

- But some source of concern with Glasgow.sortie, Alcool
- Alternative method run to test sensitivity of causal inference to imputation: random forest, package missForest





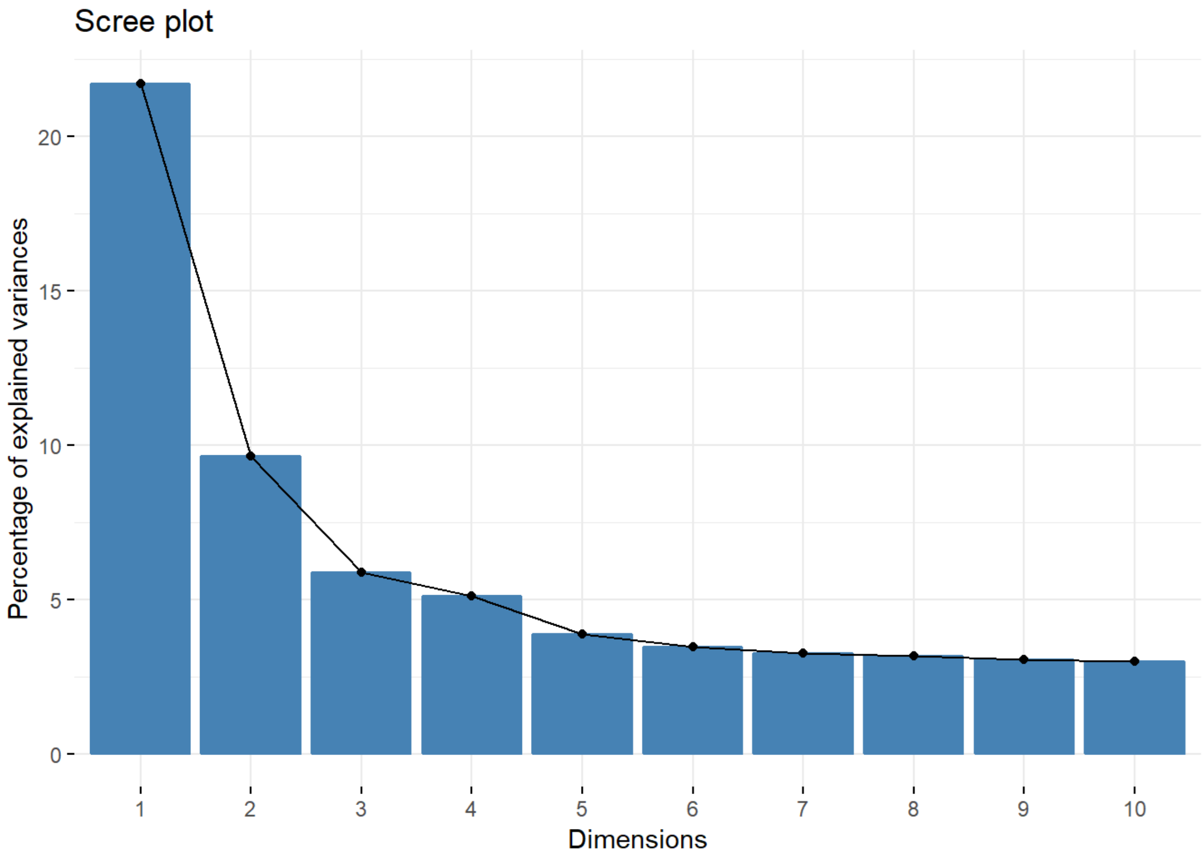


II. Descriptive analysis

II. Descriptive analysis

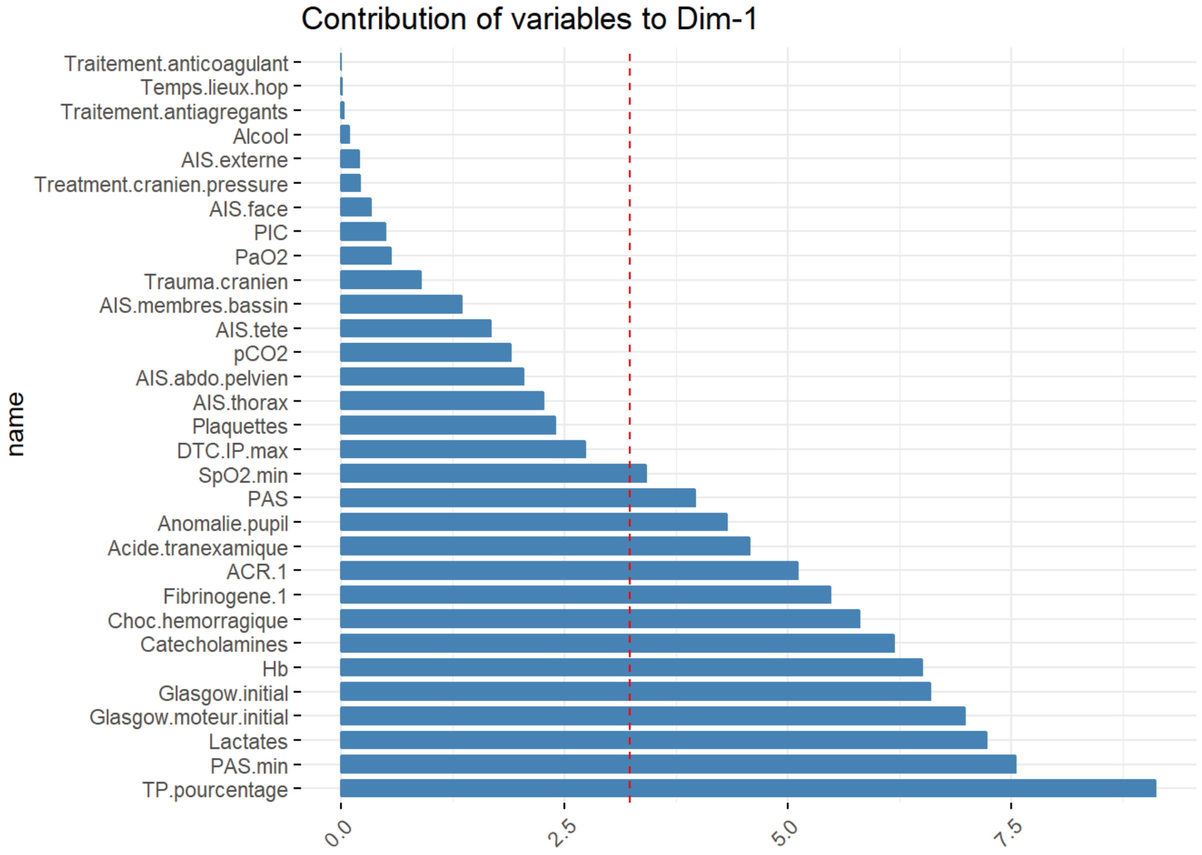
A. Factorial analysis of mixed data

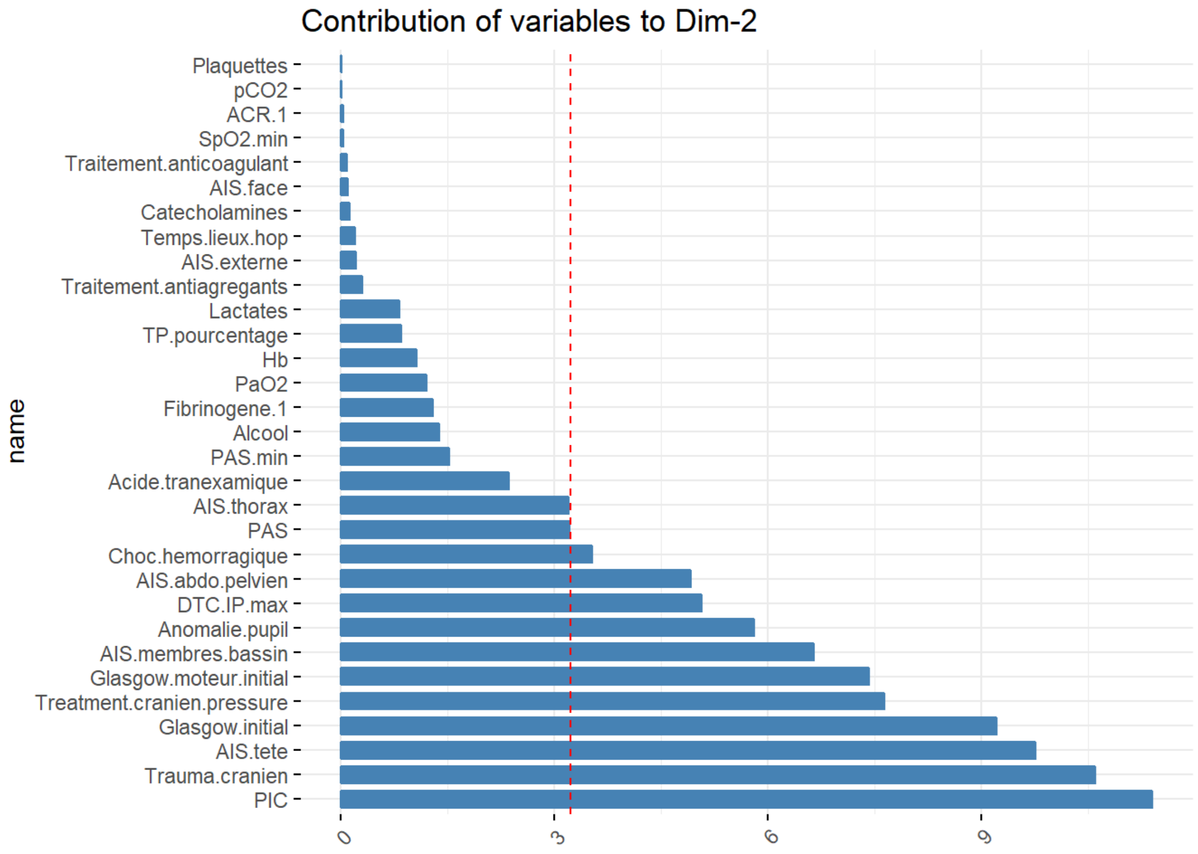
- Extracting most of the information using only 10 components
- 62% of the variance kept
- Working on quantitative variables



II. Descriptive analysis

A. Factorial analysis of mixed data



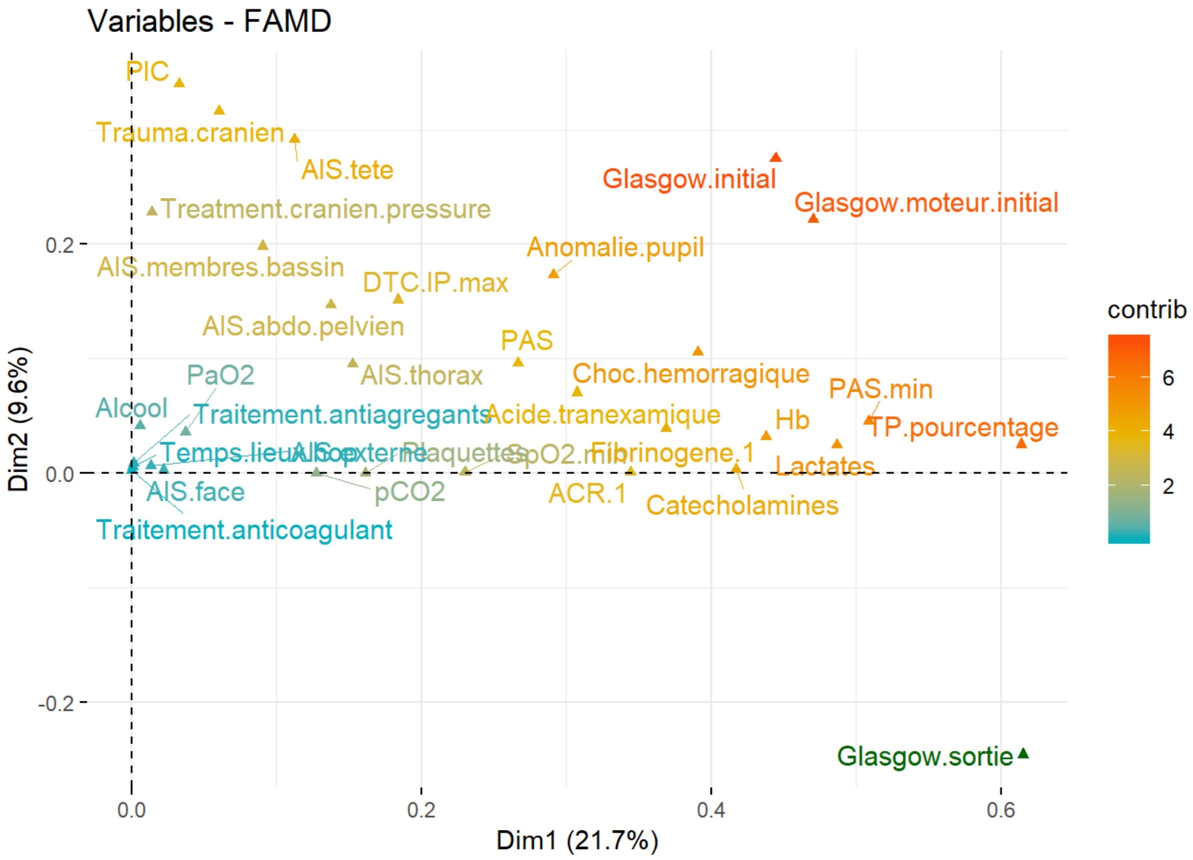


ÉCOLE POLYTECHNIQUE
UNIVERSITÉ PARIS-SACLAY

II. Descriptive analysis

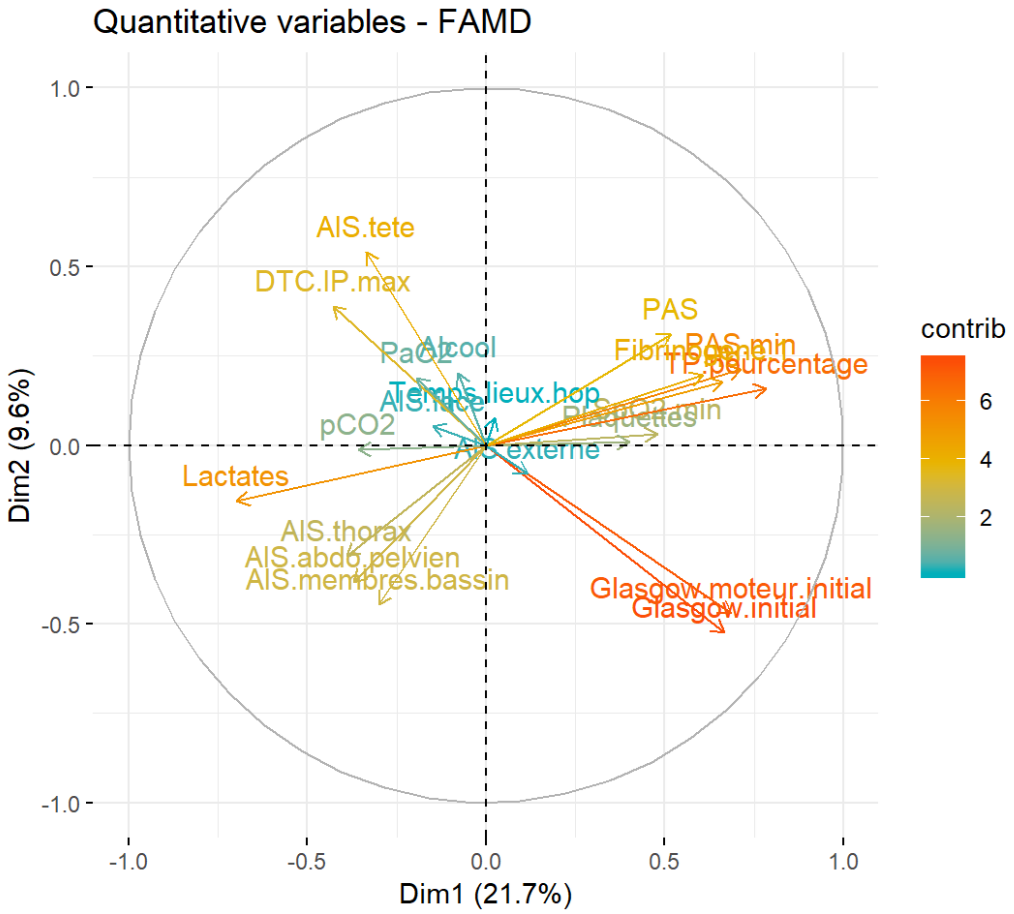
A. Factorial analysis of mixed data

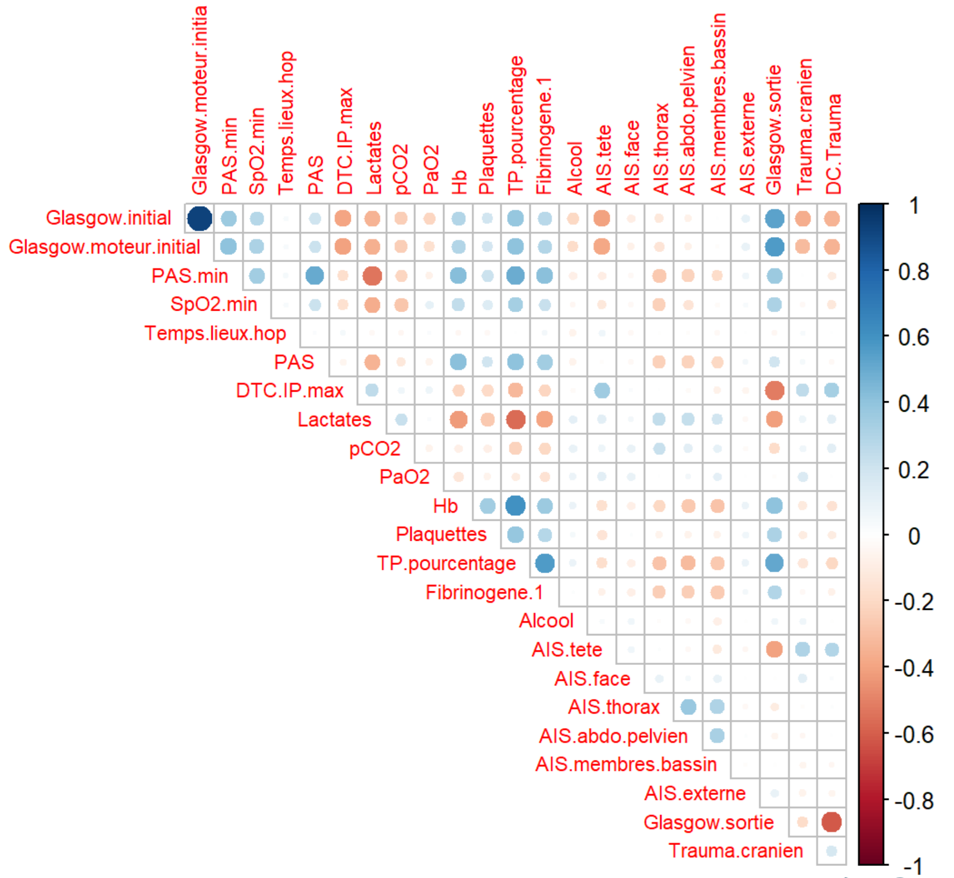
- Visualizing the contribution of each variable to the principal components
- Gives an idea on the correlation between variables



II. Descriptive analysis

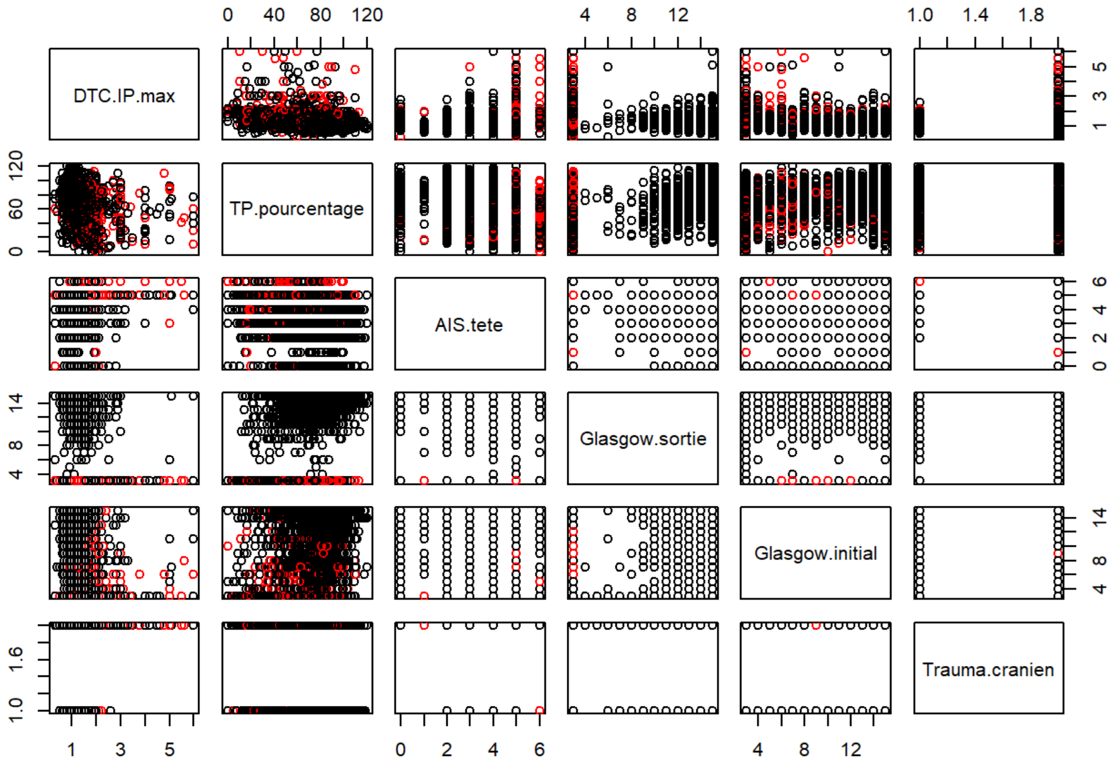
A. Factorial analysis of mixed data





II. Descriptive analysis

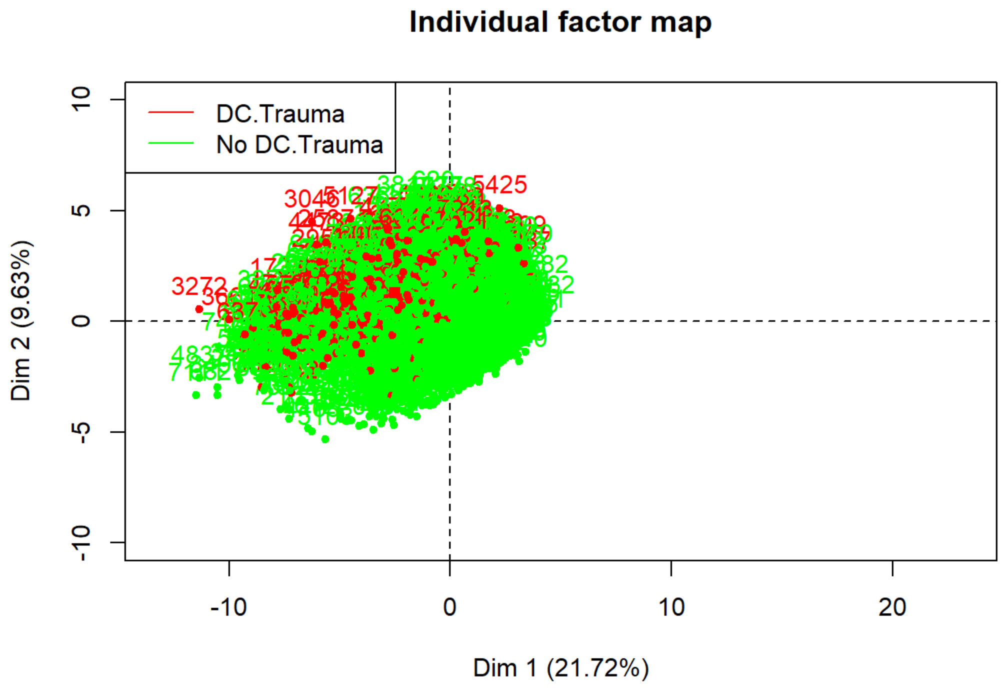
A. Factorial analysis of mixed data



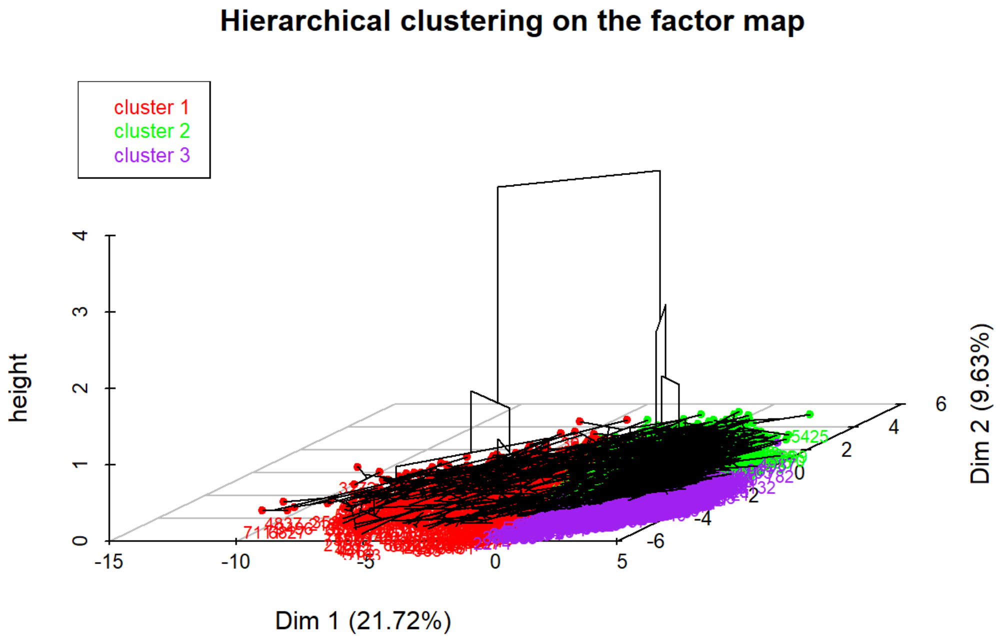
II. Descriptive analysis

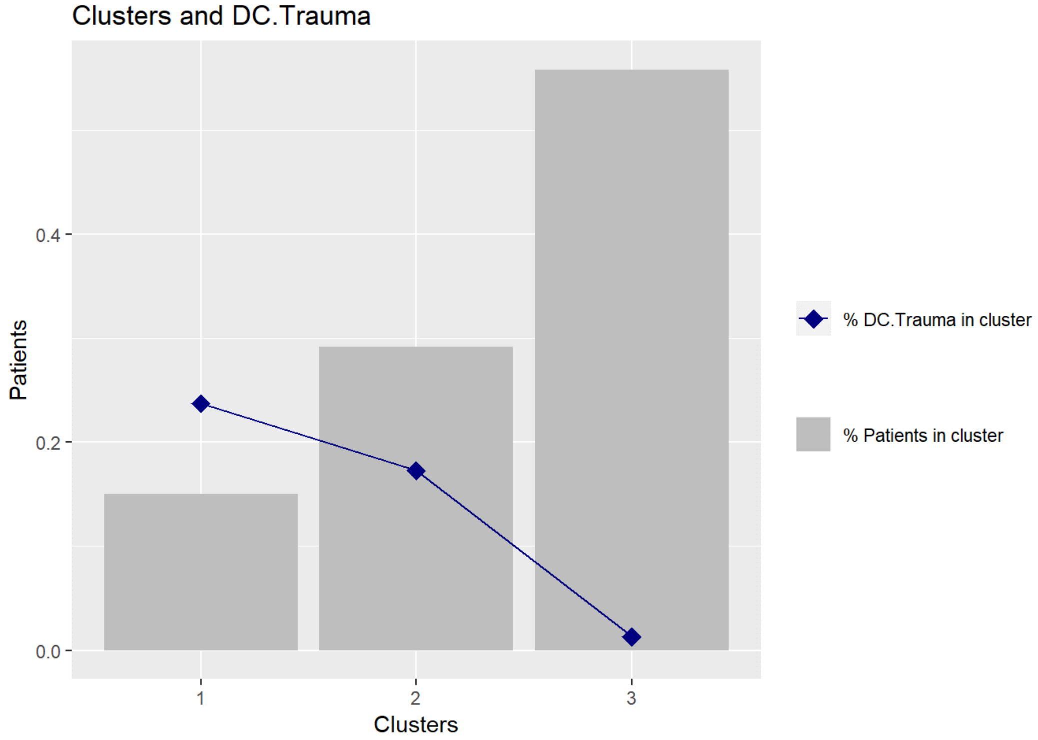
B. Hierarchical clustering

- Cluster patients to understand the link between DC.Trauma and variables
- Both categorical and quantitative variables: clustering using FAMD
- R package: *Hierarchical Clustering on Principal Components*



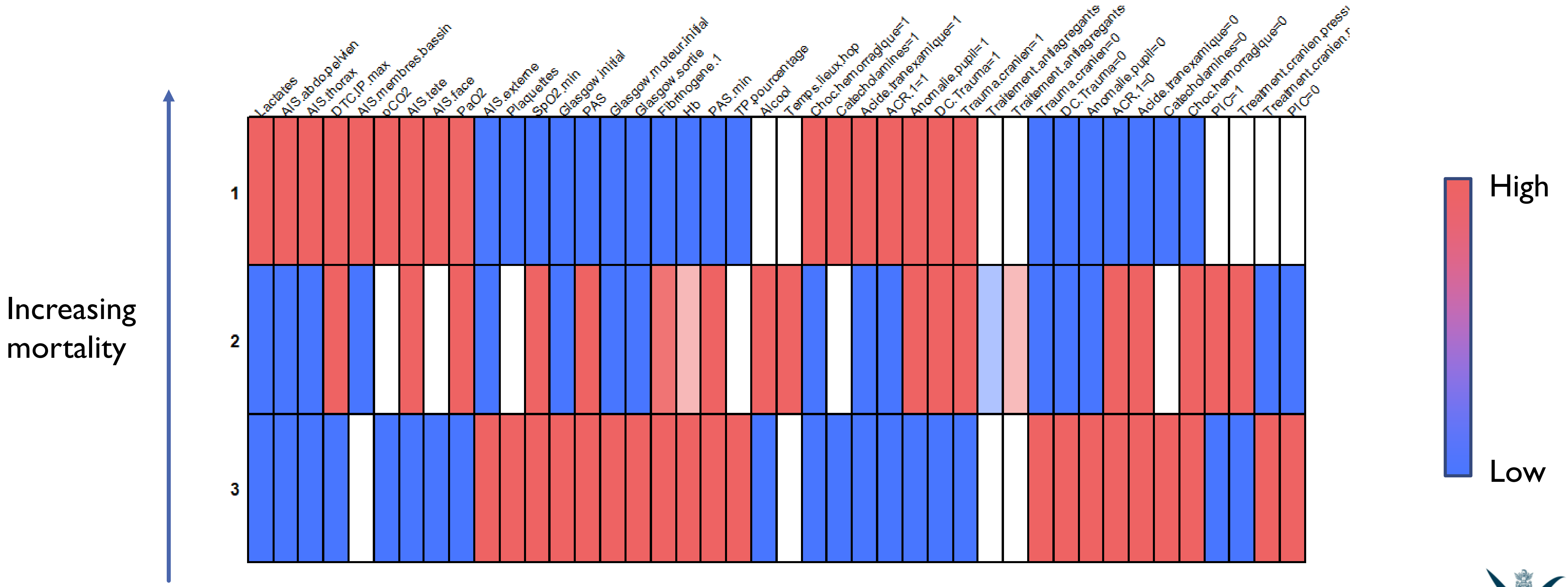
Hierarchical clustering output





II. Descriptive analysis

B. Hierarchical clustering

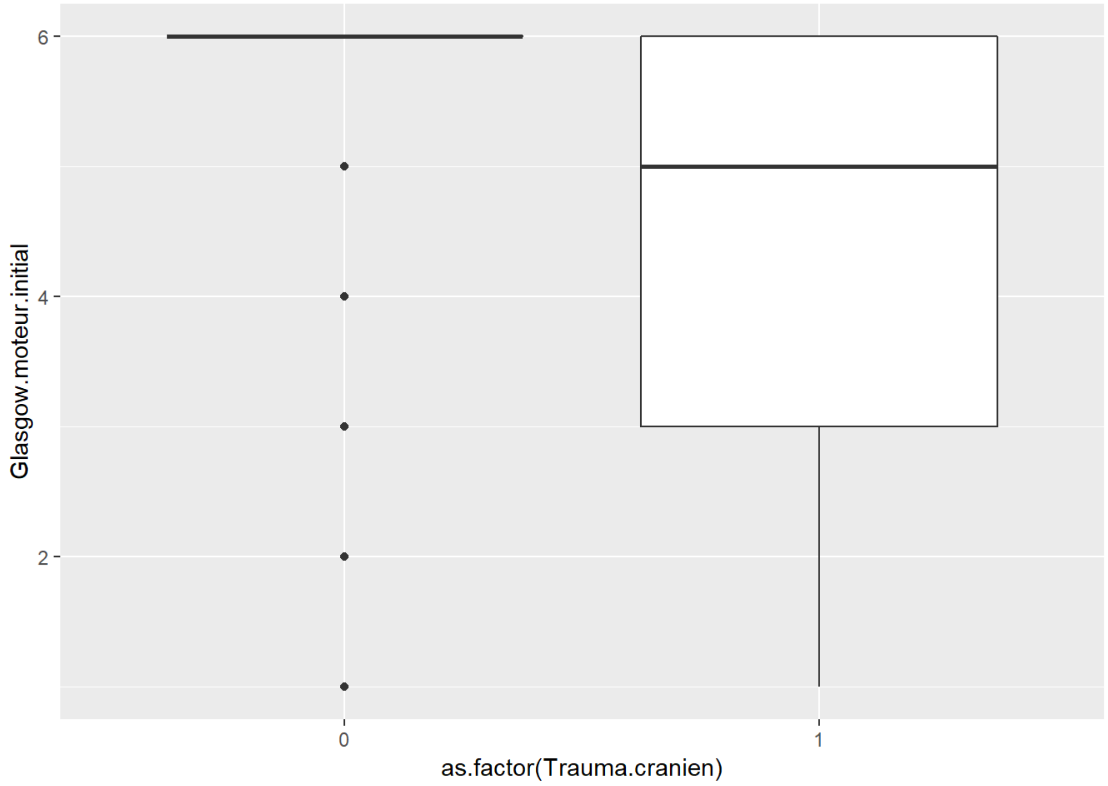


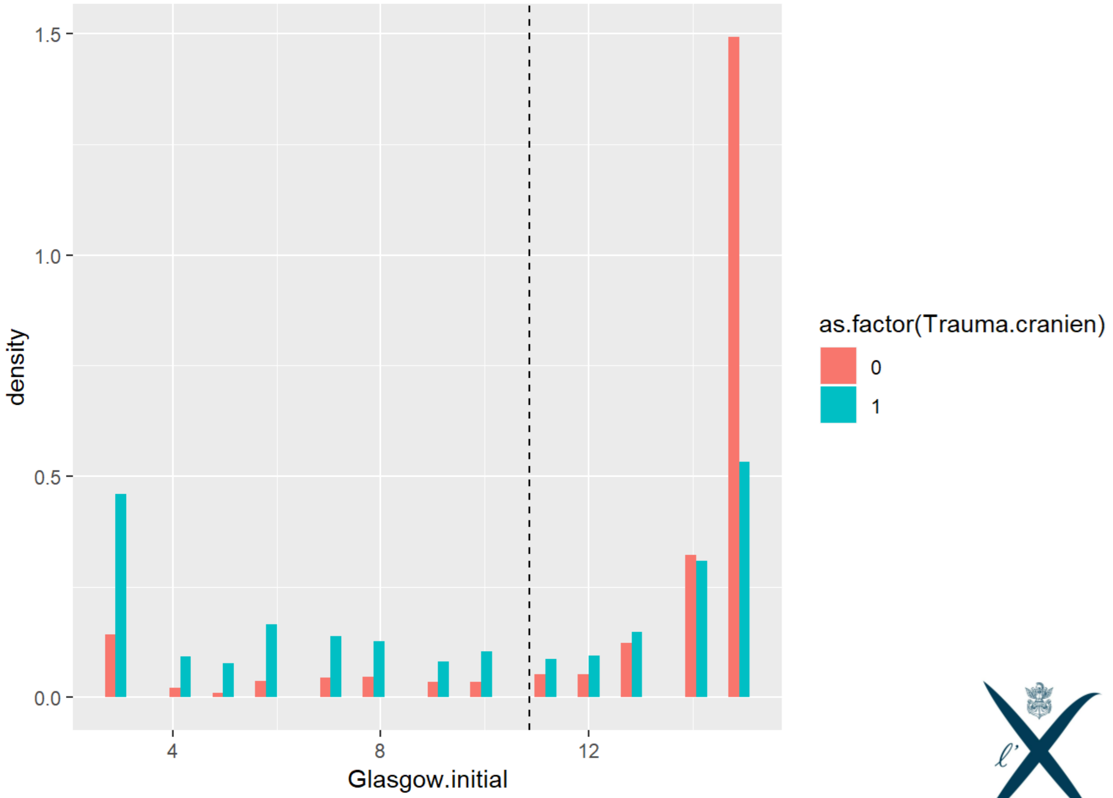
II. Descriptive analysis

C. Data visualization

- Idea : Doctors look at measurements of variables to take their decisions
- Goal : Check their assumptions and possibly find new ways to predict trauma/shock/necessity of using Acide.tranexamique

- Glasgow.initial and Cranial trauma

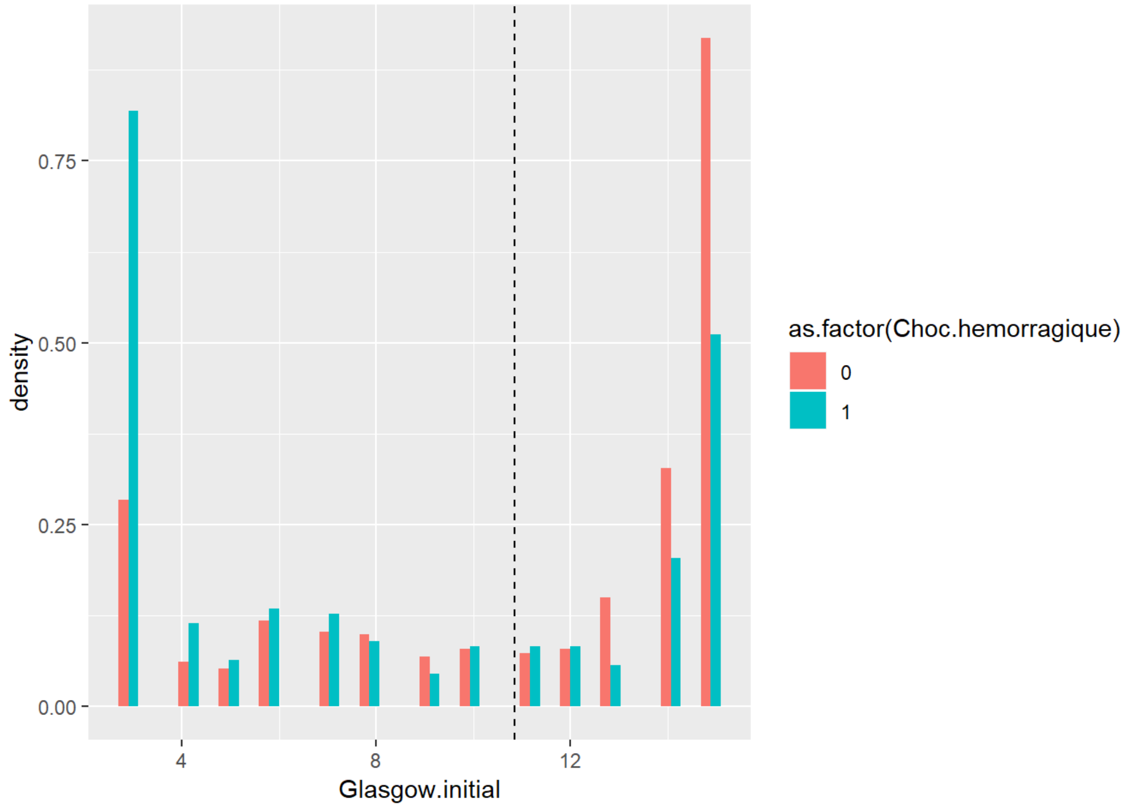




II. Descriptive analysis

C. Data visualization

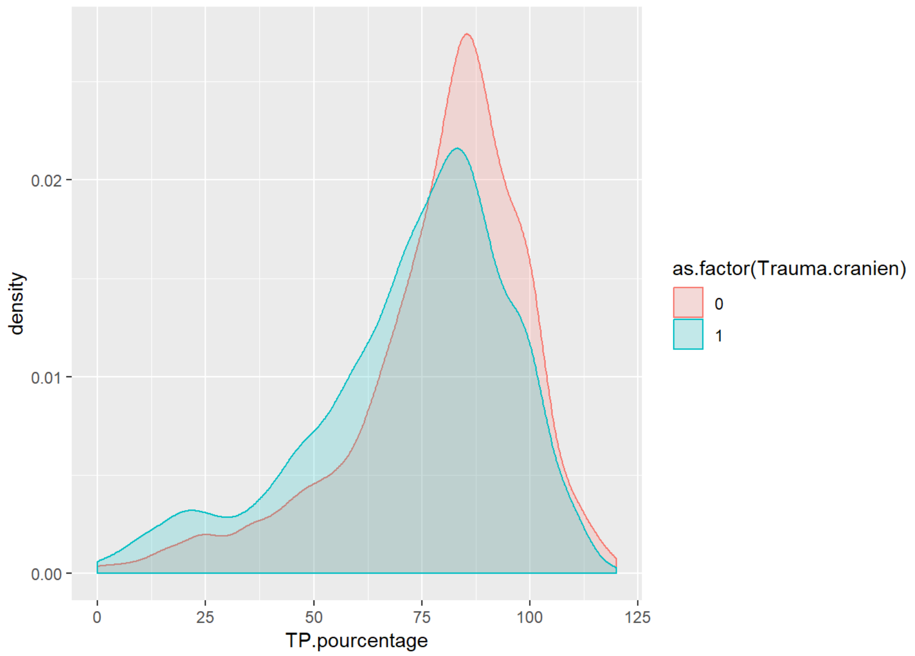
- Glasgow.initial and Hemorrhagic shock



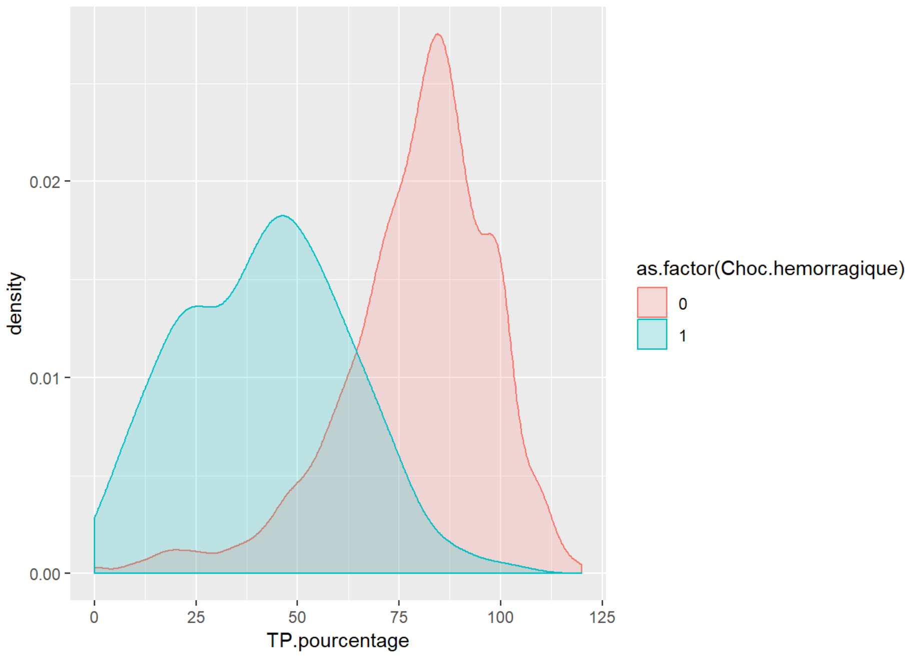
II. Descriptive analysis

C. Data visualization

- Prothrombin Ratio



Cranial Trauma

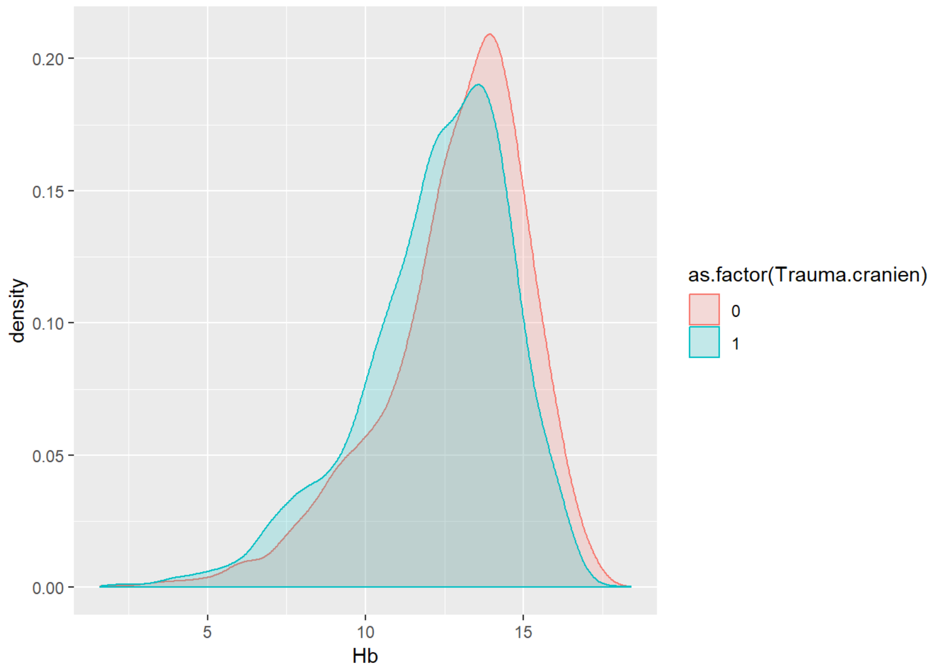


Hemorrhagic shock

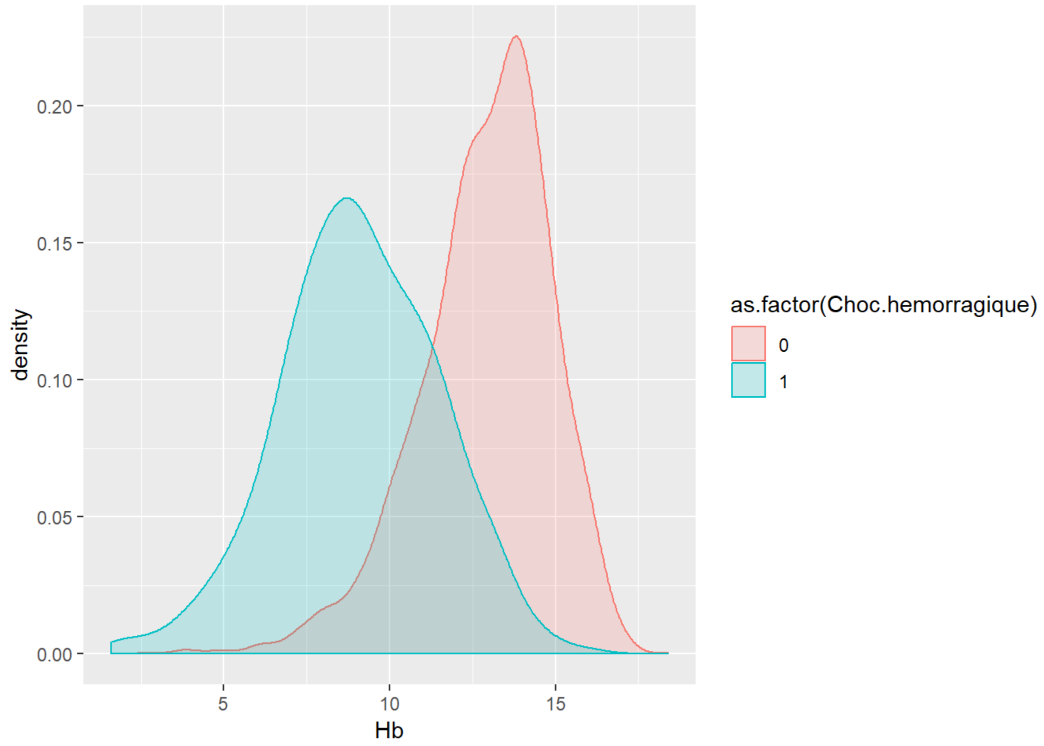
II. Descriptive analysis

C. Data visualization

- Hb Percentage



Cranial Trauma

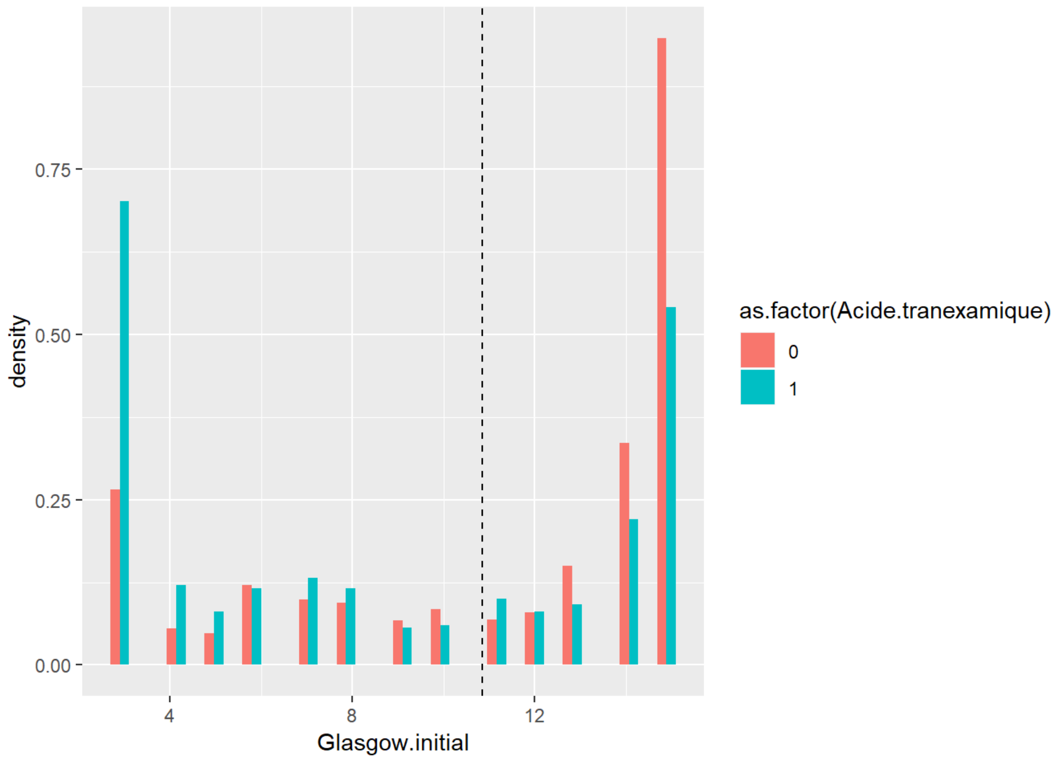


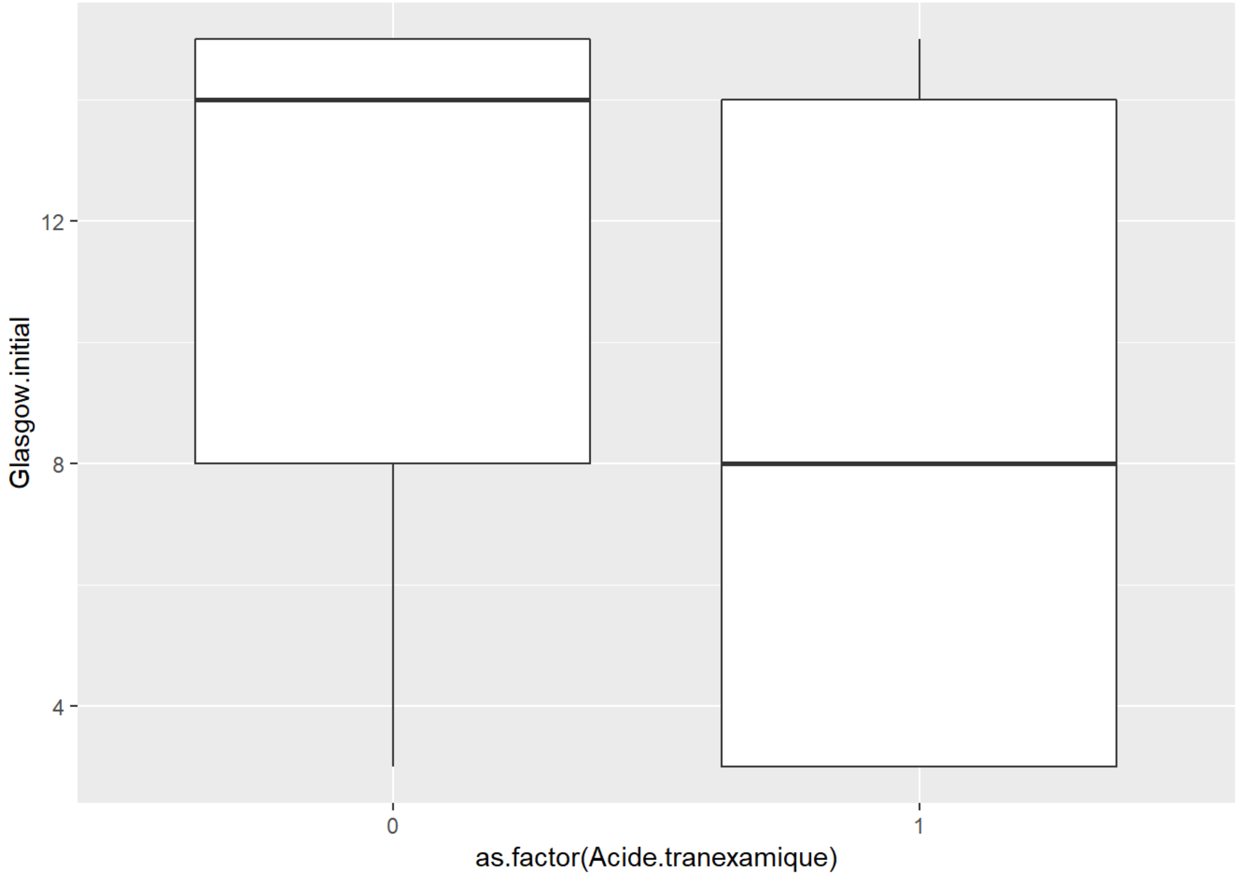
Hemorrhagic shock

II. Descriptive analysis

C. Data visualization

- Acide tranexamique

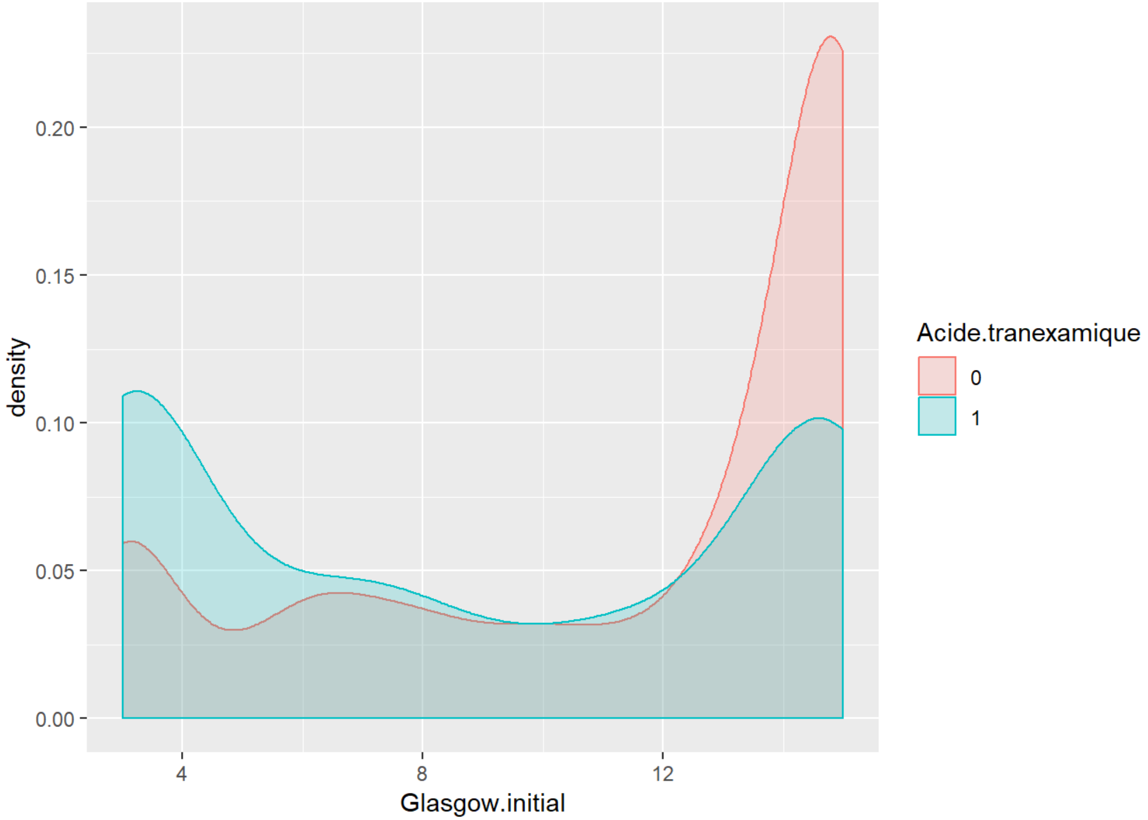




II. Descriptive analysis

C. Data visualization

- Acide tranexamique



II. Descriptive analysis

C. Data visualization

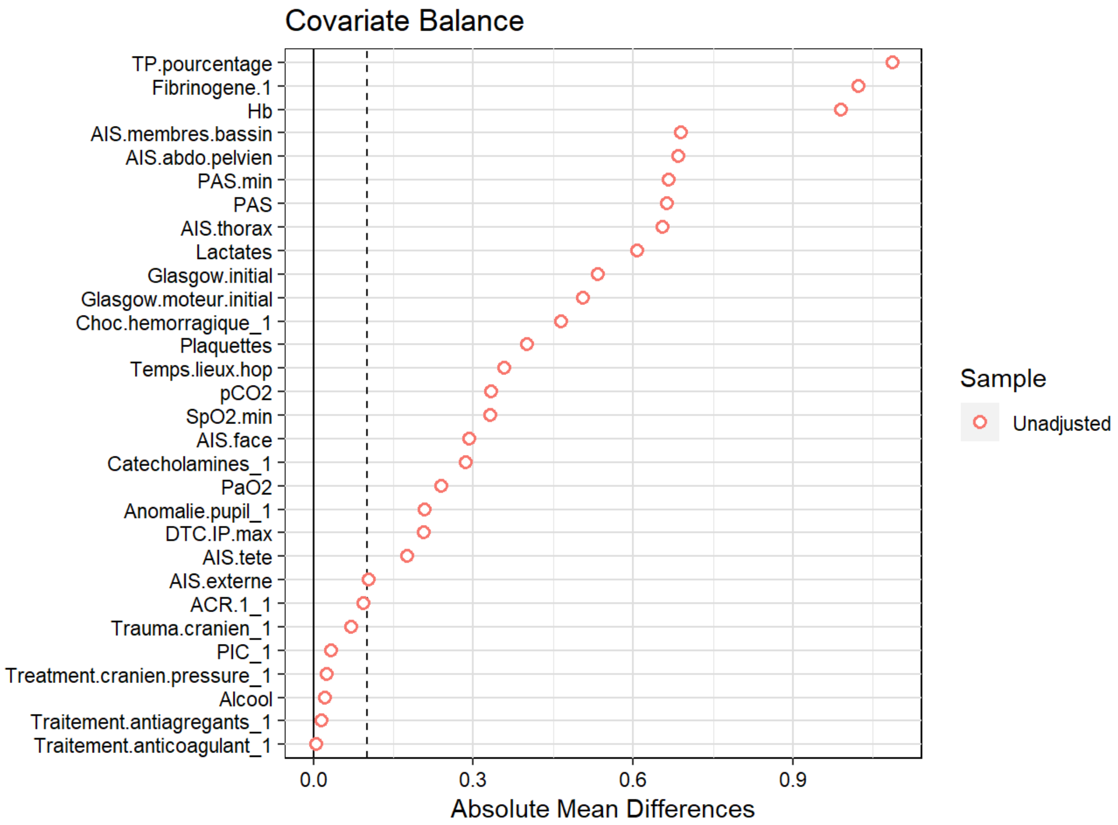
- P(Died| Treated) : 0.16
- P(Died| Not treated) : 0.08

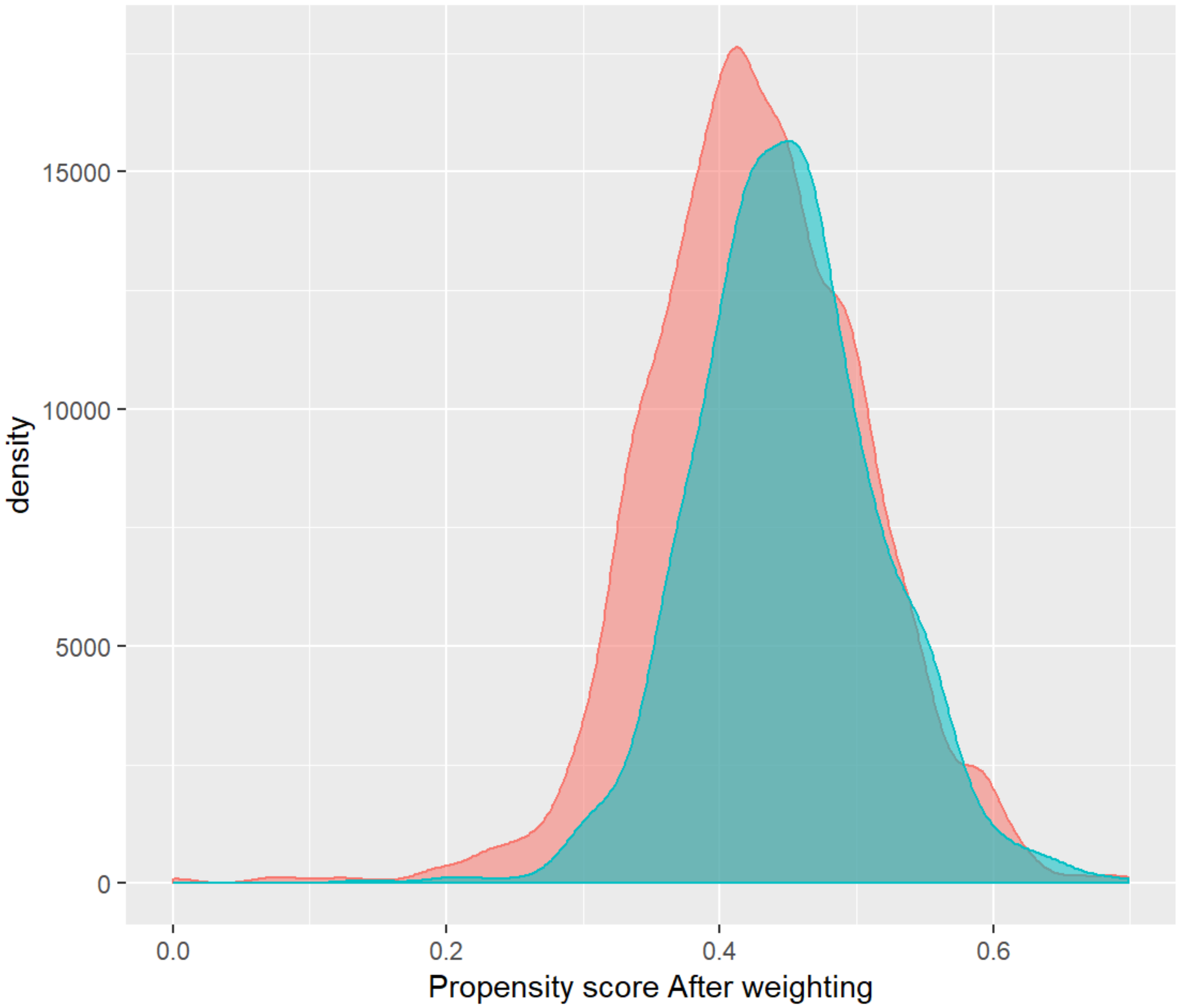
Treatment kills ?

II. Descriptive analysis

C. Data visualization

- Comparaison of the treated/nontreated patients

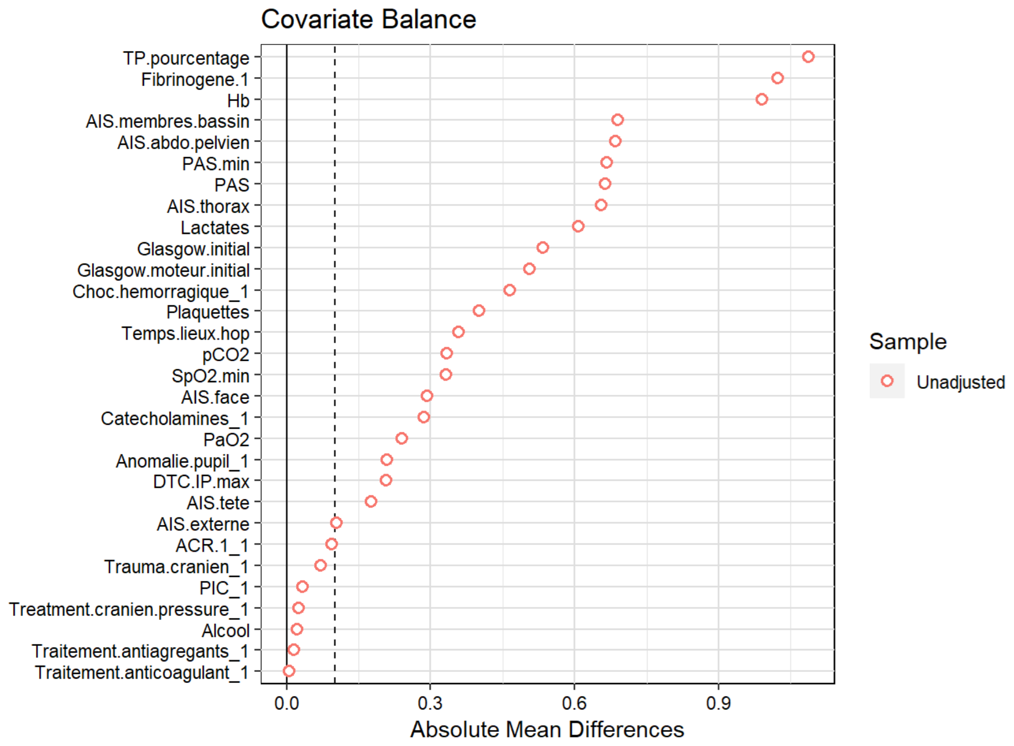




III. Causal Inference

III. Causal Inference

A. Matching

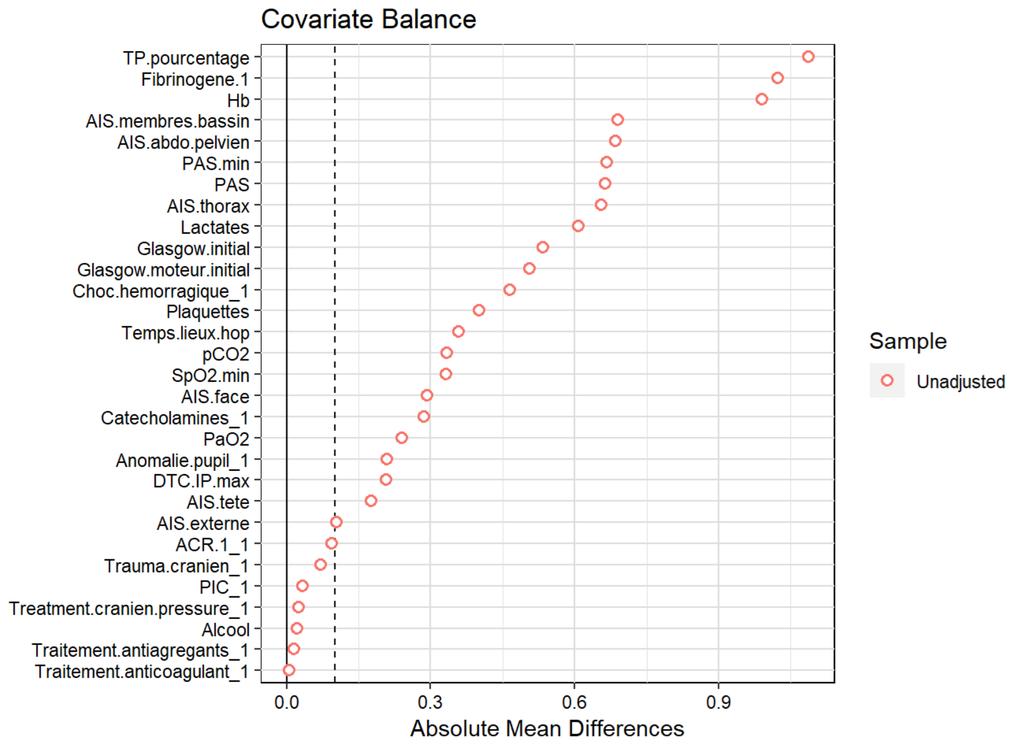


- Pair each treated patient with a similar untreated patient
- R package: *Matching*
- Then, regression: *DC.Trauma ~ Acide.tranexamique*

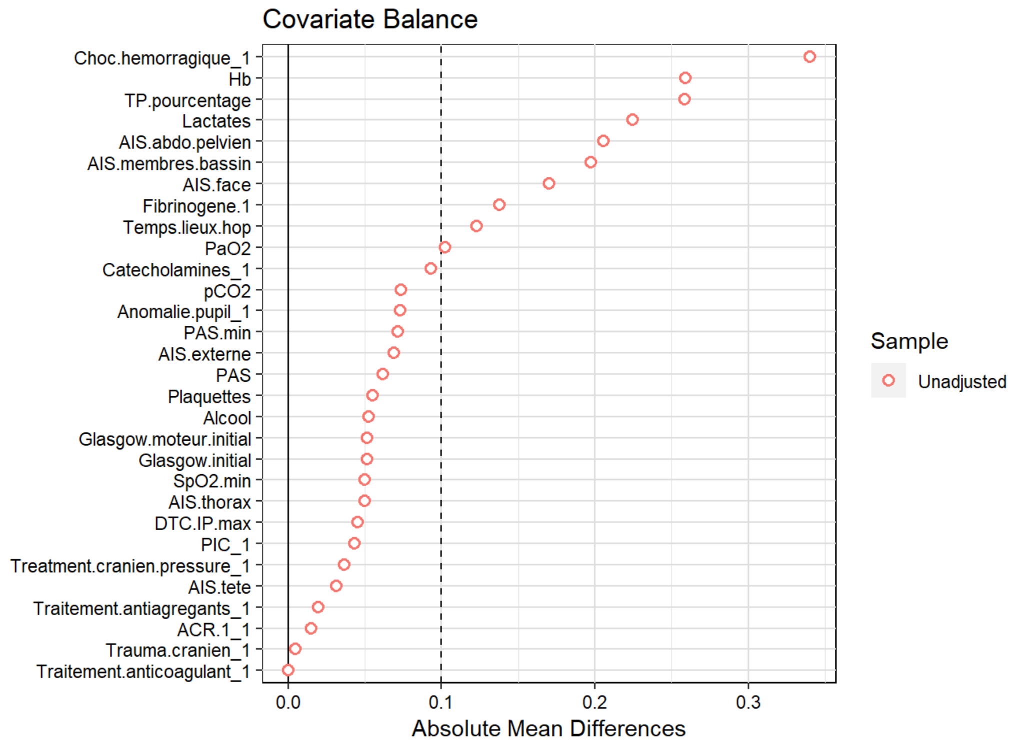
III. Causal Inference

A. Matching

Before matching



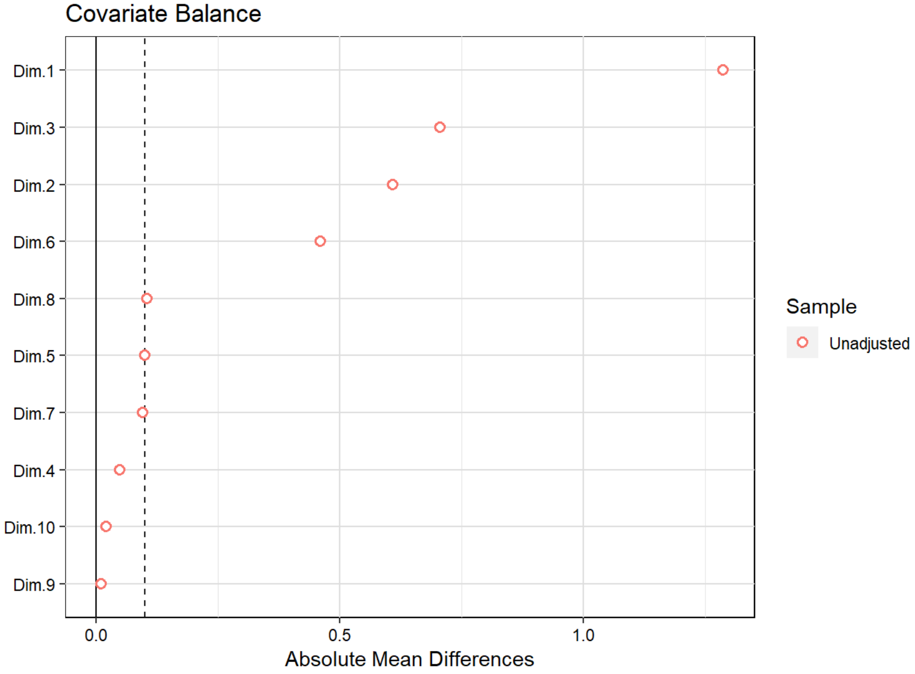
After matching



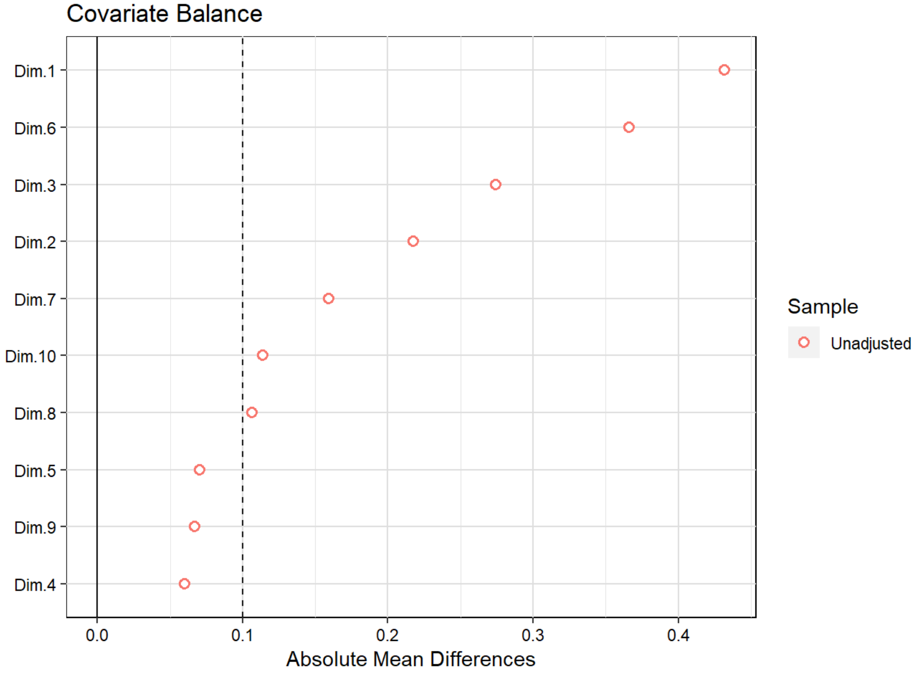
III. Causal Inference

A. Matching

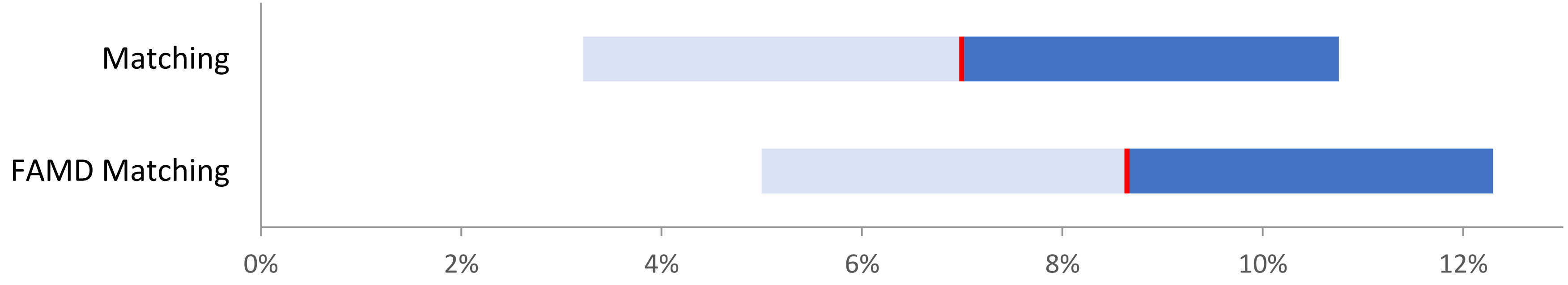
Before matching



After matching



Average treatment effect using Matching



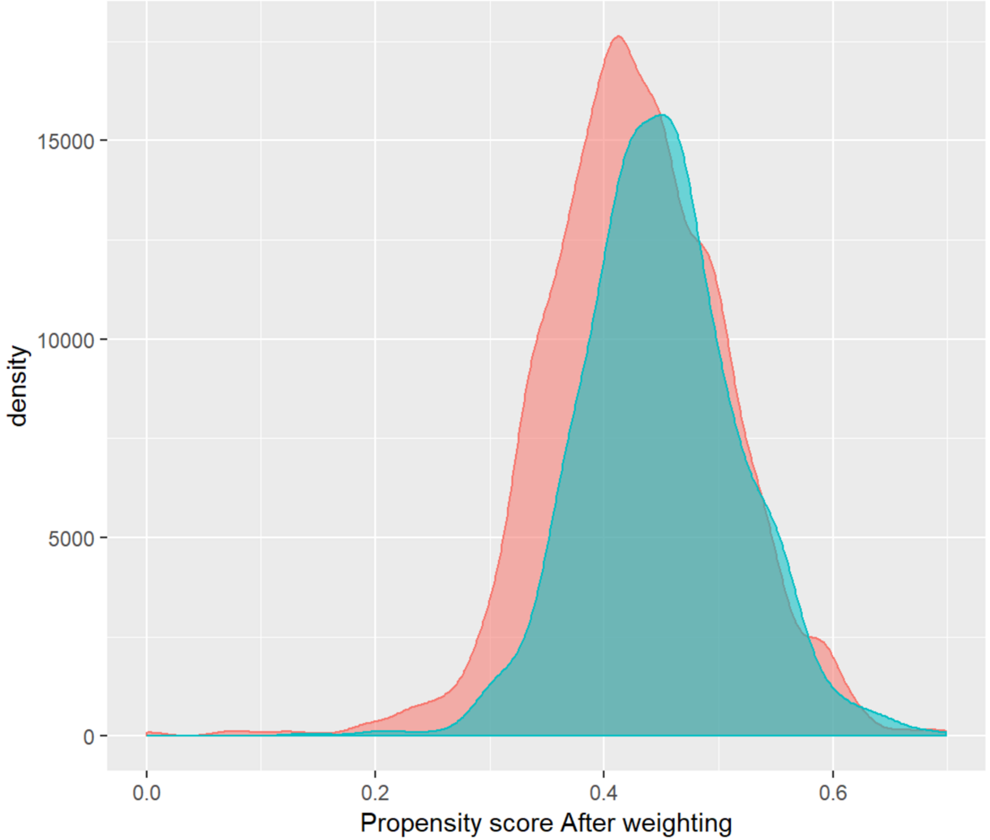

```
## [1] "95% ATE on Glasgow.sortie : [-1.26, -0.67, -0.07]"
```

```
## [1] "95% ATE on Glasgow.sortie (FAMD) : [-1.59, -1.00, -0.42]"
```

III. Causal Inference

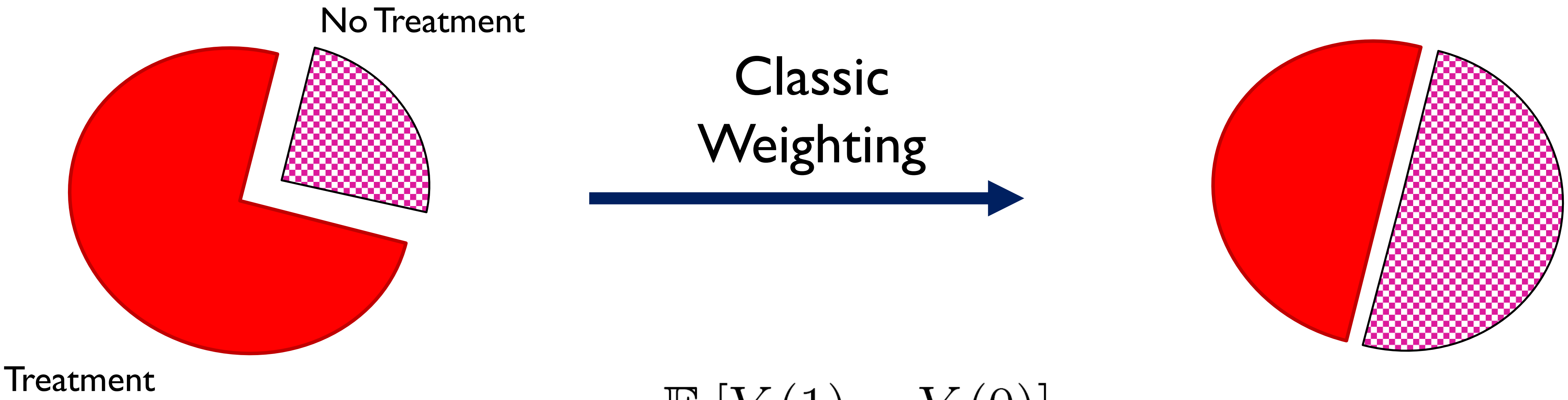
B. Inverse Propensity Weighting

- Correct the effect of confounding factors
- R package: *ipw, survey, grf*
- *IPW estimator of IPW, double-robust estimator, linear regression*



III. Causal Inference

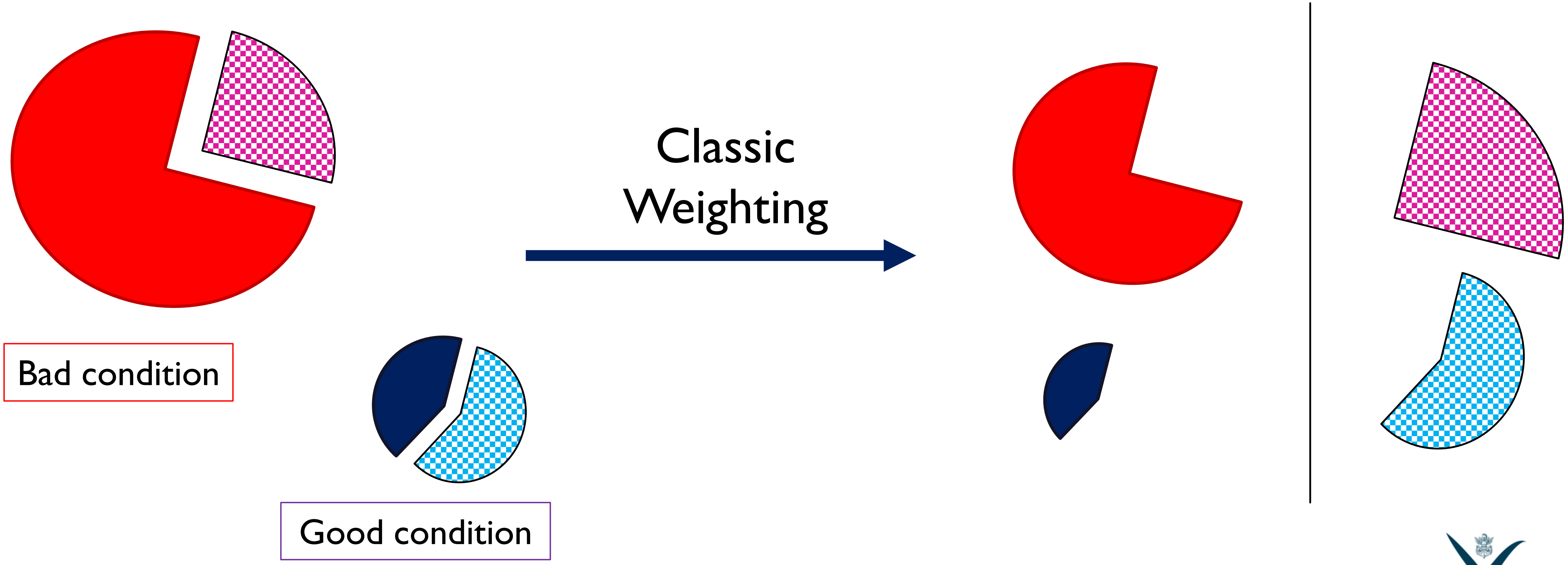
B. Inverse Propensity Weighting



$$\tau = \mathbb{E}\left[Y_i(1) - Y_i(0)\right]$$

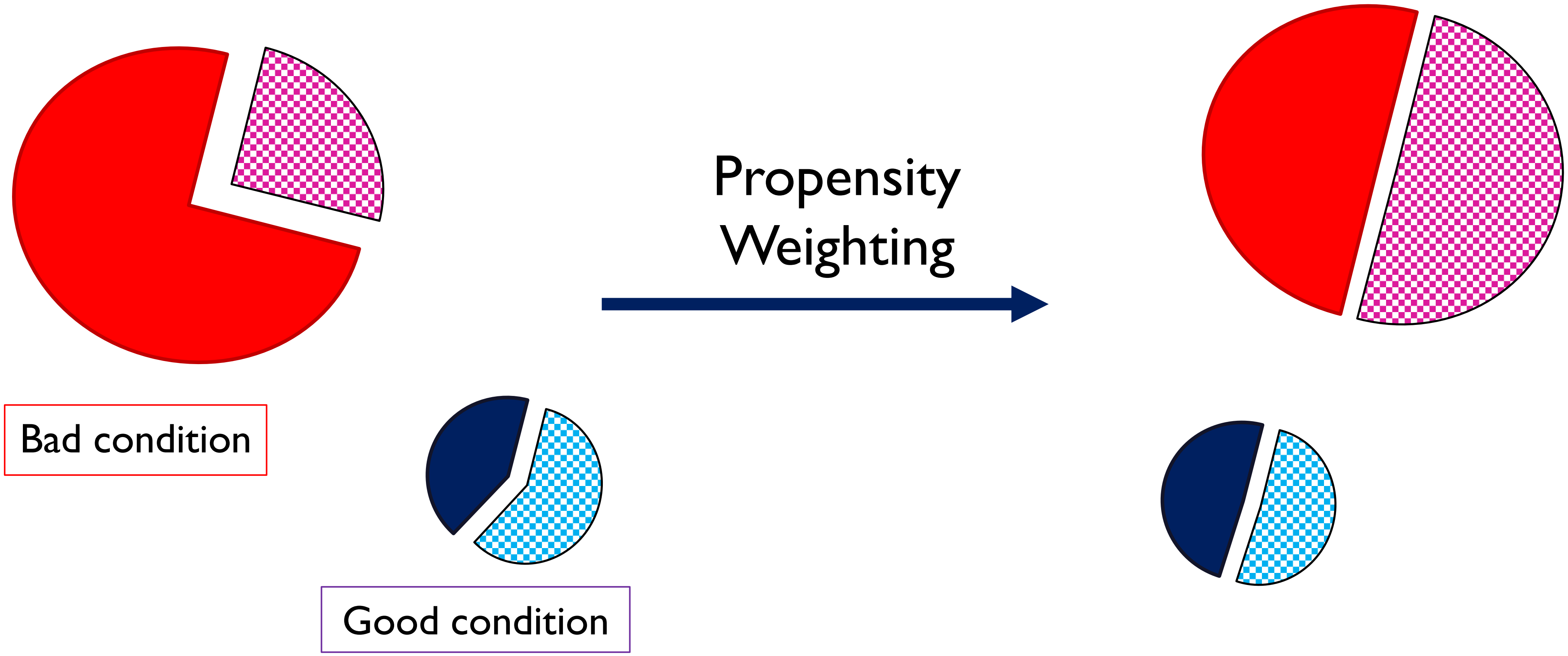
III. Causal Inference

B. Inverse Propensity Weighting



III. Causal Inference

B. Inverse Propensity Weighting



III. Causal Inference

B. Inverse Propensity Weighting

$$\tau = \mathbb{E}\left[Y_i(1) - Y_i(0)\right]$$

$$\hat{\tau}_{DM} = \frac{1}{n_1}\sum_{W_i=1} Y_i - \frac{1}{n_0}\sum_{W_i=1} Y_i$$

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

III. Causal Inference

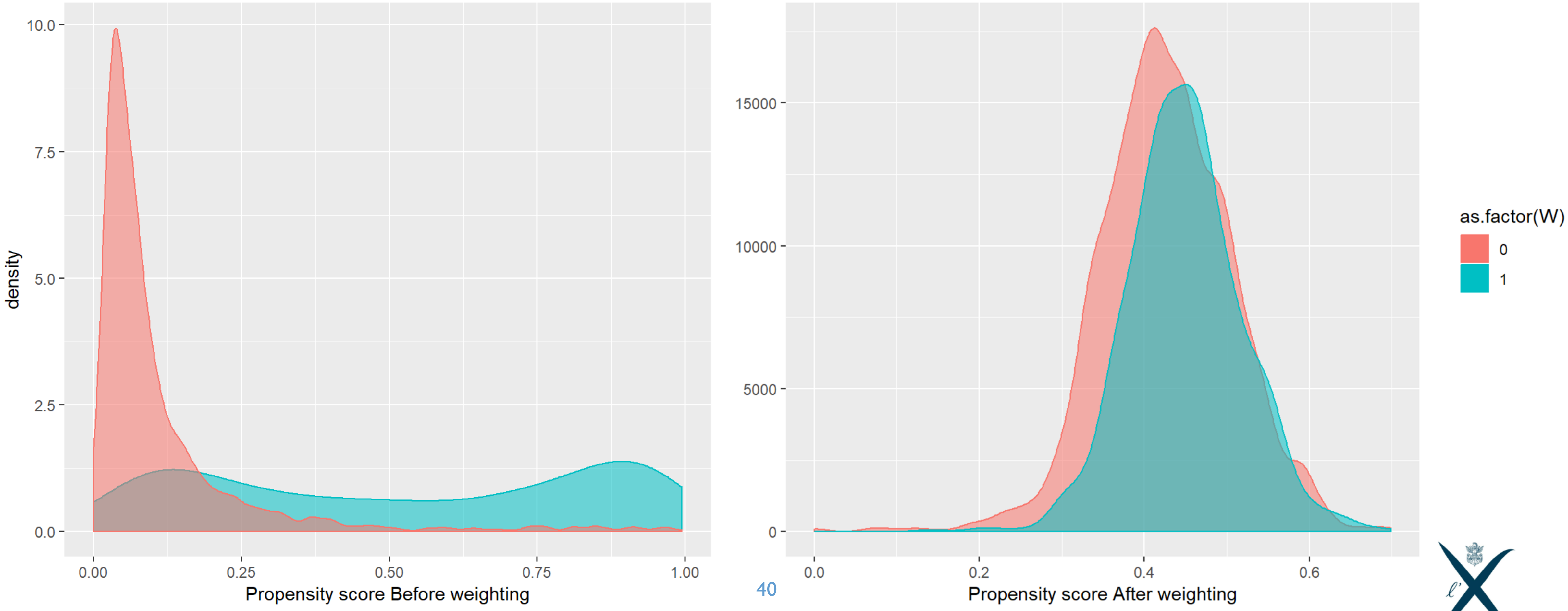
B. Inverse Propensity Weighting

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

$$e(x) = \mathbb{P}\left[W_i = 1 \,\middle|\, X_i = x \right]$$

III. Causal Inference

B. Inverse Propensity Weighting



III. Causal Inference

B. Inverse Propensity Weighting

- IPW estimator :
  - e : Logistic regression Acide tranexamique ~ X (confounder factors)
  - Computing the formula
- Double-robust estimator
- Weighted linear regression DC.Trauma ~ Treatment or not (slope)

III. Causal Inference

B. Inverse Propensity Weighting

```
ATE_table(treatment_data_important, important = TRUE)
```

```
##                    Methode ATE (%) Demi-largeur 95%
## 1                      IPW    1.29               NA
## 2      Weighted regression    8.61             7.31
## 3      Double robust (all)    1.88             1.07
## 4 Double robust (overlap)     2.95             1.53
```

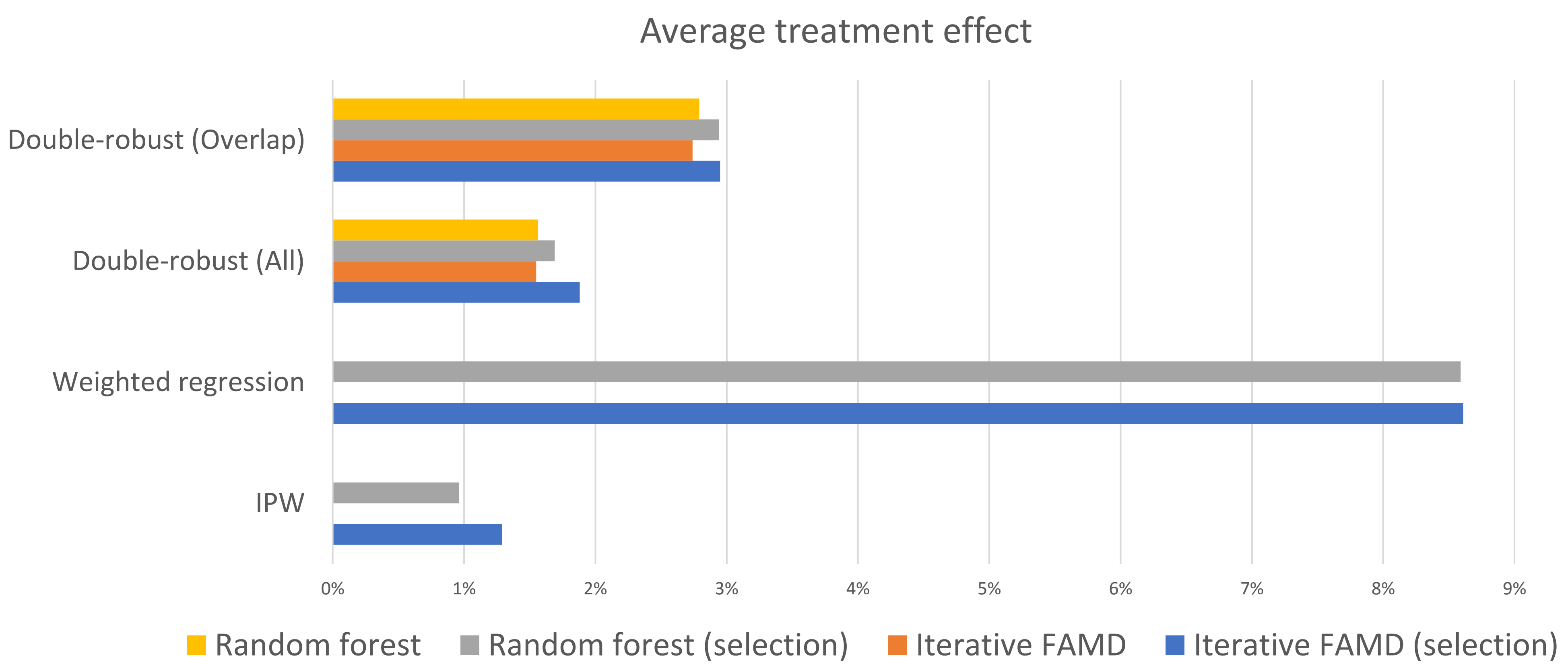
III. Causal Inference

B. Inverse Propensity Weighting

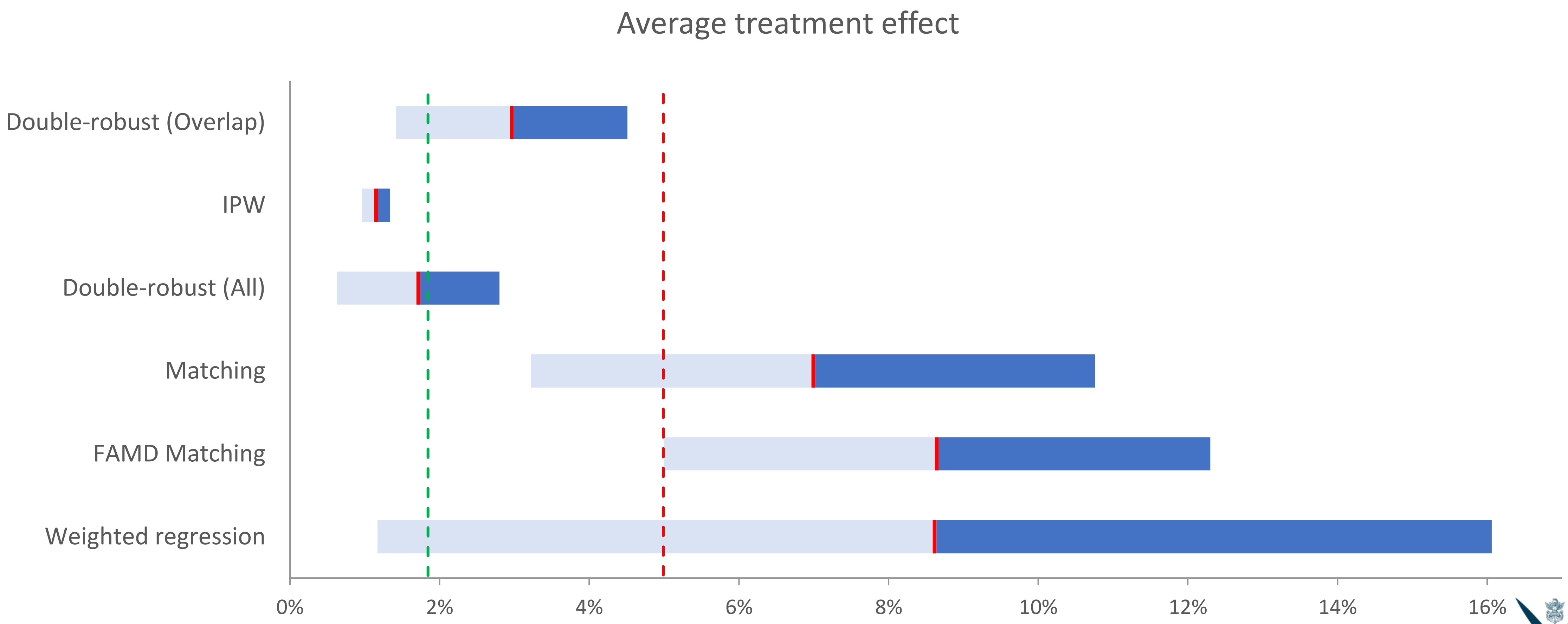
- Random forest ?
- Alcool ? DTC.IP.max ?
- Glasgow.sortie ?

III. Causal Inference

B. Inverse Propensity Weighting



Conclusion



Average treatment effect
Double-robust (Overlap)
IPW
Double-robust (All)
Matching
FAMD Matching
Weighted regression
0%
2%
4%
6%
8%
10%
12%
14%
16%