Data preprocessing for joint analysis of CRASH-2 and/or CRASH-3 with Traumabase

Imke Mayer* Other contributors[^1]

March 2021

Abstract

This notebook is a complement to the data analysis presented in the submitted article *Transporting treatment effect with missing attributes* (2021) and performs the data preprocessing for the joint analysis of CRASH-2, CRASH-3 and the Traumabase. It takes as an entry the raw data from all data sets and binds them with proper covariates. The output is the combined data with the raw Traumabase data (with missing values kept). Another similar data frame but with the imputed Traumabase is also produced.

Contents

[^1] Other contributors to this notebook through previous collaborations or active discussions: Bénédicte Colnet, Julie Josse, François-Xavier Ageron

*EHESS, imke.mayer@ehess.fr

Load libraries

```
library(readxl)  # for reading xlsx
library(dplyr)  # for data.frame handling
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(reshape2)  # for data.frame handling
library(ggplot2)  # for general plots
library(FactoMineR)  # for factorial analysis and PCA and catdes plot
library(naniar)  # for missing values plots
source("catdes_redefined.R")

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

```
# Define data directory for
# loading raw data
traumabase_rawdata_dir <- "~/Documents/TraumaMatrix/TraumaMatrixPipeline/"
crash_rawdata_dir <- "~/Documents/TraumaMatrix/CausalInference/CRASH/"

data_dir <- "./data/"  # where to store output
fig_dir <- "./figures/"
```

CRASH-2 and CRASH-3

Data loading

In this part we load the CRASH-2 and the CRASH-3 data and merge them with the code of treated or placebo that come from separate files. These files correspond to what we received from the CRASH-2 and CRASH-3 principal investigators.

```
# Load CRASH2 data
rawData_CRASH2 <- read_xlsx(paste0(crash_rawdata_dir,
    "CRASH-2_Data_anonymised_dataset_w.boxpack.xlsx"),
    na = c("", "NR", "NA", "NF",
        "IMP", "ND"), )

# Load CRASH3 data
rawData_CRASH3 <- read_xlsx(paste0(crash_rawdata_dir,
    "CRASH-3_dataset_anonymised_for_Freebird.xlsx"),
    na = c("", "NR", "NA", "NF",
        "IMP", "ND"), )
```

```
## [1] "Raw CRASH-2 data contains following number of observations: 20207"
```

```
## [1] "Raw CRASH-2 data contains following number of eligible observations: 20207 (10060 TXA and 10067
```

```
## [1] "Raw CRASH-3 data contains following number of observations: 12743"
```

```
## [1] "Raw CRASH-3 data contains following number of eligible observations: 12663"
```

We observe that in CRASH-3, a column precises which patients are eligible or not.

Treatment assignment and outcome loading

```
## [1] "Treatment code for CRASH-2 contains following number of observations: 20207"
```

```
## [1] "Merged data table number of observations: 20207"
```

```
## [1] "Treatment code for CRASH-3 contains following number of observations: 16488"
```

```
## [1] "Merged data table number of observations: 12737"
```

We also load the country information.

```
## [1] "Site and country codes for CRASH-2 contain following number of observations: 20207"
```

```
## [1] "Merged data table number of observations: 20207"
```

```
## [1] "Site and country codes for CRASH-3 contains following number of observations: 12737"
```

```
## [1] "Merged data table number of observations: 12737"
```

And add a continent variable

```
CRASH2$continent <- case_when(CRASH2$scountry %in%
    c("Belgium", "Albania", "United Kingdom",
        "Spain", "Italy", "Czech Republic",
        "Slovakia", "Serbia & Montenegro") ~
    "Europe", CRASH2$scountry %in%
    c("Georgia", "Malaysia", "Saudi Arabia",
        "Japan", "Iran", "Sri Lanka",
        "Nepal", "Bangladesh",
        "Thailand", "Iraq", "Indonesia",
        "India", "Singapore", "China") ~
    "Asia-Oceania", CRASH2$scountry %in%
    c("Jamaica", "Colombia", "El Salvador",
        "Mexico", "Canada", "Peru",
        "Argentina", "Cuba", "Ecuador") ~
    "America", CRASH2$scountry %in%
    c("Cameroon", "Nigeria", "Zambia",
        "Egypt", "Kenya", "South Africa",
        "Tunisia", "Ghana", "Tanzania") ~
    "Africa", CRASH2$scountry %in%
    c("Australia") ~ "Asia-Oceania")

CRASH3$continent <- case_when(CRASH3$country %in%
    c("United Kingdom", "Spain",
        "Albania", "Italy", "Romania",
        "Ireland", "Slovenia") ~
    "Europe", CRASH3$country %in%
    c("Georgia", "Malaysia", "United Arab Emirates",
        "Japan", "Afghanistan",
        "Pakistan", "Nepal", "Myanmar (Burma)",
        "Cambodia", "Iraq", "Indonesia") ~
    "Asia-Oceania", CRASH3$country %in%
    c("Jamaica", "Colombia", "El Salvador",
```

```
        "Mexico", "Canada") ~ "America",
    CRASH3$country %in% c("Cameroon",
        "Nigeria", "Zambia", "Egypt",
        "Kenya") ~ "Africa", CRASH3$country %in%
        c("Papua New Guinea") ~
        "Asia-Oceania")
```

Homogeneization of variable names

We rename the CRASH2 columns to be concordant with those of CRASH3.

Check with MDs for equivalence between:

- `intraCranialBleeding` (Crash3) and `bheadinj` (Crash2)
- `neuroHaemEvac`, `neuroOther` (Crash3) and `bneuro` (Crash2)
- `majorExtracranial` (Crash3) and `bbleed` (Crash2)
- `doseOne`, `doseTwo` (Crash3) and `bloading`, `bmaint` (Crash2)

```
colnames(CRASH2) <- c("code", "crash2_ientryid.x",
    "crash2_isource", "siteId",
    "sex", "age", "timeSinceInjury",
    "crash2_iinjurytype", "systolicBloodPressure",
    "crash2_irr", "crash2_icc",
    "crash2_ihr", "gcsEyeOpening",
    "gcsMotorResponse", "gcsVerbalResponse",
    "totalGcs", "timerandtodeath",
    "causeDeath", "causeDeathOther",
    "dischargeStatus", "daysrandtodischarge",
    "levelOfFunctioning", "daysIcu",
    "intraCranialBleeding", "neuroOther",
    "crash2_bchest", "crash2_babdomen",
    "crash2_bpelvis", "pe", "dvt",
    "stroke", "crash2_bbleed",
    "myocardialInfarction", "gastroIntestinal",
    "doseOne", "doseTwo", "crash2_btransf",
    "crash2_ncell", "crash2_nplasma",
    "crash2_nplatelets", "crash2_ncryo",
    "crash2_bvii", "box.x", "pack.x",
    "crash2_ientryid.y", "box.y",
    "pack.y", "treatment", "country",
    "continent")
```

Certain variables are only present in the CRASH2 trial:

```
colnames(CRASH2)[which(startsWith(colnames(CRASH2),
    "crash2_"))]
```

```
##  [1] "crash2_ientryid.x"  "crash2_isource"     "crash2_iinjurytype"
##  [4] "crash2_irr"         "crash2_icc"         "crash2_ihr"
##  [7] "crash2_bchest"      "crash2_babdomen"    "crash2_bpelvis"
## [10] "crash2_bbleed"      "crash2_btransf"     "crash2_ncell"
## [13] "crash2_nplasma"     "crash2_nplatelets"  "crash2_ncryo"
## [16] "crash2_bvii"        "crash2_ientryid.y"
```

Others only exist in the CRASH3 trial:

```r
setdiff(colnames(CRASH3), colnames(CRASH2))
```

```
##  [1] "sbpStatus"             "gcsTiming"             "pupilReact"
##  [4] "majorExtracranial"     "epidural"              "subdural"
##  [7] "subarachnoid"          "parenchymal"           "intraventricular"
## [10] "eligible"              "consent"               "causeDeathOtherMeddra"
## [13] "eyeOpening"            "communicationAbility"  "motorResponse"
## [16] "feeding"               "toileting"             "grooming"
## [19] "employability"         "walking"               "washingDressing"
## [22] "painDiscomfort"        "anxietyDepression"     "agitationAggression"
## [25] "fatigue"               "neuroHaemEvac"         "estBloodLoss"
## [28] "renalFailure"          "sepsis"                "seizure"
## [31] "otherComplication"     "MedDRALLT1"            "MedDRALLTCode1"
## [34] "MedDRALLT2"            "MedDRALLTCode2"        "MedDRALLT3"
## [37] "MedDRALLTCode3"        "MedDRALLT4"            "MedDRALLTCode4"
## [40] "MedDRALLT5"            "MedDRALLTCode5"
```

Important variables out of this list of non-available variables in CRASH-2 (for eligibility criterion)

- `pupilReact`
- `majorExtracranial`

Question: what equivalence of `majorExtracranial` should we use for CRASH-2?

- `bbleed` (operation for bleeding, yes/no),
- `btransf` (blood products transfusion, yes/no),
- `ncell` (number of units of red cell products transfused in 28 days, between 0 and 60),
- `nplasma` (number of units of fresh frozen plasma transfused in 28 days, between 0 and 60),
- `nplatelets` (number of units of platelets transfused in 28 days, between 0 and 87),
- `ncyro` (number of units of cryoprecipitate transfused in 28 days, between 0 and 61)
- `bvii` (recombinant Factor VIIa given in 28 days, yes/no)

**For now we will take `bbledd` and `btransf` as a proxy or equivalent of `majorExtracranial`.

Question: and for `pupilReact`? The only proxy we have is GCS eye opening.

Outcome and treatment

Note that the outcome is the 28-day death due to brain injury (and not all deaths).

CRASH-2

```r
# Death to binary (1=death)
CRASH2$Death <- ifelse(is.na(CRASH2$timerandtodeath),
    0, 1)

# Brain injury death related to
# binary (1=tbi-death) ---> The
# outcome of interest In
# CRASH2, cause is coded with
# integers 1 = Bleeding, 2 =
# Head injury 3 = Myocardial
# infarction 4 = Stroke 5 =
# Pulmonary embolism 6 = Multi
# organ failure 7 = Other
CRASH2$TBI_Death <- ifelse((is.na(CRASH2$causeDeath) |
```

```
    CRASH2$causeDeath != 2), 0,
    1)

# Treatment as a binary
# variable
CRASH2$treatment <- ifelse(CRASH2$treatment ==
    "Placebo", 0, 1)
```

```
table(CRASH2$Death, CRASH2$TBI_Death,
    dnn = c("Death", "Head injury Death"))
```

```
##       Head injury Death
## Death     0     1
##     0 17121     0
##     1  1861  1225
```

Around 40% of the deaths in the CRASH-2 trial are (mainly) caused by head injury.

CRASH-3

```
# Death to binary (1=death)
CRASH3$Death <- ifelse(is.na(CRASH3$timerandtodeath),
    0, 1)

# Brain injury death related to
# binary (1=tbi-death) ---> The
# outcome of interest
CRASH3$TBI_Death <- ifelse((is.na(CRASH3$causeDeath) |
    CRASH3$causeDeath != "Head injury"),
    0, 1)

# Treatment as a binary
# variable
CRASH3$treatment <- ifelse(CRASH3$treatment ==
    "Placebo", 0, 1)
```

```
table(CRASH3$Death, CRASH3$TBI_Death,
    dnn = c("Death", "Head injury Death"))
```

```
##       Head injury Death
## Death     0     1
##     0 10178     0
##     1   221  2338
```

In the CRASH-3 trial, over 90% of the deaths are (mainly) caused by head injury.

Traumabase

Data loading

```
rawData_Traumabase <- read.csv(paste0(traumabase_rawdata_dir,
    "4_computed_dataset.csv"),
    na.strings = c("", "NR", "NA",
        "NF", "IMP", "ND"), sep = ",")
```

```
print(paste0("Raw traumabase data contains following number of observations: ",
    nrow(rawData_Traumabase)))
```

```
## [1] "Raw traumabase data contains following number of observations: 20037"
```

Outcome and treatment

We also define treatment and outcome on the Traumabase.

The treatment is considered given when the column *Acide.tranexamique* is equal to "Oui", and if "No" or missing value is present it is considered as no-treatment.

The outcome in the Traumabase is brain injury related death, with column *Cause.du.décès* equals to "LATA", or "Mort encéphalique", or "Trauma cranien", or "Défaillance multi-viscérale". Note that it is not all death, and this outcome matches the definition of the CRASH-3 outcome.

```
# Traumabase outcome
rawData_Traumabase$Death <- rawData_Traumabase$Décès
rawData_Traumabase$Death <- ifelse(rawData_Traumabase$Death ==
    "Non" | is.na(rawData_Traumabase$Death),
    0, 1)

# TBI-related deth definition
# according to the doctors
rawData_Traumabase$TBI_Death <- ifelse(rawData_Traumabase$Cause.du.décès %in%
    c("LATA", "Mort encéphalique",
        "Trauma cranien", "Défaillance multi-viscérale"),
    1, 0)


# Traumabase treatment
rawData_Traumabase$treatment <- ifelse(is.na(rawData_Traumabase$Acide.tranexamique) |
    rawData_Traumabase$Acide.tranexamique ==
        "Non", 0, 1)
```

For consistency with the CRASH-2 and CRASH-3 data, we add a country and continent column to the Traumabase which are naturally constant.

```
rawData_Traumabase$country <- "France"
rawData_Traumabase$continent <- "Europe"
```

Common set of covariates

Covariates accounting for patient inclusion into CRASH-2 trial

Taken from the CRASH-2 protocol:

CRASH-2 population: Adult trauma patients with ongoing significant haemorrhage or at risk of significant haemorrhage, within 8 hours of injury, except those for whom antifibrinolytic agents are thought to be clearly indicated or clearly contra-indicated.

Inclusion criteria: All trauma patients with ongoing significant haemorrhage (systolic blood pressure less than 90 mmHg and/or heart rate more than 110 beats per minute), or who are considered to be at risk of significant haemorrhage, and are within 8 hours of the injury, are eligible for trial entry if they appear to be at least 16 years old. Although entry is allowed up to 8 hours from injury, the earlier that patients can be treated the better.

Exclusion criteria: The fundamental eligibility criterion is the responsible doctor's 'uncertainty' as to whether or not to use an antifibrinolytic agent in a particular adult with traumatic haemorrhage. Patients for whom the responsible doctor considers there is a clear indication for antifibrinolytic therapy should not be randomised. Likewise, patients for whom there is considered to be a clear contraindication to antifibrinolytic therapy (such as, perhaps, those who have clinical evidence of a thrombotic disseminated intravascular coagulation) should not be randomised. Where the responsible doctor is substantially uncertain as to whether or not to use an antifibrinolytic, all these patients are eligible for randomisation and should be considered for the trial. There are no other pre-specified exclusion criteria.

10096 patients were allocated to tranexamic acid and 10115 to placebo, of whom 10060 and 10067, respectively, were analysed.

When listing the inclusion/exclusion criteria, we find less "eligible" patients but this could be because of a wrong definition/interpretation of "risk of significant haemorrhage":

```
# inclusion and exclusion
# criteria
sum(!is.na(CRASH2$timeSinceInjury) &
    CRASH2$timeSinceInjury <= 8 &
    !is.na(CRASH2$age) & CRASH2$age >=
    16 & (!is.na(CRASH2$systolicBloodPressure) |
    !is.na(CRASH2$crash2_ihr) |
    !is.na(CRASH2$crash2_btransf)) &
    ifelse(is.na(CRASH2$systolicBloodPressure <
        90 | CRASH2$crash2_ihr >
        110 | CRASH2$crash2_btransf),
        FALSE, CRASH2$systolicBloodPressure <=
            90 | CRASH2$crash2_ihr >=
            110 | CRASH2$crash2_btransf ==
            1))
```

```
## [1] 15639
```

Covariates accounting for patient inclusion into CRASH-3 trial

Extra-cranial bleeding

In CRASH-3, one of the eligibility criteria is no major extra-cranial bleeding. The feature is called "majorExtracranial" in the CRASH3 trial with a Yes/No answer. We binarize this data with Yes corresponding to 1, and No to 0 (this is the standard procedure we apply all along this part for binary covariate).

The equivalent variable in the Traumabase is chosen based on $CGR.6h > 3$ or if variable *colloides* ou *cristallides* > 0 (corresponding to a major extracranial bleeding). These conditions determining presence or absence of an major extracranial bleeding have been decided with the Traumbase doctors.

Age

Only adults are said to be eligible in CRASH3, but we observe that children are included. We record 58 values with age lower than 18 years. Some of them are eligible, others are not. Note that we also record 12737 observations with missing data in the age column.

TBI

The Traumabase contains this feature, we just rename it and binarize it (1 for TBI, and 0 for no TBI). In the CRASH3 trial we made it correspond with intraCranialbleeding feature which as Yes, No and, No CT scan available. To conclude on an intracranial bleeding with no CT scan, we consider there is a TBI since the patient is said to be eligible in CRASH3.

GSC

The Traumabase contains the *Glasgow.initial* covariate (a discrete, range: [3, 15]), and corresponds to Initial Glasgow Coma Scale (GCS) on arrival on scene of enhanced care team and on arrival at the hospital (GCS = 3: deep coma; GCS = 15: conscious and alert). In CRASH 3 data it corresponds to 3 variables that have to be summed. It is also important to note that some Glasgow score are taken after intubation, so not initially. As only one GSC values is mentioned per observation, we keep all the values and consider it initial value.

Other covariates

In this part we also include other covariates that are in the baseline (so that probably have an impact on the treatment effect and the outcome), and other "easy" covariates to map. We include systolic blood pressure, sex, and also pupils reactivity for CRASH-3 and Glasgow.initial, sex, central_capillary, respiratory_rate and type_injury for CRASH-2.

```
# vector of variables that are
# not observed in all 3 data
# frames
consistently_missing_in_1 <- c()
consistently_missing_in_2 <- c()

# the vector that stores the
# variable name relative to
# trial inclusion
crash2_trial_eligibility <- c()
crash3_trial_eligibility <- c()


# the vector that stores
# additional common variables
# not said to be relative to
# the trial inclusion criteria,
# but still mentioned in the
# CRASH-3 table 1 results and
# CRASH-2 table 1 results
crash2_outcome_impact <- c()
crash3_outcome_impact <- c()

# Extracranial bleeding ->
# majorExtracranial
rawData_Traumabase$majorExtracranial <- ifelse((!is.na(rawData_Traumabase$CGR.6h) &
    rawData_Traumabase$CGR.6h >
        3) | (!is.na(rawData_Traumabase$Cristalloïdes) &
    rawData_Traumabase$Cristalloïdes >
        0) | (!is.na(rawData_Traumabase$Colloïdes) &
    rawData_Traumabase$Colloïdes >
        0), 1, 0)

# Suspicion of hemorrhage ->
# hemorrhage_risk
rawData_Traumabase$hemorrhage_risk <- ifelse((!is.na(rawData_Traumabase$Choc.hémorragique....4.CGR.sur.
    rawData_Traumabase$Choc.hémorragique....4.CGR.sur.6h. ==
        1) | (!is.na(rawData_Traumabase$Activation.procédure.choc.hémorragique) &
    rawData_Traumabase$Activation.procédure.choc.hémorragique ==
        1) | (!is.na(rawData_Traumabase$Colloïdes) &
```

```
    rawData_Traumabase$Colloïdes >
        0) | (!is.na(rawData_Traumabase$CGR.6h) &
    rawData_Traumabase$CGR.6h >
        3) | (!is.na(rawData_Traumabase$Cristalloïdes) &
    rawData_Traumabase$Cristalloïdes >
        0) | (!is.na(rawData_Traumabase$Colloïdes) &
    rawData_Traumabase$Colloïdes >
        0), 1, 0)

CRASH3$majorExtracranial <- ifelse(CRASH3$majorExtracranial ==
    "Yes", 1, 0)
CRASH3$hemorrhage_risk <- ifelse(CRASH3$majorExtracranial ==
    1, 1, 0)

# Temporarily we define
# majorExtracranial using
# `bbleed` and `btransf`
CRASH2$majorExtracranial <- ifelse(CRASH2$crash2_bbleed ==
    1 | CRASH2$crash2_btransf ==
    1, 1, 0)
CRASH2$hemorrhage_risk <- ifelse(CRASH2$crash2_bbleed ==
    1 | CRASH2$crash2_btransf ==
    1, 1, 0)

# store majorExtracranial
# component
crash3_trial_eligibility <- c(crash3_trial_eligibility,
    "majorExtracranial")
crash2_trial_eligibility <- c(crash2_trial_eligibility,
    "hemorrhage_risk")

# Age
rawData_Traumabase$age <- rawData_Traumabase$Age.du.patient..ans

# Note that there are two
# outliers with age>120 years.
# By manual inspection, we can
# correct these observations
rawData_Traumabase$age[which(rawData_Traumabase$age ==
    721)] <- 72
rawData_Traumabase$age[which(rawData_Traumabase$age ==
    184)] <- 18

# store age component
crash3_trial_eligibility <- c(crash3_trial_eligibility,
    "age")
crash2_trial_eligibility <- c(crash2_trial_eligibility,
    "age")

# TBI (1 for TBI, 0 if not TBI)
CRASH2$TBI <- CRASH2$intraCranialBleeding
CRASH3$TBI <- ifelse(CRASH3$intraCranialBleeding ==
    "Yes" | (CRASH3$intraCranialBleeding ==
```

```
    "No CT scan available" & CRASH3$eligible ==
    "Yes"), 1, 0)
rawData_Traumabase$TBI <- ifelse(rawData_Traumabase$Trauma.crânien..lésion.cérébrale.TDM. ==
    "Oui" | rawData_Traumabase$ISS....Head_neck >=
    2, 1, 0)

# store TBI component
crash3_trial_eligibility <- c(crash3_trial_eligibility,
    "TBI")

# GSC
CRASH2$Glasgow.initial <- CRASH2$totalGcs

CRASH3$Glasgow.initial <- as.numeric(substring(CRASH3$gcsEyeOpening,
    1, 1)) + as.numeric(substring(CRASH3$gcsMotorResponse,
    1, 1)) + as.numeric(substring(CRASH3$gcsVerbalResponse,
    1, 1))
crash3_trial_eligibility <- c(crash3_trial_eligibility,
    "Glasgow.initial")
crash2_outcome_impact <- c(crash2_outcome_impact,
    "Glasgow.initial")

# Systolic blood pressure
rawData_Traumabase$systolicBloodPressure <- rawData_Traumabase$Pression.Artérielle.Systolique..PAS..à.l

# store SBP component
crash2_trial_eligibility <- c(crash2_trial_eligibility,
    "systolicBloodPressure")

crash3_outcome_impact <- c(crash3_outcome_impact,
    "systolicBloodPressure")

# Women (1) and men (0)
CRASH2$sexe <- ifelse(CRASH2$sex ==
    2, 1, 0)
CRASH3$sexe <- ifelse(CRASH3$sex ==
    "Female", 1, 0)
rawData_Traumabase$sexe <- ifelse(rawData_Traumabase$Sexe ==
    "Féminin", 1, 0)

crash3_outcome_impact <- c(crash3_outcome_impact,
    "sexe")
crash2_outcome_impact <- c(crash2_outcome_impact,
    "sexe")

# Pupil reactivity
x <- rawData_Traumabase[, "Anomalie.pupillaire..Pré.hospitalier."]
rawData_Traumabase$pupilReact <- case_when(x ==
    "Non" ~ "Both React", x ==
    "Anisocorie (unilatérale)" ~
    "One Reacts", x == "Mydriase Bilatérale" ~
    "None React", x == "Pas précisé" ~
    "Unable to assess")
```

```
rawData_Traumabase$pupilReact_num <- case_when(rawData_Traumabase$pupilReact ==
    "Both React" ~ 2, rawData_Traumabase$pupilReact ==
    "One Reacts" ~ 1, rawData_Traumabase$pupilReact ==
    "None React" ~ 0, rawData_Traumabase$pupilReact ==
    "Unable to assess" ~ -1)

CRASH3$pupilReact_num <- case_when(CRASH3$pupilReact ==
    "Both React" ~ 2, CRASH3$pupilReact ==
    "One Reacts" ~ 1, CRASH3$pupilReact ==
    "None React" ~ 0, CRASH3$pupilReact ==
    "Unable to assess" ~ -1)
CRASH2$pupilReact <- NA
CRASH2$pupilReact_num <- NA
crash3_outcome_impact <- c(crash3_outcome_impact,
    "pupilReact_num")
consistently_missing_in_1 <- c(consistently_missing_in_1,
    "pupilReact_num")

# Heart rate
rawData_Traumabase$heart_rate <- rawData_Traumabase$Fréquence.cardiaque..FC..à.l.arrivée.du.SMUR
CRASH3$heart_rate <- NA
CRASH2$heart_rate <- CRASH2$crash2_ihr
crash2_trial_eligibility <- c(crash2_trial_eligibility,
    "heart_rate")
consistently_missing_in_1 <- c(consistently_missing_in_1,
    "heart_rate")

# Central capillary refill time
rawData_Traumabase$central_capillary <- NA
CRASH3$central_capillary <- NA
CRASH2$central_capillary <- CRASH2$crash2_icc
crash2_outcome_impact <- c(crash2_outcome_impact,
    "central_capillary")
consistently_missing_in_2 <- c(consistently_missing_in_2,
    "central_capillary")

# Respiratory rate
rawData_Traumabase$respiratory_rate <- NA
CRASH3$respiratory_rate <- NA
CRASH2$respiratory_rate <- CRASH2$crash2_irr
crash2_outcome_impact <- c(crash2_outcome_impact,
    "respiratory_rate")
consistently_missing_in_2 <- c(consistently_missing_in_2,
    "respiratory_rate")

# Type of injury
rawData_Traumabase$type_injury <- case_when(rawData_Traumabase$Mécanisme.en.cause %in%
    c("Arme blanche", "Arme à feu") ~
    2, rawData_Traumabase$Mécanisme.en.cause %in%
    c("Chute (de sa hauteur)",
        "Traumatisme par objet contondant (non pénétrant)",
        "Chute d'une hauteur") ~
    1, startsWith(rawData_Traumabase$Mécanisme.en.cause,
```

```
    "AVP") | rawData_Traumabase$Mécanisme.en.cause ==
    "Engin de déplacement personnel motorisé" ~
    3, TRUE ~ NA_real_)
CRASH3$type_injury <- NA
CRASH2$type_injury <- CRASH2$crash2_iinjurytype
crash2_outcome_impact <- c(crash2_outcome_impact,
    "type_injury")
consistently_missing_in_1 <- c(consistently_missing_in_1,
    "type_injury")

# Time since injury we consider
# that if the time indicated in
# the Traumabase is below 3,
# then it wasn't given in
# minutes but hours
rawData_Traumabase$timeSinceInjury <- pmax(ifelse(rawData_Traumabase$Délai...départ.base...arrivée.sur.l
    3, rawData_Traumabase$Délai...départ.base...arrivée.sur.les.lieux..,
    rawData_Traumabase$Délai...départ.base...arrivée.sur.les.lieux../60) +
    ifelse(rawData_Traumabase$Délai...arrivée.sur.les.lieux...arrivée.hôpital.. <
        3, rawData_Traumabase$Délai...arrivée.sur.les.lieux...arrivée.hôpital..,
        rawData_Traumabase$Délai...arrivée.sur.les.lieux...arrivée.hôpital../60),
    ifelse(rawData_Traumabase$Délai...départ.base...arrivée.sur.les.lieux.. <
        3, rawData_Traumabase$Délai...départ.base...arrivée.sur.les.lieux..,
        rawData_Traumabase$Délai...départ.base...arrivée.sur.les.lieux../60),
    ifelse(rawData_Traumabase$Délai...arrivée.sur.les.lieux...arrivée.hôpital.. <
        3, rawData_Traumabase$Délai...arrivée.sur.les.lieux...arrivée.hôpital..,
        rawData_Traumabase$Délai...arrivée.sur.les.lieux...arrivée.hôpital../60),
    na.rm = T)
crash2_trial_eligibility <- c(crash2_trial_eligibility,
    "timeSinceInjury")
crash3_outcome_impact <- c(crash3_outcome_impact,
    "timeSinceInjury")
```

We add the additional trial eligibility criteria `continent` to the sets of CRASH-2 and CRASH-3 to account for the different proportions of patients included on different continents.

```
crash2_trial_eligibility_addition <- c("continent")
crash3_trial_eligibility_addition <- c("continent")
```

Merge and store data

Note that in CRASH3, first patient could be treated in a 8h window after injury, and then finally 3h. In the final data frame we only keep these patients.

In CRASH-3 it corresponds to `timeSinceInjury` in hours. In France recommendations for doctors already state to use the tranexamic acid as soon as possible, and in a 3h window after injury.

Also, we only consider patient in the Traumabase that have TBI, as it is the criteria on which inclusion was done in CRASH3.

```
# Time between injury and
# treatment --> keep only
# patients treated within 3h
# data treatment from string to
# numeric hours and minutes
```

```
CRASH3$timeSinceInjury_h = format(as.POSIXct(CRASH3$timeSinceInjury,
    format = "%Y-%m-%d %H:%M"),
    format = "%H")
CRASH3$timeSinceInjury_h <- as.numeric(CRASH3$timeSinceInjury_h)
CRASH3$timeSinceInjury_m = format(as.POSIXct(CRASH3$timeSinceInjury,
    format = "%Y-%m-%d %H:%M"),
    format = "%M")
CRASH3$timeSinceInjury_m <- as.numeric(CRASH3$timeSinceInjury_m)

CRASH3$timeSinceInjury <- CRASH3$timeSinceInjury_h +
    CRASH3$timeSinceInjury_m/60

## selection of the pertinent
## subtable for the rest of the
## analysis as the CRASH3
## investigators change the
## protocol to keep only patient
## treated before 3h after
## injury
CRASH3_3h <- CRASH3[which(CRASH3$timeSinceInjury_h <
    3 | (CRASH3$timeSinceInjury_h ==
    3 & CRASH3$timeSinceInjury_m ==
    0)), ]

# we do the same for the CRASH2
# study even though this 3h
# threshold was not known for
# this study
CRASH2_3h <- CRASH2[which(CRASH2$timeSinceInjury <=
    3), ]

# for the Traumabase we take a
# crude proxy of the time since
# injury, using the delay
# between the departure of the
# ambulance from its base to
# the arrival at the hospital
Traumabase_3h <- rawData_Traumabase[which(rawData_Traumabase$timeSinceInjury <=
    3), ]

# only patients from the
# Traumabase with TBI
Traumabase_3h_tbionly <- Traumabase_3h[which(Traumabase_3h$TBI ==
    1), ]
Traumabase_tbionly <- rawData_Traumabase[which(rawData_Traumabase$TBI ==
    1), ]

# a few patients have no TBI in
# CRASH-3, to compare similar
# quantity we exclude them
CRASH3_3h_tbionly <- CRASH3_3h[which(CRASH3_3h$TBI ==
    1), ]
CRASH3_tbionly <- CRASH3[which(CRASH3$TBI ==
```

```
    1), ]

# we also only keep patients
# with TBI from the CRASH-2
# study
CRASH2_3h_tbionly <- CRASH2_3h[which(CRASH2_3h$TBI ==
    1), ]
CRASH2_tbionly <- CRASH2[which(CRASH2$TBI ==
    1), ]

# drop this variable as it
crash3_trial_eligibility <- setdiff(crash3_trial_eligibility,
    "TBI")

# additionally, we only
# consider patients from
# centers with sufficiently
# many trauma patients
df_3h <- Traumabase_3h_tbionly %>%
    dplyr::select(c("Numéro.de.centre")) %>%
    group_by(Numéro.de.centre) %>%
    summarise(n = n()) %>% mutate(effectifs = paste(n,
    "TBI \n patients"))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
df <- Traumabase_tbionly %>% dplyr::select(c("Numéro.de.centre")) %>%
    group_by(Numéro.de.centre) %>%
    summarise(n = n()) %>% mutate(effectifs = paste(n,
    "TBI \n patients"))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
centers.too.small <- unique(c(unlist(df[which(df$n <
    20), "Numéro.de.centre"]),
    unlist(df_3h[which(df_3h$n <
        20), "Numéro.de.centre"])))
Traumabase_3h_tbionly_goodcenters <- Traumabase_3h_tbionly[which(!(Traumabase_3h_tbionly[,
    "Numéro.de.centre"] %in% centers.too.small)),
    ]
Traumabase_tbionly_goodcenters <- Traumabase_tbionly[which(!(Traumabase_tbionly[,
    "Numéro.de.centre"] %in% centers.too.small)),
    ]
Traumabase_goodcenters <- rawData_Traumabase[which(!(rawData_Traumabase[,
    "Numéro.de.centre"] %in% centers.too.small)),
    ]
Traumabase_3h_goodcenters <- Traumabase_3h[which(!(Traumabase_3h[,
    "Numéro.de.centre"] %in% centers.too.small)),
    ]


# indicator for RCTs and RWD
Traumabase_3h_tbionly_goodcenters$V <- rep(0,
    nrow(Traumabase_3h_tbionly_goodcenters))
CRASH3_3h_tbionly$V <- rep(1, nrow(CRASH3_3h_tbionly))
```

```
CRASH2_3h_tbionly$V <- rep(1, nrow(CRASH2_3h_tbionly))
Traumabase_tbionly_goodcenters$V <- rep(0,
    nrow(Traumabase_tbionly_goodcenters))
CRASH3_tbionly$V <- rep(1, nrow(CRASH3_tbionly))
CRASH2_tbionly$V <- rep(1, nrow(CRASH2_tbionly))
Traumabase_goodcenters$V <- rep(0,
    nrow(Traumabase_goodcenters))
CRASH3$V <- rep(1, nrow(CRASH3))
CRASH2$V <- rep(1, nrow(CRASH2))

# differentiated indicator for
# RCTs and RWD
Traumabase_3h_tbionly_goodcenters$V3 <- rep(0,
    nrow(Traumabase_3h_tbionly_goodcenters))
CRASH3_3h_tbionly$V3 <- rep(2,
    nrow(CRASH3_3h_tbionly))
CRASH2_3h_tbionly$V3 <- rep(1,
    nrow(CRASH2_3h_tbionly))
Traumabase_tbionly_goodcenters$V3 <- rep(0,
    nrow(Traumabase_tbionly_goodcenters))
CRASH3_tbionly$V3 <- rep(2, nrow(CRASH3_tbionly))
CRASH2_tbionly$V3 <- rep(1, nrow(CRASH2_tbionly))
Traumabase_goodcenters$V3 <- rep(0,
    nrow(Traumabase_goodcenters))
CRASH3$V3 <- rep(2, nrow(CRASH3))
CRASH2$V3 <- rep(1, nrow(CRASH2))

# total data frame for 3h
# patients of RCTs and all
# patients of Traumabase
total_3h <- rbind(CRASH2_3h_tbionly[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "Death", "TBI_Death", "treatment",
        "V", "V3"))], CRASH3_3h_tbionly[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "Death", "TBI_Death", "treatment",
        "V", "V3"))], Traumabase_tbionly_goodcenters[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
```

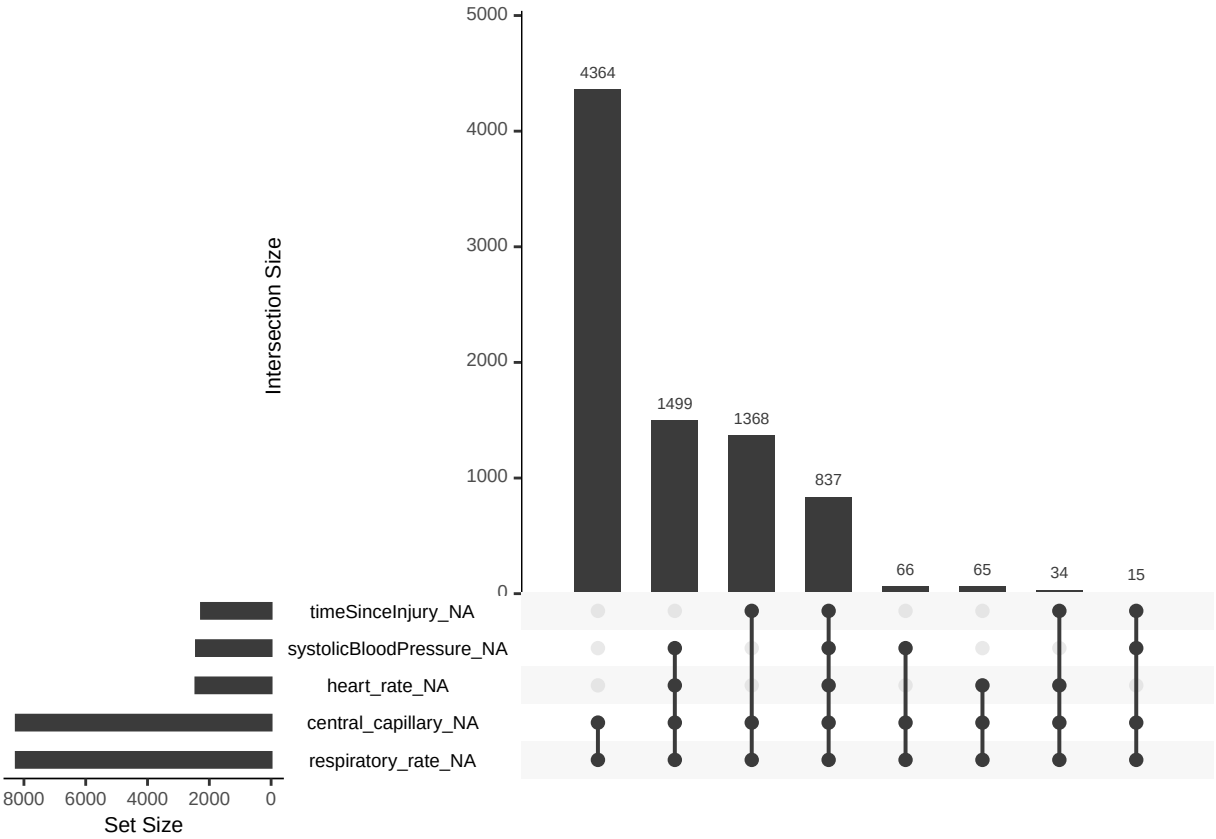
```
        "Death", "TBI_Death", "treatment",
        "V", "V3"))])

# total data frame
total <- rbind(CRASH2_tbionly[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "Death", "TBI_Death", "treatment",
        "V", "V3"))], CRASH3_tbionly[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "Death", "TBI_Death", "treatment",
        "V", "V3"))], Traumabase_tbionly_goodcenters[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "Death", "TBI_Death", "treatment",
        "V", "V3"))])

# total data frame for all
# patients of RCTs
total_allPatients <- rbind(CRASH2[,
    unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "TBI", "Death", "TBI_Death",
        "treatment", "V", "V3"))],
    CRASH3[, unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "TBI", "Death", "TBI_Death",
        "treatment", "V", "V3"))],
    Traumabase_goodcenters[, unique(c(crash2_trial_eligibility,
        crash3_trial_eligibility,
        crash2_outcome_impact,
        crash3_outcome_impact,
```

```
        crash2_trial_eligibility_addition,
        crash3_trial_eligibility_addition,
        "TBI", "Death", "TBI_Death",
        "treatment", "V", "V3"))])
```

Summaries of the different data sets

```
## [1] "Traumabase, TBI: 8248 observations"
```

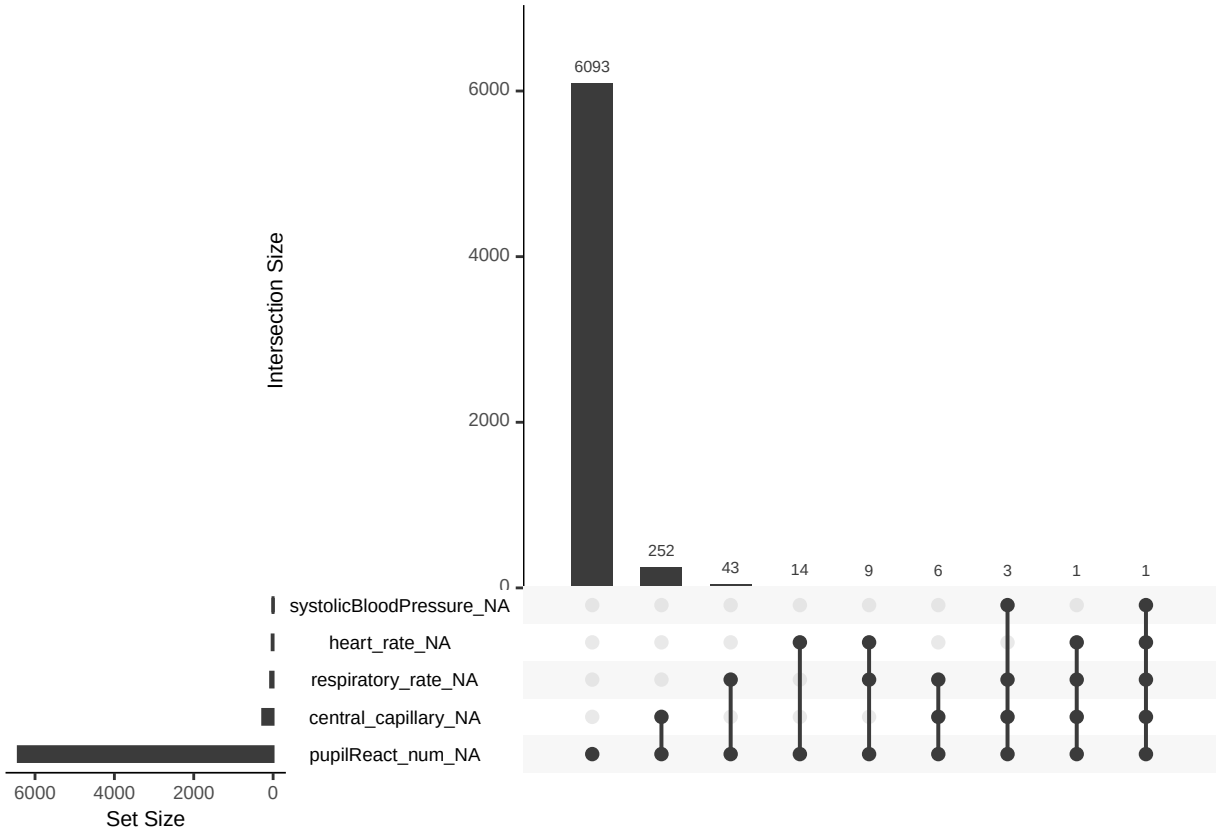
```
##  hemorrhage_risk       age        systolicBloodPressure   heart_rate
##  Min.   :0.0000   Min.   : 0.00   Min.   :  0.0         Min.   :  0.00
##  1st Qu.:0.0000   1st Qu.:26.00   1st Qu.:112.0         1st Qu.: 72.00
##  Median :1.0000   Median :40.00   Median :130.0         Median : 88.00
##  Mean   :0.6769   Mean   :43.16   Mean   :127.7         Mean   : 88.09
##  3rd Qu.:1.0000   3rd Qu.:58.00   3rd Qu.:147.0         3rd Qu.:105.00
##  Max.   :1.0000   Max.   :96.00   Max.   :256.0         Max.   :200.00
##                   NA's   :22      NA's   :2417          NA's   :2435
##  timeSinceInjury  majorExtracranial Glasgow.initial      sexe
##  Min.   : 0.000   Min.   :0.0000    Min.   : 3.00   Min.   :0.0000
##  1st Qu.: 1.083   1st Qu.:0.0000    1st Qu.: 6.00   1st Qu.:0.0000
##  Median : 1.483   Median :1.0000    Median :13.00   Median :0.0000
##  Mean   : 1.744   Mean   :0.6769    Mean   :10.61   Mean   :0.2331
##  3rd Qu.: 2.000   3rd Qu.:1.0000    3rd Qu.:15.00   3rd Qu.:0.0000
##  Max.   :48.000   Max.   :1.0000    Max.   :15.00   Max.   :1.0000
##  NA's   :2254                       NA's   :165     NA's   :63
##  central_capillary respiratory_rate  type_injury    pupilReact_num
##  Mode:logical      Mode:logical     Min.   :1.000   Min.   :-1.000
##  NA's:8248         NA's:8248        1st Qu.:1.000   1st Qu.: 2.000
##                                     Median :3.000   Median : 2.000
##                                     Mean   :2.172   Mean   : 1.641
##                                     3rd Qu.:3.000   3rd Qu.: 2.000
##                                     Max.   :3.000   Max.   : 2.000
##                                     NA's   :377     NA's   :161
##   continent             Death          TBI_Death        treatment
##  Length:8248        Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :0.00000
##                     Mean   :0.1992   Mean   :0.1711   Mean   :0.08281
##                     3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##
```




```
## [1] "CRASH-2, TBI: 6422 observations"
```

```
##  hemorrhage_risk       age        systolicBloodPressure   heart_rate
##  Min.   :0.0000   Min.   : 1.00   Min.   :  0.0         Min.   :  3.0
##  1st Qu.:0.0000   1st Qu.:24.00   1st Qu.: 84.0         1st Qu.: 90.0
##  Median :1.0000   Median :33.00   Median :100.0         Median :106.0
##  Mean   :0.5017   Mean   :36.42   Mean   :102.2         Mean   :105.3
##  3rd Qu.:1.0000   3rd Qu.:46.00   3rd Qu.:120.0         3rd Qu.:120.0
##  Max.   :1.0000   Max.   :95.00   Max.   :999.0         Max.   :220.0
##                                   NA's   :4             NA's   :25
##  timeSinceInjury majorExtracranial Glasgow.initial       sexe
##  Min.   :0.100   Min.   :0.0000    Min.   : 3.000   Min.   :0.0000
##  1st Qu.:1.000   1st Qu.:0.0000    1st Qu.: 6.000   1st Qu.:0.0000
##  Median :3.000   Median :1.0000    Median : 9.000   Median :0.0000
##  Mean   :3.316   Mean   :0.5017    Mean   : 9.318   Mean   :0.1724
##  3rd Qu.:5.000   3rd Qu.:1.0000    3rd Qu.:13.000   3rd Qu.:0.0000
##  Max.   :8.000   Max.   :1.0000    Max.   :15.000   Max.   :1.0000
##                                    NA's   :1
##  central_capillary respiratory_rate  type_injury    pupilReact_num
##  Min.   : 1.000    Min.   : 0.00    Min.   :1.000   Mode:logical
##  1st Qu.: 2.000    1st Qu.:18.00    1st Qu.:1.000   NA's:6422
##  Median : 3.000    Median :22.00    Median :1.000
##  Mean   : 3.404    Mean   :22.53    Mean   :1.383
##  3rd Qu.: 4.000    3rd Qu.:26.00    3rd Qu.:1.000
##  Max.   :30.000    Max.   :95.00    Max.   :3.000
##  NA's   :263       NA's   :63
##   continent            Death          TBI_Death        treatment
##  Length:6422        Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
```

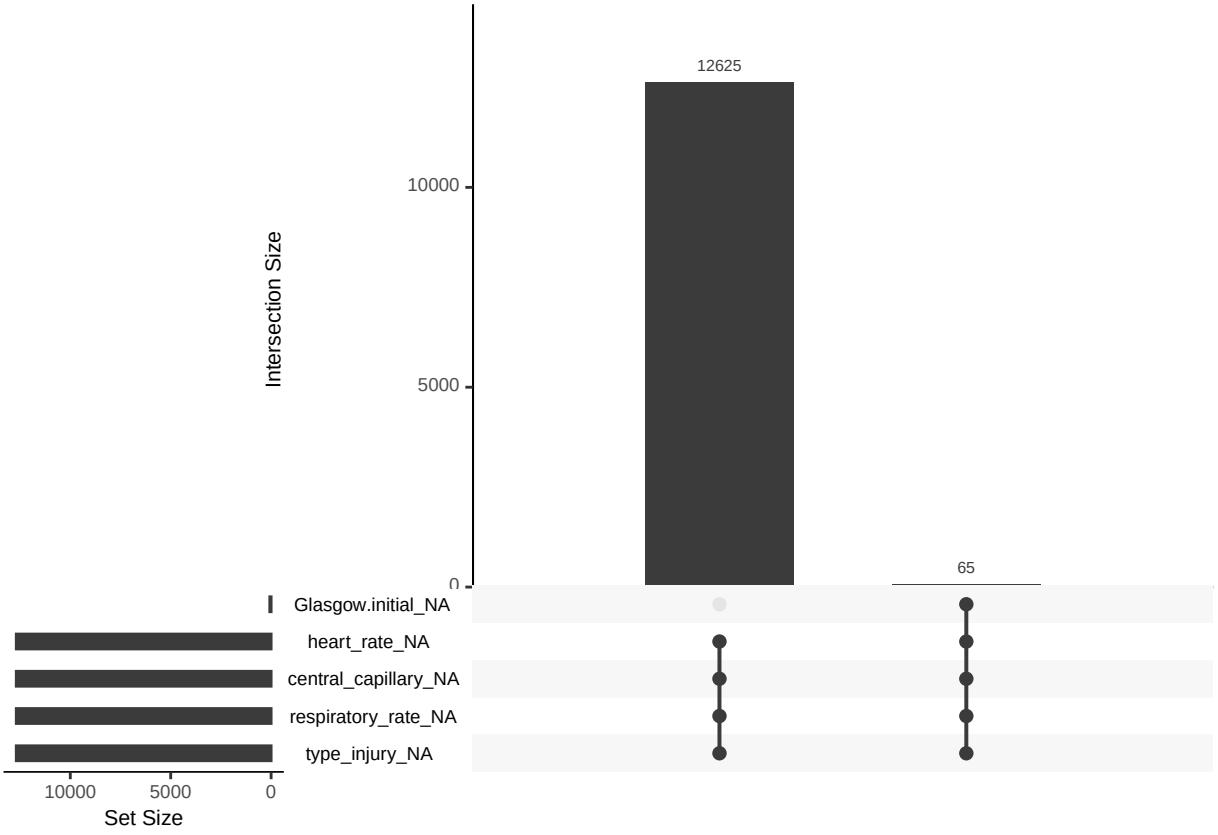
```
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :1.0000
##                     Mean   :0.2854   Mean   :0.1904   Mean   :0.5003
##                     3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```



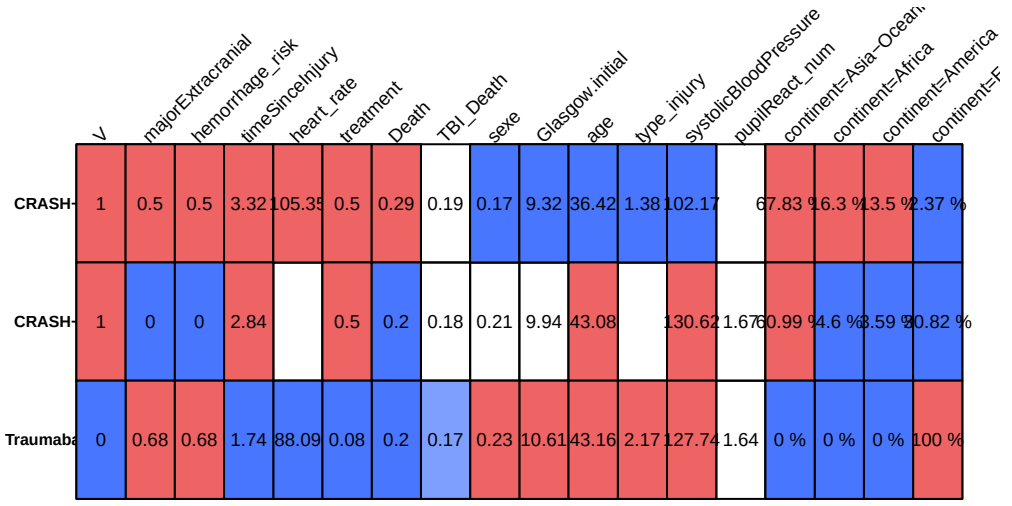

```
## [1] "CRASH-3, TBI: 12690 observations"
```

```
##  hemorrhage_risk          age         systolicBloodPressure heart_rate
##  Min.   :0.0000000   Min.   :  5.00   Min.   : 40.0         Mode:logical
##  1st Qu.:0.0000000   1st Qu.: 25.00   1st Qu.:110.0         NA's:12690
##  Median :0.0000000   Median : 40.00   Median :130.0
##  Mean   :0.0006304   Mean   : 43.08   Mean   :130.6
##  3rd Qu.:0.0000000   3rd Qu.: 58.00   3rd Qu.:145.0
##  Max.   :1.0000000   Max.   :100.00   Max.   :280.0
##                                       NA's   :40
##  timeSinceInjury  majorExtracranial   Glasgow.initial       sexe
##  Min.   : 0.000   Min.   :0.0000000   Min.   : 3.000   Min.   :0.0000
##  1st Qu.: 1.583   1st Qu.:0.0000000   1st Qu.: 7.000   1st Qu.:0.0000
##  Median : 2.417   Median :0.0000000   Median :11.000   Median :0.0000
##  Mean   : 2.838   Mean   :0.0006304   Mean   : 9.942   Mean   :0.2056
##  3rd Qu.: 3.500   3rd Qu.:0.0000000   3rd Qu.:14.000   3rd Qu.:0.0000
##  Max.   :17.583   Max.   :1.0000000   Max.   :15.000   Max.   :1.0000
##                                       NA's   :65       NA's   :1
##  central_capillary respiratory_rate type_injury    pupilReact_num
##  Mode:logical      Mode:logical     Mode:logical   Min.   :-1.000
##  NA's:12690        NA's:12690       NA's:12690     1st Qu.: 2.000
```

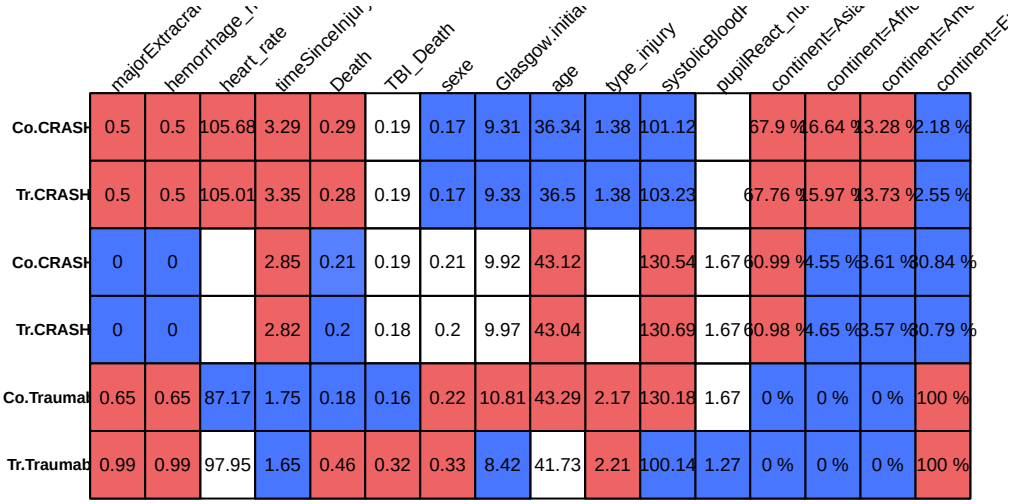
```
##                                                 Median : 2.000
##                                                 Mean   : 1.666
##                                                 3rd Qu.: 2.000
##                                                 Max.   : 2.000
##                                                 NA's   :3
##   continent             Death          TBI_Death        treatment
##  Length:12690       Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :1.0000
##                     Mean   :0.2011   Mean   :0.1837   Mean   :0.5028
##                     3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```



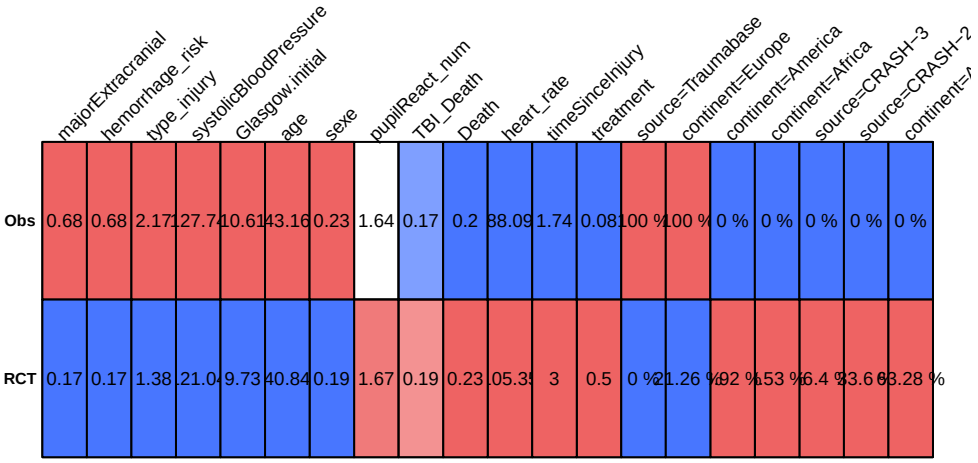

```
## pdf
##   2
```



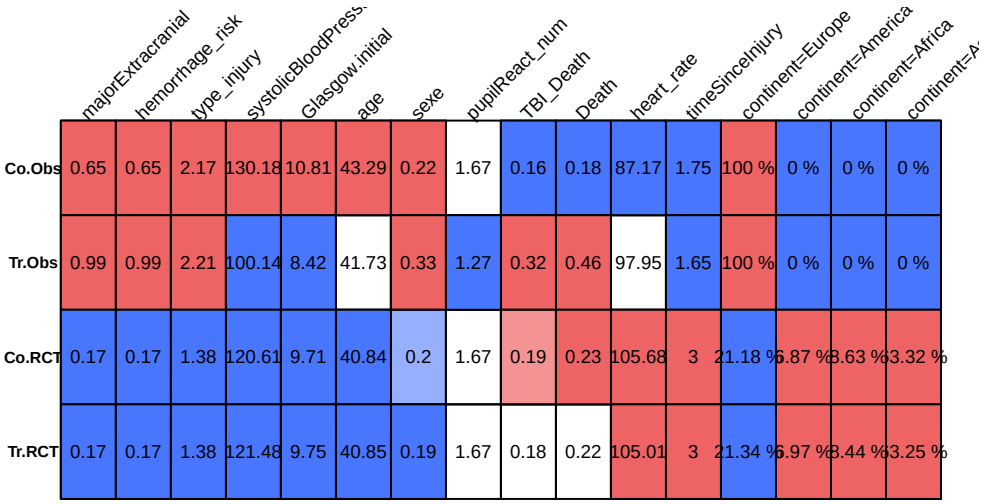

```
## pdf
##   2
```




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

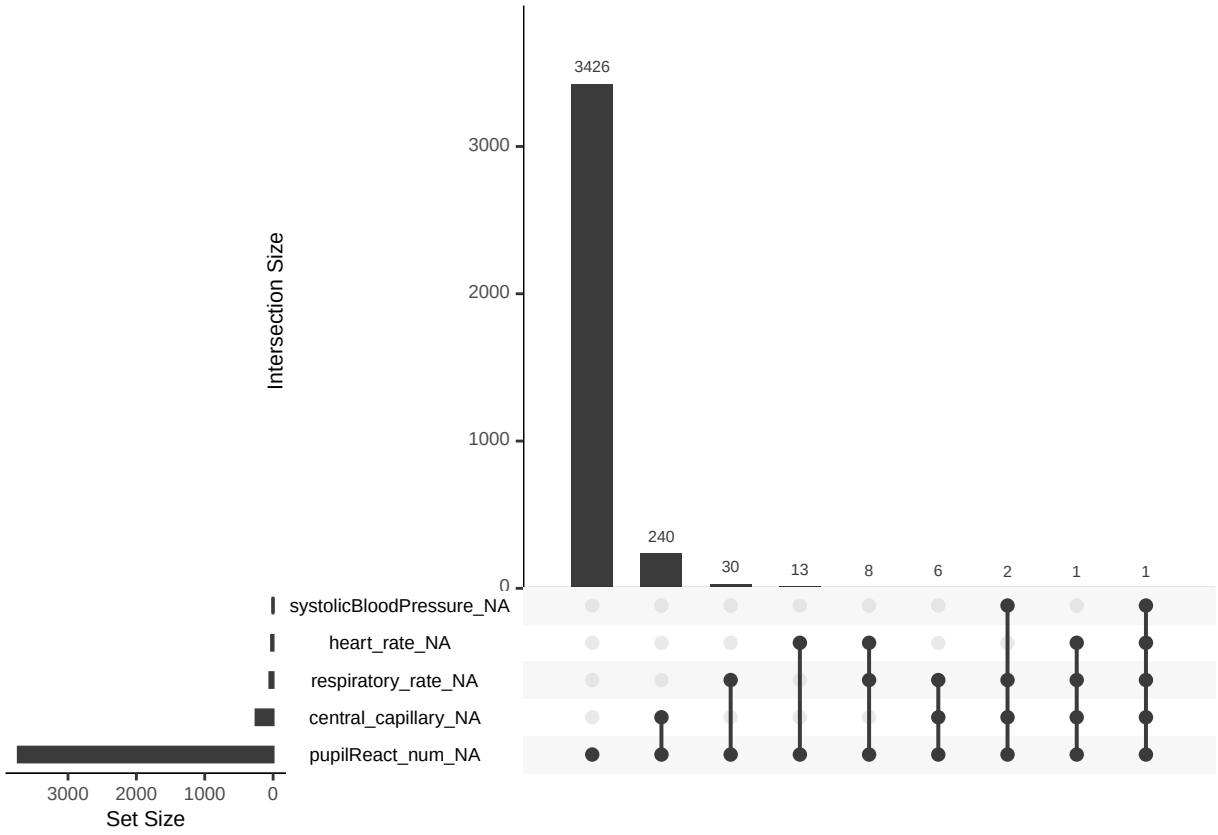



```
## pdf
##   2
```




```
## [1] "CRASH-2, TBI, within 3h: 3727 observations"
```

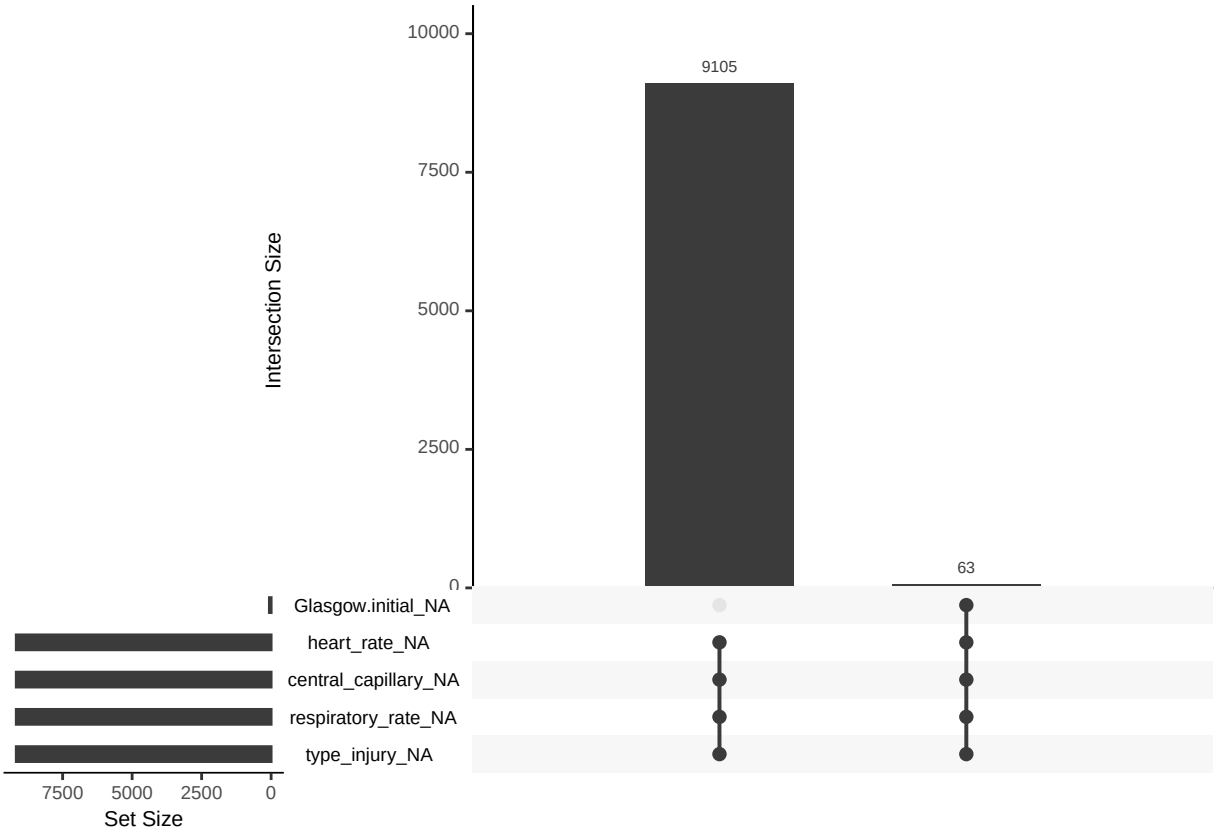
```
##  hemorrhage_risk       age        systolicBloodPressure   heart_rate
##  Min.   :0.0000   Min.   :16.00   Min.   :  0.00        Min.   :  3.0
##  1st Qu.:0.0000   1st Qu.:25.00   1st Qu.: 80.00        1st Qu.: 92.0
##  Median :0.0000   Median :33.00   Median : 95.00        Median :108.0
##  Mean   :0.4704   Mean   :36.54   Mean   : 98.79        Mean   :105.9
##  3rd Qu.:1.0000   3rd Qu.:46.00   3rd Qu.:115.00        3rd Qu.:120.0
##  Max.   :1.0000   Max.   :89.00   Max.   :999.00        Max.   :200.0
##                                   NA's   :3             NA's   :23
##  timeSinceInjury majorExtracranial Glasgow.initial       sexe
##  Min.   :0.100   Min.   :0.0000    Min.   : 3.000   Min.   :0.0000
##  1st Qu.:1.000   1st Qu.:0.0000    1st Qu.: 6.000   1st Qu.:0.0000
##  Median :2.000   Median :0.0000    Median :10.000   Median :0.0000
##  Mean   :1.718   Mean   :0.4704    Mean   : 9.418   Mean   :0.1833
##  3rd Qu.:2.000   3rd Qu.:1.0000    3rd Qu.:13.000   3rd Qu.:0.0000
##  Max.   :3.000   Max.   :1.0000    Max.   :15.000   Max.   :1.0000
##                                    NA's   :1
##  central_capillary respiratory_rate  type_injury    pupilReact_num
##  Min.   : 1.000    Min.   : 0.0     Min.   :1.000   Mode:logical
##  1st Qu.: 2.000    1st Qu.:20.0     1st Qu.:1.000   NA's:3727
##  Median : 3.000    Median :22.0     Median :1.000
##  Mean   : 3.361    Mean   :23.1     Mean   :1.445
##  3rd Qu.: 4.000    3rd Qu.:27.0     3rd Qu.:2.000
##  Max.   :30.000    Max.   :95.0     Max.   :3.000
##  NA's   :250       NA's   :48
##   continent             Death          TBI_Death        treatment
##  Length:3727        Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :1.0000
##                     Mean   :0.3155   Mean   :0.2058   Mean   :0.5007
##                     3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```



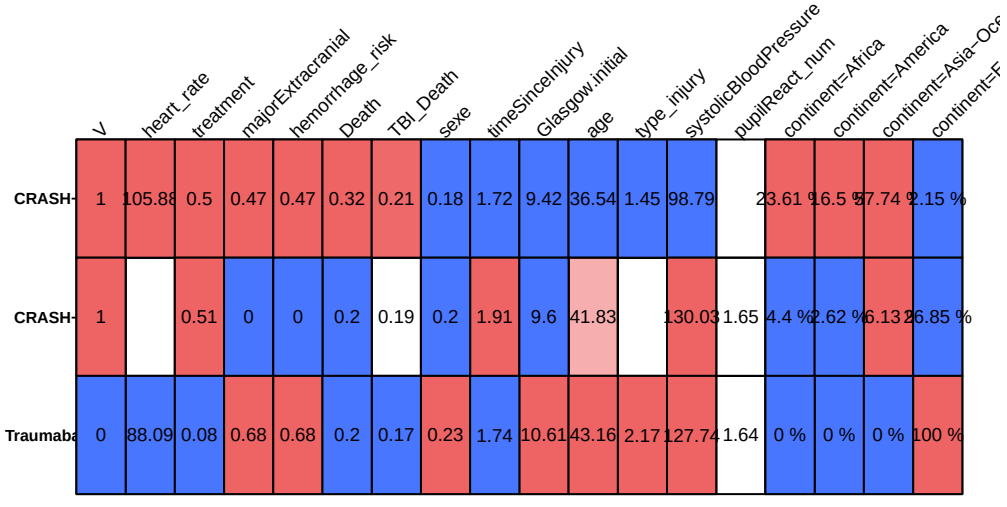

```
## [1] "CRASH-3, TBI, within 3h: 9168 observations"
```

```
##  hemorrhage_risk          age        systolicBloodPressure heart_rate
##  Min.   :0.0000000   Min.   : 5.00   Min.   : 40           Mode:logical
##  1st Qu.:0.0000000   1st Qu.:25.00   1st Qu.:110           NA's:9168
##  Median :0.0000000   Median :38.00   Median :129
##  Mean   :0.0005454   Mean   :41.83   Mean   :130
##  3rd Qu.:0.0000000   3rd Qu.:55.00   3rd Qu.:144
##  Max.   :1.0000000   Max.   :98.00   Max.   :280
##                                      NA's   :23
##  timeSinceInjury majorExtracranial   Glasgow.initial      sexe
##  Min.   :0.000   Min.   :0.0000000   Min.   : 3.0    Min.   :0.0000
##  1st Qu.:1.333   1st Qu.:0.0000000   1st Qu.: 7.0    1st Qu.:0.0000
##  Median :2.000   Median :0.0000000   Median :10.0    Median :0.0000
##  Mean   :1.907   Mean   :0.0005454   Mean   : 9.6    Mean   :0.1953
##  3rd Qu.:2.500   3rd Qu.:0.0000000   3rd Qu.:13.0    3rd Qu.:0.0000
##  Max.   :3.000   Max.   :1.0000000   Max.   :15.0    Max.   :1.0000
##                                      NA's   :63      NA's   :1
##  central_capillary respiratory_rate type_injury    pupilReact_num
##  Mode:logical      Mode:logical     Mode:logical   Min.   :-1.000
##  NA's:9168         NA's:9168        NA's:9168      1st Qu.: 2.000
##                                                    Median : 2.000
##                                                    Mean   : 1.645
##                                                    3rd Qu.: 2.000
##                                                    Max.   : 2.000
##                                                    NA's   :3
##   continent             Death          TBI_Death        treatment
##  Length:9168        Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
```

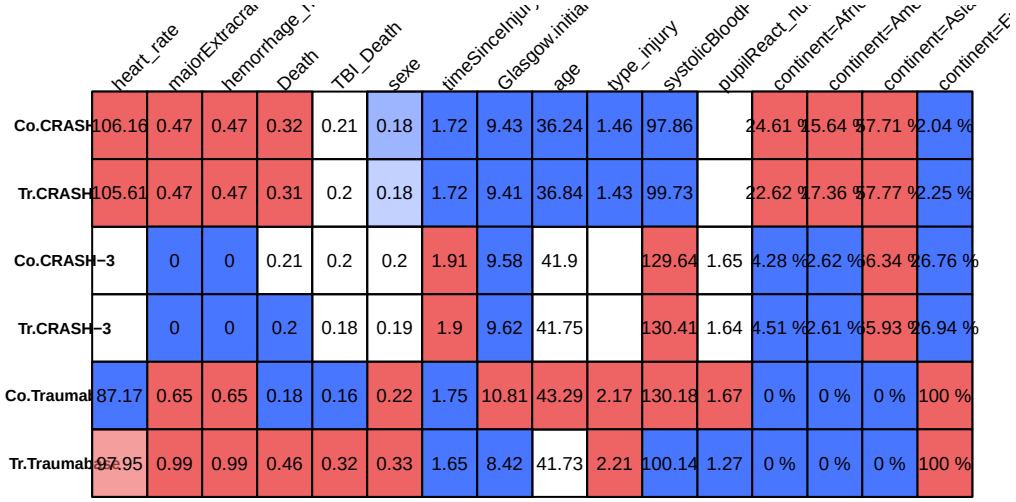
```
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :1.0000
##                     Mean   :0.2047   Mean   :0.1903   Mean   :0.5052
##                     3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##
```



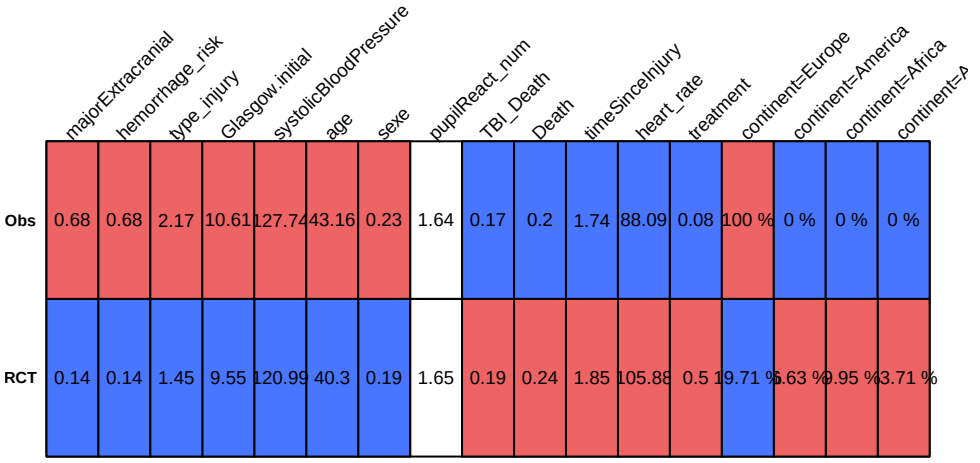

```
## pdf
##   2
```



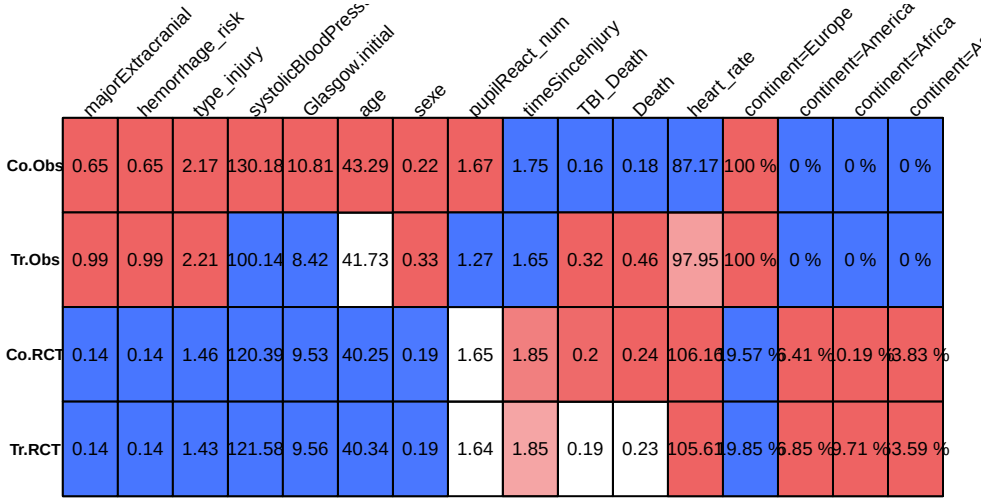

```
## pdf
##   2
```




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




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

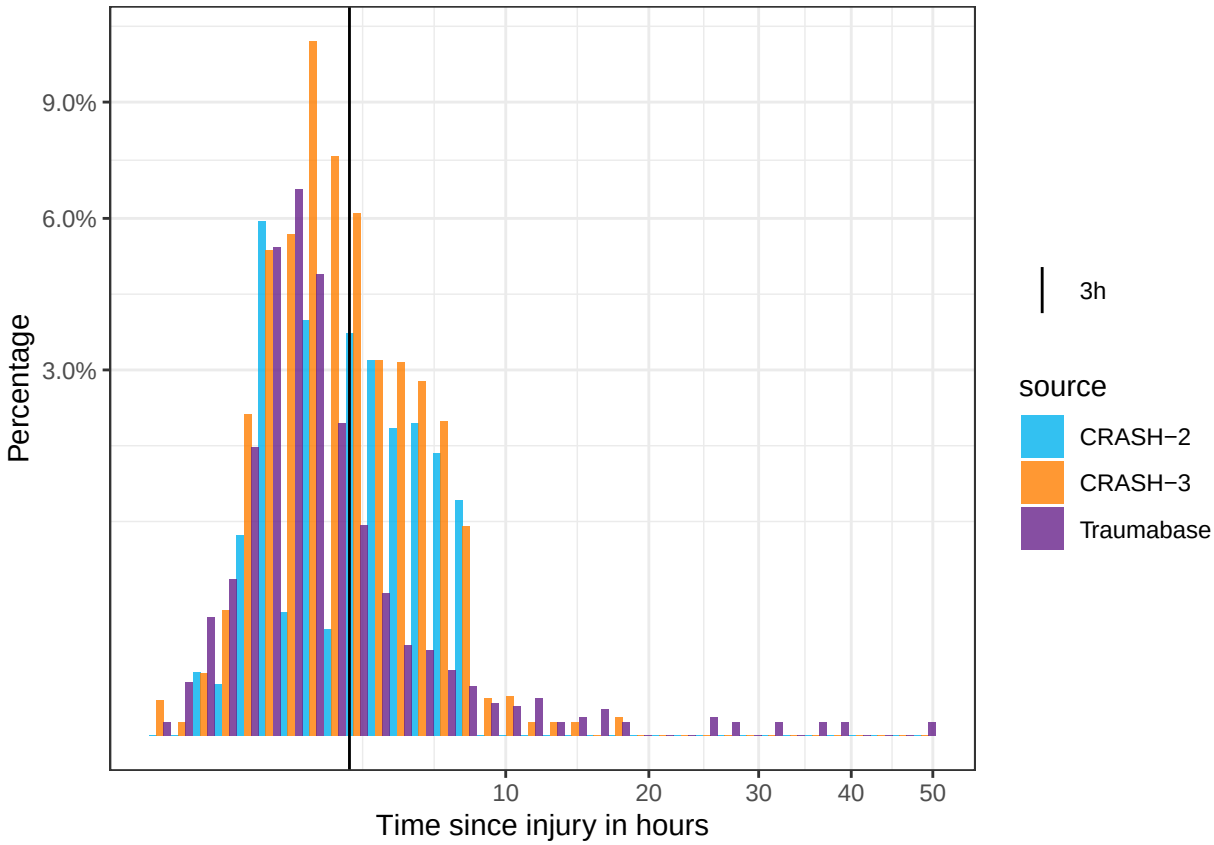



```
path_to_output <- paste0(data_dir,
    "output_preprocess_combined_crash2_crash3_TB.csv")
```

```
write.csv(total, path_to_output)

path_to_output <- paste0(data_dir,
    "output_preprocess_combined_crash2_3h_crash3_3h_TB.csv")
write.csv(total_3h, path_to_output)

path_to_output <- paste0(data_dir,
    "output_preprocess_combined_allPatients_crash2_crash3_TB.csv")
write.csv(total_allPatients, path_to_output)

path_to_output <- paste0(data_dir,
    "crash2_crash3_variables.RData")
save(crash2_trial_eligibility,
    crash3_trial_eligibility, crash2_outcome_impact,
    crash3_outcome_impact, crash2_trial_eligibility_addition,
    crash3_trial_eligibility_addition,
    file = path_to_output)
```

Short analysis of time to treatment

```
comparison_time <- data.frame(total_detail$timeSinceInjury,
    source = as.factor(total_detail$source))
ggplot(total_detail, aes(x = timeSinceInjury,
    group = source, fill = source)) +
    geom_histogram(binwidth = 0.2,
        alpha = 0.8, position = "dodge",
        aes(y = (..count..)/sum(..count..))) +
    geom_vline(aes(xintercept = 3,
        color = "3h")) + scale_color_manual("",
    values = c(`3h` = "black")) +
    xlab("Time since injury in hours") +
    ylab("Percentage") + theme_bw() +
    scale_x_sqrt() + scale_y_sqrt(labels = scales::percent) +
    scale_fill_manual(values = c(Traumabase = "darkorchid4",
        `CRASH-2` = "deepskyblue2",
        `CRASH-3` = "darkorange1"))
```




```
path_to_output <- paste0(fig_dir,
    "timeSinceInjury_crash2_crash3_TB.pdf")
ggsave(path_to_output)
```

```
## Saving 6.5 x 4.5 in image
```

The black vertical line shows the 3h threshold

Imputed data for the Traumbase

The same procedure is performed with the imputed Traumabase.

Perform imputation

Imputation is performed on the already filtered Traumabase data set.

```
# Recode values of imputed
# categorical variables and
# recast some numerical
# variables into integers
cast_types = function(i, df, data.num) {
    if (is.factor(df[, i])) {
        df[, i] = plyr::mapvalues(df[,
            i], from = levels(df[,
            i]), to = gsub(paste(i,
            "_", sep = ""), "",
            levels(df[, i])))
    } else {
```

```
        if (i %in% data.num) {
            df[, i] <- round(df[,
                i], digits = 1)
        } else {
            df[, i] <- as.integer(round(df[,
                i], digits = 0))
        }
    }
    return(df[, i])
}

vars.for.imputation <- c("Numéro.de.centre",
    "Traitement.anticoagulant",
    "Traitement.antiagrégants",
    "Glasgow.initial", "Glasgow.moteur.initial",
    "Mannitol...SSH", "Régression.mydriase.sous.osmothérapie",
    "Arrêt.cardio.respiratoire..massage.",
    "Fréquence.cardiaque..FC..à.l.arrivée.du.SMUR",
    "Cristalloïdes", "Colloïdes",
    "Hémocue.initial", "Delta.Hémocue",
    "Catécholamines", "SpO2.min",
    "Délai...arrivée.sur.les.lieux...arrivée.hôpital..",
    "Score.de.Glasgow.en.phase.hospitalière",
    "Glasgow.moteur", "Anomalie.pupillaire..Phase.hospitalière.",
    "FC.en.phase.hospitalière",
    "Doppler.TransCrânien..DTC...Index.de.Pulsatilité..IP..max",
    "FiO2", "Bloc.dans.les.premières.24h....Neurochirurgie..ex....Craniotomie.ou.DVE.",
    "Total.Score.IGS", "Osmothérapie",
    "HTIC...25.PIC.simple.sédation.",
    "Dérivation.ventriculaire.externe..DVE.",
    "Craniectomie.dé.compressive",
    "ISS....Head_neck", "ISS....Face",
    "ISS....External", "Score.ISS",
    "Activation.procédure.choc.hémorragique",
    "ISS....Selection", "age",
    "TBI", "majorExtracranial",
    "hemorrhage_risk", "systolicBloodPressure",
    "pupilReact_num", "sexe", "timeSinceInjury",
    "type_injury", "treatment")



if (file.exists(paste0(data_dir,
    "traumabase_tbideath_jointanalysis_tbi_imputed_mice.RData"))) {
    load(file = paste0(data_dir,
        "traumabase_tbideath_jointanalysis_tbi_imputed_mice.RData"))
    load(file = paste0(data_dir,
        "traumabase_jointanalysis_tbi_imputed_mice.RData"))
} else {
    m = 5

    DF_tbi <- Traumabase_tbionly_goodcenters
```

```
df.tmp <- DF_tbi[, vars.for.imputation]
df.tmp$treatment <- as.factor(df.tmp$treatment)
imp.mice.mids <- mice::mice(df.tmp,
    m = m, printFlag = F)
df.imp.mice <- list()
for (k in 1:m) {
    df.imp.mice[[k]] <- mice::complete(imp.mice.mids,
        k)
    df.imp.mice[[k]]$continent <- DF_tbi$continent
    df.imp.mice[[k]]$TBI_Death <- DF_tbi$TBI_Death
    df.imp.mice[[k]]$treatment <- as.numeric(as.character(df.imp.mice[[k]]$treatment))
    df.imp.mice[[k]]$heart_rate <- df.imp.mice[[k]]$Fréquence.cardiaque..FC..à.l.arrivée.du.SMUR
    df.imp.mice[[k]]$respiratory_rate <- NA
    df.imp.mice[[k]]$central_capillary <- NA
    df.imp.mice[[k]] <- df.imp.mice[[k]][,
        unique(c(crash2_trial_eligibility,
            crash3_trial_eligibility,
            crash2_outcome_impact,
            crash3_outcome_impact,
            crash2_trial_eligibility_addition,
            crash3_trial_eligibility_addition,
            "TBI", "TBI_Death",
            "treatment", "Numéro.de.centre",
            "ISS....Head_neck"))]
}
save(df.imp.mice, imp.mice.mids,
    file = paste0(data_dir,
        "traumabase_tbideath_jointanalysis_tbi_imputed_mice.RData"))

for (k in 1:m) {
    df.imp.mice[[k]] <- mice::complete(imp.mice.mids,
        k)
    df.imp.mice[[k]]$continent <- DF_tbi$continent
    df.imp.mice[[k]]$Death <- DF_tbi$Death
    df.imp.mice[[k]]$TBI_Death <- DF_tbi$TBI_Death
    df.imp.mice[[k]]$treatment <- as.numeric(as.character(df.imp.mice[[k]]$treatment))
    df.imp.mice[[k]]$heart_rate <- df.imp.mice[[k]]$Fréquence.cardiaque..FC..à.l.arrivée.du.SMUR
    df.imp.mice[[k]]$respiratory_rate <- NA
    df.imp.mice[[k]]$central_capillary <- NA
    df.imp.mice[[k]] <- df.imp.mice[[k]][,
        unique(c(crash2_trial_eligibility,
            crash3_trial_eligibility,
            crash2_outcome_impact,
            crash3_outcome_impact,
            crash2_trial_eligibility_addition,
            crash3_trial_eligibility_addition,
            "TBI", "Death",
            "TBI_Death", "treatment",
            "Numéro.de.centre",
            "ISS....Head_neck"))]
}
save(df.imp.mice, imp.mice.mids,
    file = paste0(data_dir,
```

```
            "traumabase_jointanalysis_tbi_imputed_mice.RData"))
}
```

Merge imputed data

```
for (k in 1:length(df.imp.mice)) {
    imputed_traumabase <- df.imp.mice[[k]]
    imputed_traumabase$V <- rep(0,
        nrow(imputed_traumabase))
    imputed_traumabase$V3 <- rep(0,
        nrow(imputed_traumabase))
    imputed_traumabase$source <- "Traumabase"
    imputed_traumabase <- imputed_traumabase[,
        names(total)]
    total_with_imputations <- total[total$V ==
        1, ]
    total_with_imputations <- rbind(total_with_imputations,
        imputed_traumabase)

    path_to_imputed <- paste0(data_dir,
        "output_preprocess_combined_crash2_crash3_TB_imputed",
        k, ".csv")
    write.csv(total_with_imputations,
        file = path_to_imputed)

    total_3h_with_imputations <- total_3h[total_3h$V ==
        1, ]
    total_3h_with_imputations <- rbind(total_3h_with_imputations,
        imputed_traumabase)
    path_to_imputed <- paste0(data_dir,
        "output_preprocess_combined_crash2_3h_crash3_3h_TB_imputed",
        k, ".csv")
    write.csv(total_3h_with_imputations,
        file = path_to_imputed)

    imputed_traumabase <- df.imp.mice[[k]]
    imputed_traumabase$V <- rep(0,
        nrow(imputed_traumabase))
    imputed_traumabase$V3 <- rep(0,
        nrow(imputed_traumabase))
    imputed_traumabase$source <- "Traumabase"
    imputed_traumabase <- imputed_traumabase[,
        names(total_allPatients)]
    total_with_imputations <- total_allPatients[total$V ==
        1, ]
    total_with_imputations <- rbind(total_with_imputations,
        imputed_traumabase)

    path_to_imputed <- paste0(data_dir,
        "output_preprocess_combined_allPatients_crash2_crash3_TB_imputed",
        k, ".csv")
    write.csv(total_with_imputations,
        file = path_to_imputed)
```

```
}
```

```
comparison_time <- data.frame(timeSinceInjury = total_with_imputations$timeSinceInjury,
    source = as.factor(total_with_imputations$V3))
levels(comparison_time$source) <- c("Traumabase",
    "CRASH-2", "CRASH-3")
ggplot(comparison_time, aes(x = timeSinceInjury,
    group = source, fill = source)) +
    geom_histogram(binwidth = 0.2,
        alpha = 0.8, position = "dodge",
        aes(y = (..count..)/sum(..count..))) +
    geom_vline(aes(xintercept = 3,
        color = "3h")) + scale_color_manual("",
    values = c(`3h` = "black")) +
    xlab("Time since injury in hours (with imputed Traumabase)") +
    ylab("Percentage") + theme_bw() +
    scale_x_sqrt() + scale_y_sqrt(labels = scales::percent) +
    scale_fill_manual(values = c(Traumabase = "darkorchid4",
        `CRASH-2` = "deepskyblue2",
        `CRASH-3` = "darkorange1"))
```


```
path_to_output <- paste0(fig_dir,
    "timeSinceInjury_crash2_crash3_TB_imputed.pdf")
ggsave(path_to_output)
```

```
## Saving 6.5 x 4.5 in image
```