Combined analysis of CRASH-3 and Traumabase

Bénédicte Colnet*and Imke Mayer†

November 2020

Abstract

This notebook accompanies the review article *Causal inference methods for combining randomizedtrials and observational studies: a review* (2020) and performs the average treatment effect estimation on the preprocessed data for the joint analysis of CRASH-3 and the Traumabase. The input is the merged table of both the randomized controlled trial and the observational study (corresponding to the output of `preprocess.Rmd`). The key functions to perform the analysis below come from the script `estimators.R`.

Contents

Preliminaries

Load libraries

*Inria, benedicte.colnet@inria.fr
†EHESS, imke.mayer@ehess.fr

```r
library(cobalt)
library(ggplot2)
library(dplyr)
library(forcats)
library(misaem)  # glm with missing data
library(boot)  # Import library for bootstrap methods
library(grf)
library(naniar)  # for missing values plots
library(FactoMineR)  # for catdes
library(assertthat)

# Set random generator seed for
# reproducible results
set.seed(123)

# Set data path Define data
# directory for loading
# pre-processed data
data_dir <- "./Data/"
# Define figure directory to
# save figures
fig_dir <- "./Figures/"
# Define results directory to
# save computation results
# (bootstrap)
results_dir <- "./Results/"

# Load estimators and auxiliary
# functions
source("./estimators.R")
source("./catdes_redefined.R")
```

Choose analysis parameters (outcome, stratum, target population, methods, number of bootstrap samples)

```r
outcome_name <- "TBI_Death"  # outcome used in all analyses (either 'Death' or 'TBI_Death')
stratum_name <- "all"  # stratum to consider (either 'all', 'mild_moderate', 'severe', 'any_non_react',
target_population <- "Traumabase_allTBI"  # patients from the Traumabase representing the target popula
use_majorExtracranial <- T  # whether to use `majorExtracranial` as a covariate or to drop it (without
methods <- c("MIA_AIPW_grf", "MICE_AIPW_grf")  # which methods to present at the end in the summary plo
nboot <- 100  # number of bootstrap samples to be used for confidence intervals
```

Recall observational results for the Traumabase

```r
results_rwe
```

```
##     Context         Model Stratum          ATE        STD      CI_inf
## 11      RWD MICE_AIPW_glm     all  0.034736887 0.06360091 -0.08992090
## 12      RWD MICE_AIPW_grf     all -0.001393437 0.02314503 -0.04675769
## 14      RWD  MIA_AIPW_grf     all -0.003688357 0.02544694 -0.05356436
## 111     RWD  MICE_IPW_glm     all  0.150025034 0.04164762  0.06839569
## 121     RWD  MICE_IPW_grf     all  0.110382102 0.02560812  0.06019019
```

```
## 13      RWD   MIA_IPW_grf     all  0.092570846 0.02696681  0.03971591
##         CI_sup
## 11  0.15939467
## 12  0.04397082
## 14  0.04618764
## 111 0.23165437
## 121 0.16057402
## 13  0.14542579
```

CRASH-3 analysis

In this part we load the CRASH3 data and reproduce the results in the publication with the risk ratio (RR). We also provide the results with the ATE to fit the framework of the review.

Note that in CRASH3 several subgroups are considered depending on the severity of the trauma. The biggest treatment effect is observed for the mild to moderate group (where the Glasgow score is 9 or more).

The outcome is the 28-day death due to brain injury (same output is taken in the Traumabase).

Load the pre-processed CRASH-3 and Traumabase data

To pre-process the CRASH-3 and the Traumabase data, first run the notebook `preprocess.Rmd`.

```
# Load incomplete combined data
total <- read.csv(paste0(data_dir,
    "output_preprocess_combined_crash3_TB.csv"),
    row.names = 1)

# Load imputed combined data
# (only Traumabase is imputed)
total_imputed <- read.csv(paste0(data_dir,
    "output_preprocess_combined_crash3_TB_imputed.csv"),
    row.names = 1)

N <- nrow(total)
n <- nrow(total[total$V == 1, ])
m <- nrow(total[total$V == 0, ])

if (target_population == "Traumabase_minorExtracranialTBI" &
    "majorExtracranial" %in% colnames(total)) {
    total <- total[which(total$majorExtracranial ==
        0), ]
    total_imputed <- total_imputed[which(total_imputed$majorExtracranial ==
        0), ]
    total <- dplyr::select(total,
        -c("majorExtracranial"))
    total_imputed <- dplyr::select(total_imputed,
        -c("majorExtracranial"))
}

if (!use_majorExtracranial & "majorExtracranial" %in%
    colnames(total)) {
    total <- dplyr::select(total,
        -c("majorExtracranial"))
    total_imputed <- dplyr::select(total_imputed,
```

```
        -c("majorExtracranial"))
}
```

Check correct outcome

```
## [1] TRUE
```

Recover RCT results presented in CRASH-3 paper

To recover the exact same results as presented in the CRASH-3 paper, we exclude patients with minimal GCS (equal to 3), or bilateral non-reactive pupils (*mydriasis*).

```
## risk_placebo     risk_TXA           RR     lower_ci     upper_ci
##   14.0294511   12.5356125    0.8935212    0.7784344    1.0086080
```

```
##           ATE      lower_ci      upper_ci
## -0.0149383860 -0.0302115118  0.0003347398
```

Last, we take the patients corresponding to the pre-specified stratum

```
## risk_placebo     risk_TXA           RR     lower_ci     upper_ci
##   19.6428571   18.4369603    0.9386089    0.8541480    1.0230698
```

```
##          ATE     lower_ci     upper_ci
## -0.012058969 -0.028137233  0.004019295
```

```
##          ATE     lower_ci     upper_ci
## -0.011066123 -0.025048882  0.002916636
```

Final data set overview

Size

The final size of the data.frame is 17416, with

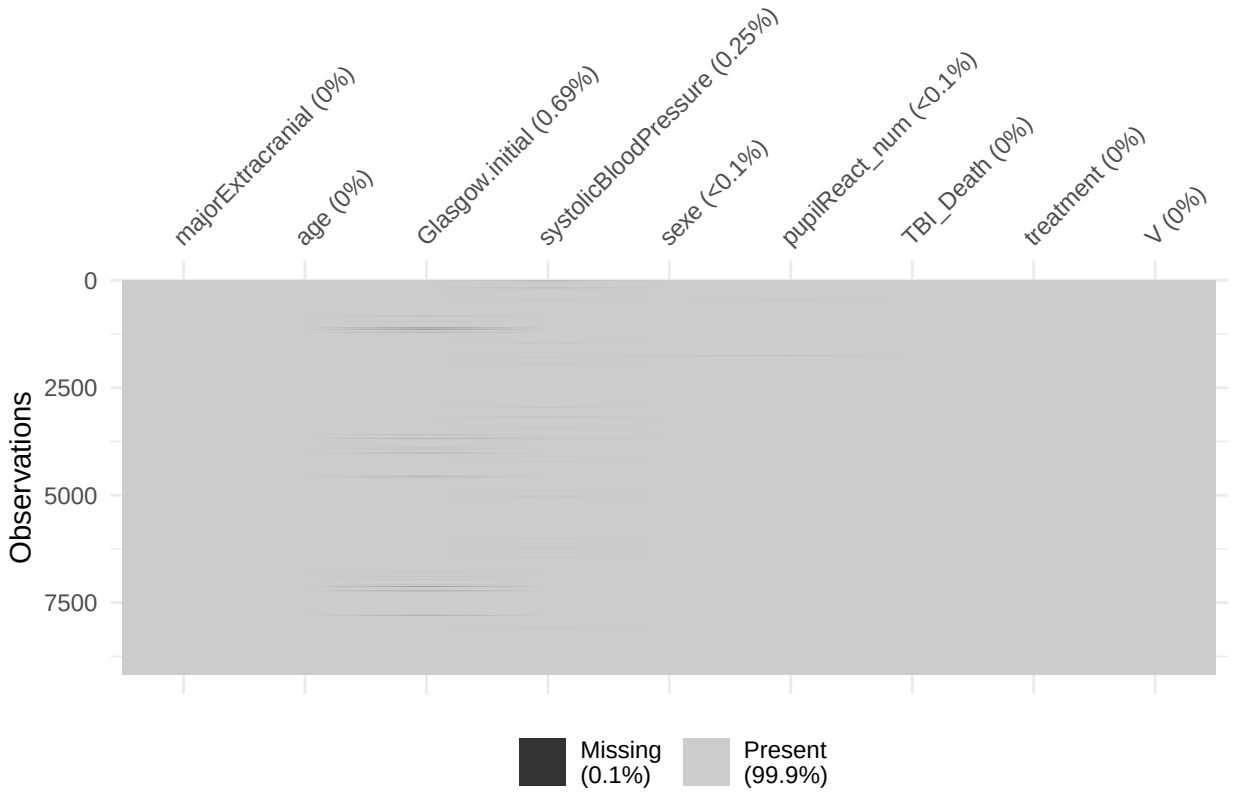
- 9168 observations from CRASH-3, and
- 8248 observations from the Traumabase.

```
##      Treated
## Study    0    1
##     0 7565  683
##     1 4536 4632
```

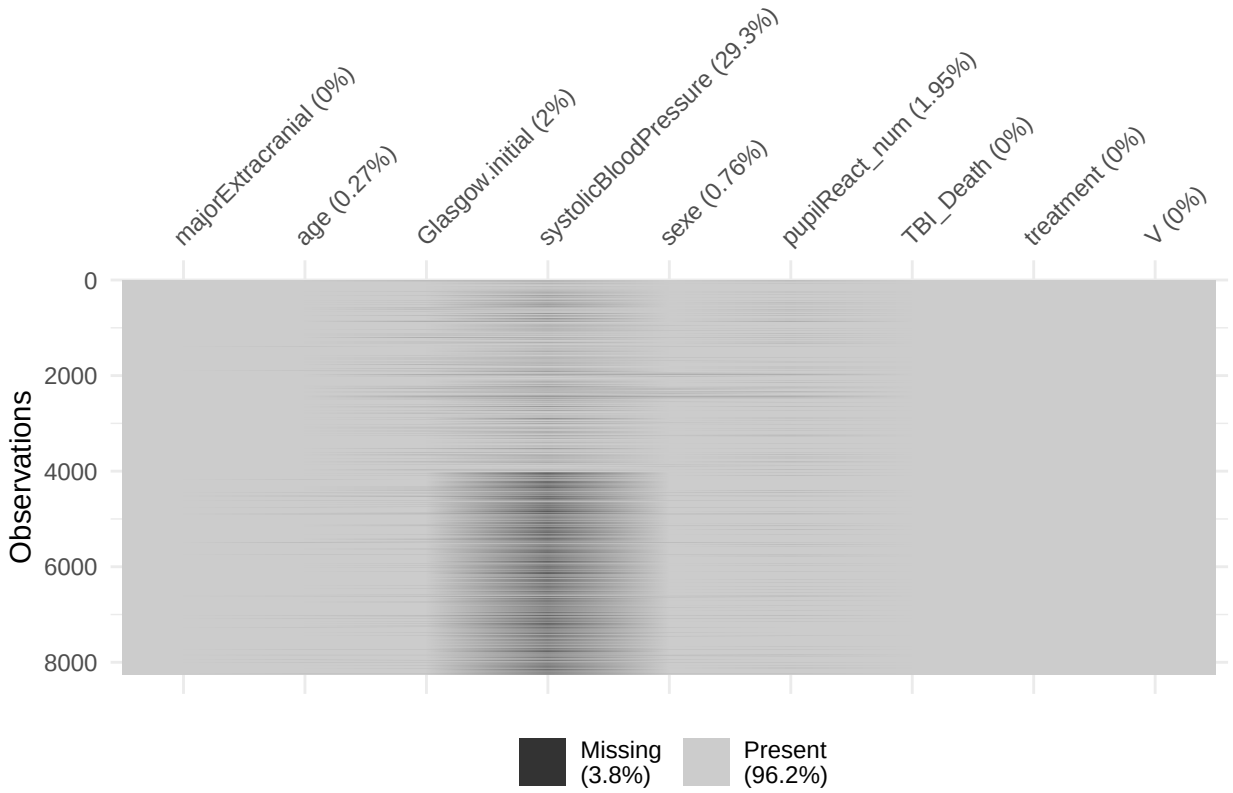
```
##      Died
## Study    0    1
##     0 6837 1411
##     1 7423 1745
```

Missing values

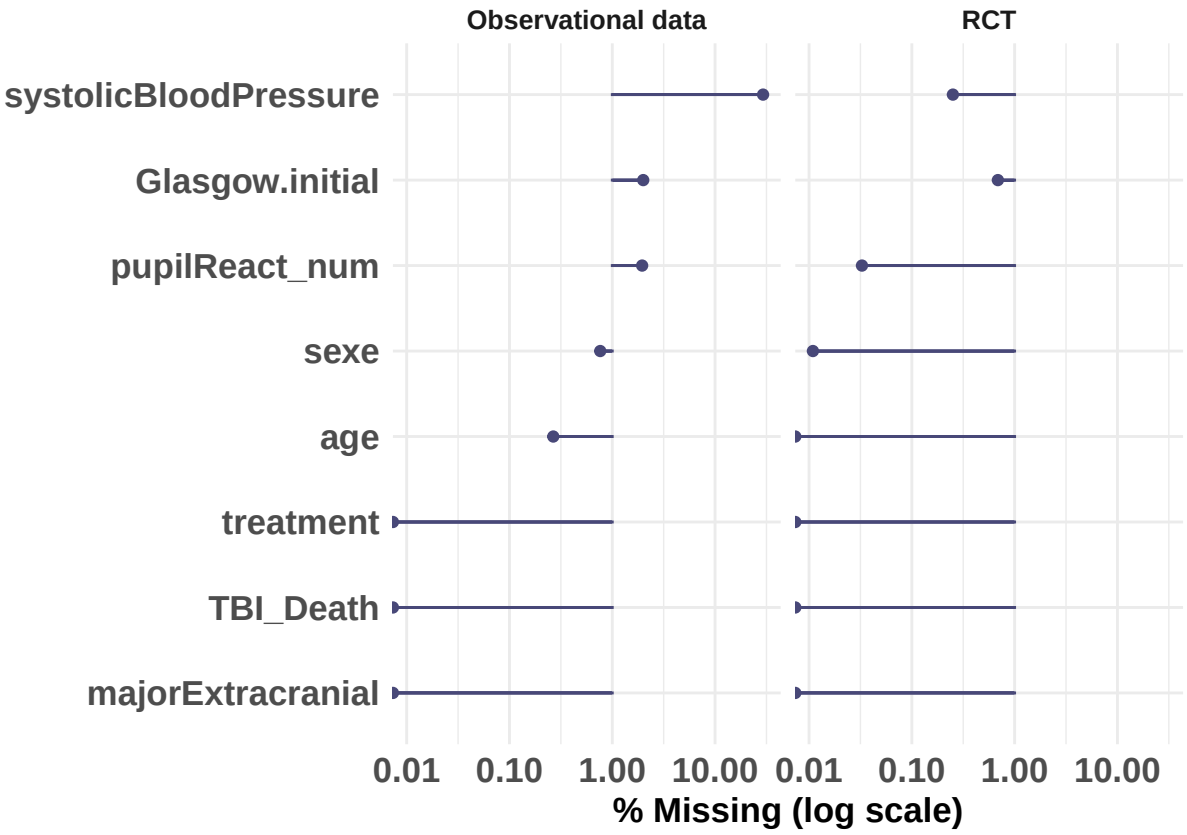
First, note that the RCT contains nearly no missing values.



The Traumabase subset taken contains missing values, it explains why the estimators for transporting the ATE have to be adapted to take into account these missing values.



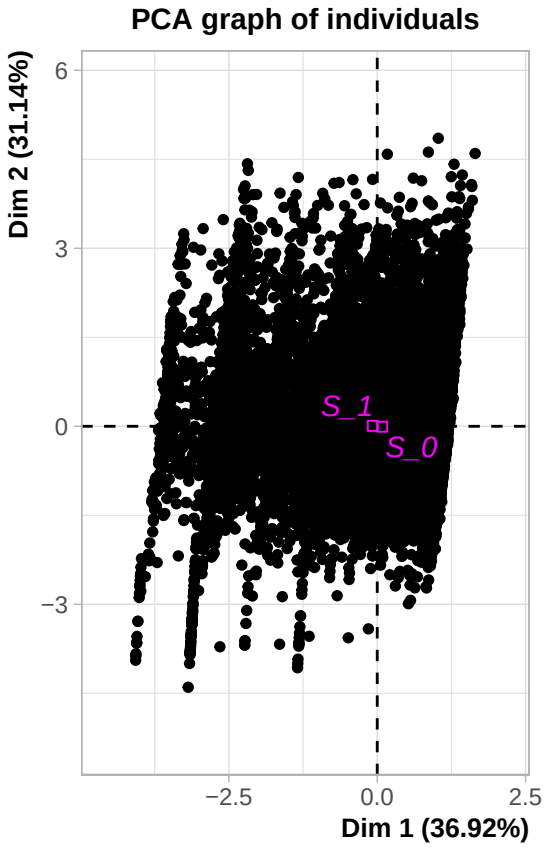
Alternatively we can plot the barplots of percentage of missing values

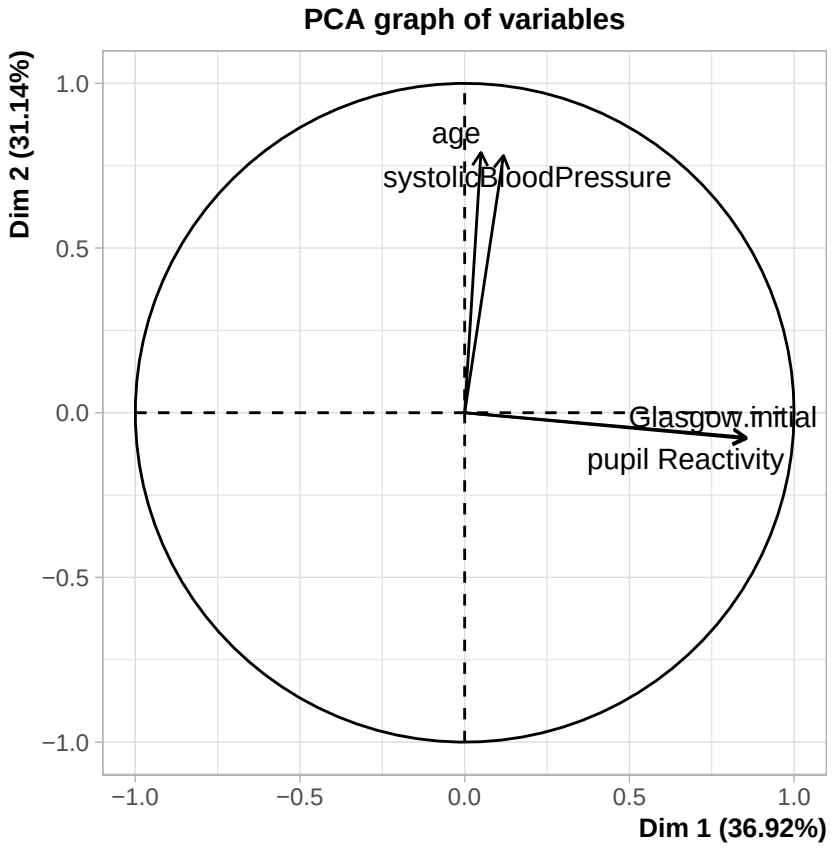


ACP

```
## pdf
##   2
```

```
## pdf
##   2
```





Analysis

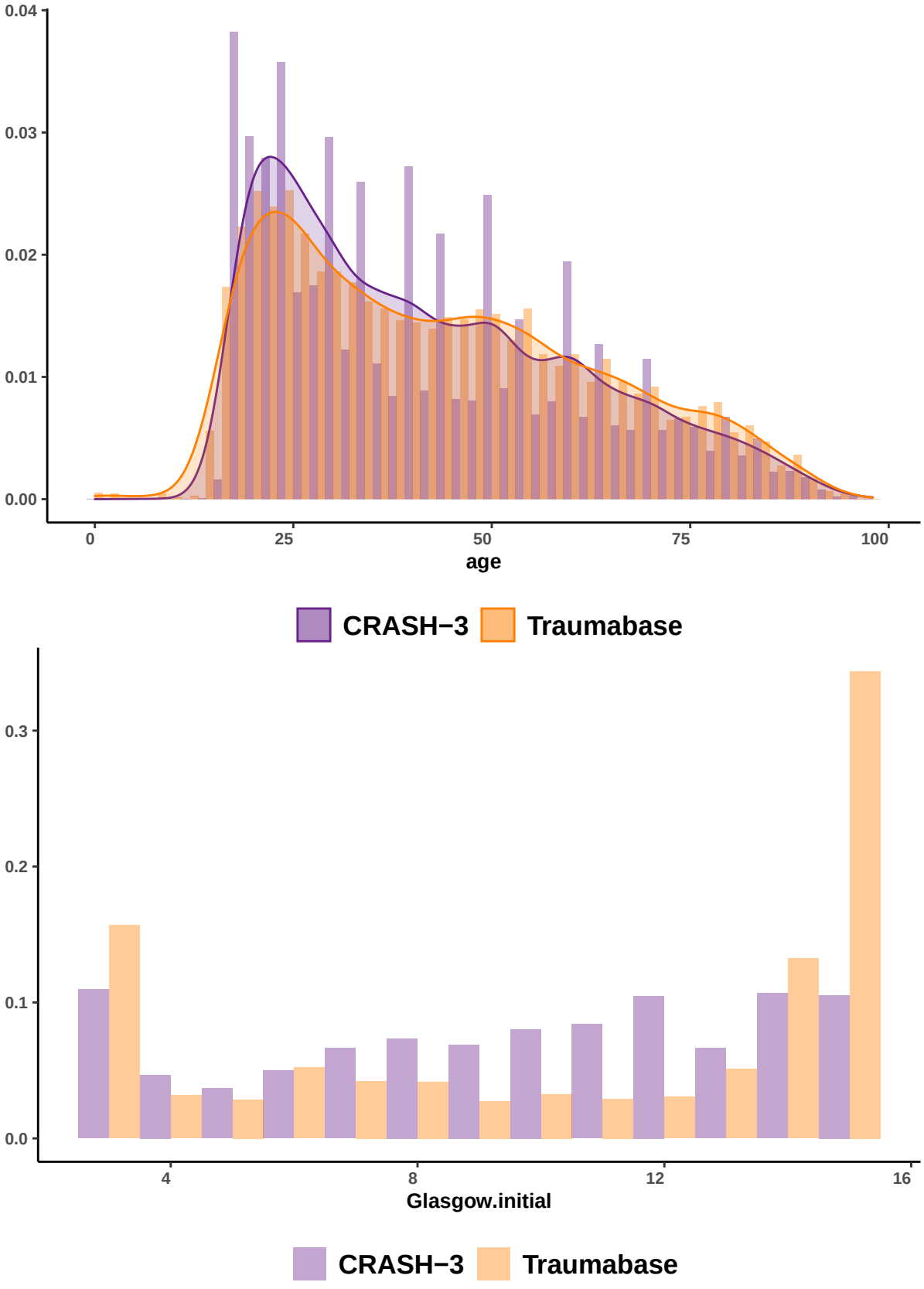
ATE using only the Traumabase data

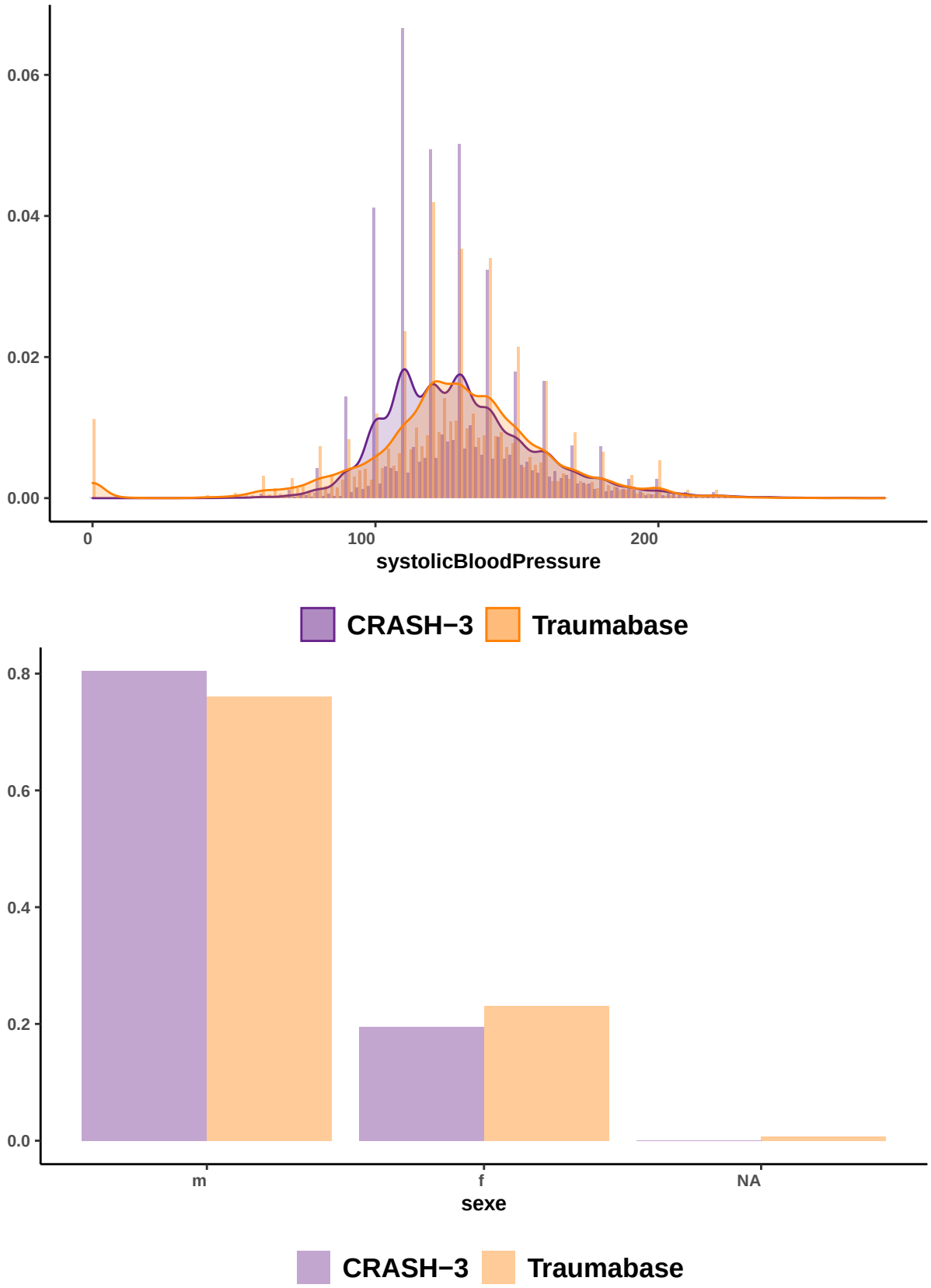
We can first use a naive difference in means, for which we can conclude that TXA increases death (treatment bias, confounding bias, Simpson's paradox).

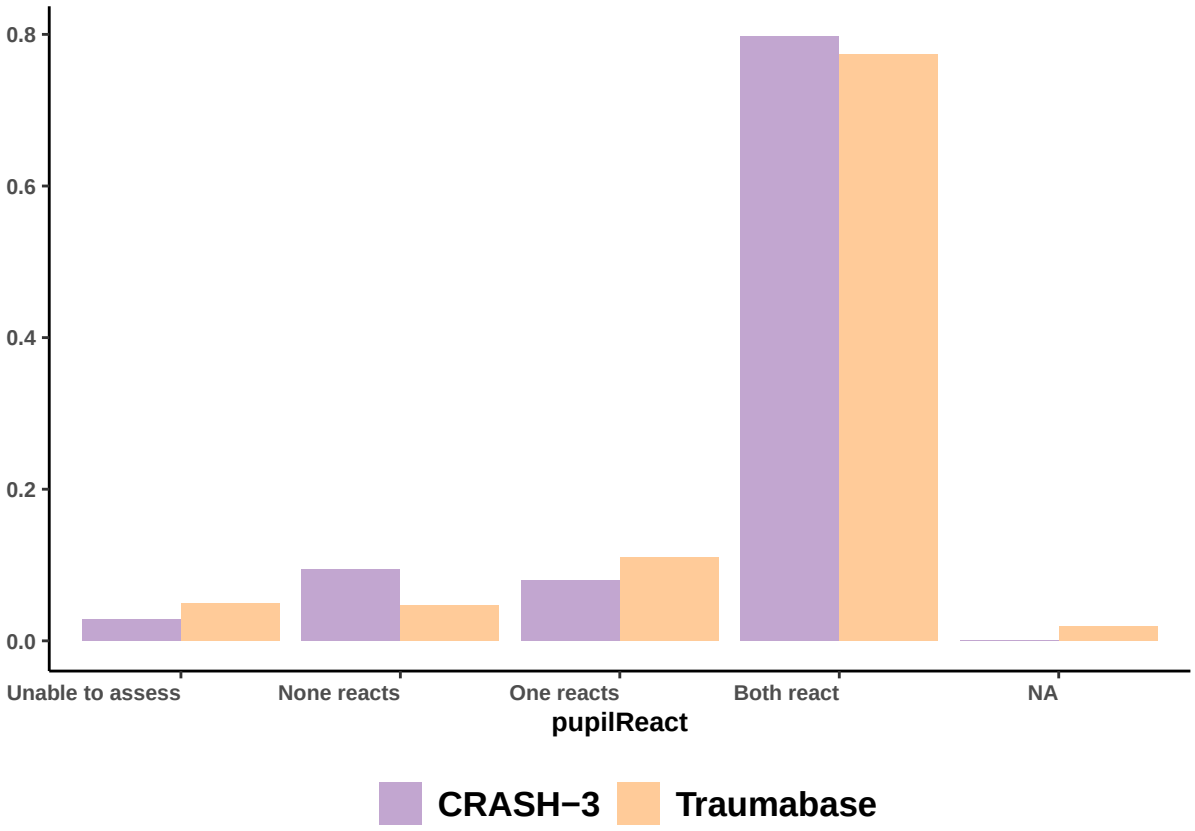
```
## risk_placebo     risk_TXA           RR     lower_ci     upper_ci
##    15.756775    32.064422     2.034961     1.913998     2.155924
```

```
##       ATE  lower_ci  upper_ci
## 0.1630765 0.1270732 0.1990797
```

Distributional shift visualization



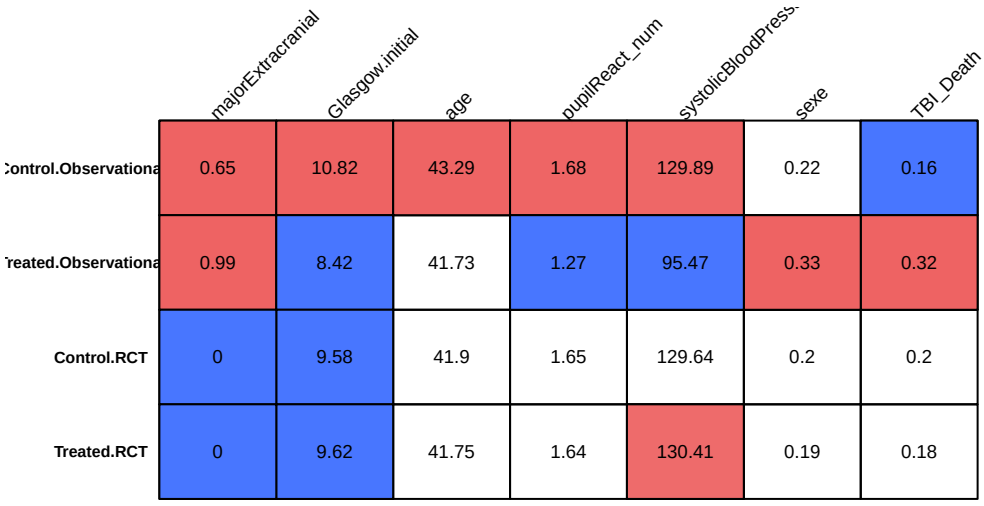

0.06
0.04
0.02
0.00
0
100
200
systolicBloodPressure
CRASH−3
Traumabase
0.8
0.6
0.4
0.2
0.0
m
f
NA
sexe
CRASH−3
Traumabase




```
## pdf
##   2
```

	majorExtracranial	Glasgow.initial	age	pupilReact_num	systolicBloodPress	sexe	TBI_Death
control.Observationa	0.65	10.81	43.29	1.67	130.18	0.22	0.16
reated.Observationa	0.99	8.42	41.73	1.27	100.14	0.33	0.32
Control.RCT	0	9.58	41.9	1.65	129.64	0.2	0.2
Treated.RCT	0	9.62	41.75	1.64	130.41	0.19	0.18

```
## pdf
##   2
```



ATE transport from CRASH3 to the Traumabase

Note that when using the original Traumabase, the standard estimators (IPSW, G-formula, AIPSW) need to be adapted to handle missing values that are not missing completely at random (MCAR).

We propose two ways of addressing this handling of missing values:

- Logistic regression via Expectation Maximization (EM) that explicitly handles missing values that are missing at random (MAR)
- Generalized random forests that consider that missing values are potentially informative, this is achieved through the *missing incorporated in attributes* (MIA) criterion

Point estimates

We start by applying all estimators (implemented in the `estimators.R` script) on the `total` data.frame.

```
## Iteration of SAEM:
## 50 Iteration of SAEM:
## 50 IPSW startIteration of SAEM:
## 50 AIPSW start
```

```
## IPSW startAIPSW start
```

```
##      ipsw_hat ipsw.norm_hat  gformula_hat     aipsw_hat
##    0.08749251    0.03944023    0.15093523    0.08682439
```

```
##      ipsw_hat ipsw.norm_hat  gformula_hat     aipsw_hat
##   0.324628153   0.157280540  -0.005716057   0.088107606
```

Confidence interval estimation (Bootstrap)

The confidence intervals are estimated via non-parametric stratified bootstrap.

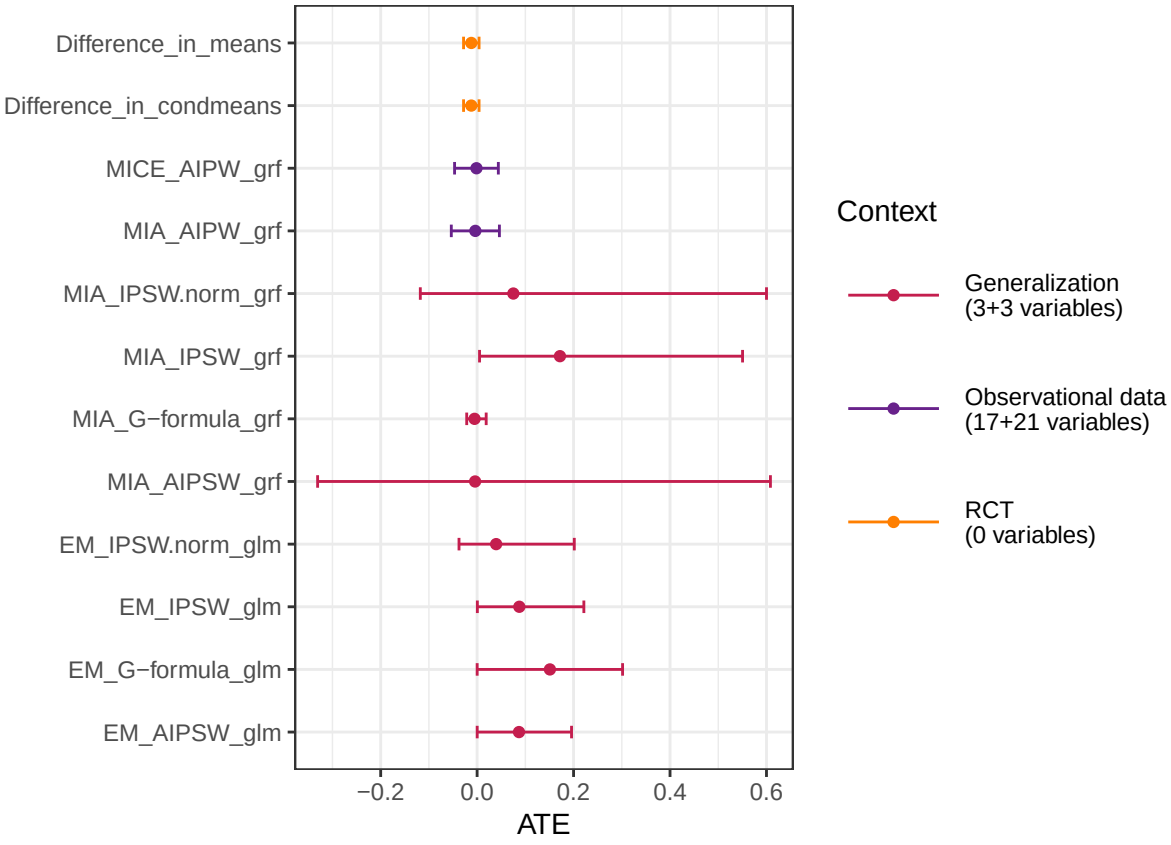
```
stratified_bootstrap <- function(DF,
    nboot = 100, estimator, method,
    outcome_name = "TBI_Death",
    complete_cases = FALSE, ampute = FALSE,
    verbose = FALSE) {

    estimands <- c()
```

```
    ct_fail <- 0
    if (verbose)
        cat("Iteration ")
    for (i in 1:nboot) {
        if (verbose)
            cat(paste0(i, " "))

        # random resamples from RCT
        n = nrow(DF[DF$V == 1,
            ])
        index_RCT = sample(1:n,
            n, replace = TRUE)

        # random resamples from RWD
        m = nrow(DF[DF$V == 0,
            ])
        index_RWD = sample(1:m,
            m, replace = TRUE)

        # new data set
        RCT_RWD <- rbind(DF[which(DF$V ==
            1), ][index_RCT, ],
            DF[which(DF$V == 0),
                ][index_RWD, ])

        # ampute values to keep similar
        # fraction of NA in RWD part of
        # the data
        if (ampute) {
            prop_miss_RWD <- sapply(DF[DF$V ==
                0, ], function(x) mean(is.na(x)))
            for (j in 1:ncol(DF)) {
                prop_miss_boot <- mean(is.na(RCT_RWD[which(RCT_RWD$V ==
                  0), j]))
                if (prop_miss_RWD[j] >
                  0.1 & prop_miss_RWD[j] >
                  prop_miss_boot) {
                  idx_miss <- which(is.na(RCT_RWD[which(RCT_RWD$V ==
                    0), j]))
                  idx_new_miss <- sample(m -
                    length(idx_miss),
                    floor(m * (prop_miss_RWD[j] -
                      prop_miss_boot)),
                    replace = F)
                }
            }
        }

        # estimation
        estimand <- NULL

        try(estimand <- unlist(estimator(RCT_RWD,
            outcome_name = outcome_name,
```

```
            method = method, complete_cases = complete_cases)))
        if (!is.null(estimand)) {
            estimands <- rbind(estimands,
                data.frame(t(estimand)))
        } else {
            cat(paste0(i, "-> fail, "))
            ct_fail <- ct_fail +
                1
        }
    }
    if (as.character(substitute(estimator)) ==
        "compute_ipsw") {
        estimands <- data.frame(estimands)
        colnames(estimands) <- paste0(c("IPSW_",
            "IPSW.norm_"), method)
    }
    if (as.character(substitute(estimator)) ==
        "compute_all") {
        estimands <- data.frame(estimands)
        colnames(estimands) <- paste0(c("IPSW_",
            "IPSW.norm_", "G-formula_",
            "AIPSW_"), method)
    }
    print(paste0("Number of failed iterations: ",
        ct_fail))
    return(estimands)
}
```

Plot of the final results

On incomplete Traumabase



On imputed Traumabase