Validation of the St. Gallen Algorithm for Treatment of Proximal Humerus Fractures

Angela Moosman, MD, Moritz Tannast, MD, Joseph M. Schwab, MD

**Authors: Angela Moosman, MD, Moritz Tannast, MD, Joseph M. Schwab, MD**

# Data Import, Tidying, and Validation

The first thing we need to do is to import our excel spreadsheet and create our data\_frame(s).

library(readxl, quietly = TRUE)  
library(dplyr, quietly = TRUE)

##   
## Attaching package: 'dplyr'

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

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

SGA\_df <- read\_excel(  
 "Schwab Database SGA.xlsx",  
 col\_types = c(  
 "numeric",  
 "date",  
 "numeric",  
 "text",  
 "text",  
 "numeric",  
 "numeric",  
 "date",  
 "text",  
 "numeric",  
 "text",  
 "numeric",  
 "date",  
 "numeric",  
 "numeric",   
 "numeric",  
 "numeric"  
 )  
)  
  
SGA\_df <- rename(SGA\_df,  
 event = `unplanned secondary surgery (0-12 months postop)`,  
 time = `follow up (months)`,  
 group = `following algorithm`)  
SGA\_df$group <- factor(  
 SGA\_df$group,  
 levels = c(0, 1),  
 labels = c("Not following SGA", "Following SGA")  
 )  
  
SGA\_df

## # A tibble: 320 × 17  
## `Study Number` D.o.B `Age at diagnosis` Gender side Neer  
## <dbl> <dttm> <dbl> <chr> <chr> <dbl>  
## 1 1 1939-04-06 00:00:00 74 F R 4  
## 2 2 1965-11-29 00:00:00 48 F L 4  
## 3 3 1947-03-15 00:00:00 73 M R 3  
## 4 4 1966-01-14 00:00:00 47 F R 4  
## 5 5 1963-12-17 00:00:00 52 F L 3  
## 6 6 1975-02-06 00:00:00 38 F L 4  
## 7 7 1952-11-06 00:00:00 68 M R 4  
## 8 8 1960-06-21 00:00:00 57 M L 2  
## 9 9 1964-11-21 00:00:00 52 F L 3  
## 10 10 1939-06-23 00:00:00 79 F R 2  
## # ℹ 310 more rows  
## # ℹ 11 more variables: DTI <dbl>, `Date of Surgery/diagnosis` <dttm>,  
## # `done treatment` <chr>, `pathway allocation` <dbl>, `suggested TTT` <chr>,  
## # group <fct>, `Last follow up` <dttm>, time <dbl>, event <dbl>,  
## # `year of surgery/diganosis` <dbl>, year <dbl>

Since the original SGA was created based on secondary surgery within 12 months, we will create a second dataframe where any events that occurred after twelve months are converted to non-events, and all time to follow-up are capped at 12 months

SGA\_12mo\_df <- SGA\_df  
SGA\_12mo\_df$event[SGA\_12mo\_df$event == 1 & SGA\_12mo\_df$time > 12] <- 0  
SGA\_12mo\_df$time[SGA\_12mo\_df$event == 0 & SGA\_12mo\_df$time > 12] <- 12  
SGA\_12mo\_df

## # A tibble: 320 × 17  
## `Study Number` D.o.B `Age at diagnosis` Gender side Neer  
## <dbl> <dttm> <dbl> <chr> <chr> <dbl>  
## 1 1 1939-04-06 00:00:00 74 F R 4  
## 2 2 1965-11-29 00:00:00 48 F L 4  
## 3 3 1947-03-15 00:00:00 73 M R 3  
## 4 4 1966-01-14 00:00:00 47 F R 4  
## 5 5 1963-12-17 00:00:00 52 F L 3  
## 6 6 1975-02-06 00:00:00 38 F L 4  
## 7 7 1952-11-06 00:00:00 68 M R 4  
## 8 8 1960-06-21 00:00:00 57 M L 2  
## 9 9 1964-11-21 00:00:00 52 F L 3  
## 10 10 1939-06-23 00:00:00 79 F R 2  
## # ℹ 310 more rows  
## # ℹ 11 more variables: DTI <dbl>, `Date of Surgery/diagnosis` <dttm>,  
## # `done treatment` <chr>, `pathway allocation` <dbl>, `suggested TTT` <chr>,  
## # group <fct>, `Last follow up` <dttm>, time <dbl>, event <dbl>,  
## # `year of surgery/diganosis` <dbl>, year <dbl>

# Kaplan-Meier Survival Curves

Now we will generate Kaplan-Meier Survival Curves for the overall database.

library(survival, quietly = TRUE)  
library(ggsurvfit, quietly = TRUE)

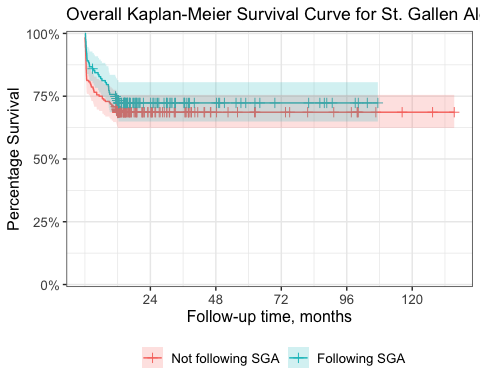
## Loading required package: ggplot2

library(survminer, quietly = TRUE)

##   
## Attaching package: 'survminer'

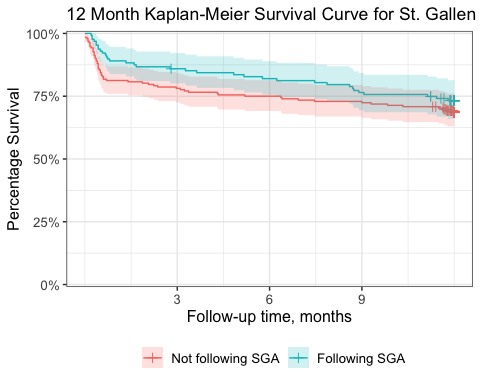
## The following object is masked from 'package:survival':  
##   
## myeloma

library(Rcpp, quietly = TRUE)  
  
# Overall KM Survival Curve  
fit\_SGA <- survfit2(Surv(time, event) ~ group, data = SGA\_df)  
fit\_SGA %>% ggsurvfit() +  
 labs(title = "Overall Kaplan-Meier Survival Curve for St. Gallen Algorithm",  
 y = "Percentage Survival",  
 x = "Follow-up time, months") +  
 add\_censor\_mark() +  
 add\_confidence\_interval() +  
 theme(legend.position = "bottom") +  
 scale\_y\_continuous(  
 limits = c(0, 1),  
 labels = scales::percent,  
 expand = c(0.01, 0)  
 ) +  
 scale\_x\_continuous(  
 breaks = c(24, 48, 72, 96, 120)  
 )



Now we can do the same Kaplan-Meier Curve for the data that has been capped at 12 months. We will make some slight changes to the x axis to reflect the shorter timeframe.

# 12-month KM Survival Curve  
fit\_SGA\_12mo <- survfit2(Surv(time, event) ~ group, data = SGA\_12mo\_df)  
fit\_SGA\_12mo %>% ggsurvfit() +  
 labs(title = "12 Month Kaplan-Meier Survival Curve for St. Gallen Algorithm",  
 y = "Percentage Survival",  
 x = "Follow-up time, months") +  
 add\_censor\_mark() +  
 add\_confidence\_interval() +  
 theme(legend.position = "bottom") +  
 scale\_y\_continuous(  
 limits = c(0, 1),  
 labels = scales::percent,  
 expand = c(0.01, 0)  
 ) +  
 scale\_x\_continuous(limits = c(0, 12),  
 breaks = c(3, 6, 9))



# Logrank tests for comparison of Survival Curves

Next we can perform Logrank tests on both the overall data, and the 12-month capped data.

surv\_diff <- survdiff(Surv(time, event) ~ group, data = SGA\_df)  
surv\_diff

## Call:  
## survdiff(formula = Surv(time, event) ~ group, data = SGA\_df)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## group=Not following SGA 192 60 55.4 0.387 0.932  
## group=Following SGA 128 35 39.6 0.541 0.932  
##   
## Chisq= 0.9 on 1 degrees of freedom, p= 0.3

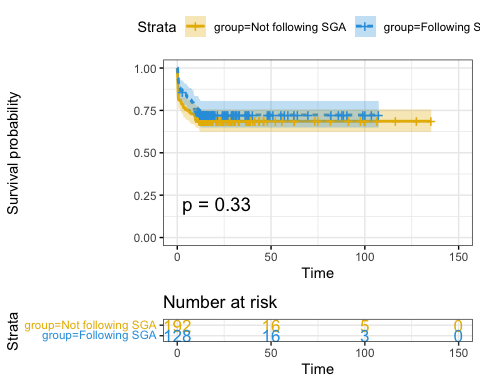
surv\_diff\_12mo <- survdiff(Surv(time, event) ~ group, data = SGA\_12mo\_df)  
surv\_diff\_12mo

## Call:  
## survdiff(formula = Surv(time, event) ~ group, data = SGA\_12mo\_df)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## group=Not following SGA 192 60 54.8 0.493 1.19  
## group=Following SGA 128 34 39.2 0.689 1.19  
##   
## Chisq= 1.2 on 1 degrees of freedom, p= 0.3

Finally we can plot the overall survival again with some more information.

ggsurvplot(  
 fit\_SGA,  
 pval = TRUE,  
 conf.int = TRUE,  
 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.



And here is that curve plotted for the 12-month capped data

ggsurvplot(  
 fit\_SGA\_12mo,  
 pval = TRUE,  
 conf.int = TRUE,  
 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.

