1 - SURVIVAL ANALYSIS

{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(gridExtra)
library(grid)
library(cmprsk)
library(mstate)
library(cobalt)

rm(list = ls()) # To select all loaded objects and delete them
setwd(dirname(rstudioapi::getActiveDocumentContext()$path)) # setting
up working directory in the location of the .Rmd file

load("./1_datamanaged_files/datamanaged.Rdata") # loading data.managed
data
```

Loading fonctions

```
# home-made function to force writing with two decimals
formattable = function(nbr){return(formatC(nbr,format = "f",digits =
nombreapresvirgule))}
nombreapresvirgule <- 2

# Home-made Fonction to write the p value (by denis)
writepvalue = function(pvalue) {
  if (is.na(pvalue)) {result <- NA} else {
    if(pvalue < 0.001) {
```

```
      result <- "<0.001"
    } else if (pvalue <0.01) {
      result <- formatC( pvalue ,format = "f",digits = 3)
    }
    else {
      (result <- formatC( pvalue ,format = "f",digits = 2) )
      i = 1
      while(result == 0.05) {
        result <- formatC( pvalue ,format = "f",digits = 2 + i)
        i = i + 1
      }}
    return(result)
  }
}
options(scipen = 999)
```

Mini Exploration

```
uniqueN(BARI_DATA[]$patient_id) # number of patients (< than number of
TC)
uniqueN(BARI_DATA[]$TC_id) # unmber of TC

plot <- qplot(x = BARI_DATA[]$time_on_drug)+
  geom_vline(xintercept = 365, size = 1.2, alpha = 0.5)+
  geom_text(aes(x = 365 + 40, label="1 Year", y=20), colour="white",
angle=0)+

  geom_vline(xintercept = 2*365, size = 1.2, alpha = 0.5)+
  geom_text(aes(x = 2*365 + 40, label="2 Year", y=20), colour="black",
angle=0)+

  geom_vline(xintercept = mean(BARI_DATA$time_on_drug), color = "red",
size = 1.2)+
  geom_text(aes(x = mean(BARI_DATA$time_on_drug) + 40, label="Mean",
y=20), colour="red", angle=0)+

  geom_vline(xintercept = median(BARI_DATA$time_on_drug), color =
"green", size = 1.2)+
  geom_text(aes(x = median(BARI_DATA$time_on_drug) - 40,
label="Median", y=20), colour="green", angle=0)+

  labs(x = "Duration of TC (days)", y = "Number of TC", title =
"Repartition of the duration of included TC (all groups)")+

  theme_pubclean()

plot

mean(BARI_DATA$time_on_drug)
```

```
median(BARI_DATA$time_on_drug)

# Nb : Research protocol said we wanted a follow-up duration of
"average of 18 months"
mean(BARI_DATA[cohort == "BARI"]$time_on_drug)/30
mean(BARI_DATA[cohort == "TNFi"]$time_on_drug)/30
mean(BARI_DATA[cohort == "OMA"]$time_on_drug)/30 # looks more like 9
months..

# ok, 24-month follow-up will be complicated
uniqueN(BARI_DATA[time_on_drug > 2*365, TC_id]) # number if TC with
duration > 24 months
```

1. [0] Table 1 BARI vs TNFi et OMA bDMARDs

Common table with all the data

Showing NA to have complete counts and accurate % in each category

```
BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
in drug < 0
BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]

myVars2 <- c("gender", "age_base","disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
"line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
"PREDNISON_STEROID_dose", "dose", "initiation_year",
"time_on_drug","HAQ_score_base")

catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
"obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
"time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
"seropositivity_base")

nonnormalVars2 <- c()
tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
catVars2, strata = "cohort", test = F, includeNA = T)
tab1exp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
```

saving table 1 NA

```
write.xlsx(tab1exp2, file =
"./3_clean_output/BARI_3_groups_table1_NA.xlsx")
```

Without NA to obtain adequate p values

```
BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
in drug < 0
```

```
BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]

myVars2 <- c("gender", "age_base","disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
"line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
"PREDNISON_STEROID_dose", "dose", "initiation_year",
"time_on_drug","HAQ_score_base")

catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
"obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
"time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
"seropositivity_base")

nonnormalVars2 <- c()
tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
catVars2, strata = "cohort", test = T, includeNA = F)
tab1exp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
```

Saving table 1

```
write.xlsx(tab1exp2, file =
"./3_clean_output/BARI_3_groups_table1.xlsx")
```

But BMJ-Open reviewer 2 aksed for p-values in Table 1 that account for patients providing multiple TCs. Here is how to proceed (it is a bit less conservative):

```
library(lme4)
library(lmerTest)

# Two glmer() models have to be compared, to assess the impact of
grouping, for each baseline variable).

# gender
gender.tab1 <- glmer(gender ~ cohort + (1|patient_id), data =
BARI_DATA , family = "binomial")
gender.null <- glmer(gender ~ (1|patient_id), data = BARI_DATA ,
family = "binomial")
anova(gender.tab1, gender.null)

# Age_base
age_base.tab1 <- lmer(age_base ~ cohort + (1|patient_id), data =
BARI_DATA )
age_base.null <- lmer(age_base ~ (1|patient_id), data = BARI_DATA)
anova(age_base.tab1, age_base.null)

# Disease duration
disease_duration_base_years.tab1 <- lmer(disease_duration_base_years ~
cohort + (1|patient_id), data = BARI_DATA )
```

```
disease_duration_base_years.null <- lmer(disease_duration_base_years ~
(1|patient_id), data = BARI_DATA)
anova(disease_duration_base_years.tab1,
disease_duration_base_years.null)


# CDAI raw
CDAI0_raw.tab1 <- lmer(CDAI0_raw ~ cohort + (1|patient_id), data =
BARI_DATA )
CDAI0_raw.null <- lmer(CDAI0_raw ~ (1|patient_id), data = BARI_DATA)
anova(CDAI0_raw.tab1, CDAI0_raw.null)


# CDAI (imputed)
CDAI0.tab1 <- lmer(CDAI0 ~ cohort + (1|patient_id), data = BARI_DATA )
CDAI0.null <- lmer(CDAI0 ~ (1|patient_id), data = BARI_DATA)
anova(CDAI0.tab1, CDAI0.null)


# obesity
obese_base.tab1 <- glmer(obese_base ~ cohort + (1|patient_id), data =
BARI_DATA , family = "binomial")
obese_base.null <- glmer(obese_base ~ (1|patient_id), data = BARI_DATA
, family = "binomial")
anova(obese_base.tab1, obese_base.null)


# smoker base - 1st level vs second level
smoker_base.tab1 <- glmer(smoker_base ~ cohort + (1|patient_id), data
= BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "FORMER_SMOKER")]  ,
family = "binomial")
smoker_base.null <- glmer(smoker_base ~ (1|patient_id), data =
BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "FORMER_SMOKER")]  ,
family = "binomial")
anova(smoker_base.tab1, smoker_base.null)


# smoker base - 2nd level versus third
smoker_base.tab1 <- glmer(smoker_base ~ cohort + (1|patient_id), data
= BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "NEVER_SMOKER")]  ,
family = "binomial")
smoker_base.null <- glmer(smoker_base ~ (1|patient_id), data =
BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "NEVER_SMOKER")]  ,
family = "binomial")
anova(smoker_base.tab1, smoker_base.null)


# seropositivity
seropositivity_base.tab1 <- glmer(seropositivity_base ~ cohort + (1|
```

```
patient_id), data = BARI_DATA , family = "binomial")
seropositivity_base.null <- glmer(seropositivity_base ~ (1|
patient_id), data = BARI_DATA , family = "binomial")
anova(seropositivity_base.tab1, seropositivity_base.null)


# Concomittant csDMARD
TC_with_csDMARD.tab1 <- glmer(TC_with_csDMARD ~ cohort + (1|
patient_id), data = BARI_DATA , family = "binomial")
TC_with_csDMARD.null <- glmer(TC_with_csDMARD ~ (1|patient_id), data =
BARI_DATA , family = "binomial")
anova(TC_with_csDMARD.tab1, TC_with_csDMARD.null)


# Line of therapy
line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
(1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "2nd")]
, family = "binomial")
line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "2nd")] ,
family = "binomial")
anova(line_of_therapy.tab1, line_of_therapy.null)


# Line of therapy
line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
(1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "3rd")]
, family = "binomial")
line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "3rd")] ,
family = "binomial")
anova(line_of_therapy.tab1, line_of_therapy.null)


# Line of therapy
line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
(1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st",
"4th_or_later")] , family = "binomial")
line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
patient_id), data = BARI_DATA[line_of_therapy %in% c("1st",
"4th_or_later")] , family = "binomial")
anova(line_of_therapy.tab1, line_of_therapy.null)


# N_prev_tsDMARD
N_prev_tsDMARD.tab1 <- glmer(as.factor(N_prev_tsDMARD) ~ cohort + (1|
patient_id), data = BARI_DATA , family = "binomial")
N_prev_tsDMARD.null <- glmer(as.factor(N_prev_tsDMARD) ~ (1|
patient_id), data = BARI_DATA , family = "binomial")
```

```
anova(N_prev_tsDMARD.tab1, N_prev_tsDMARD.null)


# Concomittant prednisone
PREDNISON_STEROID.tab1 <- glmer(PREDNISON_STEROID ~ cohort + (1|
patient_id), data = BARI_DATA , family = "binomial")
PREDNISON_STEROID.null <- glmer(PREDNISON_STEROID ~ (1|patient_id),
data = BARI_DATA , family = "binomial")
anova(PREDNISON_STEROID.tab1, PREDNISON_STEROID.null)


# Dose of PREDNISONE
PREDNISON_STEROID_dose.tab1 <- lmer(PREDNISON_STEROID_dose ~ cohort +
(1|patient_id), data = BARI_DATA )
PREDNISON_STEROID_dose.null <- lmer(PREDNISON_STEROID_dose ~ (1|
patient_id), data = BARI_DATA)
anova(PREDNISON_STEROID_dose.tab1, PREDNISON_STEROID_dose.null)
```

Other various computations for Table 1 —-

```
uniqueN(BARI_DATA$patient_id)

mean(BARI_DATA[, time_on_drug])

median(BARI_DATA[cohort == "Bari" , time_on_drug])
median(BARI_DATA[cohort == "OMA" , time_on_drug])
median(BARI_DATA[cohort == "TNFi" , time_on_drug])

median(BARI_DATA[cohort == "OMA"]$time_on_drug)

mean(BARI_DATA[,disease_duration_base_years], na.rm = T)

table(is.na(BARI_DATA$CDAI0_raw), BARI_DATA$cohort) # number of
missing CDAI0_raw...
table(is.na(BARI_DATA$CDAI0), BARI_DATA$cohort) # number of missing
CDAI0... (after imputation)

table(is.na(BARI_DATA$CDAI12_raw), BARI_DATA$cohort) # number of
missing CDAI12_raw...
table(is.na(BARI_DATA$CDAI12), BARI_DATA$cohort) # number of missing
CDAI12 after imputation

hist(BARI_DATA$CDAI0_raw)
hist(BARI_DATA$CDAI0)

summary(BARI_DATA[cohort=="BARI", c("gender",
"age_base","disease_duration_base_years", "CDAI0_raw", "CDAI0",
"obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
"TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
"PREDNISON_STEROID",  "PREDNISON_STEROID_dose", "dose",
```

```
"initiation_year", "time_on_drug","HAQ_score_base")]) # to see NA
values for all variables

summary(BARI_DATA[cohort=="TNFi", c("gender",
"age_base","disease_duration_base_years", "CDAI0_raw", "CDAI0",
"obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
"TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
"PREDNISON_STEROID",  "PREDNISON_STEROID_dose", "dose",
"initiation_year", "time_on_drug","HAQ_score_base")]) # to see NA
values for all variables

summary(BARI_DATA[cohort=="OMA", c("gender",
"age_base","disease_duration_base_years", "CDAI0_raw", "CDAI0",
"obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
"TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
"PREDNISON_STEROID",  "PREDNISON_STEROID_dose", "dose",
"initiation_year", "time_on_drug","HAQ_score_base")]) # to see NA
values for all variables

table(is.na(BARI_DATA$disease_duration_base_years), BARI_DATA$cohort)
# number of missing disease duration...

table(is.na(BARI_DATA$age_base), BARI_DATA$cohort) # number of steroid
doses missing

table(is.na(BARI_DATA$PREDNISON_STEROID_dose), BARI_DATA$cohort) #
number of missing baseline steroids
```

Imputation using MICE —- BARI vs TNFi et OMA bDMARDs —-

Common imputation step with all data

```
BARI <- BARI_DATA[,c("TC_id", "patient_id", "stop_DMARD",
"stop_reasons", "age_base", "concomitant_csDMARD",
"concomitant_csDMARD_type", "TC_with_csDMARD",  "PREDNISON_STEROID",
"CDAI0", "disease_duration_base_years", "time_on_drug", "bmi_base",
"smoker_base", "line_of_therapy", "obese_base", "gender", "cohort",
"adverse_event_reported", "seropositivity_base")] # choose variables
of interest

BARI$smoker_base <- as.factor(BARI$smoker_base) # put labels as factor
BARI$line_of_therapy <- as.factor(BARI$line_of_therapy)
BARI$gender <- as.factor(BARI$gender)
BARI$concomitant_csDMARD <- as.factor(BARI$concomitant_csDMARD)
BARI$PREDNISON_STEROID <- as.factor(BARI$PREDNISON_STEROID)
BARI$cohort <- as.factor(BARI$cohort)

# Imputation

if(!file.exists("./2_cached_files/imputed_data")){ # to avoid re-
```

```
computing if alsready done
  imputed_data <- mice(BARI, m=50, method="pmm", maxit=25, seed=500)
  save(imputed_data, file = "./2_cached_files/imputed_data")
} else {
  load("./2_cached_files/imputed_data")
}

# Subsettings

BARI1 <- BARI[cohort %in% c("BARI", "TNFi")] # creating subset for
BARI vs TNFi comparaison
BARI1[,cohort := as.factor(as.character(cohort))]

imputed_data1 <- complete(imputed_data,"long", include=T) # to put in
long format and categorize variables
imputed_data1 <- imputed_data1[imputed_data1$cohort %in% c("BARI",
"TNFi"),] # to keep only BARI and TNFi rows
imputed_data1$cohort <- as.factor(as.character(imputed_data1$cohort))
imputed_data1 <- as.mids(imputed_data1) # re concateneting in previous
format, to use fit.mult.impute


BARI2 <- BARI[cohort %in% c("BARI", "OMA")] # creating subset for BARI
vs OMA comparaison
BARI2[,cohort := as.factor(as.character(cohort))]

imputed_data2 <- complete(imputed_data,"long", include=T) # to put in
long format and categorize variables
imputed_data2 <- imputed_data2[imputed_data2$cohort %in% c("BARI",
"OMA"),] # to keep only BARI and OMA rows
imputed_data2$cohort <- as.factor(as.character(imputed_data2$cohort))
imputed_data2 <- as.mids(imputed_data2) # re concateneting in previous
format, to use fit.mult.impute
```

1. [1] SURVIVAL ANALYSIS (drug discontinuation)

Exploration

```
table(BARI_DATA$cohort, BARI_DATA$stop_DMARD)
table(BARI_DATA$cohort, BARI_DATA$stop_reasons)
```

Checking adequacy of COX models —-

For BARI vs TNFi

```
# categorization for linearity checking
test1 <- complete(imputed_data1,"long", include=T) # to put in long
format and categorize variables
```

```
test1$agecat <- cut(test1$age_base, 4)
test1$bmicat <- cut(test1$bmi_base, 4)
test1$cdaicat <- cut(test1$CDAI0, 4)
test1$duracat <- cut(test1$disease_duration_base_years, 4)

test1 <- as.mids(test1, .imp=1, .id=2) # re concateneting in previous
format, to use fit.mult.impute

# linearity checking

BARI1.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
as.factor(cohort)+
                                       I(agecat)+
                                       I(bmicat)+
                                       TC_with_csDMARD+
                                       PREDNISON_STEROID+
                                       I(cdaicat)+
                                       I(duracat)+
                                       C(smoker_base, base=3)+
                                       line_of_therapy+
                                       gender+
                                       seropositivity_base+
                                       cluster(patient_id),
                                     fitter = coxph, xtrans = test1,
data = BARI1)

summary(BARI1.adj.mi.test)
rm(BARI1.adj.mi.test)

# Log-linearity of coefficients ?

# Coefs age are between 0.15 and 0.25, let's assume it's ok
# Hum bmi coefs are not so log-linear, rather close to 0
# For CDAI also
# Looks ok for disease_duration_base_years.

# --> Let's keep all variable in the continuous format

# Proportionality test of hazards on raw data

test1ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
as.factor(cohort)+
                              cluster(patient_id),
                         data= BARI1)
cox.zph(test1ph) # it's ok

# Hazard proportionality test on imputed data sets

test1 <- complete(test1,"long",include=T) # To reset the charges to
```

```
long format
test1 <- test1[test1$.imp==1 | test1$.imp==2 |  test1$.imp==3 |
test1$.imp==4 | test1$.imp==5 ,] # To select only 5 data sets

test1ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
~ as.factor(cohort)+
                          I(age_base/10)+
                          bmi_base+
                          TC_with_csDMARD+
                          PREDNISON_STEROID+
                          CDAI0+
                          I(disease_duration_base_years/10)+
                          C(smoker_base, base=3)+
                          line_of_therapy+
                          gender+
                          seropositivity_base+
                          cluster(patient_id),
                        data = test1)

cox.zph(test1ph.adj.mi)

schonfeldall <- cox.zph(test1ph.adj.mi) # Test cox.zph may not be ok,
but it's because of the multiple imputation (often the case with a lot
of data)
for (i in 1:11){
    plot(schonfeldall[i]) # so we should go to a visual testing --> ok
  }

rm(schonfeldall, test1ph.adj.mi, test1)
```

For BARI vs OMA

```
# categorization for linearity checking
test2 <- complete(imputed_data2,"long", include=T) # to put in long
format and categorize variables

test2$agecat <- cut(test2$age_base, 4)
test2$bmicat <- cut(test2$bmi_base, 4)
test2$cdaicat <- cut(test2$CDAI0, 4)
test2$duracat <- cut(test2$disease_duration_base_years, 4)

test2 <- as.mids(test2, .imp=1, .id=2) # re concateneting in previous
format, to use fit.mult.impute

# linearity checking

BARI2.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
as.factor(cohort)+
                                       I(agecat)+
                                       I(bmicat)+
```

```
                                          TC_with_csDMARD+
                                          PREDNISON_STEROID+
                                          I(cdaicat)+
                                          I(duracat)+
                                          C(smoker_base, base=3)+
                                          line_of_therapy+
                                          gender+
                                          seropositivity_base+
                                          cluster(patient_id),
                                        fitter = coxph, xtrans = test2,
data = BARI2)

summary(BARI2.adj.mi.test)
rm(BARI2.adj.mi.test)

# Log-linearity of coeficients ?

# Coefs age are around -0.4, let's assume it's ok
# Hum bmi coefs are discusable
# For CDAI it's ok
# More or less ok for disease_duration_base_years.

# --> Let's keep all variable in the continuous format

# Proportionality test of hazards on raw data

test2ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
as.factor(cohort)+
                              cluster(patient_id),
                         data= BARI2)
cox.zph(test2ph) # it's ok


# Hazard proportionality test on imputed data sets

test2 <- complete(test2,"long",include=T) # To put imputed data in
ling format
test2 <- test2[test2$.imp==1 | test2$.imp==2 |  test2$.imp==3 |
test2$.imp==4 | test2$.imp==5 ,] # To select only 5 datasets

test2ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
~ as.factor(cohort)+
                           I(age_base/10)+
                           bmi_base+
                           TC_with_csDMARD+
                           PREDNISON_STEROID+
                           CDAI0+
                           I(disease_duration_base_years/10)+
                           C(smoker_base, base=3)+
```

```
                              line_of_therapy+
                              gender+
                              seropositivity_base+
                              cluster(patient_id),
                            data=test2)

cox.zph(test2ph.adj.mi)

schonfeldall <- cox.zph(test2ph.adj.mi) # Test cox.zph may not be ok,
but it's because of the multiple imputation (often the case with a lot
of data)
for (i in 1:11){
    plot(schonfeldall[i]) # so we should go to a visual testing --> ok
  }

rm(schonfeldall, test2ph.adj.mi, test2)
```

BARI vs TNFi —-

COX model

Final Cox Model

```
BARI1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
cohort +
                                    I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base+
                                    cluster(patient_id),
                                 fitter = coxph, xtrans = imputed_data1,
data = BARI1)

summary(BARI1.adj.mi)
```

Creation of HR table and p-values

```
ploufrows <- names(BARI1.adj.mi$coefficients)
ploufcols  <- c("HR","95%CI","p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI1.adj.mi)
```

```
for(row in ploufrows)
{
  coxtable[row,"HR"] <-
formattable(plouf$coefficients[row,"exp(coef)"])
  coxtable[row,"95%CI"] <-
paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",
formattable(plouf$conf.int[row,"upper .95"]))
  coxtable[row,"p"] <-  writepvalue(plouf$coefficients[row,"Pr(>|
z|)"])
}

write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR.xlsx") #
saving excel file
```

Forest plot

```
meanall  <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
textall <- c("TNFi (vs BARI)", "Age (decades)", "BMI", "Concomitant
csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
"Disease duration (decades)", "Current smoker (vs non-smoker)",
"Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
"Seropositivity (RF or ACPA)")

dfall <- data.frame(textall, meanall, lowerall, upperall)
dfall$textall <- factor(dfall$textall,
                        levels = textall)

HR_plot_1 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
lowerall, ymax = upperall))+

  geom_pointrange(size=0.5)+
  geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
  geom_hline(yintercept =1, linetype=2)+

  xlab('')+ ylab(" ")+
  ggtitle("BARI vs TNFi")+

  scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
+
  facet_wrap(~textall,nrow=16, strip.position= "right", scales =
"free_y") +

  theme_pubclean()+
  theme(strip.text.y = element_blank(),
        strip.background = element_blank(),
        axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", colour = "black"),
        legend.position="bottom", plot.margin =
```

```
unit(c(1,3,2,1),"lines"))+

    coord_flip()

HR_plot_1

# adding some manual annotation
grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
unit(0.05, "npc"), gp = gpar(fontface = "bold"))
grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
unit(0.05, "npc"), gp = gpar(fontface = "bold"))
```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R

BARI vs TNFi

```
surv_object1 <- Surv(time = BARI1$time_on_drug, event =
BARI1$stop_DMARD) # indiquate time on drug and stop variable
summary(coxph(surv_object1 ~ cohort, data=BARI1))
fit1 <- survfit(surv_object1 ~ cohort, data = BARI1) # this function
creates the data for Kaplan Meyer
fit1
survplot_1 <- ggsurvplot(fit1, data = BARI1,  # plot
           pval = T,
           pval.method = TRUE,
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "TNFi"),
           xlab = "Time (days)",
           xlim = c(0, 700),
           censor = FALSE,
           title = "Non-adjusted model of drug discontinuation by type
of treatment",
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           #palette = c("grey78", "grey10"),
           palette = c("red2", "green3"), # specify colors
           risk.table = T)
survplot_1
summary(fit1, times = 365)
summary(fit1, times = 730)
```

Saving the plot curv object for Lilly

```
plot_BARI_vs_TNFi_data <- survplot_1$data.survplot
write.xlsx(plot_BARI_vs_TNFi_data, file =
"./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
ed.xlsx", row.names = F)
```

Home-made attempt to obtain adjusted curves based on imputed data

```
dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
T)
dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1$.imp != 0,]

BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
                           I(age_base/10)+
                           bmi_base+
                           TC_with_csDMARD+
                           PREDNISON_STEROID+
                           CDAI0+
                           I(disease_duration_base_years/10)+
                           C(smoker_base, base=3)+
                           line_of_therapy+
                           gender+
                           seropositivity_base+
                           cluster(patient_id)+
                           strata(cohort),dummy_cox_impute1), data =
dummy_cox_impute1)


survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
variable = "cohort",
           xlab = "Time (days)",
           title = "Multivariable Cox model of drug discontinuation by
type of treatment - BARI vs TNFi",
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "TNFi"),
           censor = FALSE,
           xlim = c(0, 700),
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           # palette = c("grey78", "grey10")
           palette = c("red2", "green3") # to change colors
           )

# adding some legends
survplot_1_adj <- survplot_1_adj +
    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_1_adj
# summary(BARI_fit1) # to see detailed surv probabilities at given
timepoints
```

```
summary(BARI_fit1, times = 365)
summary(BARI_fit1, times = 730)
```

Saving the plot curv object for Lilly

```
plot_BARI_vs_TNFi_data_adj <- survplot_1_adj$data.survplot
write.xlsx(plot_BARI_vs_TNFi_data_adj, file =
"./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_adj.xlsx",
row.names = F)
```

Sensitivity analysis with package RiskRegression (AIPTW)

```
# Rappel: imputed_data1 = BARI vs TNFi
#         imputed_data2 = BARI vs OMA
library(riskRegression)

# I select only one imputed dataset. Would be even better to find a
way to pool/average the results from the 50 imputed datasets, but it
does not seem doable by default
test.data <- complete(imputed_data1, 1)

# First, we specify the treatment model (propensity score model)
# Logistic regression where the treatment group is the dependent
variable.

m.treatment <- glm(cohort~I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base,
                                    data = test.data, family =
"binomial" )

# Then we specify both the "event model" and the "censoring model".
Both are cox model

m.event <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
                                    I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
```

```
                                    seropositivity_base,
                                    data = test.data, x = TRUE, y =
TRUE)

m.censor <- coxph(Surv(time_on_drug,stop_DMARD==0) ~ cohort +
                                    I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base
                                    , x =TRUE, y = TRUE,
                                    data = test.data)


# And we measure the average treatment effect using function "ate",
specifying the time at which we want to compute the ATE

out <- ate(event = m.event ,
             treatment = m.treatment,
             censor = m.censor,
             data = test.data,
             cause = 1,
             estimator = "AIPTW",
             times = seq(from = 0, to = 500, by = 5))

dt.out <- as.data.table(out)
```

Diagnostics asked by Lily statistician

```
library(cobalt)

# First, the distribution of propensity scores
test.data$pscores <- m.treatment$fitted.values
test.data %>% setDT()

pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
= cohort)) +
  geom_density(alpha = .47) +
  xlab("Estimated Probability of being assigned BARI") +
  ylab("Density") +
    theme_minimal()+
    theme(axis.ticks.y = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          text = element_text(size = 16),
```

```
          axis.title.x = element_text(hjust = 0.2, size = 16))
pscore_plot # overlap


## Computing the weights
test.data$weights <- ifelse(test.data$cohort == "TNFi",
1/test.data$pscores, 1/(1-test.data$pscores))

# Selecting only our covariates of interest (the ones in the ps model)
COVS <- subset(test.data, select = c(cohort,age_base,
                                     bmi_base,
                                     TC_with_csDMARD,
                                     PREDNISON_STEROID,
                                     CDAI0,
                                     disease_duration_base_years,
                                     smoker_base,
                                     line_of_therapy,
                                     gender,
                                     seropositivity_base))

# To get the SMD & variance ratios before/after weighting
# bal.tab(COVS, treat = test.data$cohort, thresholds = 0.1)
# bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
thresholds = 0.1)
# bal.tab(COVS, treat = test.data$cohort,  v.threshold = 2)
# bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
v.threshold = 2)


# But plotting is clearer:
love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
stats =  c("mean.diffs"), thresholds = c(m = .1), var.order =
"adjusted")

# We can also plot variance ratios for continuous variables
love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
stats =  c("variance.ratios"))

# propensity scores enhance the balance overall, except for the CDAI0.
However, this is the reason we use the AIPTW. The remaining imbalance
is accounted for by the outcome model (outcome model is the cox
regression), and the misspecification of the outcome model is
mitigated by the balancing done by propensity score.
```

First plot to get the difference in average treatment effect in percentage

```
plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.4)+
```

```
  geom_line(aes(y = estimate, color = level), size = 1)+
  #geom_vline(xintercept = 90)+

  scale_colour_manual(values = c("lightblue","darkseagreen3"))+
  scale_fill_manual(values = c("lightblue","darkseagreen3"))+
  theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
= scales::percent)+

  xlab("Days since intiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:")+
  labs(group = "Groups:")+
  theme_bw(base_size = 14)+
  theme(axis.title.x = element_text(margin = margin(t = .3,unit =
"cm")),
        axis.title.y = element_text(margin = margin(r = .3,unit =
"cm")))

plot.ate.diff
```

Second plot to get the ratio in average treatment effect

```
plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.4)+
  geom_line(aes(y = estimate, color = level), size = 2)+

  theme_minimal()+
  theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
  scale_x_continuous(breaks=seq(0,500,50))+
  scale_y_continuous(limits = c(0.9,4.5))+

  xlab("Days since intiation of treatment")+
  ylab("Ratio in Average Treatment Effect")+
  labs(colour="treatment", fill = "treatment")

plot.ate.ratio
```

We can also consider the AIPTW estimate at a specific time point. For example at365-day.

```
r.one <- dt.out[type == "diffRisk" & time == 365, .
(estimate,lower,upper,p.value)]
r.two <- dt.out[type == "ratioRisk" & time == 365, .
(estimate,lower,upper,p.value)]

ploufrows <- c("Difference in average treatment effect","Ratio in
average treatment effect")
ploufcols  <- c("Estimate","95%CI","p")
```

```
table <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(table) <- ploufrows
colnames(table) <- ploufcols

library(formattable)
table[1,"Estimate"] <-  paste0(formattable(r.one$estimate*100),"%")
table[1,"95%CI"] <-
paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
table[1,"p"] <-  writepvalue(r.one$p.value)
table[2,"Estimate"] <-  paste0(r.two$estimate)
table[2,"95%CI"] <-  paste0(r.two$lower,"-",r.two$upper)
table[2,"p"] <-  writepvalue(r.two$p.value)

table

# Interpretation: If every patient had received BARI, the 365-day risk
of treatment discontinuation would have been 19.34% (points) lower
compared to when every patient had received TNFi.
```

BARI vs OMA —-

COX model

```
BARI2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
cohort+
                                  I(age_base/10)+
                                  bmi_base+
                                  TC_with_csDMARD+
                                  PREDNISON_STEROID+
                                  I(CDAI0/10)+
                                  I(disease_duration_base_years/10)+
                                  C(smoker_base, base=3)+
                                  line_of_therapy+
                                  gender+
                                  seropositivity_base+
                                  cluster(patient_id),
                                fitter = coxph, xtrans = imputed_data2,
data = BARI2)

summary(BARI2.adj.mi)
```

Creation of HR table and p-values (denis)

```
ploufrows <- names(BARI2.adj.mi$coefficients)
ploufcols  <- c("HR","95%CI","p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI2.adj.mi)
```

```
for(row in ploufrows)
{
  coxtable[row,"HR"] <-
formattable(plouf$coefficients[row,"exp(coef)"])
  coxtable[row,"95%CI"] <-
paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
ouf$conf.int[row,"upper .95"]))
  coxtable[row,"p"] <-  writepvalue(plouf$coefficients[row,"Pr(>|
z|)"])
}

write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR.xlsx") #
saving excel file
```

Forest plot

```
meanall  <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
textall <- c("OMA (vs BARI)", "Age (decades)", "BMI", "Concomitant
csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
"Disease duration (decades)", "Current smoker (vs non-smoker)",
             "Former smoker (vs non-smoker)", "2nd line therapy (vs
1st)", "3rd line therapy (vs 1st)", "4th or later line (vs 1st)",
"Female gender", "Seropositivity (RF or ACPA)")

dfall <- data.frame(textall, meanall, lowerall, upperall)
dfall$textall <- factor(dfall$textall,
                        levels = textall)

HR_plot_2 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
lowerall, ymax = upperall))+

  geom_pointrange(size=0.5)+
  geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
  geom_hline(yintercept =1, linetype=2)+

  xlab('')+ ylab(" ")+
  ggtitle("BARI vs OMA")+

  scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
+
  facet_wrap(~textall,nrow=16, strip.position= "right", scales =
"free_y") +

  theme_pubclean()+
  theme(strip.text.y = element_blank(),
        strip.background = element_blank(),
        axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", colour = "black"),
        legend.position="bottom", plot.margin =
```

```
unit(c(1,3,2,1),"lines"))+

    coord_flip()

HR_plot_2

# adding some manual annotation
grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
unit(0.05, "npc"), gp = gpar(fontface = "bold"))
grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
unit(0.05, "npc"), gp = gpar(fontface = "bold"))
```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R)

BARI vs OMA

```
surv_object2 <- Surv(time = BARI2$time_on_drug, event =
BARI2$stop_DMARD)
fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
creates the data for Kaplan Meyer
survplot_2 <- ggsurvplot(fit2, data = BARI2,  # plot
           pval = T,
           pval.method = TRUE,
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "OMA"),
           xlab = "Time (days)",
           xlim = c(0, 700),
           censor = FALSE,
           title = "Non-adjusted model of drug discontinuation by type
of treatment",
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           #palette = c("grey78", "grey50"),
           palette = c("red2", "blue3"), # to put colors
           risk.table = T)
survplot_2
summary(fit2, times = 365)
summary(fit2, times = 730)
```

Saving the plot curv object for Lilly

```
plot_BARI_vs_OMA_data <- survplot_2$data.survplot
write.xlsx(plot_BARI_vs_OMA_data, file =
"./3_clean_output/Lilly_curves_excel/plot_BARI_vs_OMA_data_non_adjuste
d.xlsx", row.names = F)
```

Home-made attempt to obtain adjusted cuves based on imputed data :

```
dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
T)
dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]

BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
                           I(age_base/10)+
                           bmi_base+
                           TC_with_csDMARD+
                           PREDNISON_STEROID+
                           CDAI0+
                           I(disease_duration_base_years/10)+
                           C(smoker_base, base=3)+
                           line_of_therapy+
                           gender+
                           seropositivity_base+
                           cluster(patient_id)+
                           strata(cohort),dummy_cox_impute2), data =
dummy_cox_impute2)

survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
variable = "cohort",
           xlab = "Time (days)",
           title = "Multivariable Cox model of drug discontinuation by
type of treatment - BARI vs OMA",
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "OMA bDMARDs"),
           censor = FALSE,
           xlim = c(0, 700),
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           # palette = c("grey78", "grey50")
           palette = c("red2", "blue3") # to change colors
           )

# adding some legends
survplot_2_adj <- survplot_2_adj +
    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_2_adj
summary(BARI_fit2, times = 365) # to see detailed surv probabilities
at given timepoints
summary(BARI_fit2, times = 730)
```

Saving the plot curv object for Lilly

```
plot_BARI_vs_OMA_data_adj <- survplot_2_adj$data.survplot
write.xlsx(plot_BARI_vs_OMA_data_adj, file =
"./3_clean_output/Lilly_curves_excel/plot_BARI_vs_OMA_data_adj.xlsx",
row.names = F)
```

Sensitivity analysis with package RiskRegression (AIPTW)

```
# I select only one imputed dataset. Would be good to find a way to
pool the results from the 50 datasets imputed.
test.data2 <- complete(imputed_data2,1)

# First, we specify the treatment model (propensity score model)
# Logistic regression where the treatment group is the dependent
variable.

m.treatment2 <- glm(cohort~I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base,
                                    data = test.data2, family =
"binomial" )

# Then we specify both the "event model" and the "censoring model".
Both are cox model

m.event2 <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
                                    I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base,
                                    data = test.data2, x = TRUE, y =
TRUE)

m.censor2 <- coxph(Surv(time_on_drug,stop_DMARD==0) ~ cohort +
                                    I(age_base/10)+
                                    bmi_base+
                                    TC_with_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
```

```
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base
                                    , x =TRUE, y = TRUE,
                                    data = test.data2)

# And we measure the average treatment effect using function "ate",
specifying the times at which we want to compute the ATE

out2 <- ate(event = m.event2 ,
            treatment = m.treatment2,
            censor = m.censor2,
            data = test.data2,
            cause = 1,
            estimator = "AIPTW",
            times = seq(from = 0, to = 500, by = 5))

dt.out2 <- as.data.table(out2)
```

Diagnostics asked by Lily statistician

```
library(cobalt)

# First, the distribution of propensity scores
test.data2$pscores <- m.treatment2$fitted.values
test.data2 %>% setDT()

pscore_plot2 <- ggplot(test.data2,aes(x = pscores, color = cohort,
fill = cohort)) +
  geom_density(alpha = .47) +
  xlab("Estimated Probability of being assigned BARI") +
  ylab("Density") +
    theme_minimal()+
    theme(axis.ticks.y=element_blank(),
          panel.grid.minor=element_blank(),
          legend.title=element_blank(),
          text = element_text(size = 16),
          axis.title.x =element_text(hjust = 0.2, size = 16))
pscore_plot2
# Good overlap

## Computing the weights
test.data2$weights <- ifelse(test.data2$cohort == "OMA",
1/test.data2$pscores, 1/(1-test.data2$pscores))

# Selecting only our covariates of interest (the ones in the ps model)
COVS_2 <- subset(test.data2, select = c(cohort,age_base,
                                    bmi_base,
```

```
                                    TC_with_csDMARD,
                                    PREDNISON_STEROID,
                                    CDAI0,
                                    disease_duration_base_years,
                                    smoker_base,
                                    line_of_therapy,
                                    gender,
                                    seropositivity_base))

# To get the SMD & variance ratios before/after weighting
# bal.tab(COVS_2, treat = test.data2$cohort, thresholds = 0.1)
# bal.tab(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, thresholds = 0.1)
# bal.tab(COVS_2, treat = test.data2$cohort,  v.threshold = 2)
# bal.tab(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, v.threshold = 2)
#

#But plotting it is better:
love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, stats =  c("mean.diffs"), thresholds = c(m = .1),
var.order = "adjusted")

# We can also plot variance ratios for continuous variables
love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights,stats =  c("variance.ratios"))

# propensity scores enhance the balance overall, except for the CDAI0.
However, this is the reason we use the AIPTW. The remaining imbalance
is accounted for by the outcome model (outcome model is the cox
regression), and the misspecification of the outcome model is
mitigated by the balancing done by propensity score.
```

First plot to get the difference in average treatment effect in percentage

```
plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.3)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
= scales::percent)+

  xlab("Days since intiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:", title = "Absolute risk of
treatment discontinuation by type of treatment - BARI vs TNFi")+
```

```
  labs(group = "Groups:")

plot.ate.diff2
```

Second plot to get the ratio in average treatment effect

```
plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.3)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  theme_minimal()+
  theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
  scale_x_continuous(breaks=seq(100,400,50))+
  scale_y_continuous(limits = c(0.8,3))+

  xlab("Days since intiation of treatment")+
  ylab("Ratio in Average Treatment Effect")+
  labs(colour="treatment", fill = "treatment")

plot.ate.ratio2
```

We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```
r.one <- dt.out2[type == "diffRisk" & time == 365, .
(estimate,lower,upper,p.value)]
r.two <- dt.out2[type == "ratioRisk" & time == 365, .
(estimate,lower,upper,p.value)]

ploufrows <- c("Difference in average treatment effect","Ratio in
average treatment effect")
ploufcols  <- c("Estimate","95%CI","p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols

library(formattable)
coxtable[1,"Estimate"] <-  paste0(formattable(r.one$estimate*100),"%")
coxtable[1,"95%CI"] <-
paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
coxtable[1,"p"] <-  writepvalue(r.one$p.value)
coxtable[2,"Estimate"] <-  paste0(r.two$estimate)
coxtable[2,"95%CI"] <-  paste0(r.two$lower,"-",r.two$upper)
coxtable[2,"p"] <-  writepvalue(r.two$p.value)

coxtable

# Interpretation: If every patient had received BARI, the 365-day risk
```

```
of treatment discontinuation would have been xx% (points) lower
compared to when every patient had received TNFi.
```

.

[3] 1st LINE vs 1st LINE analysis

Common Table 1

Table 1 with NA, to have exact counts and proportions

```
BARI_first <- BARI_DATA
BARI_first <- BARI_first[line_of_therapy == "1st"] # selection of TC
first line

myVars2 <- c("gender", "age_base","disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
"line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
"PREDNISON_STEROID_dose", "dose", "initiation_year",
"time_on_drug","HAQ_score_base")

catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
"obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
"time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
"seropositivity_base")

nonnormalVars <- c()

tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
catVars2, strata = "cohort", test = F, includeNA = T)
tab1exp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
```

saving

```
write.xlsx(tab1exp, file = "./3_clean_output/BARI 3 groups first line
table1 NA.xlsx")
```

Table 1 without NA to have adequate p values to interpret

```
BARI_first <- BARI_DATA
BARI_first <- BARI_first[line_of_therapy == "1st"] # selection TC
first line

myVars2 <- c("gender", "age_base","disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
```

```
"line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
"PREDNISON_STEROID_dose", "dose", "initiation_year",
"time_on_drug","HAQ_score_base")

catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
"obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
"time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
"seropositivity_base")

nonnormalVars <- c()

tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
catVars2, strata = "cohort", test = T, includeNA = F)
tab1exp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
```

Saving

```
write.xlsx(tab1exp, file = "./3_clean_output/BARI 3 groups first line
table1.xlsx")

summary(BARI_first[cohort=="BARI", c("TC_id", "patient_id",
"stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
"concomitant_csDMARD_type", "TC_with_csDMARD",  "PREDNISON_STEROID",
"CDAI0","CDAI0_raw", "disease_duration_base_years", "time_on_drug",
"bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
"cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
to see NA values for all variables

summary(BARI_first[cohort=="TNFi", c("TC_id", "patient_id",
"stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
"concomitant_csDMARD_type", "TC_with_csDMARD",  "PREDNISON_STEROID",
"CDAI0","CDAI0_raw", "disease_duration_base_years", "time_on_drug",
"bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
"cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
to see NA values for all variables

summary(BARI_first[cohort=="OMA", c("TC_id", "patient_id",
"stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
"concomitant_csDMARD_type", "TC_with_csDMARD",  "PREDNISON_STEROID",
"CDAI0","CDAI0_raw", "disease_duration_base_years", "time_on_drug",
"bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
"cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
to see NA values for all variables
```

Non-adjusted Survival curves

BARI vs TNFi

```
BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
```

```
surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
BARI_first1$stop_DMARD) # indiquate stop variable and time_on_drug
summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # function
which creates Kaplan-meier data
survplot_first1 <- ggsurvplot(fit3, data = BARI_first1,   # plot
           pval = T,
           pval.method = TRUE,
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "TFNi"),
           xlab = "Time (days)",
           xlim = c(0, 700),
           censor = FALSE,
           title = "Non-adjusted model of drug discontinuation by type
of treatment",
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           # palette = c("grey78", "grey50", "grey10"),
           palette = c("red2", "green3"), # to get colors
           risk.table = T
           )

survplot_first1
table(BARI_first1$cohort)
summary(fit3)

rm(surv_object3, fit3)
```

BARI vs OMA

```
BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
c("BARI", "OMA")] # selection des TC TNFi

surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
BARI_first2$stop_DMARD) # indiquate stop variable and time_on_drug
summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # function
which creates Kaplan-meier data
survplot_first2 <- ggsurvplot(fit3, data = BARI_first2,   # plot
           pval = T,
           pval.method = TRUE,
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "OMA"),
           xlab = "Time (days)",
           xlim = c(0, 700),
           censor = FALSE,
           title = "Non-adjusted model of drug discontinuation by type
of treatment",
```

```
            surv.median.line = "v",
            linetype = 1,
            size = 1.5,
            ggtheme = theme_minimal(),
            # palette = c("grey78", "grey50", "grey10"),
            palette = c("red2", "blue3"), # to get colors
            risk.table = T
            )

survplot_first2
table(BARI_first2$cohort)
summary(fit3)

rm(surv_object3, fit3)
```

Adjusted survival analyses

BARI vs TNFi

Verification (quick)

```
# Test of proportionality of hazards on raw data
test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
as.factor(cohort)+
                             cluster(patient_id),
                       data= BARI_first1)
cox.zph(test_first_ph)
```

Adjusted Cox-model

```
imputed_data1_first <- complete(imputed_data1,"long",include=T) # to
put in the long format
imputed_data1_first <- filter(imputed_data1_first, line_of_therapy ==
"1st")  # only keep 1st line imputed TC
imputed_data1_first <- as.mids(imputed_data1_first) # put back in
previous format, to use fit.mult.impute

BARI_first1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
cohort+
                                    I(age_base/10)+
                                    bmi_base+
                                    concomitant_csDMARD+
                                    PREDNISON_STEROID+
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base+
                                    cluster(patient_id),
                                  fitter = coxph, xtrans =
```

```
imputed_data1_first, data = BARI_first1)
summary(BARI_first1.adj.mi)
```

Creation of HR table with p-values

```
ploufrows <- names(BARI_first1.adj.mi$coefficients)
ploufcols  <- c("HR","95%CI","p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI_first1.adj.mi)

for(row in ploufrows)
{
  coxtable[row,"HR"] <-
formattable(plouf$coefficients[row,"exp(coef)"])
  coxtable[row,"95%CI"] <-
paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
ouf$conf.int[row,"upper .95"]))
  coxtable[row,"p"] <-  writepvalue(plouf$coefficients[row,"Pr(>|
z|)"])
}

write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR 1st
lines.xlsx") # save in excel format
```

Adjusted curves with imputed data

```
dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
include = T)
dummy_cox_impute_first1 <-
dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]

BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
                             I(age_base/10)+
                             bmi_base+
                             concomitant_csDMARD+
                             PREDNISON_STEROID+
                             CDAI0+
                             I(disease_duration_base_years/10)+
                             C(smoker_base, base=3)+
                             line_of_therapy+
                             gender+
                             seropositivity_base+
                             cluster(patient_id)+
                             strata(cohort),dummy_cox_impute_first1),
data = dummy_cox_impute_first1)


survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
```

```
dummy_cox_impute_first1, variable = "cohort",
           xlab = "Time (days)",
           title = "Multivariable Cox model of drug discontinuation by
type of treatment - 1st line vs 1st line",
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "TNFi"),
           censor = FALSE,
           xlim = c(0, 700),
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           #palette = c("grey78", "grey10")
           palette = c("red2", "green3"), # to get colors
           )

survplot_first1_adj <- survplot_first1_adj +
    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_first1_adj
table(BARI_first1$cohort)
rm(dummy_cox_impute_first1, BARI_first1_fit)
```

BARI vs OMA

Verification (quick)

```
# Test of proportionality of hazards on raw data
test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
as.factor(cohort)+
                            cluster(patient_id),
                       data= BARI_first2)
cox.zph(test_first_ph)
```

Adjusted Cox-model

```
imputed_data2_first <- complete(imputed_data2,"long",include=T) # to
put in the long format
imputed_data2_first <- filter(imputed_data2_first, line_of_therapy ==
"1st")  # only keep 1st line imputed TC
imputed_data2_first <- as.mids(imputed_data2_first) # put back in
previous format, to use fit.mult.impute

BARI_first2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
cohort+
                                      I(age_base/10)+
                                      bmi_base+
                                      concomitant_csDMARD+
                                      PREDNISON_STEROID+
```

```
                                    I(CDAI0/10)+
                                    I(disease_duration_base_years/10)+
                                    C(smoker_base, base=3)+
                                    line_of_therapy+
                                    gender+
                                    seropositivity_base+
                                    cluster(patient_id),
                                  fitter = coxph, xtrans =
imputed_data2_first, data = BARI_first2)
summary(BARI_first2.adj.mi)
```

Creation of HR table with p-values

```
ploufrows <- names(BARI_first2.adj.mi$coefficients)
ploufcols  <- c("HR","95%CI","p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI_first2.adj.mi)

for(row in ploufrows)
{
  coxtable[row,"HR"] <-
formattable(plouf$coefficients[row,"exp(coef)"])
  coxtable[row,"95%CI"] <-
paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
ouf$conf.int[row,"upper .95"]))
  coxtable[row,"p"] <-  writepvalue(plouf$coefficients[row,"Pr(>|
z|)"])
}

write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR 1st
lines.xlsx") # save in excel format
```

Adjusted curves with imputed data

```
dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
include = T)
dummy_cox_impute_first2 <-
dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]

BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
                             I(age_base/10)+
                             bmi_base+
                             concomitant_csDMARD+
                             PREDNISON_STEROID+
                             CDAI0+
                             I(disease_duration_base_years/10)+
                             C(smoker_base, base=3)+
                             line_of_therapy+
```

```
                                gender+
                                seropositivity_base+
                                cluster(patient_id)+
                                strata(cohort),dummy_cox_impute_first2),
data = dummy_cox_impute_first2)


survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
dummy_cox_impute_first2, variable = "cohort",
           xlab = "Time (days)",
           title = "Multivariable Cox model of drug discontinuation by
type of treatment - 1st line vs 1st line",
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "OMA bDMARDs"),
           censor = FALSE,
           xlim = c(0, 700),
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           #palette = c("grey78", "grey50")
           palette = c("red2", "blue3"), # to get colors
           )

survplot_first2_adj <- survplot_first2_adj +
    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_first2_adj
table(BARI_first2$cohort)
rm(dummy_cox_impute_first2, BARI_first2_fit)
```

1. [4] LACK of EFFICACYY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```
BARI_comp <- copy(BARI_DATA)

#General

BARI_comp[stop_reasons == "ADVERSE_EVENT",status := 1]
BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
status := 3]
```

```
BARI_comp[stop_reasons == "CONTINUE", status := 0]
BARI_comp$cohort <- as.factor(BARI_comp$cohort)

library(reshape)

BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
ci_BARI <- Cuminc(time = "time_on_drug",status = "status", data =
BARI_comp_B)
ci_BARI <- ci_BARI[,-c(2,6,7,8,9)]
ci_long_BARI <- reshape2::melt(ci_BARI,id.vars = "time")

BARI_comp_T <- BARI_comp[cohort %in% c("TNFi")] #TNFi only
ci_TNFi <- Cuminc(time = "time_on_drug",status = "status", data =
BARI_comp_T)
ci_TNFi <- ci_TNFi[,-c(2,6,7,8,9)]
ci_long_TNFi <- reshape2::melt(ci_TNFi,id.vars = "time")

ci_long_BARI$cohort <- 0
ci_long_TNFi$cohort <- 1
ci_long <- rbind(ci_long_BARI,ci_long_TNFi)
ci_long$cohort <- as.factor(ci_long$cohort)

plot2 <- ggplot(data = ci_long, aes(x = time,
                                    y = value,
                                    linetype =
interaction(cohort,variable),
                                    col =
interaction(cohort,variable))) +
  geom_line(size = 0.75) +
  scale_color_manual(name = "",
                     values
=c("#08306B","#08306B","#238B45","#238B45","#FD8D3C","#FD8D3C"),
                     labels = c("Adverse Event (BARI)","Adverse Event
(TNFi)","Ineffectiveness (BARI)","Inefectiveness (TNFi)","Other
(BARI)","Other (TNFi)"))+
  scale_linetype_manual(name="",
                        values = c(1,3,1,3,1,3),
                        labels = c("Adverse Event (BARI)","Adverse
Event (TNFi)","Ineffectiveness (BARI)","Inefectiveness (TNFi)","Other
(BARI)","Other (TNFi)"))+
  scale_x_continuous(name = "Time", limits = c(1,365)) +
  scale_y_continuous(name = "Cumulative incidence", limits =
c(0.0,0.3)) +
  theme_bw()+
  theme(strip.text.y = element_blank(),
        strip.background = element_blank(),
        axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", colour = "black"),
        legend.position="right", plot.margin =
unit(c(1,3,2,1),"lines"))+
```

```
  #ggtitle("Cumulative incidence functions")+
  theme(plot.title = element_text(hjust = 0.5))

plot2
```

Adjusting variables

```
# Covariates of interest for Cox

covs <-
c("cohort","age_base","bmi_base","TC_with_csDMARD","PREDNISON_STEROID"
,"CDAI0","disease_duration_base_years","smoker_base","line_of_therapy"
,"gender","seropositivity_base")
```

Cause-specific hazard model

```
# Rappel: imputed_data1 = BARI vs TNFi
#         imputed_data2 = BARI vs OMA

# Transition matrix definition
tmat <- trans.comprisk(2, names = c("event-free","ae","lae"))
tmat

imputed_data1_long <- complete(imputed_data1, action = "long") %>%
setDT()
imputed_data1_long[,stop_ae := fifelse(stop_reasons ==
"ADVERSE_EVENT",1,0)]
imputed_data1_long[,stop_lae := fifelse(stop_reasons ==
"NOT_EFFECTIVE",1,0)]
imputed_data1_long[,stop_other := fifelse(stop_reasons == "OTHER" |
stop_reasons == "REMISSION",1,0)]
#[,continue := fifelse(stop_reasons == "CONTINUE",1,0)]
imputed_data1_long[,continue := fifelse(stop_reasons == "OTHER" |
stop_reasons == "REMISSION" | stop_reasons == "CONTINUE",1,0)]

M <- imputed_data1$m

mice_fit <- lapply(1:M,function(m){

  # subset
  data_sub <- imputed_data1_long[.imp == m]

  mst_hosp <- msprep(time =
c("time_on_drug","time_on_drug","time_on_drug"),
                   status = c("continue","stop_ae","stop_lae"),
                   data = as.data.frame(data_sub),
                   trans = tmat,
                   keep = covs)
  # get covariates
tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)
```

```
  # fit
coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
                                paste0(tmp_cov,collapse = " + "),
                                "+ strata(trans)")),
              data = tmp,
              method = "breslow")

}) %>%
  as.mira()

est <- pool(mice_fit)

# Transition 1 = Adverse Event. Hazard Ratio vs VARI
# Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
# estimate = Hazard ratio

summary(est, conf.int = T, exponentiate = T)

# Conclusion
# => The hazard ratio of lack of efficacy (lae) for TNFi is 65% higher
than for BARI. Significant.
# => No difference between TNFi and BARI for Adverse Event (ae)
```

Clean table with confidence intervals & p-values

```
# Hazard ratios
ploufrows <- as.character(summary(est)$term)

ploufcols  <- c("HR","95%CI","p")
coxtable_csh <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable_csh) <- ploufrows
colnames(coxtable_csh) <- ploufcols
plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()

for(row in ploufrows)
{
  coxtable_csh[row,"HR"] <-  formattable(plouf[term %in% row,
estimate])
  coxtable_csh[row,"95%CI"] <-  paste0(formattable(plouf[term %in%
row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
  coxtable_csh[row,"p"] <-  writepvalue(plouf[term %in% row,
p.value])}

output <- coxtable_csh
row.names(output)[1:2] <- c("TNFi Adverse Event (vs BARI)", "TNFi Lack
of Eff (vs BARI)")

# Transition 1 = Adverse Event. Hazard Ratio vs VARI
```

```
# Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
output
```

Subdistribution hazard model (Fine-Gray)

```
# Status variable
imputed_data1_long[stop_reasons == "ADVERSE_EVENT",status := 1]
imputed_data1_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
imputed_data1_long[stop_reasons == "OTHER" | stop_reasons ==
"REMISSION", status := 3]
imputed_data1_long[stop_reasons == "CONTINUE", status := 0]

## ATTENTION levels() re-ecrit juste l'étiquette!! Change pas la
donnée !!! Donc ça re écrit les lablels

imputed_data1_long$line_of_therapy <-
as.factor(imputed_data1_long$line_of_therapy)
imputed_data1_long$seropositivity_base <-
as.factor(imputed_data1_long$seropositivity_base)

levels(imputed_data1_long$cohort) <- c("0","1")
levels(imputed_data1_long$line_of_therapy) <- c("0","1","2","3")
levels(imputed_data1_long$gender) <- c("0","1")
levels(imputed_data1_long$smoker_base) <- c("2","1","0")
levels(imputed_data1_long$smoker_base)
levels(imputed_data1_long$seropositivity_base) <- c("0","1")

M <- imputed_data1$m

# First loop to get estimates for event = 1: ADVERSE EVENT

mice_fit <- lapply(1:M,function(m){
    # subset
  BARI_toto <- imputed_data1_long[.imp == m]
    # subdistribution hazard model
  shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
BARI_toto[,..covs],failcode = 1,cencode = 0)

}) %>%
  as.mira()
est <- pool(mice_fit)
summary(est, conf.int = T, exponentiate = T)

# Second loop to get estimates for event = 2: LACK OF EFFICACY

mice_fit2 <- lapply(1:M,function(m){
    # subset
  BARI_toto <- imputed_data1_long[.imp == m]
    #subdistribution hazard model
  shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
```

```
BARI_toto[,..covs],failcode = 2,cencode = 0)

}) %>%
  as.mira()

est2 <- pool(mice_fit2)
summary(est2, conf.int = T, exponentiate = T)

# Conclusions:
# No significant difference in incidence of adverse event between TNFi
and BARI
# Increased incidence of lack of efficacy for TNFi compared to BARI.
# CAREFUL: using a Fine-Gray model allows us to make claim about the
association between a covariate and the direction of the increase in
incidence, but we can't quantify the magnitude of the increase in
incidence.
```

Clean tables with Hazard ratios with confidence intervals & p-values

```
# Adverse event
ploufrows <- as.character(summary(est)$term)
ploufcols  <- c("HR","95%CI","p")
coxtable_ae <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable_ae) <- ploufrows
colnames(coxtable_ae) <- ploufcols
plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()

for(row in ploufrows)
{
  coxtable_ae[row,"HR"] <-  formattable(plouf[term %in% row,
estimate])
  coxtable_ae[row,"95%CI"] <-  paste0(formattable(plouf[term %in% row,
`2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
  coxtable_ae[row,"p"] <-  writepvalue(plouf[term %in% row, p.value])}

row.names(coxtable_ae)[1] <- c("TNFi vs BARI Advsere Events")

# Lack of efficacy
ploufrows <- as.character(summary(est2)$term)
ploufcols  <- c("HR","95%CI","p")
coxtable_lae <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable_lae) <- ploufrows
colnames(coxtable_lae) <- ploufcols
plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()

for(row in ploufrows)
{
  coxtable_lae[row,"HR"] <-  formattable(plouf[term %in% row,
```

```
estimate])
  coxtable_lae[row,"95%CI"] <-  paste0(formattable(plouf[term %in%
row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
  coxtable_lae[row,"p"] <-  writepvalue(plouf[term %in% row,
p.value])}

row.names(coxtable_lae)[1] <- c("TNFi vs BARI Lack of Eff")

# output
coxtable_ae
coxtable_lae

write.xlsx(coxtable_ae, file="./3_clean_output/BARI vs TNFi HR
competing risk Fine-Gray AE.xlsx") # saving excel file
write.xlsx(coxtable_lae, file="./3_clean_output/BARI vs TNFi HR
competing risk Fine-Gray LAE.xlsx") # saving excel file
```

(BARI vs OMA)

Cumulative incidence function

```
BARI_comp <- copy(BARI_DATA)

#General

BARI_comp[stop_reasons == "ADVERSE_EVENT",status := 1]
BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
status := 3]
BARI_comp[stop_reasons == "CONTINUE", status := 0]
BARI_comp$cohort <- as.factor(BARI_comp$cohort)

library(reshape)

BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
ci_BARI <- Cuminc(time = "time_on_drug",status = "status", data =
BARI_comp_B)
ci_BARI <- ci_BARI[,-c(2,6,7,8,9)]
ci_long_BARI <- reshape2::melt(ci_BARI,id.vars = "time")

BARI_comp_O <- BARI_comp[cohort %in% c("OMA")] #OMA only
ci_OMA <- Cuminc(time = "time_on_drug",status = "status", data =
BARI_comp_O)
ci_OMA <- ci_OMA[,-c(2,6,7,8,9)]
ci_long_OMA <- reshape2::melt(ci_OMA,id.vars = "time")

ci_long_BARI$cohort <- 0
ci_long_OMA$cohort <- 1
ci_long_2 <- rbind(ci_long_BARI,ci_long_OMA)
ci_long_2$cohort <- as.factor(ci_long_2$cohort)
```

```
plot3 <- ggplot(data = ci_long_2, aes(x = time,
                                      y = value,
                                      linetype =
interaction(cohort,variable),
                                      col =
interaction(cohort,variable))) +
  geom_line(size = 0.75) +
  scale_color_manual(name = "",
                     values =
c("#08306B","#08306B","#238B45","#238B45","#FD8D3C","#FD8D3C"),
                     labels = c("Adverse Event (BARI)","Adverse Event
(OMA)","Ineffectiveness (BARI)","Inefectiveness (OMA)","Other
(BARI)","Other (OMA)"))+
  scale_linetype_manual(name="",
                        values = c(1,3,1,3,1,3),
                        labels = c("Adverse Event (BARI)","Adverse
Event (OMA)","Ineffectiveness (BARI)","Inefectiveness (OMA)","Other
(BARI)","Other (OMA)"))+
  scale_x_continuous(name = "Time", limits = c(1,365)) +
  scale_y_continuous(name = "Cumulative incidence", limits =
c(0.0,0.3)) +
  theme_bw()+
  theme(strip.text.y = element_blank(),
        strip.background = element_blank(),
        axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", colour = "black"),
        legend.position="right", plot.margin =
unit(c(1,3,2,1),"lines"))+
  #ggtitle("Cumulative incidence functions")+
  theme(plot.title = element_text(hjust = 0.5))

plot3
```

Adjusting variables

```
# Covariates of interest for Cox

covs <-
c("cohort","age_base","bmi_base","TC_with_csDMARD","PREDNISON_STEROID"
,"CDAI0","disease_duration_base_years","smoker_base","line_of_therapy"
,"gender","seropositivity_base")
```

Cause-specific hazard model

```
# Rappel: imputed_data1 = BARI vs TNFi
#         imputed_data2 = BARI vs OMA

# Transition matrix definition
library(mstate)
tmat <- trans.comprisk(2, names = c("event-free","ae","lae"))
tmat
```

```
imputed_data2_long <- complete(imputed_data2, action = "long") %>%
setDT()
imputed_data2_long[,stop_ae := fifelse(stop_reasons ==
"ADVERSE_EVENT",1,0)]
imputed_data2_long[,stop_lae := fifelse(stop_reasons ==
"NOT_EFFECTIVE",1,0)]
imputed_data2_long[,stop_other := fifelse(stop_reasons == "OTHER" |
stop_reasons == "REMISSION",1,0)]
#[,continue := fifelse(stop_reasons == "CONTINUE",1,0)]
imputed_data2_long[,continue := fifelse(stop_reasons == "OTHER" |
stop_reasons == "REMISSION" | stop_reasons == "CONTINUE",1,0)]

M <- imputed_data2$m

mice_fit <- lapply(1:M,function(m){

  # subset
  data_sub <- imputed_data2_long[.imp == m]

  mst_hosp <- msprep(time =
c("time_on_drug","time_on_drug","time_on_drug"),
                   status = c("continue","stop_ae","stop_lae"),
                   data = as.data.frame(data_sub),
                   trans = tmat,
                   keep = covs)
  # get covariates
tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)

  # fit
coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
                                paste0(tmp_cov,collapse = " + "),
                                "+ strata(trans)")),
             data = tmp,
             method = "breslow")

}) %>%
  as.mira()

est <- pool(mice_fit)
summary(est, conf.int = T, exponentiate = T)

# Transition 1 = Adverse Event
# Transition 2 = Lack of Efficacy

# Conclusion
# => No difference between OMA and BARI for Adverse Event (ae) and for
Lack of Event (lae)
```

Cleaner table with Hazard ratios with confidence intervals & p-values

```
ploufrows <- as.character(summary(est)$term)
ploufcols  <- c("HR","95%CI","p")
coxtable_csh2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable_csh2) <- ploufrows
colnames(coxtable_csh2) <- ploufcols
plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()

for(row in ploufrows)
{
  coxtable_csh2[row,"HR"] <-  formattable(plouf[term %in% row,
estimate])
  coxtable_csh2[row,"95%CI"] <-  paste0(formattable(plouf[term %in%
row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
  coxtable_csh2[row,"p"] <-  writepvalue(plouf[term %in% row,
p.value])}

row.names(coxtable_csh2)[1:2] <- c("OMA vs BARI Adverse event", "OMA
vs BARI Lack of Eff" )
coxtable_csh2
```

Subdistribution hazard model (Fine-Gray)

```
# Status variable
imputed_data2_long[stop_reasons == "ADVERSE_EVENT",status := 1]
imputed_data2_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
imputed_data2_long[stop_reasons == "OTHER" | stop_reasons ==
"REMISSION", status := 3]
imputed_data2_long[stop_reasons == "CONTINUE", status := 0]

imputed_data2_long$line_of_therapy <-
as.factor(imputed_data2_long$line_of_therapy)
imputed_data2_long$seropositivity_base <-
as.factor(imputed_data2_long$seropositivity_base)
levels(imputed_data2_long$cohort) <- c("0","1")
levels(imputed_data2_long$line_of_therapy) <- c("0","1","2","3")
levels(imputed_data2_long$gender) <- c("0","1")
levels(imputed_data2_long$smoker_base) <- c("2","1","0")
levels(imputed_data2_long$smoker_base)
levels(imputed_data2_long$seropositivity_base) <- c("0","1")

M <- imputed_data2$m

# First loop to get estimates for event = 1: ADVERSE EVENT

mice_fit <- lapply(1:M,function(m){
    # subset
  BARI_toto <- imputed_data2_long[.imp == m]
```

```
  #subdistribution hazard model
  shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
BARI_toto[,..covs],failcode = 1,cencode = 0)

}) %>%
  as.mira()
est <- pool(mice_fit)
summary(est, conf.int = T, exponentiate = T)

# Second loop to get estimates for event = 2: LACK OF EFFICACY
mice_fit2 <- lapply(1:M,function(m){

  # subset
  BARI_toto <- imputed_data2_long[.imp == m]

  # subdistribution hazard model
  shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
BARI_toto[,..covs],failcode = 2,cencode = 0)

}) %>%
  as.mira()
est2 <- pool(mice_fit2)
summary(est2, conf.int = T, exponentiate = T)

# Conclusions:
# No significant difference in incidence of "adverse event" and "lack
of efficacy" between TNFi and BARI

# CAREFUL: using a Fine-Gray model allows us to make claim about the
association between a covariate and the direction of the increase in
incidence, but we can't quantify the magnitude of the increase in
incidence.
```

Cleaner Table with Hazard ratios with confidence intervals & p-values

```
# Adverse event
ploufrows <- as.character(summary(est)$term)
ploufcols  <- c("HR","95%CI","p")
coxtable_ae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable_ae2) <- ploufrows
colnames(coxtable_ae2) <- ploufcols
plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()

for(row in ploufrows)
{
  coxtable_ae2[row,"HR"] <-  formattable(plouf[term %in% row,
estimate])
  coxtable_ae2[row,"95%CI"] <-  paste0(formattable(plouf[term %in%
row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
```

```
  coxtable_ae2[row,"p"] <-  writepvalue(plouf[term %in% row,
p.value])}

row.names(coxtable_ae2)[1] <- c("OMA vs BARI Advsere Events")

# Lack of efficacy
ploufrows <- as.character(summary(est2)$term)
ploufcols  <- c("HR","95%CI","p")
coxtable_lae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable_lae2) <- ploufrows
colnames(coxtable_lae2) <- ploufcols
plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()

for(row in ploufrows)
{
  coxtable_lae2[row,"HR"] <-  formattable(plouf[term %in% row,
estimate])
  coxtable_lae2[row,"95%CI"] <-  paste0(formattable(plouf[term %in%
row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
  coxtable_lae2[row,"p"] <-  writepvalue(plouf[term %in% row,
p.value])}

row.names(coxtable_lae2)[1] <- c("OMA vs BARI Lack of Eff")

#Output
coxtable_ae2
coxtable_lae2

write.xlsx(coxtable_ae2, file="./3_clean_output/BARI vs OMA HR
competing risk Fine-Gray AE.xlsx") # saving excel file
write.xlsx(coxtable_lae2, file="./3_clean_output/BARI vs OMA HR
competing risk Fine-Gray LAE.xlsx") # saving excel file
```

1. Saving

```
save.image(file="./3_clean_output/full_workspaces/workspace_1.RData")
```

2 - LDA and REM ANALYSIS

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

load("./1_datamanaged_files/datamanaged.Rdata")
```

This code aims at providing estimates for the remssion rates of the different treatments groups REM = REMmission LDA = Low Disease Activity

Both outcome are base on the CDAI CDAI = Clinical Disease Activity Index

CDAI is an index computed by the physician, which scores the severity of the disease.

1. [0] Exploration

See all available raw CDAI measures :

```
BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

nrow(BARI_DATA)
summary(BARI_DATA[, .(CDAI0_raw, CDAI12_raw)])
```

1. [1] CARRAC (confirm covariates for confounding and for attrition)

For LDA with updated function

```
library(modules)
source_comp_eff <- modules::use("ETAPE_2_supp_code.R")

LDA_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD","PREDNISON_STEROID",
                  "line_of_therapy","CDAI0"),
  attrition = c("TC_with_csDMARD","PREDNISON_STEROID",
                "line_of_therapy","CDAI0", "stop_reasons" ),
  seed = 123)

LDA_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD","PREDNISON_STEROID",
                  "line_of_therapy","CDAI0"),
  attrition = c("TC_with_csDMARD","PREDNISON_STEROID",
                "line_of_therapy","CDAI0", "stop_reasons"),
  seed = 123)

LDA_BARI_TNF
LDA_BARI_OMA
```

For REM with updated function

```
REM_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 2.8,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD","PREDNISON_STEROID",
                  "line_of_therapy","CDAI0"),
  attrition = c("TC_with_csDMARD","PREDNISON_STEROID",
                "line_of_therapy","CDAI0", "stop_reasons" ),
   seed = 123)

REM_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
```

```
  var = "CDAI12",
  thres = 2.8,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD","PREDNISON_STEROID",
                  "line_of_therapy","CDAI0"),
  attrition = c("TC_with_csDMARD","PREDNISON_STEROID",
                "line_of_therapy","CDAI0", "stop_reasons"),
   seed = 123)

REM_BARI_TNF
REM_BARI_OMA
```

This methods was developed by Mongin et al,
https://ard.bmj.com/content/early/2022/01/12/annrheumdis-2021-221477

Pooled table

```
table <- rbind(LDA_BARI_TNF, LDA_BARI_OMA, REM_BARI_TNF, REM_BARI_OMA)

write.xlsx(table, file = "./3_clean_output/table_LDA_REM_CARRAC.xlsx",
row.names = F)
```

1. Saving

```
save.image(file="./3_clean_output/full_workspaces/workspace_2.RData")
```

2 - LDA and REM supp CODE

10/11/2020

```
{r setup, include=FALSE} import("data.table") import("plyr")
import("data.table") import("mice") import("ipw") import("survey")
import("geepack") import("futile.logger") import("emmeans")
import("stats") import("survival")
```

function to perform checks on data

```{r setup, include=FALSE} check_data = function(datain, var = “CDAI_fu”, ttt_var = “ttt”, ref_ttt = “ttt_ref”, ID_ttt = NULL, othervar = c())

{

data <- setDT(copy(datain))

vartocheck <- Reduce(union,list(var,ttt_var,othervar)) notindata <- setdiff(vartocheck,names(data))

if(length(notindata)>0){ stop(paste0(“the variables”,paste0(notindata,collapse = “,”),” are not in the dataplease correct”)) }

# force ttt as var name setnames(data,ttt_var,“ttt”)

if( data[,uniqueN(ttt)]>2){ stop(“there are more than two treatments. The analysis has been implemented only for 2 treatments”) }

if(!any(data$ttt == ref_ttt)){ stop(paste0(“The variable”,ttt_var,” does not contain any “,ttt_ref,” value”)) }

data[,ttt := relevel(as.factor(ttt),ref_ttt)] if(is.null(ID_ttt))
{ data[,ID_ttt := .I] }else{ setnames(data,ID_ttt,“ID_ttt”) data[,N := .N,by = ID_ttt] if(any(data$N>1)){ stop(“there are”,data[N>1,uniqueN(ID_ttt)],” treatment course which have more than one entry in the table. Each row should be an unique treatment”) } }
return(data) }

adjusted_model = function(data, weights = NULL, covariates = NULL){

# transform char to factor to_fact <- data[,lapply(.SD,class)] %>% transpose(keep.names = “var”) %>% .[V1 == “character”,var]

data[,c(to_fact) := lapply(.SD,factor),.SDcols = to_fact]

#droplevels facto_vars <- data[,lapply(.SD,class)] %>% transpose(keep.names = “var”) %> % .[V1 == “factor”,var] data[,c(facto_vars) := lapply(.SD,droplevels),.SDcols = facto_vars]
```

define forumla formula <- as.formula(paste0("LDA ~",paste0(c("ttt",covariates),collapse = " + ")))

if(!is.null(covariates)){ # fit fit <- geeglm(formula, data = data, id = ID_ttt, family = gaussian) }else{ fit <- geeglm(LDA ~ ttt , data = data, weights = weights, id = ID_ttt, family = gaussian) }

fitsummary <- summary(fit) # create table with difference between the two treatments diff <- data.table(ttt = "diff", LDA = fitsummary$coefficients[2,"Estimate"], LDA_var = fitsummary$coefficients[2,"Std.err"]^2, LDA_sup = fitsummary$coefficients[2,"Estimate"] + 1.96*fitsummary$coefficients[2,"Std.err"], LDA_inf = fitsummary$coefficients[2,"Estimate"] - 1.96*fitsummary$coefficients[2,"Std.err"], methods = "CC_adjusted")

marginal effects: margi_df <- emmeans(fit, "ttt") %>% as.data.table()

margi_df[,methods := "CC_adjusted"] setnames(margi_df,"emmean","LDA") margi_df[,LDA_inf := LDA - 1.96*SE] margi_df[,LDA_sup := LDA + 1.96*SE] margi_df[,LDA_var := SE^2]

output <- rbind(diff,margi_df[,.(ttt,LDA,LDA_sup,LDA_inf,LDA_var,methods)])

return(list(output = output,fit = fit)) }

```
# Not adjusted complete case imputation
```

```{r setup, include=FALSE}

export("CC_raw")
CC_raw <- function(datain,
                   # data
                   var = "CDAI_fu",
                   # variable measuring effctiveness
                   thres = 10,
                   # threshold for remission or LDA
                   ttt_var = "ttt",
                   ref_ttt = "ttt_ref")
  # variable name containing the treatment
  {
  data <- check_data(datain,var,ttt,ref_ttt)
  # raw proportion
  raw_prop <- data[!is.na(get(var)),
                   .(LDA = sum(get(var)<=thres)/.N,
                     methods = "CC_raw",
                     N = .N),
                   by = ttt]
```

```
  # calculation of the Standard error
  raw_prop[,c("LDA_inf","LDA_sup") := lapply(c(-1.96,1.96),function(z)
{
    LDA + z*sqrt(LDA*(1-LDA)/N)
  })]

  # difference between treatments
  diff_tmp <- raw_prop[,.(ttt = "diff",
                          LDA = LDA[ttt == "ttt_1"]-LDA[ttt ==
"ttt_ref" ],
                    methods = methods[1] ,
                    SE = (sum(1/N))/2 +  1.96*sqrt(sum( LDA*(1-LDA)/N
))))]

  diff_tmp[,LDA_inf := LDA - SE]
  diff_tmp[,LDA_sup := LDA + SE]

  # bind outputs
  output <- rbind(diff_tmp[,.(ttt,LDA,LDA_inf,LDA_sup,methods)],
                  raw_prop[,-"N"])
  # change name back
  setnames(output,"ttt",ttt_var)
  return(output)
}
```

Adjusted complete case imputation

```{r setup, include=FALSE} export("CC_adjusted") CC_adjusted = function(datain, var = "CDAI_fu", thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", covariates = c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0") ) # variable name containing the treatment { data <- check_data(datain,var,ttt_var,ref_ttt) data[,LDA := get(var) <= thres] output <- adjusted_model(data = data[!is.na(get(var))], covariates = covariates)$output

output[,methods := "CC_adjusted"] # change name back setnames(output,"ttt",ttt_var) return(output) }

```
LOCF imputation


```{r setup, include=FALSE}
export("LOCF")

LOCF <- function(datain,
                 var = "CDAI_fu",
                 var_before = "CDAI_beforefu",
```

```
                  thres = 10,
                  ttt_var = "ttt",
                  ref_ttt = "ttt_ref",
                  covariates =
c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0")
){
  data <- copy(datain)

  data <- check_data(datain,var,ttt_var,ref_ttt)
  data[is.na(get(var)),c(var) := get(var_before)]
  data[,LDA := get(var) <= thres]

  output <- adjusted_model(data = data,
                           covariates = covariates)$output

  output[,methods := "LOCF"]
  # change name back
  setnames(output,"ttt",ttt_var)
  return(output)
}
```

Lundex imputation

```{r setup, include=FALSE} export("Lundex") Lundex <- function(datain, var = "CDAI_fu", thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", treatment_duration = "treatment_duration", stop_var = "stopany", covariates = c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0"), boot_num = 1000) {

data <- check_data(datain,var,ttt_var,ref_ttt) data[,LDA := get(var) <= thres] #### bootstrap for SE data[,tmp := 1] # replicated data for bootstrap replicateddata <- data[CJ(tmp = 1,boot = 1:boot_num),on = "tmp",allow.cartesian=TRUE] # sample with replacement for each boot sampled_idx <- replicateddata[,.I[sample(1:.N,replace = T)],by = boot]$V1 bootstrapdata <- replicateddata[sampled_idx]

# raw proportions raw_prop <- bootstrapdata[!is.na(get(var)), { adjusted_model(data = .SD)$output %>% .[ttt != "diff",.(ttt,LDA_raw = LDA)] }, by = .(boot)]

# surv analysis for each bootstraped dataset surv_formula <- as.formula(paste0("Surv(",treatment_duration,",",stop_var,") ~ ttt"))

surv_coeff <- bootstrapdata[, { temp.km <- survfit(surv_formula, data = .SD) list(surv = summary(temp.km, times = 1)$surv,  ttt = gsub("ttt=","",unique(summary(temp.km) $strata))) }, by = boot]

# LDA: LDA raw * surv coeff tmp_bootstrap <- merge(raw_prop,surv_coeff,by = c("boot","ttt")) tmp_bootstrap[,LDA := LDA_raw*surv]
```

difference between treatments diff_boot <- tmp_bootstrap[,.(ttt = “diff”, LDA = LDA[ttt != ref_ttt] - LDA[ttt == ref_ttt]), by = boot]

tot_bottstrap <- rbind(diff_boot[, .(ttt, LDA, boot)], tmp_bootstrap[, .(ttt, LDA, boot)])

calculate the mean and the SE: output <- tot_bottstrap[, .(LDA = mean(LDA), LDA_sup = quantile(LDA, 0.975), LDA_inf = quantile(LDA, 0.025)), by = ttt] # change name back output[,methods := “LUNDEX”]

setnames(output,“ttt”,ttt_var) return(output) }

```
# non-responder imputation
```

```{r setup, include=FALSE}
export("NRI")
NRI = function(datain,
               var="CDAI_fu",
               thres = 10,
               ttt_var = "ttt",
               ref_ttt = "ttt_ref",
               covariates =
c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0")
)
  # variable name containing the treatment
{
  data <- check_data(datain,var,ttt_var,ref_ttt)
  data[,LDA := get(var) <= thres]
  data[is.na(LDA),LDA := 0] # missing are non responders
  output <- adjusted_model(data = data,
                           covariates = covariates)$output

  # change name back
  setnames(output,"ttt",ttt_var)
  output[,methods := "NRI"]
  return(output)
}
```

Inverse probability weighting imputation

```{r setup, include=FALSE} export(“IPW”) IPW <- function(datain, var = “CDAI_fu”, thres = 10, ttt_var = “ttt”, ref_ttt = “ttt_ref”, counfunders = c(“Disease_duration”,“concomitantCsDMARD”,“Prev_bDMARD3”,“CDAI0”), attrition = c(“Disease_duration”,“concomitantCsDMARD”, “Prev_bDMARD3”,“CDAI0”,“stopreason”)) {

data <- check_data(datain, var, ttt_var, ref_ttt, othervar = c(counfunders,attrition))
```

data[,ttt2 := as.numeric(ttt != ref_ttt)] # weight for confounding formula_coeff <- paste0("~",paste0(counfunders,collapse = "+")) function_call <- paste0('IPWT <- ipwpoint(exposure = ttt2 , family = "binomial", link = "logit", numerator = ~ 1, denominator =',formula_coeff,', data = data, trunc = 0.01)') eval(parse(text = function_call)) data$sw <- IPWT$ipw.weights

weights for attrition formula_attr <- paste0("~",paste0(attrition,collapse = "+")) data[,MISS := as.numeric(is.na(get(var)))] function_call <- paste0('IPCT <- ipwpoint(exposure = MISS , family = "binomial", link = "logit", numerator = ~ 1, denominator =',formula_attr,', data = data)') eval(parse(text = function_call)) data $swc <- IPCT$ipw.weights

dataNoNA <- na.omit(data[,.(ttt,get(var),sw,swc,ID_ttt) %>% setNames(c("ttt",var,"sw","swc","ID_ttt"))]) dataNoNA[,LDA := as.numeric(get(var) <= thres)]

output <- adjusted_model(data = dataNoNA, weights = dataNoNA$sw*dataNoNA$swc) $output

output[,methods := "IPW"]

change name back setnames(output,"ttt",ttt_var) return(output)

}

```
# Confounder-Adjusted Response Rate with Attrition Correction (CARRAC)
imputation


```{r setup, include=FALSE}
export("CARRAC")
CARRAC <- function(datain,
                   var = "CDAI_fu",
                   thres = 10,
                   ttt_var = "ttt",
                   ref_ttt = "ttt_ref",
                   counfunders =
c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0"),
                   attrition =
c("Disease_duration","concomitantCsDMARD",

"Prev_bDMARD3","CDAI0","stopreason"),
                   seed = NA) {

  data <- check_data(datain,var,ttt_var,ref_ttt)
  dataS <- data[,.SD,.SDcols =
c("ID_ttt",var,"ttt",union(counfunders,attrition))]
```
```

```
  impute_data <- mice(
    dataS,
    m = 10,
    method = "pmm",
    maxit = 5,
    printFlag = F,seed = seed
  )
  # open the data
  impute_data_complete <- setDT(complete(impute_data,action = "long"))
  # calculate LDA
  impute_data_complete[,LDA := get(var) <= thres]

  # get LDA and error for each imputation
  res_mice <- lapply(seq(1:impute_data$m),function(imp){

    adjusted_model(data = impute_data_complete[.imp == imp],
                   covariates = counfunders)$output

  }) %>% rbindlist()

  res_mice_2 <- lapply(seq(1:impute_data$m),function(imp){

    adjusted_model(data = impute_data_complete[.imp == imp],
                   covariates = counfunders)$fit

  })

  test <- pool(res_mice_2)
  df_pval <- summary(test)  %>% as.data.table()
  p.output <- df_pval[grepl("ttt",term),p.value]

# pooling
  pool_res <- res_mice[,.(
    LDA_mi = mean(LDA),
    w = mean(LDA_var),
    m = .N,
    b = 1/(.N-1)*sum( (LDA-mean(LDA))^2 )
              ),by = ttt]

  pool_res[,LDA_var := w + (1+1/m)*b]
  pool_res[,LDA_sd := sqrt(LDA_var)]

  # mean, 95% CI
  output <- pool_res[,.(ttt,
                        LDA_mi,
                        LDA_mi + 1.96*LDA_sd,
                        LDA_mi-1.96*LDA_sd) %>%
                       setNames(c("ttt","LDA","LDA_sup","LDA_inf"))]

  output[,methods := "CARRAC"]
```

```
  output[ttt == "diff",p := p.output]

  # change name back
  setnames(output,"ttt",ttt_var)

  return(output)
}
```

3 - FINAL FIGURES CODE

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(patchwork) # package to compose multiplots !
library(ggpubr)
library(grid)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

load("./3_clean_output/full_workspaces/workspace_1.RData")
load("./3_clean_output/full_workspaces/workspace_2.RData")
load("./3_clean_output/full_workspaces/workspace_3.RData")
```

1. Common theme

```
theme_benoit = function(){
theme_pubclean()+
  theme(panel.grid.major.x = element_line(linetype = "dotted", colour
= "grey50"),
        panel.grid.major.y = element_line(linetype = "dotted", colour
= "grey50"),
        axis.title.y = element_text(margin = margin(r = .2, unit =
"cm")),
        axis.title.x = element_text(margin = margin(t = .2, unit =
```

```
"cm")),
        plot.title = element_text(margin = margin(b = .5, unit =
"cm")))
}
```

1. [0] Mini explation

TC lenght

```
BARI_DATA[,time_on_drug_year := time_on_drug/365.25]

p1 <- ggplot(BARI_DATA)+
  geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
1/12)+

  scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
  labs(x = "Duration of observation (years)",
       y = "Number of TC",
       title = "Time of observation for all included TC")+
  ylim(-25,NA)+
  theme_benoit()
p1

p2 <- ggplot(BARI_DATA)+
    geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
"grey80")+
    theme_void()

plot_mini_exploration <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)
plot_mini_exploration
```

Saving plot

```
png("./3_clean_output/figures/PLOT_Exploration_TC_duration.png",
    width = 7,
    height = 5,
    units = "in",
    res = 300) # opening graphic device
plot_mini_exploration
dev.off() # closing graphic device
```

TC lenght for BARI only

```
data_sub <- BARI_DATA[cohort == "BARI"]

p1 <- ggplot(data_sub)+
  geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
1/13, fill = "red3")+

  scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
```

```
  labs(x = "Duration of observation (years)",
       y = "Number of TC",
       title = "A - BARI")+
  ylim(-11,50)+
  theme_benoit()
p1

p2 <- ggplot(data_sub)+
    geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
"grey80")+
    theme_void()

plot_mini_exploration_bari <- p1 +
inset_element(p2,0.01,0.05,0.99,0.2)
plot_mini_exploration_bari
```

TC lenght for TNFi only

```
data_sub <- BARI_DATA[cohort == "TNFi"]

p1 <- ggplot(data_sub)+
  geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
1/13, fill = "green2")+

  scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
  labs(x = "Duration of observation (years)",
       y = "Number of TC",
       title = "B - TNFi")+
  ylim(-11,50)+
  theme_benoit()
p1

p2 <- ggplot(data_sub)+
    geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
"grey80")+
    theme_void()

plot_mini_exploration_tnfi <- p1 +
inset_element(p2,0.01,0.05,0.99,0.2)
plot_mini_exploration_tnfi
```

TC lenght for OMA only

```
data_sub <- BARI_DATA[cohort == "OMA"]

p1 <- ggplot(data_sub)+
  geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
1/13, fill = "blue2")+

  scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
```

```
  labs(x = "Duration of observation (years)",
       y = "Number of TC",
       title = "C - OMA")+
  ylim(-11,50)+
  theme_benoit()
p1

p2 <- ggplot(data_sub)+
    geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
"grey80")+
    theme_void()

plot_mini_exploration_oma <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)
plot_mini_exploration_oma
```

multiplot

```
multi_plot <- plot_mini_exploration_bari + plot_mini_exploration_tnfi
+ plot_mini_exploration_oma
multi_plot

median(BARI_DATA[cohort == "BARI", time_on_drug])
median(BARI_DATA[cohort == "TNFi", time_on_drug])
median(BARI_DATA[cohort == "OMA", time_on_drug])
```

Saving plot

```
png("./3_clean_output/figures/
PLOT_Exploration_TC_duration_3_groups.png",
    width = 9,
    height = 5,
    units = "in",
    res = 300) # opening graphic device
multi_plot
dev.off() # closing graphic device
```

1. [1] Survival analysis

Forest plot BARI vs TNFi + BARI vs OMA

```
meanall  <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
textall <- c("Treatment (vs BARI)", "Age (decades)", "BMI",
"Concomitant csDMARD", "Concomitant glucocorticoid", "CDAI score (10
pts)", "Disease duration (decades)", "Current smoker (vs non-smoker)",
"Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
"Seropositivity (RF or ACPA)")
```

```
dfall1 <- data.table(textall, meanall, lowerall, upperall)
dfall1[,ttt := "TNFi"]

meanall  <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
dfall2 <- data.table(textall, meanall, lowerall, upperall)
dfall2[, ttt := "OMA"]

dfall <- rbind(dfall1,dfall2)
dfall$textall <- factor(dfall$textall, levels = rev(textall))

text_high <- textGrob("\u2192 Reduces \ndrug maintenance",
gp=gpar(fontsize=8, fontface="bold"))
text_low <- textGrob("\u2190 Improves \ndrug maintenance",
gp=gpar(fontsize=8, fontface="bold"))

HR_plot <- ggplot(data=dfall,
                    aes(x = textall,
                        y = meanall,
                        ymin = lowerall,
                        ymax = upperall,
                        color = ttt))+
  geom_hline(yintercept =1, linetype=2)+
  geom_point(size=2,position = position_dodge(width = .7))+
  geom_errorbar(position = position_dodge(width = .7))+
  labs(x = "",y = "",color = "")+
  scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
+
  theme(axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", color = "black"),
        legend.position="top",
        legend.key = element_blank(),
        plot.margin = unit(c(1,3,2,1),"lines"))+
  coord_flip(clip = "off")+
  annotation_custom(text_high,
           xmin=-0.64,xmax=-0.64,ymin=.2,ymax=.2)+
  annotation_custom(text_low,
           xmin=-0.64,xmax=-0.64,ymin=-.15,ymax=-.15)+
  theme_pubclean()+
  scale_color_manual(breaks = c("OMA","TNFi"),values =
c("blue3","green3"),labels = c("OMA","TNFi"))

HR_plot
```

Saving the plot in PNG file

```
png("./3_clean_output/figures/PLOT FOREST BARI vs TNFi vs OMA HR.png",

    width = 7,
```

```
    height = 5.5,
    units = "in",
    res = 300)

HR_plot

dev.off() # closing graphic device
```

BARI vs TNFi

Non-adjusted Kaplan-Meier curves

BARI vs TNFi

```
BARI1[,time_on_drug_year := time_on_drug/365.25]

surv_object1 <- Surv(time = BARI1$time_on_drug_year, event =
BARI1$stop_DMARD) # indiquate time on drug and stop variable
fit1 <- survfit(surv_object1 ~ cohort, data = BARI1)

survplot_1 <- ggsurvplot(fit1, data = BARI1,  # plot
                         pval = T,
                         pval.method = TRUE,
                         legend.title = "Groups :",
                         legend.labs = c("BARI", "TNFi"),
                         xlab = "Time (years)",
                         xlim = c(0, 2.5),
                         censor = FALSE,
                         title = "Non-adjusted model of drug
discontinuation \nby type of treatment",
                         surv.median.line = "v",
                         linetype = 1,
                         size = 1.5,
                         #palette = c("grey78", "grey10"),
                         palette = c("red3", "green2"), # pour mettre
les couleurs
                         ggtheme = theme_benoit(),
                         risk.table = T)

values <- summary(fit1)$table[,"median"]
df <- data.frame(y = .1,x = values+.2,label =
as.character(round(values,2)))

survplot_1$plot <- survplot_1$plot +
  geom_text(data = df,aes(x,y,label = label), color = c("red3",
"green2"), size = 5)


print(survplot_1)
```

Saving surplot

```
png("./3_clean_output/figures/PLOT BARI vs TNFi curves non adjusted
COLOR.png",
     width = 7,
       height = 7, units = "in",
       res = 300) # opening graphic device
survplot_1

dev.off() # closing graphic device
```

Saving the plot curv object for Lilly

```
plot_BARI_vs_TNFi_data <- survplot_1$data.survplot
write.xlsx(plot_BARI_vs_TNFi_data, file =
"./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
ed.xlsx", row.names = F)
```

Home-made attempt to obtain adjusted curves based on imputed data

```
dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
T)
dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1$.imp != 0,]
dummy_cox_impute1$time_on_drug_year <-
dummy_cox_impute1$time_on_drug/365.25

BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug_year, event =
stop_DMARD) ~ cohort+
                           I(age_base/10)+
                           bmi_base+
                           TC_with_csDMARD+
                           PREDNISON_STEROID+
                           CDAI0+
                           I(disease_duration_base_years/10)+
                           C(smoker_base, base=3)+
                           line_of_therapy+
                           gender+
                           seropositivity_base+
                           cluster(patient_id)+
                           strata(cohort),dummy_cox_impute1), data =
dummy_cox_impute1)


survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
variable = "cohort",
                           xlab = "Time (years)",
                           title = "A - BARI vs TNFi",
                           legend.title = "Groups :",
                           legend.labs = c("BARI", "TNFi"),
                           censor = FALSE,
                           xlim = c(0, 2.5),
```

```
                              surv.median.line = "v",
                              linetype = 1,
                              size = 1.5,
                              ggtheme = theme_benoit(),
                              # palette = c("grey78", "grey10")
                              palette = c("red2", "green3") )+
  labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,\n
concomitant glucocorticoid, baseline CDAI, disease duration
(decades),\n smoking status, line of therapy, gender, seropositivity")

# adding days label
values <- summary(BARI_fit1)$table[,"median"]
df <- data.frame(y = .1,x = values+.1,label =
as.character(paste(round(values*365.25,2), "\n days")))
df[1,2] <- 1.82

survplot_1_adj$plot$labels$y <- "Proportion still on drug" # to change
the label

survplot_1_adj$plot <- survplot_1_adj$plot +
  geom_text(data = df,aes(x,y,label = label), color = c("red3",
"green3"), size = 5)

# adding HR et p val label
HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
round(exp(BARI1.adj.mi$coefficients[1]), 2), "\n", "p =",
round(summary(BARI1.adj.mi)$coefficients[1,"Pr(>|z|)"], 4)  ) )

survplot_1_adj$plot <- survplot_1_adj$plot +
  geom_text(data = HR,aes(x,y,label = label) , size = 5)

# final print
survplot_1_adj
```

Saving the survival plot in PNG file

```
png("./3_clean_output/figures/PLOT BARI vs TNFi curves adjusted.png",

    width = 7,
    height = 5,
    units = "in",
    res = 300)

survplot_1_adj

dev.off() # closing graphic device
```

Sensitivity analysis (RiskRegression Package)

First plot to get the difference in average treatment effect in percentage

```
dt.out$time_years <- dt.out$time/365.25

plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x =
time_years, group = level))+
  geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.4)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  scale_colour_manual(values = c("red2","green3"))+
  scale_fill_manual(values = c("red2","green3"))+
  theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,2.5,0.25)) +
scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+

  xlab("Years since intiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:", title = "A - BARI vs TNFi")
+
  labs(group = "Groups:")+
  theme_benoit()+
  theme(axis.title.x = element_text(margin = margin(t = .3,unit =
"cm")),
        axis.title.y = element_text(margin = margin(r = .3,unit =
"cm")))

plot.ate.diff
```

Saving Plot

```
png("./3_clean_output/figures/PLOT BARI vs TNFi curves AIPTW.png",
width = 1300, height = 650, res = 120) # opening graphic device

plot.ate.diff

dev.off() # closing graphic device
```

Second plot to get the ratio in average treatment effect

```
plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.4)+
  geom_line(aes(y = estimate, color = level), size = 2)+

  theme_benoit()+
  theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
  scale_x_continuous(breaks=seq(0,500,50))+
```

```
  scale_y_continuous(limits = c(0.9,4.5))+

  xlab("Days since intiation of treatment")+
  ylab("Ratio in Average Treatment Effect")+
  labs(colour="treatment", fill = "treatment")

plot.ate.ratio
```

BARI vs OMA

Non-adjusted Kaplan-Meier curves

BARI vs OMA

```
BARI2[,time_on_drug_year := time_on_drug/365.25]

surv_object2 <- Surv(time = BARI2$time_on_drug_year, event =
BARI2$stop_DMARD)
fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
creates the data for Kaplan Meyer
survplot_2 <- ggsurvplot(fit2, data = BARI2,  # plot
           pval = T,
           pval.method = TRUE,
           legend.title = "Groups :",
           legend.labs = c("BARI", "OMA"),
           xlab = "Time (days)",
           xlim = c(0, 2.5),
           censor = FALSE,
           title = "Non-adjusted model of drug discontinuation by type
of treatment",
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_benoit(),
           #palette = c("grey78", "grey50"),
           palette = c("red3", "blue2"), # to put colors
           risk.table = T)
survplot_2
```

Saving surplot

```
png("./3_clean_output/PLOT BARI vs OMA curves non adjusted COLOR.png",
width = 1000, height = 600, res = 100) # opening graphic device

survplot_2

dev.off() # closing graphic device
```

Saving the plot curv object for Lilly

```
plot_BARI_vs_OMA_data <- survplot_2$data.survplot
write.xlsx(plot_BARI_vs_OMA_data, file =
"./3_clean_output/Lilly_curves_excel/plot_BARI_vs_OMA_data_non_adjuste
d.xlsx", row.names = F)
```

Home-made attempt to obtain adjusted cuves based on imputed data :

```
dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
T)
dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
dummy_cox_impute2$time_on_drug_year <-
dummy_cox_impute2$time_on_drug/365.25

BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug_year, event =
stop_DMARD) ~ cohort+
                             I(age_base/10)+
                             bmi_base+
                             TC_with_csDMARD+
                             PREDNISON_STEROID+
                             CDAI0+
                             I(disease_duration_base_years/10)+
                             C(smoker_base, base=3)+
                             line_of_therapy+
                             gender+
                             seropositivity_base+
                             cluster(patient_id)+
                             strata(cohort),dummy_cox_impute2), data =
dummy_cox_impute2)

survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
variable = "cohort",
           xlab = "Time (years)",
           title = "B - BARI vs OMA",
           legend.title = "Groups :",
           legend.labs = c("BARI", "OMA"),
           censor = FALSE,
           xlim = c(0, 2.5),
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_benoit(),
           # palette = c("grey78", "grey50")
           palette = c("red2", "blue3") # to change colors
           )+
  labs(caption = "Adjusted for : age, BMI, concomitant csDMARD, \n
concomitant glucocorticoid, baseline CDAI, disease duration
(decades),\n smoking status, line of therapy, gender, seropositivity")

# adding Days label
values <- summary(BARI_fit2)$table[,"median"]
df <- data.frame(y = .1,x = values+.1,label =
```

```
as.character(paste(round(values*365.25,2), "\n days")))
df[1,2] <- 1.82

survplot_2_adj$plot$labels$y <- "Proportion still on drug" # to change
the label

survplot_2_adj$plot <- survplot_2_adj$plot +
  geom_text(data = df,aes(x,y,label = label), color = c("red3",
"blue2"), size = 5)

# adding HR et pval label
HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
round(exp(BARI2.adj.mi$coefficients[1]), 2),  "\n", "p =",
round(summary(BARI2.adj.mi)$coefficients[1,"Pr(>|z|)"], 4)  ) )

survplot_2_adj$plot <- survplot_2_adj$plot +
  geom_text(data = HR,aes(x,y,label = label) , size = 5)



# final print
survplot_2_adj

summary(BARI_fit2, times = 1) # to see detailed surv probabilities at
given timepoints
```

Saving the survival plot in PNG file

```
png("./3_clean_output/PLOT BARI vs OMA curves adjusted.png", width =
1000, height = 600, res = 100) # opening graphic device

survplot_2_adj

dev.off() # closing graphic device
```

Sensitivity analysis (RiskRegression package)

First plot to get the difference in average treatment effect in percentage

```
dt.out2$time_years <- dt.out2$time/365.25

plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x =
time_years, group = level))+
  geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.3)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  theme_benoit() + theme(legend.spacing.x = unit(0.2, 'cm'),
```

```
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,2.5,0.25)) +
scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+

  xlab("Years since intiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:", title = "B - BARI vs OMA")+
  labs(group = "Groups:")+

  scale_colour_manual(values = c("red2","blue3"))+
  scale_fill_manual(values = c("red2","blue3"))

plot.ate.diff2
```

Saving Plot

```
png("./3_clean_output/PLOT BARI vs OMA curves AIPTW.png", width =
1300, height = 650, res = 120) # opening graphic device

plot.ate.diff2

dev.off() # closing graphic device
```

Second plot to get the ratio in average treatment effect

```
plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.3)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  theme_benoit()+
  theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
  scale_x_continuous(breaks=seq(100,400,50))+
  scale_y_continuous(limits = c(0.8,3))+

  xlab("Days since intiation of treatment")+
  ylab("Ratio in Average Treatment Effect")+
  labs(colour="treatment", fill = "treatment")

plot.ate.ratio2
```

Multipanel plots

To update using patchwork

For the paper Non adjuted curves

```
# Creating list object

plots <- list()
```

```
plots[[1]] <- survplot_1
plots[[2]] <- survplot_2

# Nice function

multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
```

Option 2 putting all data on one panel Kaplan Meier

BARI vs TNFi vs OMA

```
BARI_DATA[,time_on_drug_year := time_on_drug/365.25]

surv_object3 <- Surv(time = BARI_DATA$time_on_drug_year, event =
BARI_DATA$stop_DMARD)
fit3 <- survfit(surv_object3 ~ cohort, data = BARI_DATA) # this
function creates the data for Kaplan Meyer
survplot_3 <- ggsurvplot(fit3, data = BARI_DATA,  # plot
           pval = F,
           pval.method = F,
           legend.title = "Groups",
           legend.labs = c("BARI", "TNFi", "OMA"),
           xlab = "Time (years)",
           xlim = c(0, 2.5),
           censor = FALSE,
          # title = "Non-adjusted drug discontinuation by type of
treatment (Kaplan-Meier)",
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_benoit(),
           palette = c("red3", "green2", "blue2"), # to put colors
           risk.table = T)

values <- summary(fit3)$table[,"median"]
df <- data.frame(y = .1,x = values+.1,label =
as.character(paste(round(values*365.25,2), "\n days")))
df[3,2] <- 1.72

survplot_3$plot <- survplot_3$plot +
  geom_text(data = df,aes(x,y,label = label), color = c("red3",
"green2", "blue2"), size = 5)

survplot_3$plot$labels$y <- "Proportion still on drug" # to change the
label
survplot_3
```

Saving surplot

```
png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA curves non
adjusted COLOR.png", width = 800, height = 600, res = 100) # opening
```

```
graphic device

survplot_3

dev.off() # closing graphic device
```

Adjusted curves

```
# Creating list object

plots <- list()
plots[[1]] <- survplot_1_adj
plots[[2]] <- survplot_2_adj

# Nice function

multi_plot_cox <- arrange_ggsurvplots(plots, print = T, ncol = 2)

png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA curves
adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
graphic device

multi_plot_cox

dev.off() # closing graphic device
```

All curves

```
# Creating list object

plots <- list()
plots[[1]] <- survplot_1
plots[[3]] <- survplot_2
plots[[2]] <- survplot_1_adj
plots[[4]] <- survplot_2_adj

# Nice function

multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2, nrow =
2)

# but does not display properly now.. :(
```

AIPTW absolute risk of treatment discontinuation biplot

```
plot.ate.diff + plot.ate.diff2

png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA AIPTW curves
adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
graphic device
```

```
plot.ate.diff + plot.ate.diff2

dev.off() # closing graphic device
```

Diagnostic multipanel plots

Asked by Lilly statistician to show balance in this analysis.

```
pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
= cohort)) +
  geom_density(alpha = .47) +

    theme_minimal()+
    theme(axis.ticks.y = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          text = element_text(size = 16),
          axis.title.x = element_text(hjust = 0.2, size = 16))+

    scale_colour_manual(values = c("red2","green3"))+
    scale_fill_manual(values = c("red2","green3"))+

    xlab("Probability of being assigned BARI or TNFi") +
    ylab("Density") +
    labs(title = "A1 - BARI vs TNFi")

pscore_plot # overlap
```

```
pscore_plot2 <- ggplot(test.data2, aes(x = pscores, color = cohort,
fill = cohort)) +
  geom_density(alpha = .47) +

    theme_minimal()+
    theme(axis.ticks.y = element_blank(),
          panel.grid.minor = element_blank(),
          legend.title = element_blank(),
          text = element_text(size = 16),
          axis.title.x = element_text(hjust = 0.2, size = 16))+

    scale_colour_manual(values = c("red2","blue3"))+
    scale_fill_manual(values = c("red2","blue3"))+

    xlab("Probability of being assigned BARI or OMA") +
    ylab("Density") +
    labs(title = "A2 - BARI vs OMA")

pscore_plot2
# Good overlap
```

```
library(cobalt)

# BARI vs TNFi
B1 <- love.plot(COVS, treat = test.data$cohort, weights =
test.data$weights, stats =  c("mean.diffs"), thresholds = c(m = .1),
var.order = "adjusted", title = "B1 - BARI vs TNFi", color =
c("#FD8D3C", "#08306B"), themes = theme_pubclean() )

# BARI vs OMA
B2 <- love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, stats =  c("mean.diffs"), thresholds = c(m = .1),
var.order = "adjusted", title = "B2 - BARI vs OMA", color =
c("#FD8D3C", "#08306B") , themes = theme_pubclean() )

one <- (  pscore_plot + B1)

two <- ( pscore_plot2 + B2 )

png("./3_clean_output/figures/AIPTW diagnositc COLOR.png", width =
1300, height = 900, res = 100) # opening graphic device

one / two

dev.off() # closing graphic device
```

1. [3] Fist line analysis

Non-adjusted Survival curves

BARI vs TNFi

```
BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])

surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
BARI_first1$stop_DMARD) # indiquate stop variable and time_on_drug
summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # function
which creates Kaplan-meier data
survplot_first1 <- ggsurvplot(fit3, data = BARI_first1,   # plot
           pval = T,
           pval.method = TRUE,
           legend.title = "Groups :",
           legend.labs = c("BARI", "TFNi"),
           xlab = "Time (days)",
           xlim = c(0, 700),
           censor = FALSE,
           title = "A - BARI vs TNFi",
           surv.median.line = "v",
           linetype = 1,
```

```
            size = 1.5,
            ggtheme = theme_benoit(),
            # palette = c("grey78", "grey50"),
            palette = c("red2", "green3"), # to get colors
            risk.table = T
            )

survplot_first1$plot$labels$y <- "Proportion still on drug" # to
change the label
survplot_first1

rm(surv_object3, fit3)
```

saving plot curves

```
png("./3_clean_output/PLOT BARI vs TNFi first line curves non adjusted
COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
device

survplot_first1

dev.off() # closing graphic device
```

BARI vs OMA

```
BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
c("BARI", "OMA")] # selection des TC TNFi

surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
BARI_first2$stop_DMARD) # indiquate stop variable and time_on_drug
summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # function
which creates Kaplan-meier data
survplot_first2 <- ggsurvplot(fit3, data = BARI_first2,   # plot
            pval = T,
            pval.method = TRUE,
            legend.title = "Groups :",
            legend.labs = c("BARI", "OMA"),
            xlab = "Time (days)",
            xlim = c(0, 700),
            censor = FALSE,
            title = "B - BARI vs OMA",
            surv.median.line = "v",
            linetype = 1,
            size = 1.5,
            ggtheme = theme_benoit(),
            # palette = c("grey78", "grey50", "grey10"),
            palette = c("red2", "blue3"), # to get colors
            risk.table = T
            )
```

```
survplot_first2$plot$labels$y <- "Proportion still on drug" # to
change the label
survplot_first2

rm(surv_object3, fit3)
```

saving plot curves

```
png("./3_clean_output/PLOT BARI vs OMA first line curves non adjusted
COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
device

survplot_first2

dev.off() # closing graphic device
```

Adjusted curves with imputed data (BARI vs TNFi)

```
dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
include = T)
dummy_cox_impute_first1 <-
dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]

BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
                             I(age_base/10)+
                             bmi_base+
                             concomitant_csDMARD+
                             PREDNISON_STEROID+
                             CDAI0+
                             I(disease_duration_base_years/10)+
                             C(smoker_base, base=3)+
                             line_of_therapy+
                             gender+
                             seropositivity_base+
                             cluster(patient_id)+
                             strata(cohort),dummy_cox_impute_first1),
data = dummy_cox_impute_first1)


survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
dummy_cox_impute_first1, variable = "cohort",
           xlab = "Time (days)",
           title = "Multivariable Cox model of drug discontinuation by
type of treatment - 1st line vs 1st line",
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "TNFi"),
           censor = FALSE,
           xlim = c(0, 700),
           surv.median.line = "v",
```

```
            linetype = 1,
            size = 1.5,
            ggtheme = theme_minimal(),
            #palette = c("grey78", "grey10")
            palette = c("red2", "green3"), # to get colors
            )

survplot_first1_adj <- survplot_first1_adj +
    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_first1_adj
table(BARI_first1$cohort)
rm(dummy_cox_impute_first1, BARI_first1_fit)
```

Saving the survival plot in PNG file

```
png("./3_clean_output/PLOT BARI vs TNFi first line curves adjusted
COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
device

survplot_first1_adj

dev.off() # closing graphic device
```

Adjusted curves with imputed data (BARI vs OMA)

```
dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
include = T)
dummy_cox_impute_first2 <-
dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]

BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
                             I(age_base/10)+
                             bmi_base+
                             concomitant_csDMARD+
                             PREDNISON_STEROID+
                             CDAI0+
                             I(disease_duration_base_years/10)+
                             C(smoker_base, base=3)+
                             line_of_therapy+
                             gender+
                             seropositivity_base+
                             cluster(patient_id)+
                             strata(cohort),dummy_cox_impute_first2),
data = dummy_cox_impute_first2)


survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
```

```
dummy_cox_impute_first2, variable = "cohort",
           xlab = "Time (days)",
           title = "Multivariable Cox model of drug discontinuation by
type of treatment - 1st line vs 1st line",
           legend.title = "Groups :",
           legend.labs = c("Baricitinib", "OMA bDMARDs"),
           censor = FALSE,
           xlim = c(0, 700),
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_minimal(),
           #palette = c("grey78", "grey50")
           palette = c("red2", "blue3"), # to get colors
           )

survplot_first2_adj <- survplot_first2_adj +
    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_first2_adj
table(BARI_first2$cohort)
rm(dummy_cox_impute_first2, BARI_first2_fit)
```

Saving the survival plot in PNG file

```
png("./3_clean_output/PLOT BARI vs OMA first line curves adjusted
COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
device

survplot_first2_adj

dev.off() # closing graphic device
```

Multipanel plots

Non adjusted curves

```
plots <- list()
plots[[1]] <- survplot_first1
plots[[2]] <- survplot_first2

# Nice function

multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)

png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA 1st Line
curves non-adjusted COLOR.png", width = 1000, height = 600, res = 100)
# opening graphic device
```

```
multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)

dev.off() # closing graphic device
```

All in one BARI vs TNFi vs OMA

```
BARI_first[,time_on_drug_year := time_on_drug/365.25]

surv_object4 <- Surv(time = BARI_first$time_on_drug_year, event =
BARI_first$stop_DMARD)
fit4 <- survfit(surv_object4 ~ cohort, data = BARI_first) # this
function creates the data for Kaplan Meyer

survplot_4 <- ggsurvplot(fit4, data = BARI_first,  # plot
           pval = F,
           pval.method = F,
           legend.title = "Groups",
           legend.labs = c("BARI", "TNFi", "OMA"),
           xlab = "Time (years)",
           xlim = c(0, 2.5),
           censor = FALSE,
          # title = "Non-adjusted drug discontinuation by type of
treatment (Kaplan-Meier)",
           surv.median.line = "v",
           linetype = 1,
           size = 1.5,
           ggtheme = theme_benoit(),
           palette = c("red3", "green2", "blue2"), # to put colors
           risk.table = T)

values <- summary(fit4)$table[,"median"]
df <- data.frame(y = .2,x = values+.2,label =
as.character(paste(round(values*365.25,2), "\n days")))
df <- df[2,]

survplot_4$plot <- survplot_4$plot +
  geom_text(data = df,aes(x,y,label = label), color = c("green2"),
size = 5)

survplot_4$plot$labels$y <- "Proportion still on drug" # to change the
label
survplot_4
```

Saving surplot

```
png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA first curves
non adjusted COLOR.png", width = 800, height = 600, res = 100) #
opening graphic device
```

```
survplot_4

dev.off() # closing graphic device
```

1. [4] LACK of EFFICACYY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```
ci_long$time_months <- ci_long$time/365.25*12

plot2 <-  ggplot(data = ci_long, aes(x = time_months,
                                     y = value,
                                     linetype = variable  ,
                                     col = cohort )) +
  geom_line(size = 0.75)+

  scale_color_manual(breaks = c(0,1),
                     values = c("red3","green2"),
                     labels = c("BARI","TNFi"))+
  scale_linetype_manual(breaks = c("CI.1","CI.2"),
                        values = c("solid","dashed", "dotted"),
                        labels = c("Adverse Event","Ineffectiveness"))
+ # not showing the "other category"
  scale_x_continuous(name = "Time (months)",
                     breaks = c(0,3,6,9,12),
                     limits = c(0,12)) +
  scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
+
  theme_benoit()+
  theme(legend.box = "horizontal",
        legend.position = c(0.05,1),
        legend.justification = c(0,1),
        legend.background = element_blank(),
        legend.key = element_blank(), #no legend key background
        legend.key.width = grid::unit(2, "lines"))+ # longer line in
legend, to see properly the dashed
  labs(color = "", linetype = "", title = "A - BARI vs TNFi")

plot2

ggsave(filename = "PLOT BARI vs TNFi cumulative incidence.png",plot =
plot2, path = "./3_clean_output/", device = "png", width = 829, height
= 550, units = "px", scale = 3.2)
```

(BARI vs OMA)

Cumulative incidence function

```
ci_long_2$time_months <- ci_long_2$time/365.25*12

plot3 <-  ggplot(data = ci_long_2, aes(x = time_months,
                                       y = value,
                                       linetype = variable  ,
                                       col = cohort )) +
  geom_line(size = 0.75)+

  scale_color_manual(breaks = c(0,1),
                     values = c("red3","blue2"),
                     labels = c("BARI","OMA"))+
  scale_linetype_manual(breaks = c("CI.1","CI.2"), # not showing the
"other" category
                        values = c("solid","dashed", "dotted"),
                        labels = c("Adverse Event","Ineffectiveness"))
+
  scale_x_continuous(name = "Time (months)",
                     breaks = c(0,3,6,9,12),
                     limits = c(0,12)) +
  scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
+
  theme_benoit()+
  theme(legend.box = "horizontal",
        legend.position = c(0.05,1),
        legend.justification = c(0,1),
        legend.background = element_blank(),
        legend.key = element_blank(), #no legend key background
        legend.key.width = grid::unit(2, "lines"))+ # longer line in
legend, to see properly the dashed
  labs(color = "", linetype = "", title = "B - BARI vs OMA")

plot3

ggsave(filename = "./3_clean_output/figures/PLOT BARI vs OMA
cumulative incidence.png", plot3, height = 4, width = 6, units =
"in",dpi = 300)
```

Multipanel

```
plot2_3 <- plot2 + plot3

ggsave(filename = "./3_clean_output/figures/PLOT BARI vs TNFi and BARI
vs OMA cumulative incidence.png", plot2_3, height = 4, width = 8,
units = "in",dpi = 300)
```

1. [6] LDA - REM

Exploration

See all available raw CDAI measures : :)

```
BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

plot_data <- copy(BARI_long[!is.na(TC_id) & TC_id %in%
BARI_DATA$TC_id])
plot_data <- merge(plot_data, BARI_DATA[,.(TC_id, cohort)], by =
"TC_id")

CDAI_plot <- ggplot(data = plot_data,
                    aes(x = time, y = CDAI, fill = cohort) )+

  annotate("rect",xmin = 0.875,xmax = 1.125,
                  ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+
  annotate("rect",xmin = -.05,xmax = 0+1.5/12,
                  ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+

  geom_point(data = plot_data[cohort != "BARI"], alpha = 0.2, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+
  geom_point(data = plot_data[cohort == "BARI"], alpha = 0.25, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+


  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort !=
"BARI"], alpha = 0.2, size = 2, shape = 21, show.legend = F )+
  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort ==
"BARI"], alpha = 0.25, size = 2, shape = 21 , show.legend = F, )+

  geom_smooth(alpha = 0.1, size = 1, aes(color = cohort), show.legend
= F)+

  coord_cartesian(xlim = c(0,2.5))+
    labs(title = "CDAI across time type of treamtent (all TC)",
         x = "Time (years since TC initiation)",
         y = "CDAI score",
         color = "",
         fill = "")+
  theme_benoit()+
  theme(legend.position = c(1,1),
        legend.justification = c(1,1))+
  guides(color = guide_legend(override.aes = list(linetype = NA,size =
3)))

CDAI_plot
```

To me this figure is the best results to be discussed regarding REM and LDA

saving plot

```
png("./3_clean_output/figures/PLOT CDAI across time raw.png",
    width = 8,
    height = 6,
    units = "in",
    res = 120) # opening graphic device

CDAI_plot

dev.off() # closing graphic device
```

CARRAC histogram

Building large format data table from the CARRAC output

```
# Extracting LDA BARI
LDA_BARI <- rbind(LDA_BARI_TNF[2,1:4], LDA_BARI_OMA[2,1:4]) # I have
one estimation per comparison
LDA_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=
mean(LDA_inf)] # averaging
LDA_BARI <- LDA_BARI[1]

# REM BARI
REM_BARI <- rbind(REM_BARI_TNF[2,1:4], REM_BARI_OMA[2,1:4]) # I have
one estimation per comparison
REM_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=
mean(LDA_inf)] # averaging
REM_BARI <- REM_BARI[1]

# IDem for TNFi and OMA
LDA_TNFi <- LDA_BARI_TNF[3,1:4]
REM_TNFi <- REM_BARI_TNF[3,1:4]

LDA_OMA <- LDA_BARI_OMA[3,1:4]
REM_OMA <- REM_BARI_OMA[3,1:4]

# Binding together
LDA <- rbind(LDA_BARI, LDA_TNFi, LDA_OMA)
setnames(LDA, c("ttt", "LDA", "LDA_sup", "LDA_inf")) #putting right
labels
REM <- rbind(REM_BARI, REM_TNFi, REM_OMA)
setnames(REM, c("ttt", "REM", "REM_sup", "REM_inf"))

histo_carrac <- cbind(LDA, REM[,-1])
```

plotting

```
carrac_plot <- ggplot(data = histo_carrac, aes(x = ttt, group = ttt))
+

  theme_pubclean()+

  geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
  geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
ymax=REM_sup*100), width=0.2, size=1, color="grey50")+

  geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
= 0.5)+
  geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +

  geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
"grey65", alpha = 0.5)+
  geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +

  theme(strip.text.y = element_blank(),
        strip.background = element_blank(),
        axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", colour = "black"),
        legend.position="bottom", plot.margin =
unit(c(1,3,2,1),"lines"))+

  scale_y_continuous(limits = c(0,82))+

  labs(y = "(% of TC)", x = "Treatment group", title = "A - REM and
LDA rates \nby type of treatment \n(CARRAC)")

carrac_plot
```

also Saving CARRAC plot only

```
png("./3_clean_output/figures/PLOT BARI 3 CARRAC ONLY.png", width =
350, height = 600, res = 100) # opening graphic device

ggplot(data = histo_carrac, aes(x = ttt, group = ttt)) +

  theme_pubclean()+

  geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
  geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
ymax=REM_sup*100), width=0.2, size=1, color="grey50")+

  geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
= 0.5)+
  geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
```

```
  geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
"grey65", alpha = 0.5)+
  geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +

  theme(strip.text.y = element_blank(),
        strip.background = element_blank(),
        axis.line.x = element_line(size = 0.5),
        axis.text = element_text(face = "bold", colour = "black"),
        legend.position="bottom", plot.margin =
unit(c(1,3,2,1),"lines"))+

  scale_y_continuous(limits = c(0,82))+

  labs(y = "(% of TC)", x = "Treatment group", title = "REM and LDA
rates \nby type of treatment \n(CARRAC)")

dev.off() # closing graphic device
```