Kaplan Meier Plots

2024-03-14

# Loading R packages

setwd("C:/Users/swj88/Documents/Github/maicplus")  
devtools::load\_all()  
#> ℹ Loading maicplus  
#library(maicplus)

Additional suggested packages for this vignette:

library(survminer) # this is used for ggplot version of KM plots

# Introduction

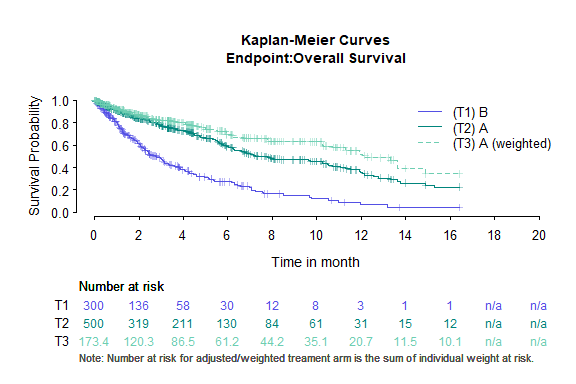
To plot Kaplan Meier curves, we need to first obtain pseudo comparator IPD by digitizing Kaplan-Meier curves from the comparator study. For more information on how to digitize Kaplan-Meier curves, refer to Guyot et al. and Liu et al. [1,2].

## Unanchored case

# Read in relevant ADaM data and rename variables of interest  
adsl <- read.csv(system.file("extdata", "adsl.csv",  
 package = "maicplus",  
 mustWork = TRUE  
))  
adtte <- read.csv(system.file("extdata", "adtte.csv",  
 package = "maicplus",  
 mustWork = TRUE  
))  
adtte$TIME <- adtte$AVAL  
adtte$EVENT <- adtte$EVNT  
  
### AgD  
# Baseline aggregate data for the comparator population  
target\_pop <- read.csv(system.file("extdata", "aggregate\_data\_example\_1.csv",  
 package = "maicplus", mustWork = TRUE  
))  
# for time-to-event endpoints, pseudo IPD from digitalized KM  
pseudo\_ipd <- read.csv(system.file("extdata", "psuedo\_IPD.csv",  
 package = "maicplus",  
 mustWork = TRUE  
))  
pseudo\_ipd$ARM <- "B"  
  
#### prepare data  
target\_pop <- process\_agd(target\_pop)  
adsl <- dummize\_ipd(adsl, dummize\_cols = c("SEX"), dummize\_ref\_level = c("Female"))  
use\_adsl <- center\_ipd(ipd = adsl, agd = target\_pop)  
  
#### derive weights  
match\_res <- estimate\_weights(  
 data = use\_adsl,  
 centered\_colnames = grep("\_CENTERED$", names(use\_adsl)),  
 start\_val = 0,  
 method = "BFGS"  
)  
#> initial value 500.000000   
#> iter 10 value 196.020742  
#> iter 20 value 173.488262  
#> final value 173.420820   
#> converged

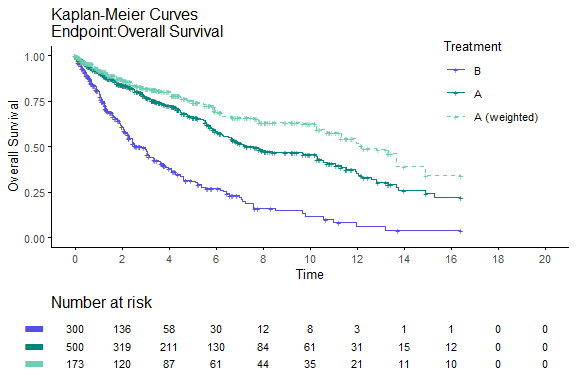
Here is a Kaplan-Meier plot using base R graphics. Note that number at risk for weighted treatment arm is the sum of individual weights at risk.

kmplot(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 trt\_var\_ipd = "ARM",  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = NULL,  
 km\_conf\_type = "log-log",  
 time\_scale = "month",  
 time\_grid = seq(0, 20, by = 2),  
 use\_colors = NULL,  
 use\_line\_types = NULL,  
 use\_pch\_cex = 0.65,  
 use\_pch\_alpha = 100  
)



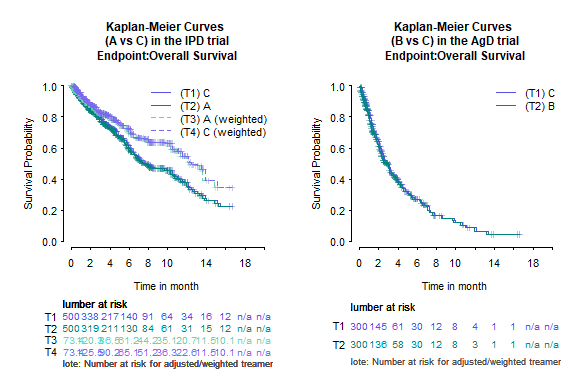
There is also a ggplot option for Kaplan-Meier curves using survminer R package.

kmplot2(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = NULL,  
 trt\_var\_ipd = "ARM",  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 km\_conf\_type = "log-log",  
 time\_scale = "month",  
 break\_x\_by = 2,  
 xlim = c(0, 20)  
)

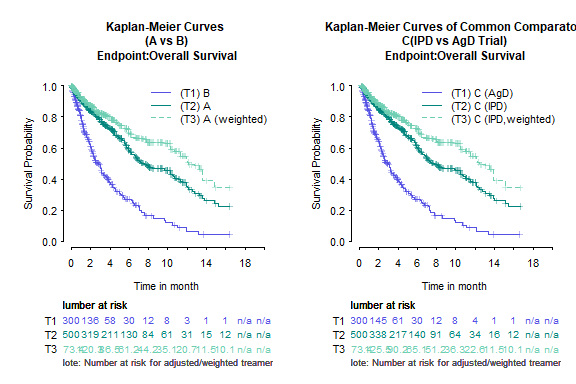


## Anchored case

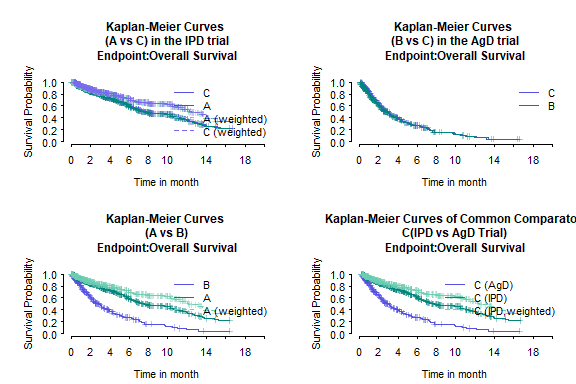
# Read in relevant ADaM data and rename variables of interest  
adsl <- read.csv(system.file("extdata", "adsl.csv",  
 package = "maicplus",  
 mustWork = TRUE  
))  
adtte <- read.csv(system.file("extdata", "adtte.csv",  
 package = "maicplus",  
 mustWork = TRUE  
))  
adtte$TIME <- adtte$AVAL  
adtte$EVENT <- adtte$EVNT  
adtte2 <- adtte  
adtte2$ARM <- "C"  
adtte2$TIME <- adtte2$TIME + 7  
adtte <- rbind(adtte, adtte2)  
  
### AgD  
# Baseline aggregate data for the comparator population  
target\_pop <- read.csv(system.file("extdata", "aggregate\_data\_example\_1.csv",  
 package = "maicplus", mustWork = TRUE  
))  
# for time-to-event endpoints, pseudo IPD from digitalized KM  
pseudo\_ipd <- read.csv(system.file("extdata", "psuedo\_IPD.csv",  
 package = "maicplus",  
 mustWork = TRUE  
))  
pseudo\_ipd$ARM <- "B"  
pseudo\_ipd2 <- pseudo\_ipd  
pseudo\_ipd2$ARM <- "C"  
pseudo\_ipd2$Time <- pseudo\_ipd2$Time + 5  
pseudo\_ipd <- rbind(pseudo\_ipd, pseudo\_ipd2)  
  
#### prepare data  
target\_pop <- process\_agd(target\_pop)  
adsl <- dummize\_ipd(adsl, dummize\_cols = c("SEX"), dummize\_ref\_level = c("Female"))  
use\_adsl <- center\_ipd(ipd = adsl, agd = target\_pop)  
  
#### derive weights  
match\_res <- estimate\_weights(  
 data = use\_adsl,  
 centered\_colnames = grep("\_CENTERED$", names(use\_adsl)),  
 start\_val = 0,  
 method = "BFGS"  
)  
#> initial value 500.000000   
#> iter 10 value 196.020742  
#> iter 20 value 173.488262  
#> final value 173.420820   
#> converged  
  
# plot by trial  
kmplot(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 trt\_var\_ipd = "ARM",  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = "C",  
 km\_conf\_type = "log-log",  
 km\_layout = "by\_trial",  
 time\_scale = "month",  
 time\_grid = seq(0, 20, by = 2),  
 use\_colors = NULL,  
 use\_line\_types = NULL,  
 use\_pch\_cex = 0.65,  
 use\_pch\_alpha = 100  
)



# plot by arm  
kmplot(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 trt\_var\_ipd = "ARM",  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = "C",  
 km\_conf\_type = "log-log",  
 km\_layout = "by\_arm",  
 time\_scale = "month",  
 time\_grid = seq(0, 20, by = 2),  
 use\_colors = NULL,  
 use\_line\_types = NULL,  
 use\_pch\_cex = 0.65,  
 use\_pch\_alpha = 100  
)

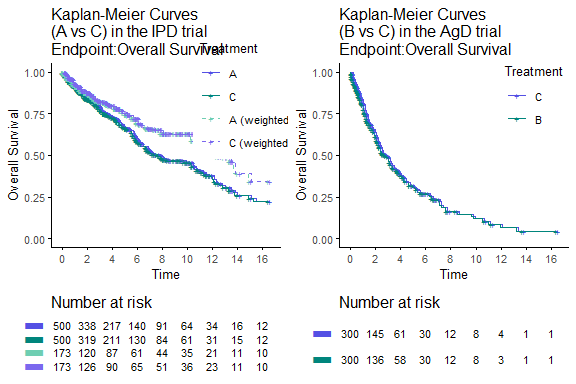


# plot all  
kmplot(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 trt\_var\_ipd = "ARM",  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = "C",  
 km\_conf\_type = "log-log",  
 km\_layout = "all",  
 time\_scale = "month",  
 time\_grid = seq(0, 20, by = 2),  
 use\_colors = NULL,  
 use\_line\_types = NULL,  
 use\_pch\_cex = 0.65,  
 use\_pch\_alpha = 100  
)

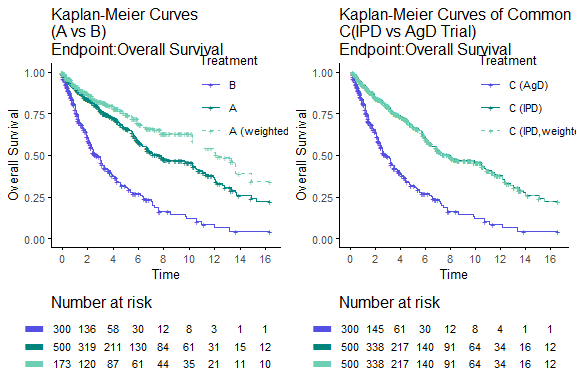


Again there is also a ggplot option

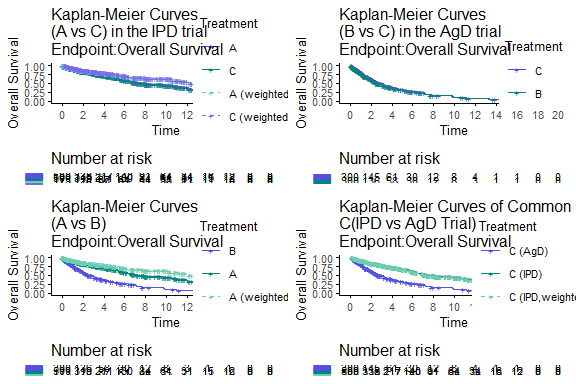
# plot by trial  
kmplot2(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = "C",  
 trt\_var\_ipd = "ARM",  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 km\_conf\_type = "log-log",  
 km\_layout = "by\_trial",  
 time\_scale = "month",  
 break\_x\_by = 2  
)



# plot by arm  
kmplot2(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = "C",  
 trt\_var\_ipd = "ARM",  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 km\_conf\_type = "log-log",  
 km\_layout = "by\_arm",  
 time\_scale = "month",  
 break\_x\_by = 2  
)



# plot all  
kmplot2(  
 weights\_object = match\_res,  
 tte\_ipd = adtte,  
 tte\_pseudo\_ipd = pseudo\_ipd,  
 trt\_ipd = "A",  
 trt\_agd = "B",  
 trt\_common = "C",  
 trt\_var\_ipd = "ARM",  
 trt\_var\_agd = "ARM",  
 endpoint\_name = "Overall Survival",  
 km\_conf\_type = "log-log",  
 km\_layout = "all",  
 time\_scale = "month",  
 break\_x\_by = 2,  
 xlim = c(0, 20)  
)



# References

[1] Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: Reconstructing the data from published kaplan-meier survival curves. BMC Medical Research Methodology 2012;12:9.

[2] Liu N, Yanhong Z, Lee JJ. IPDfromKM: Reconstruct individual patient data from published kaplan-meier survival curves. BMC Medical Research Methodology 2021;21:111.