

# Non-Parametric estimate & test

Yixuan Jiao

2023-10-03

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.2      v purrr   1.0.1
## v tibble  3.2.1      v dplyr   1.1.1
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.2      v forcats 1.0.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(survival)
library(survminer)
```

```
## Loading required package: ggpubr
##
## Attaching package: 'survminer'
##
## The following object is masked from 'package:survival':
##
##     myeloma
```

```
library(patchwork)
library(biostat3)
```

```
## Loading required package: MASS
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:patchwork':
##
##     area
##
## The following object is masked from 'package:dplyr':
##
##     select
##
## Attaching package: 'biostat3'
##
```

```
## The following object is masked from 'package:survival':
##
##   colon
```

```
aids <- read_csv("data/AIDS_Clinical_Trials_Group175.csv")
```

```
## New names:
## Rows: 2139 Columns: 25
## -- Column specification
## ----- Delimiter: "," dbl
## (25): ...1, time, trt, age, wtkg, hemo, homo, drugs, karnof, oprior, z30...
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * '' -> '...1'
```

```
aids %>% head()
```

```
## # A tibble: 6 x 25
##   ...1 time  trt  age  wtkg  hemo  homo  drugs  karnof  oprior  z30  zprior
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl> <dbl>  <dbl>
## 1     0  948    2   48  89.8    0    0    0    100    0    0    1
## 2     1 1002    3   61  49.4    0    0    0    90    0    1    1
## 3     2  961    3   45  88.5    0    1    1    90    0    1    1
## 4     3 1166    3   47  85.3    0    1    0   100    0    1    1
## 5     4 1090    0   43  66.7    0    1    0   100    0    1    1
## 6     5 1181    1   46  88.9    0    1    1   100    0    1    1
## # i 13 more variables: preanti <dbl>, race <dbl>, gender <dbl>, str2 <dbl>,
## #   strat <dbl>, symptom <dbl>, treat <dbl>, offtrt <dbl>, cd40 <dbl>,
## #   cd420 <dbl>, cd80 <dbl>, cd820 <dbl>, cid <dbl>
```

```
#KM
```

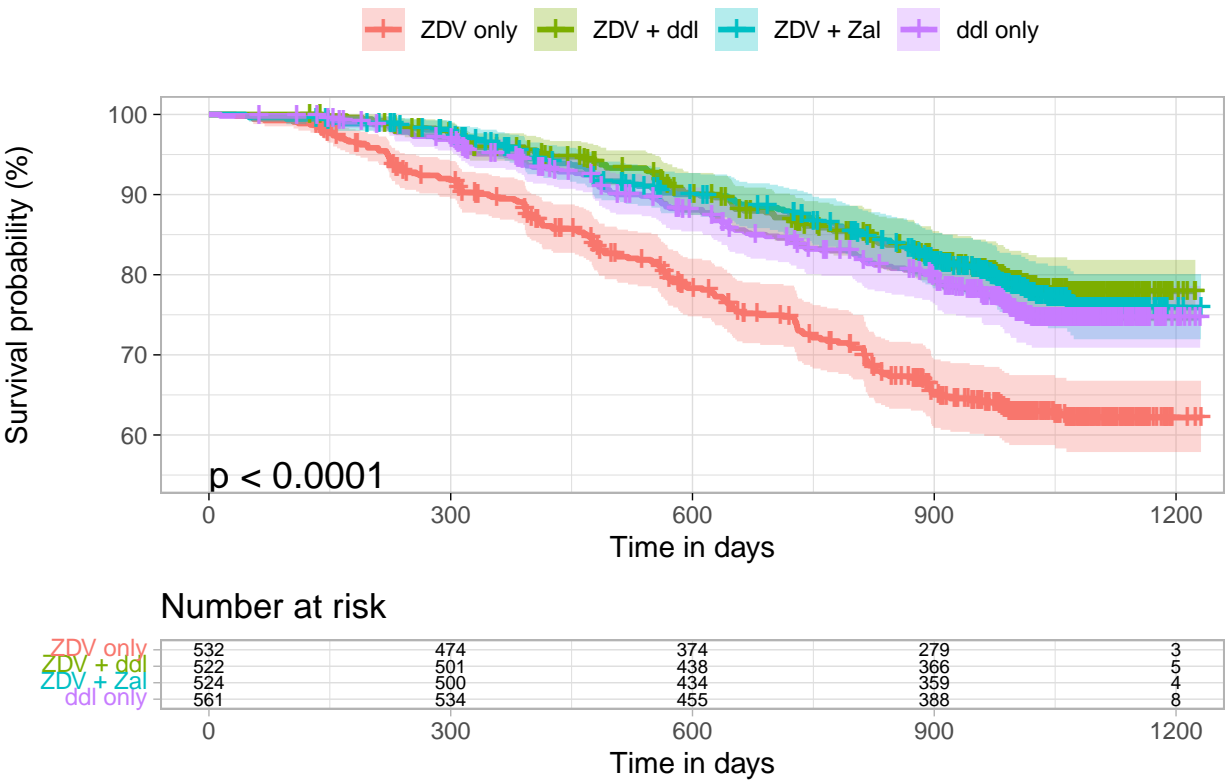
```
km_fit_trt <- survfit(Surv(time, cid) ~ trt, data = aids)
pkm <- km_fit_trt %>% ggsurvplot(data = aids,
  fun = "pct", #can be replaced by cum hazard
  conf.int = TRUE,
  risk.table = TRUE,
  pval = TRUE,
  pval.coord = c(0,55),
  fontsize = 2.5,
  ggtheme = theme_light(),
  xlab = "Time in days",
  title = "Figure 4: Kaplan-Meier Survival Function Estimate",
  legend.title = "",
  legend.labs = c("ZDV only", "ZDV + ddl", "ZDV + Zal", "ddl only"),
  ylim = c(55, 100))
```

```
hkm <- km_fit_trt %>% ggsurvplot(data = aids,
  fun = "cumhaz", #can be replaced by cum hazard
  conf.int = TRUE,
  risk.table = TRUE,
  pval = TRUE,
  fontsize = 2.5,
```

pkm

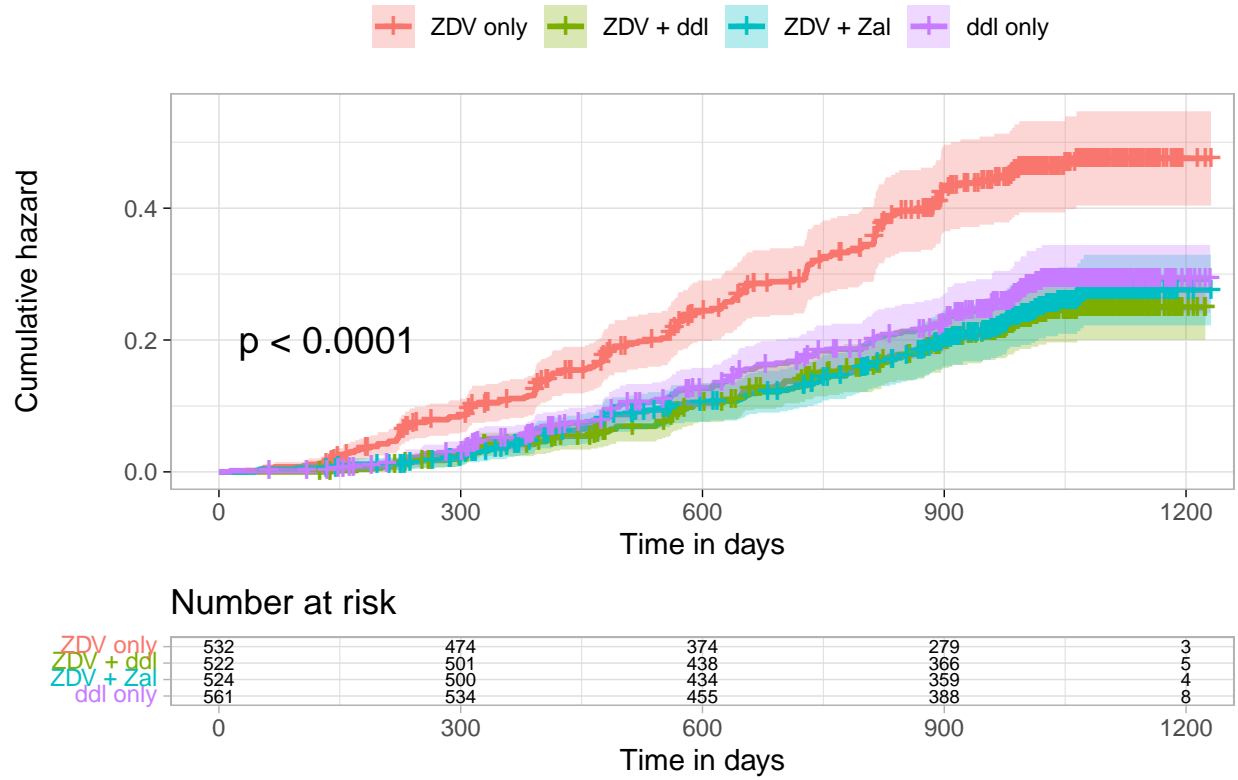
```
ggtheme = theme_light(),
xlab = "Time in days",
title = "Figure 5: Kaplan-Meier Cumulative Hazard Function",
legend.title = "",
legend.labs = c("ZDV only", "ZDV + ddl", "ZDV + Zal", "ddl only"))
```

Figure 4: Kaplan–Meier Survival Function Estimate



hkm

Figure 5: Kaplan–Meier Cumulative Hazard Function



```
quantile(km_fit_trt, probs = c(0.1, 0.15, 0.2))$quantile %>% knitr::kable(caption = "t1")
```

Table 1: t1

	10	15	20
trt=0	347	468	569
trt=1	626	822	986
trt=2	610	806	972
trt=3	537	672	898

```
quantile(km_fit_trt, probs = c(0.1, 0.15, 0.2))$quantile %>% knitr::kable(caption = "t2")
```

Table 2: t2

	10	15	20
trt=0	347	468	569
trt=1	626	822	986
trt=2	610	806	972
trt=3	537	672	898

Table 3: Quantile Survival Time (in days) by Treatments

Treatment	90 <sup>th</sup>	85 <sup>th</sup>	80 <sup>th</sup>
ZDV only	347 (284,406)	468 (394,557)	569 (484,649)
ZDV + ddl	626 (559,721)	822 (691,929)	986 (876,NA)
ZDV + Zal	610 (476,748)	806 (720,910)	972 (867,NA)
ddl only	537 (476,631)	672 (613,813)	898 (760,994)

```
survdif(Surv(time, cid) ~ trt, data = aids)
```

```
## Call:
## survdiff(formula = Surv(time, cid) ~ trt, data = aids)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## trt=0 532      181      116   37.030    47.67
## trt=1 522      103      134    6.988     9.40
## trt=2 524      109      132    4.158     5.58
## trt=3 561      128      139    0.933     1.27
##
##  Chisq= 49.2  on 3 degrees of freedom, p= 1e-10
```

```
aids %>%
  mutate(trt = case_when(trt == "0" ~ "ZDV only",
                        trt == "1" ~ "ZDV + ddl",
                        trt == "2" ~ "ZDV + Zal",
                        trt == "3" ~ "ddl only")) %>%
  pairwise_survdif(Surv(time, cid) ~ trt, data = ., p.adjust.method = "BH") %>%
  broom::tidy()
```

```
## # A tibble: 6 x 3
##   group1   group2   p.value
##   <chr>    <chr>    <dbl>
## 1 ZDV + ddl ddl only 0.278
## 2 ZDV + Zal ddl only 0.478
## 3 ZDV + Zal ZDV + ddl 0.636
## 4 ZDV only  ddl only 0.00000750
## 5 ZDV only  ZDV + ddl 0.0000000364
## 6 ZDV only  ZDV + Zal 0.000000242
```

Table 4: Pairwise Log-rank Tests by Treatments

group1	group2	P-value
ZDV + ddl	ddl only	0.2784441
ZDV + Zal	ddl only	0.4776330
ZDV + Zal	ZDV + ddl	0.6362565
ZDV only	ddl only	0.0000075*
ZDV only	ZDV + ddl	0.0000000*
ZDV only	ZDV + Zal	0.0000002*