# SURVIVAL ANALYSIS - TCGA PRAD CANCER

Kelvin Ofori-Minta University of Texas at El Paso (UTEP)

June 11, 2022

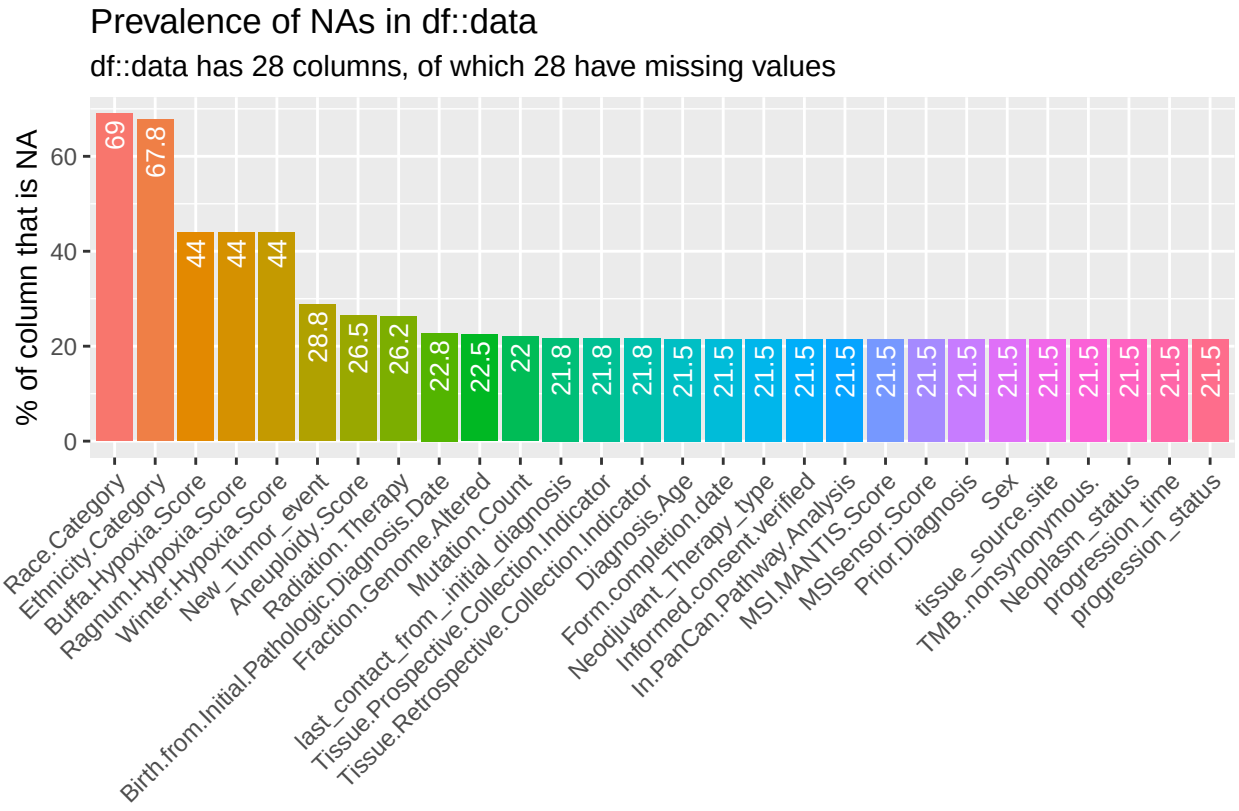
## Contents

# 1 Loading and Cleaning Data

```
data <- read.csv("neoplasm.status_tumorfree.csv", header = T, stringsAsFactors = F,
                 na.strings = "NA")
#selecting columns of interest
#data<-s[, c(4,7,8,12,13,19,20,21,23,27:32,44:47,50,53:55,58,62,40:42)]
# write.csv(data,"C://Users//Kelvin//Desktop//Spring 2022//research with Dr. Leung//survival//
```

## 1.1 Inspecting dataframe for missing values

```
require(inspectdf)
show_plot(inspect_na(data))
```




```
missing = inspect_na(data)
missing[ , 3] = round(missing[ ,3], 2)
names(missing) = c("variable", "count", "proportion")
require(kableExtra)
# missing<-as.matrix.data.frame(missing)
kable(missing)
```

| variable | count | proportion |
|---|---|---|
| Race.Category | 276 | 69.00 |
| Ethnicity.Category | 271 | 67.75 |
| Buffa.Hypoxia.Score | 176 | 44.00 |
| Ragnum.Hypoxia.Score | 176 | 44.00 |
| Winter.Hypoxia.Score | 176 | 44.00 |
| New_Tumor_event | 115 | 28.75 |
| Aneuploidy.Score | 106 | 26.50 |
| Radiation.Therapy | 105 | 26.25 |
| Birth.from.Initial.Pathologic.Diagnosis.Date | 91 | 22.75 |
| Fraction.Genome.Altered | 90 | 22.50 |
| Mutation.Count | 88 | 22.00 |
| last_contact_from_.initial_diagnosis | 87 | 21.75 |
| Tissue.Prospective.Collection.Indicator | 87 | 21.75 |
| Tissue.Retrospective.Collection.Indicator | 87 | 21.75 |
| Diagnosis.Age | 86 | 21.50 |
| Form.completion.date | 86 | 21.50 |
| Neodjuvant_Therapy_type | 86 | 21.50 |
| Informed.consent.verified | 86 | 21.50 |
| In.PanCan.Pathway.Analysis | 86 | 21.50 |
| MSI.MANTIS.Score | 86 | 21.50 |
| MSIsensor.Score | 86 | 21.50 |
| Prior.Diagnosis | 86 | 21.50 |
| Sex | 86 | 21.50 |
| tissue_source.site | 86 | 21.50 |
| TMB..nonsynonymous. | 86 | 21.50 |
| Neoplasm_status | 86 | 21.50 |
| progression_time | 86 | 21.50 |
| progression_status | 86 | 21.50 |

```
# as.data.frame.matrix(missing)
# kable(as.da(missing))
```

### 1.1.1 Rename long variables

```
"TMB-H means that the tumor has a high number of mutations. Doctors have found that
certain immunotherapy drugs are more likely to work
against TMB-H cancers. This is because the immune
system may be able to find and attack cancer cells with high
TMB more easily."
```

```
## [1] "TMB-H means that the tumor has a high number of mutations. Doctors have found that\nce
```

```
"Person neoplasm status..... You are correct, IMO:  tumor free does not mean normal, but rathe
```

```
## [1] "Person neoplasm status..... You are correct, IMO:  tumor free does not mean normal, bu
```

### 1.1.2 Re-coding variables

```
# newdata$Neodjuvant_Therapy_type <- factor(newdata$Neodjuvant_Therapy_type,
#                                           levels=c("No","Yes"),
#                                           labels=c("No","Yes")) all were "no"


data$In.PanCan.Pathway.Analysis<-factor(data$In.PanCan.Pathway.Analysis,
                                                levels=c("No","Yes"),
                                                labels=c("No","Yes"))


data$Prior.Diagnosis<-factor(data$Prior.Diagnosis,
                                    levels=c("No","Yes"),
                                    labels=c("No","Yes"))


data$tissue_source.site<-factor(data$tissue_source.site,
                                       levels = c("university","Biotech & Pharma","Hospital","Resea
                                                                       labels=c("university","biotech_pharma


data$New_Tumor_event <- factor(data$New_Tumor_event,
                                        levels=c("No","Yes"),
                                        labels=c("No","Yes"))


data$Radiation.Therapy <- factor(data$Radiation.Therapy,
                                        levels=c("No","Yes"),
                                        labels=c("No","Yes"))

#all white , no adjuvant therapy
str(data)
```

```
## 'data.frame':    400 obs. of  28 variables:
##  $ Diagnosis.Age                                 : int  51 57 47 52 70 NA 69 57 57 56 ...
##  $ Aneuploidy.Score                              : int  0 0 14 0 2 NA 1 0 1 1 ...
##  $ Buffa.Hypoxia.Score                           : int  -27 -29 -39 -25 NA NA -25 NA NA NA ..
##  $ last_contact_from_.initial_diagnosis          : int  621 1701 1373 671 1378 NA 863 1364 127
##  $ Birth.from.Initial.Pathologic.Diagnosis.Date: int  -18658 -20958 -17365 -19065 -25904 NA
##  $ Ethnicity.Category                            : chr  NA NA NA NA ...
##  $ Form.completion.date                          : chr  "3/29/2014" "3/30/2014" "3/29/2014" "3
##  $ Fraction.Genome.Altered                       : num  0.03 0.0211 0.1418 0.0092 0.0756 ...
##  $ Neodjuvant_Therapy_type                       : chr  "No" "No" "No" "No" ...
##  $ Informed.consent.verified                     : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ In.PanCan.Pathway.Analysis                    : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 NA
##  $ MSI.MANTIS.Score                              : num  0.326 0.329 0.315 0.314 0.336 ...
##  $ MSIsensor.Score                               : num  0 0 0 0 0.03 NA 0 0 0.06 0.04 ...
```

```
##  $ Mutation.Count                                : int  22 27 39 24 30 NA 34 14 20 21 ...
##  $ New_Tumor_event                               : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 NA
##  $ Prior.Diagnosis                               : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 NA
##  $ Race.Category                                 : chr  NA NA NA NA ...
##  $ Radiation.Therapy                             : Factor w/ 2 levels "No","Yes": 1 1 2 1 1 NA
##  $ Ragnum.Hypoxia.Score                          : int  -24 -24 -22 -22 NA NA -22 NA NA NA ..
##  $ Sex                                           : chr  "Male" "Male" "Male" "Male" ...
##  $ Tissue.Prospective.Collection.Indicator       : chr  "No" "No" "No" "No" ...
##  $ Tissue.Retrospective.Collection.Indicator     : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ tissue_source.site                            : Factor w/ 4 levels "university","biotech_pl
##  $ TMB..nonsynonymous.                           : num  0.7 0.9 1.3 0.8 1 ...
##  $ Winter.Hypoxia.Score                          : int  -34 -26 -42 -36 NA NA -26 NA NA NA ..
##  $ Neoplasm_status                               : chr  "tumor_free" "tumor_free" "tumor_free"
##  $ progression_time                              : num  20.4 55.9 45.1 22.1 45.3 ...
##  $ progression_status                            : int  1 1 1 1 1 NA 2 1 1 1 ...
```

## 2 KM Curve - survival probability with Radiation Therapy

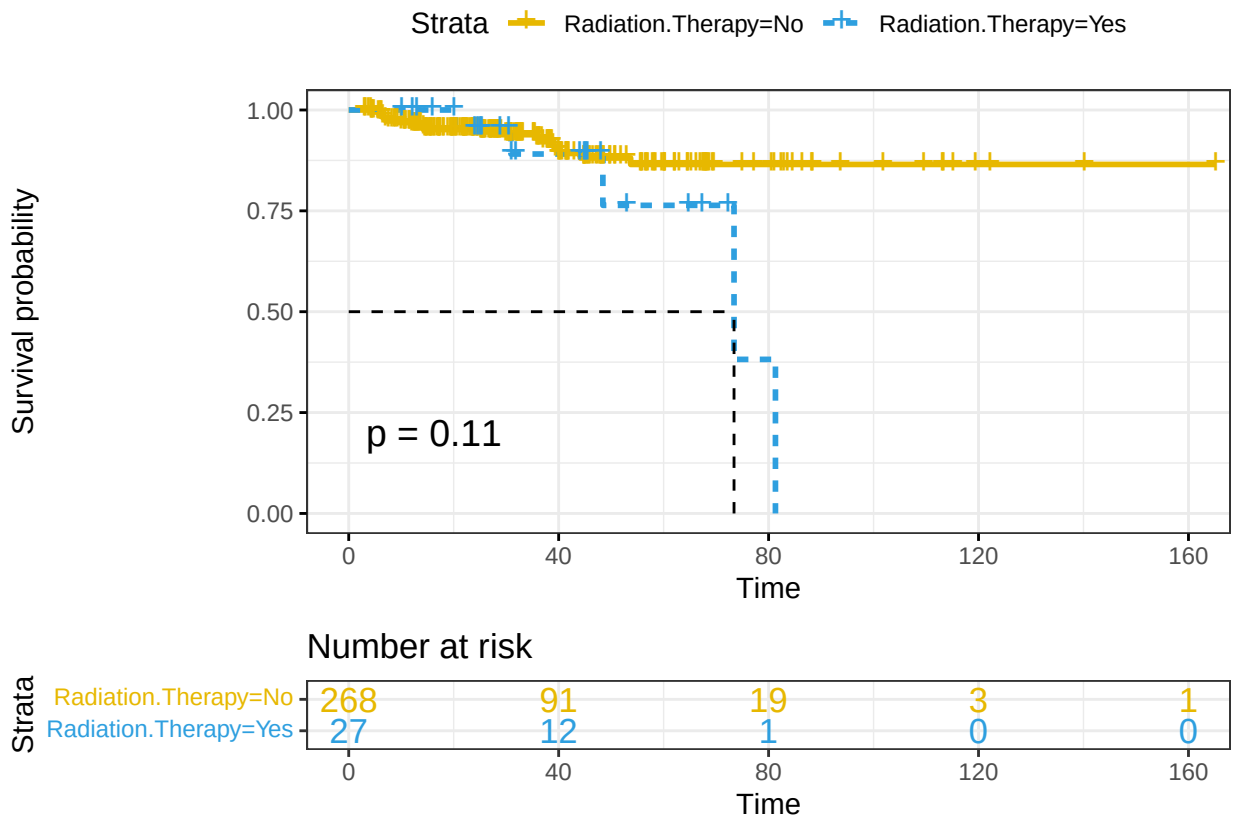
```
library("survival")
library("survminer")
ndata<-data
fit2a <- survfit(Surv(ndata$progression_time, ndata$progression_status) ~ ndata$Radiation.Thera
print(fit2a)
```

```
## Call: survfit(formula = Surv(ndata$progression_time, ndata$progression_status) ~
##     ndata$Radiation.Therapy, data = ndata)
##
##    105 observations deleted due to missingness
##                                n events median 0.95LCL 0.95UCL
## ndata$Radiation.Therapy=No  268     21     NA      NA      NA
## ndata$Radiation.Therapy=Yes  27      5   73.4    73.4      NA
```

```
summary(fit2a)$table
```

```
##                              records n.max n.start events    *rmean *se(rmean)
## ndata$Radiation.Therapy=No       268   268     268     21 146.80042   4.050186
## ndata$Radiation.Therapy=Yes       27    27      27      5  68.12697   5.279824
##                               median  0.95LCL 0.95UCL
## ndata$Radiation.Therapy=No        NA       NA      NA
## ndata$Radiation.Therapy=Yes 73.41289 73.41289      NA
```

```
ggsurvplot(fit2a,
          #legend.labs=c("tumor_free", "with_tumor"),
          pval = TRUE, conf.int = F,
          risk.table = TRUE, # Add risk table
          risk.table.col = "strata", # Change risk table color by groups
          linetype = "strata", # Change line type by groups
          surv.median.line = "hv", # Specify median survival
          ggtheme = theme_bw(), # Change ggplot2 theme
          palette = c("#E7B800", "#2E9FDF"))
```




```
#1 - censored & 2- progression
#1 - tumor_free & 2 with tumor .....neoplasm status
#1 - NO & 2-YES .....TREATMENT CODE
```

# 3 KM Curve- survival probability:Censored cases of Radiation therapy

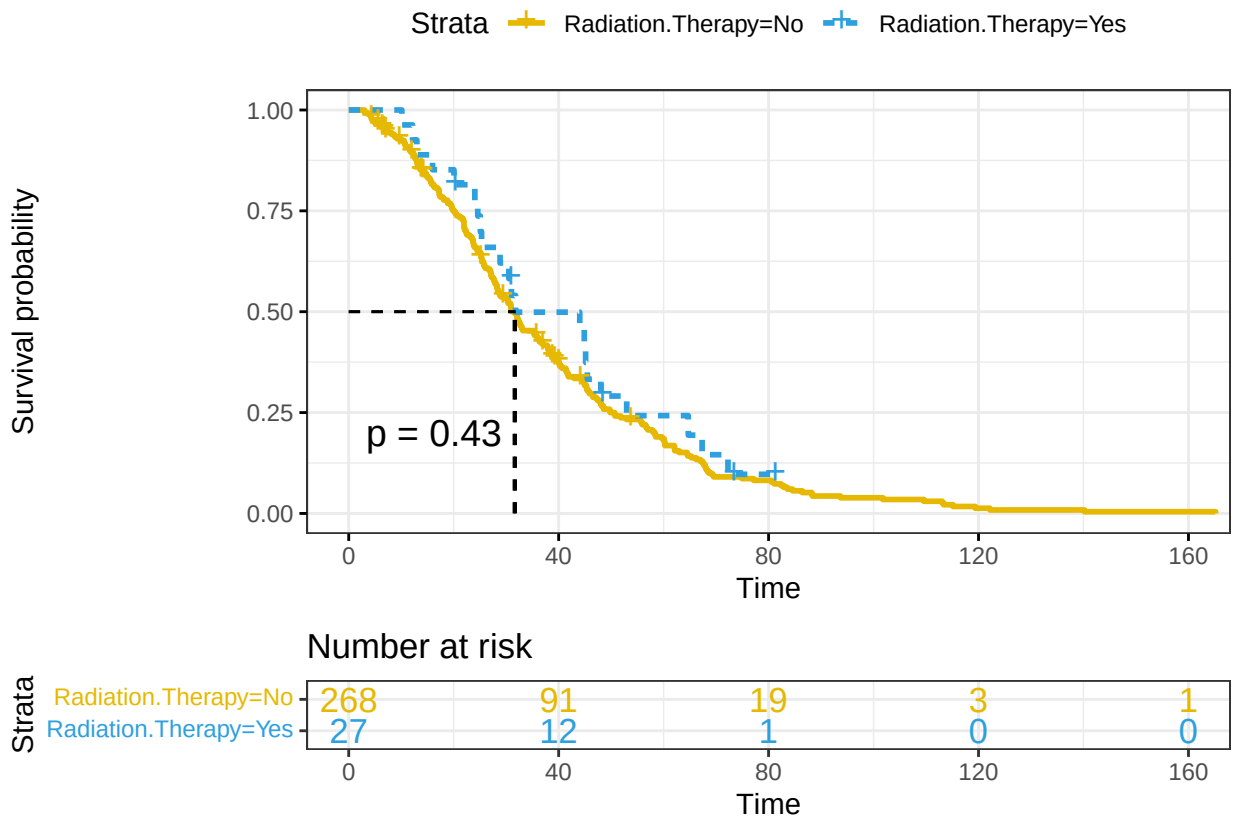
```
library("survival")
library("survminer")

fit2b <- survfit(Surv(ndata$progression_time, ndata$progression_status==1) ~ ndata$Radiation.T
               data = ndata)
print(fit2b)
```

```
## Call: survfit(formula = Surv(ndata$progression_time, ndata$progression_status ==
##     1) ~ ndata$Radiation.Therapy, data = ndata)
##
##    105 observations deleted due to missingness
##                                n events median 0.95LCL 0.95UCL
## ndata$Radiation.Therapy=No   268    247   31.5    28.3      36
## ndata$Radiation.Therapy=Yes   27     22   31.8    28.8      53
```

```
summary(fit2b)$table
```

```
##                               records n.max n.start events   *rmean *se(rmean)
## ndata$Radiation.Therapy=No        268   268     268    247 38.43101   1.687711
## ndata$Radiation.Therapy=Yes        27    27      27     22 49.35452   8.646873
##                                median  0.95LCL  0.95UCL
## ndata$Radiation.Therapy=No  31.49555 28.27366 36.03248
## ndata$Radiation.Therapy=Yes 31.79143 28.83256 52.96380
```

```
ggsurvplot(fit2b,
          #legend.labs=c("tumor_free", "with_tumor"),
          pval = TRUE, conf.int = F,
          risk.table = TRUE, # Add risk table
          risk.table.col = "strata", # Change risk table color by groups
          linetype = "strata", # Change line type by groups
          surv.median.line = "hv", # Specify median survival
          ggtheme = theme_bw(), # Change ggplot2 theme
          palette = c("#E7B800", "#2E9FDF"))
```




```
#1 - censored & 2- progression
#1 - tumor_free & 2 with tumor .....neoplasm status
#1 - NO & 2-YES .....TREATMENT CODE
```

# 4 KM Curve- survival probability:Censored cases of New tumor event

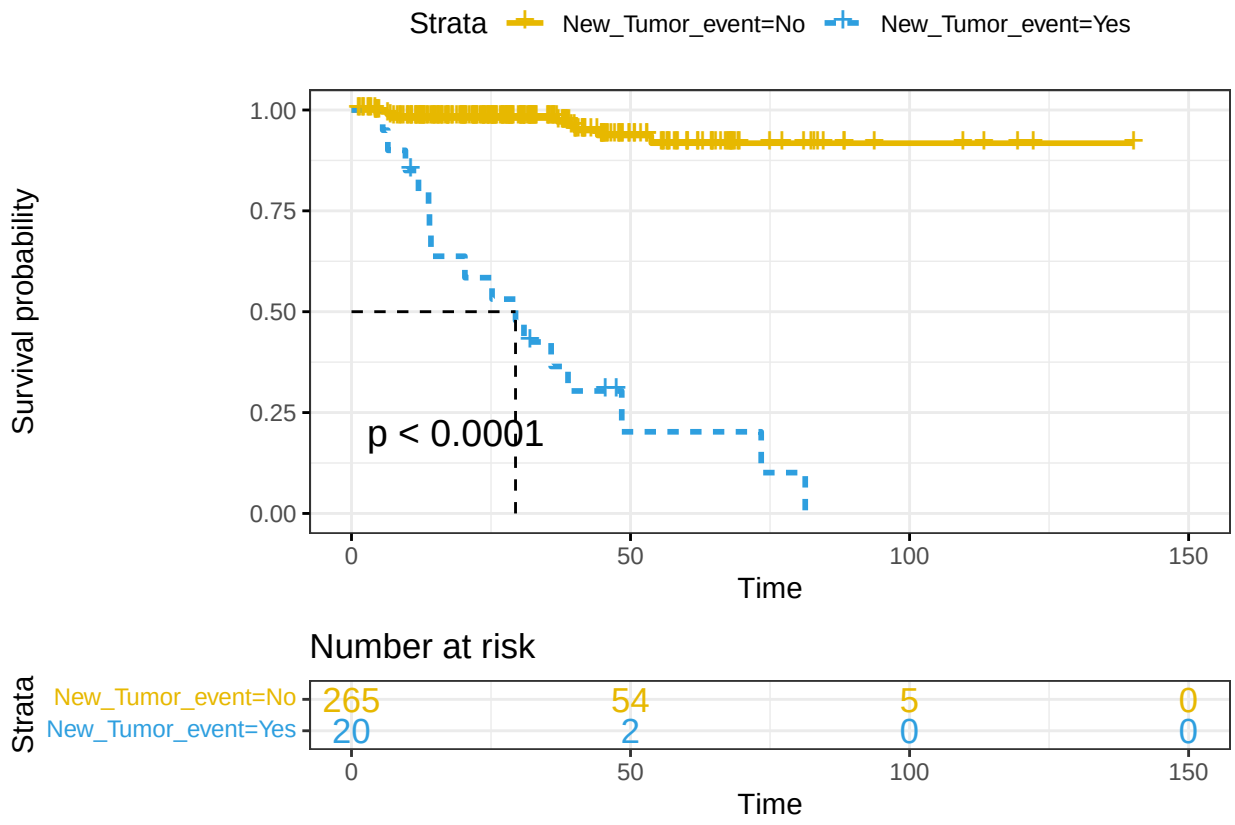
```
library("survival")
library("survminer")

fit2c <- survfit(Surv(ndata$progression_time, ndata$progression_status==2) ~ ndata$New_Tumor_ev
               data = ndata)
print(fit2c)
```

```
## Call: survfit(formula = Surv(ndata$progression_time, ndata$progression_status ==
##     2) ~ ndata$New_Tumor_event, data = ndata)
##
##    115 observations deleted due to missingness
##                              n events median 0.95LCL 0.95UCL
## ndata$New_Tumor_event=No   265     10     NA      NA      NA
## ndata$New_Tumor_event=Yes   20     16   29.4    14.2      NA
```

```
summary(fit2c)$table
```

```
##                             records n.max n.start events    *rmean *se(rmean)
## ndata$New_Tumor_event=No        265   265     265     10 131.54551   2.877115
## ndata$New_Tumor_event=Yes        20    20      20     16  34.66143   5.994247
##                              median  0.95LCL 0.95UCL
## ndata$New_Tumor_event=No         NA       NA      NA
## ndata$New_Tumor_event=Yes 29.39146 14.23546      NA
```

```
ggsurvplot(fit2c,
          #legend.labs=c("tumor_free", "with_tumor"),
          pval = TRUE, conf.int = F,
          risk.table = TRUE, # Add risk table
          risk.table.col = "strata", # Change risk table color by groups
          linetype = "strata", # Change line type by groups
          surv.median.line = "hv", # Specify median survival
          ggtheme = theme_bw(), # Change ggplot2 theme
          palette = c("#E7B800", "#2E9FDF"))
```




```
#1 - censored & 2- progression
#1 - tumor_free & 2 with tumor .....neoplasm status
#1 - NO & 2-YES .....TREATMENT CODE
```

# 5 KM Curve- survival probability:Uncensored cases of Prior Diagnosis

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

fit2d <- survfit(Surv(ndata$progression_time, ndata$progression_status==2) ~ ndata$Prior.Diagno
               data = ndata)
print(fit2d)
```

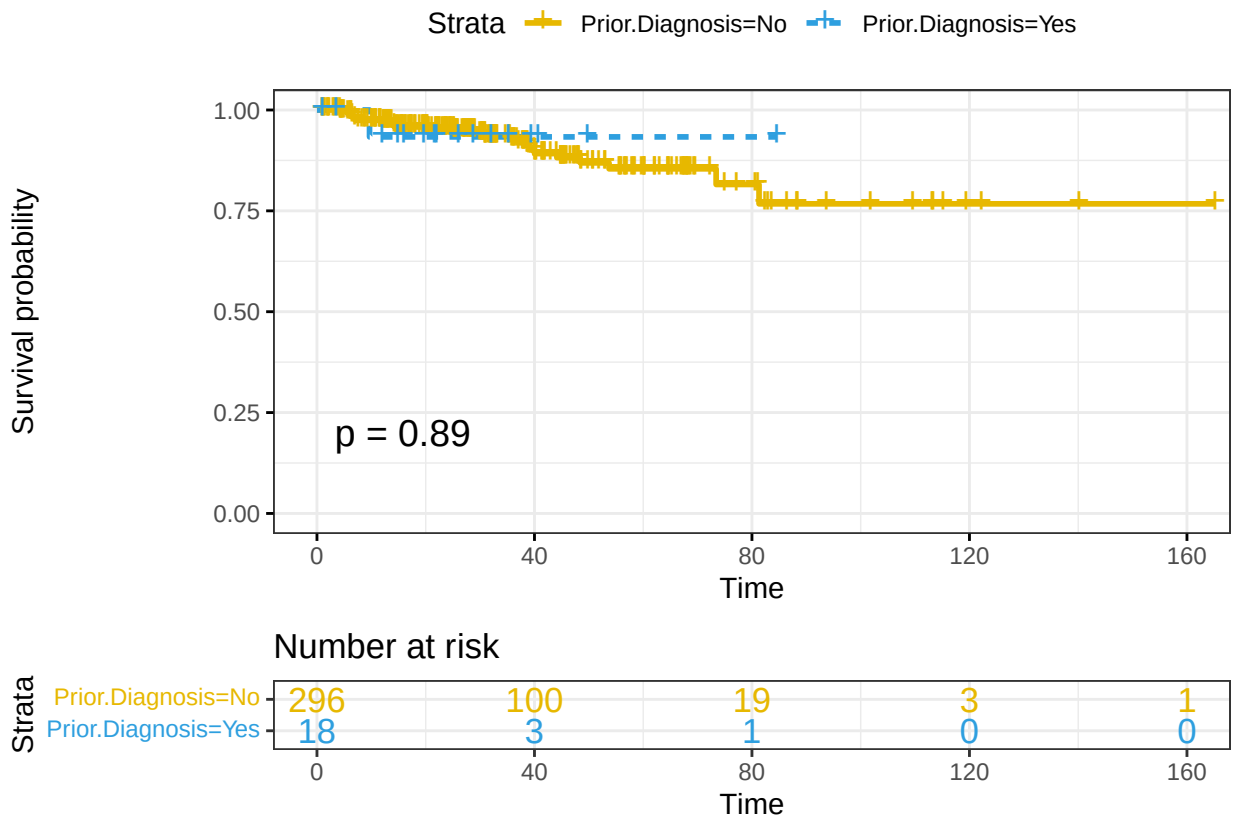
```
## Call: survfit(formula = Surv(ndata$progression_time, ndata$progression_status ==
##     2) ~ ndata$Prior.Diagnosis, data = ndata)
##
##    86 observations deleted due to missingness
##                              n events median 0.95LCL 0.95UCL
## ndata$Prior.Diagnosis=No  296     25     NA      NA      NA
## ndata$Prior.Diagnosis=Yes  18      1     NA      NA      NA
```

```
summary(fit2d)$table
```

```
##                             records n.max n.start events   *rmean *se(rmean)
## ndata$Prior.Diagnosis=No        296   296     296     25 138.0113    6.37083
## ndata$Prior.Diagnosis=Yes        18    18      18      1 154.8038   10.01548
##                             median 0.95LCL 0.95UCL
## ndata$Prior.Diagnosis=No        NA      NA      NA
## ndata$Prior.Diagnosis=Yes       NA      NA      NA
```

```
ggsurvplot(fit2d,
          #legend.labs=c("tumor_free", "with_tumor"),
          pval = TRUE, conf.int = F,
          risk.table = TRUE, # Add risk table
          risk.table.col = "strata", # Change risk table color by groups
          linetype = "strata", # Change line type by groups
          surv.median.line = "hv", # Specify median survival
          ggtheme = theme_bw(), # Change ggplot2 theme
          palette = c("#E7B800", "#2E9FDF"))
```

```
## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, : Median
## survival not reached.
```




```
#1 - censored & 2- progression
```