

Supplementary Appendix: R-Code and Additonal Analyses

Contents

Introduction	1
Data Preparation	1
Loading Packages	1
Loading Dataset	2
Recoding the variables	2
Descriptive Analysis	3
Quantitative Variables	3
Categorical Variables	4
Association between categorical variables and outcome	4
Association between continuous variables and outcome	5
Association between Early Steroid Therapy, Severity and Outcome	6
Cox Regression Analysis	7
Data Preparation	7
Univariate Cox Model	8
Multivariate Analysis	9
Visualization of the Cox Proportional Model	10
Additional Analyses	13
Exploration of FiO2	13
Initiation of steroid therapy after cytokine storm by outcome	15
Ventillator Free Days	16

Introduction

This document is the supplementary appendix which describes the R-code and provides additional tables and analyses. All the necessary data, code for performing the statsitcal analysis is provided in supporting files.

Data Preparation

Loading Packages

The following code loads the necessary packages required to perform the analyses.

```
library(readxl)
library(dplyr)
library(ggplot2)
```

```

library(gridExtra)
library(tidyverse)
library(plotly)
library(elrm)
library(ggridges)
library(reshape2)
library(lubridate)
library(gtsummary)
library(tidyr)
library(psych)
library(gridExtra)
library(tidyverse)
library(ggpubr)
library(wesanderson)
library(magrittr)
library(tinytex)
library(lubridate)

```

Loading Dataset

The data is loaded into the environment using the function `read_excel()` from the `readxl` package. The dataset contains information of 45 study participants admitted to the intensive care unit (ICU) at All India Institute of Medical Sciences, Bhopal, India (AIIMS Bhopal).

```

##import the dataset
df <- read_excel("Steroids_final.xlsx", sheet = "Sheet3")

```

Recoding the variables

The variables include sociodemographic information, clinical and biochemical parameters, variables regarding the severity of the condition, respiratory parameters and the outcome of the patient. The following code recodes and manipulates the variables for ease of use and plotting. New variables are created based on existing variables for the purpose of data analysis and data visualization.

```

df$outcome <- ifelse(df$outcome == 1, "Survived", "Died")
df$outcome <- factor(df$outcome, levels = c("Survived", "Died"))
df$app_steroid <- factor(ifelse(df$app_steroid == 1, "Optimal", "Suboptimal"))
df$age_60_above <- ifelse(df$age > 60, 1, 0)
df$age_60_above <- factor(ifelse(df$age_60_above == 1, ">60 Years", "\u2264 60 Years"))
df$comorbidity <- ifelse(df$dm == "Diabetic" | df$htn == "Hypertensive", "Present", "Absent")
df$comorbidity <- factor(df$comorbidity, levels = c("Present", "Absent"))
df$multimorbidity <- ifelse(df$dm == "Diabetic" & df$htn == "Hypertensive", "Present", "Absent")
df$multimorbidity <- factor(df$multimorbidity, levels = c("Present", "Absent"))
df$steroid_cat <- ifelse(df$steroid_cat == 1, "Early High",
                        ifelse(df$steroid_cat == 2, "Early Low", "Late"))
df$steroid_cat <- factor(df$steroid_cat)
df$Infection_cat <- ifelse(df$Infection_cat == 0, "Sterile", "Infected")
df$Infection_cat <- factor(df$Infection_cat)
df$sex <- ifelse(df$sex == "F", "Female", "Male")
df$sex <- factor(df$sex)
df$dm <- ifelse(df$dm == 1, "Diabetic", "Non-Diabetic")
df$dm <- factor(df$dm)

```

```

df$htn <- ifelse(df$htn == 1, "Hypertensive", "Non-Hypertensive")
df$htn <- factor(df$htn)
df$cad <- ifelse(df$cad == 1, "Present", "Absent")
df$cad <- factor(df$cad, levels = c("Present", "Absent"))
df$ckd <- ifelse(df$ckd == 1, "Present", "Absent")
df$ckd <- factor(df$ckd, levels = c("Present", "Absent"))
df$date_symptom <- as.Date(df$date_symptom)
df$date_cough <- as.Date(df$date_cough)
df$date_sob <- as.Date(df$date_sob)
df$date_testing <- as.Date(df$date_testing)
df$date_cytokine_storm <- as.Date(df$date_cytokine_storm)
df$date_steroid <- as.Date(df$date_steroid)
df$date_ward_shift <- as.Date(df$date_ward_shift)
df$date_discharge <- as.Date(df$date_discharge)
df$date_death <- as.Date(df$date_death)
df$date_intubation <- as.Date(df$date_intubation)
df$pf_storm <- as.numeric(df$pf_storm)
df$arads_cat <- as.factor(df$arads_cat)
df$early_high <- factor(ifelse(df$steroid_cat == "Early High", "Early High", "Others"))
df$pfr_150 <- ifelse(df$pf_d1>150, "PFR >150", "PFR \u2264 150")
df$pfr_100 <- ifelse(df$pf_d1>100, "PFR >100", "PFR \u2264 100")
df$res_support <- factor(ifelse(df$niv==1|df$high_flow==1, "Recived", "Not Recived"),
                          levels = c("Recived", "Not Recived"))

```

Descriptive Analysis

Quantitative Variables

The table below summarizes the quantitative variables describing their measures of central tendency and dispersion (mean, median, SD and IRQ). The summaries are grouped over the outcome (Survival and Death).

Variable	Outcome	N	Mean	SD	Median	Q1	Q3
age	Survived	64	53.16	12.34	55.5	45.00	61.00
saturation_admision	Survived	64	79.39	13.96	82.5	70.00	89.00
pf_d1	Survived	64	155.48	54.31	145.0	108.25	200.25
fio2_storm	Survived	62	0.53	0.19	0.5	0.40	0.60
lympho_1	Survived	60	1060.83	533.21	1005.0	643.75	1444.75
steroid_total	Survived	64	1460.78	726.78	1700.0	832.50	1900.00
steroid_duration	Survived	64	7.77	3.68	7.0	5.00	10.00
steroid_first_day	Survived	64	415.78	242.40	500.0	123.75	500.00
age	Died	50	63.32	12.04	64.0	58.25	69.75
saturation_admision	Died	50	72.94	14.66	74.5	60.00	84.75
pf_d1	Died	50	131.68	53.26	125.0	91.00	157.50
fio2_storm	Died	50	0.57	0.21	0.5	0.40	0.70
lympho_1	Died	40	761.65	418.60	657.0	495.00	921.00
steroid_total	Died	50	1459.80	708.98	1585.0	906.25	1892.50
steroid_duration	Died	50	7.42	4.23	7.0	5.00	10.00
steroid_first_day	Died	50	325.10	265.23	312.5	90.00	500.00

Categorical Variables

The table below summarizes the categorical variables describing their frequency and proportion. The summaries are grouped over the outcome (Survival and Death).

Characteristic	Survived, N = 64	Died, N = 50
Age Category (years)		
60 Years	47 (73%)	19 (38%)
>60 Years	17 (27%)	31 (62%)
Gender		
Female	18 (28%)	7 (14%)
Male	46 (72%)	43 (86%)
Diabetes		
Diabetic	32 (50%)	27 (54%)
Non-Diabetic	32 (50%)	23 (46%)
Hypertension		
Hypertensive	30 (47%)	26 (52%)
Non-Hypertensive	34 (53%)	24 (48%)
ARDS Category		
1	10 (16%)	3 (6.0%)
2	36 (56%)	29 (58%)
3	18 (28%)	18 (36%)
Infection Status		
Infected	7 (11%)	16 (32%)
Sterile	57 (89%)	34 (68%)
Comorbidity		
Present	0 (0%)	0 (0%)
Absent	64 (100%)	50 (100%)
Multimorbidity		
Present	0 (0%)	0 (0%)
Absent	64 (100%)	50 (100%)
Steroid Therapy		
Early High	45 (70%)	12 (24%)
Early Low	16 (25%)	19 (38%)
Late	3 (4.7%)	19 (38%)
Optimal Steroids		
Optimal	58 (91%)	13 (26%)
Suboptimal	6 (9.4%)	37 (74%)
Respiratory Support		
Recived	44 (70%)	43 (86%)
Not Received	19 (30%)	7 (14%)
Unknown	1	0

Association between categorical variables and outcome

Characteristic	Survived, N = 64 ¹	Died, N = 50 ¹	p-value ²
Gender			0.11
Female	18 (28%)	7 (14%)	
Male	46 (72%)	43 (86%)	
Diabetes			0.8

Diabetic	32 (50%)	27 (54%)	
Non-Diabetic	32 (50%)	23 (46%)	
Hypertension			0.7
Hypertensive	30 (47%)	26 (52%)	
Non-Hypertensive	34 (53%)	24 (48%)	
Coronary Artery Disease			>0.9
Present	6 (9.4%)	4 (8.0%)	
Absent	58 (91%)	46 (92%)	
Chronic Kidney Disease			>0.9
Present	2 (3.1%)	2 (4.0%)	
Absent	62 (97%)	48 (96%)	
ARDS Category			0.2
1	10 (16%)	3 (6.0%)	
2	36 (56%)	29 (58%)	
3	18 (28%)	18 (36%)	
Infection Status			0.011
Infected	7 (11%)	16 (32%)	
Sterile	57 (89%)	34 (68%)	
Steroid Therapy			<0.001
Early High	45 (70%)	12 (24%)	
Early Low	16 (25%)	19 (38%)	
Late	3 (4.7%)	19 (38%)	
Optimal Steroids			<0.001
Optimal	58 (91%)	13 (26%)	
Suboptimal	6 (9.4%)	37 (74%)	
Age Category (years)			<0.001
60 Years	47 (73%)	19 (38%)	
>60 Years	17 (27%)	31 (62%)	
Respiratory Support			0.072
Recived	44 (70%)	43 (86%)	
Not Received	19 (30%)	7 (14%)	
Unknown	1	0	

¹Statistics presented: n (%)

²Statistical tests performed: chi-square test of independence; Fisher's exact test

Association between continuous variables and outcome

Characteristic	Survived, N = 64 ¹	Died, N = 50 ¹	p-value ²
Age (years)	53 (12)	63 (12)	<0.001
PFR at Day 1	155 (54)	132 (53)	0.021
O2 Saturation on Admission	79 (14)	73 (15)	0.019
Lymphocyte Count on Day 1	1061 (533)	762 (419)	0.002
Unknown	4	10	
Total dose of Steroids	1461 (727)	1460 (709)	>0.9
Duration of Steroids	7.8 (3.7)	7.4 (4.2)	0.6
Dose of Steroids on first day	416 (242)	325 (265)	0.063

¹Statistics presented: mean (SD)

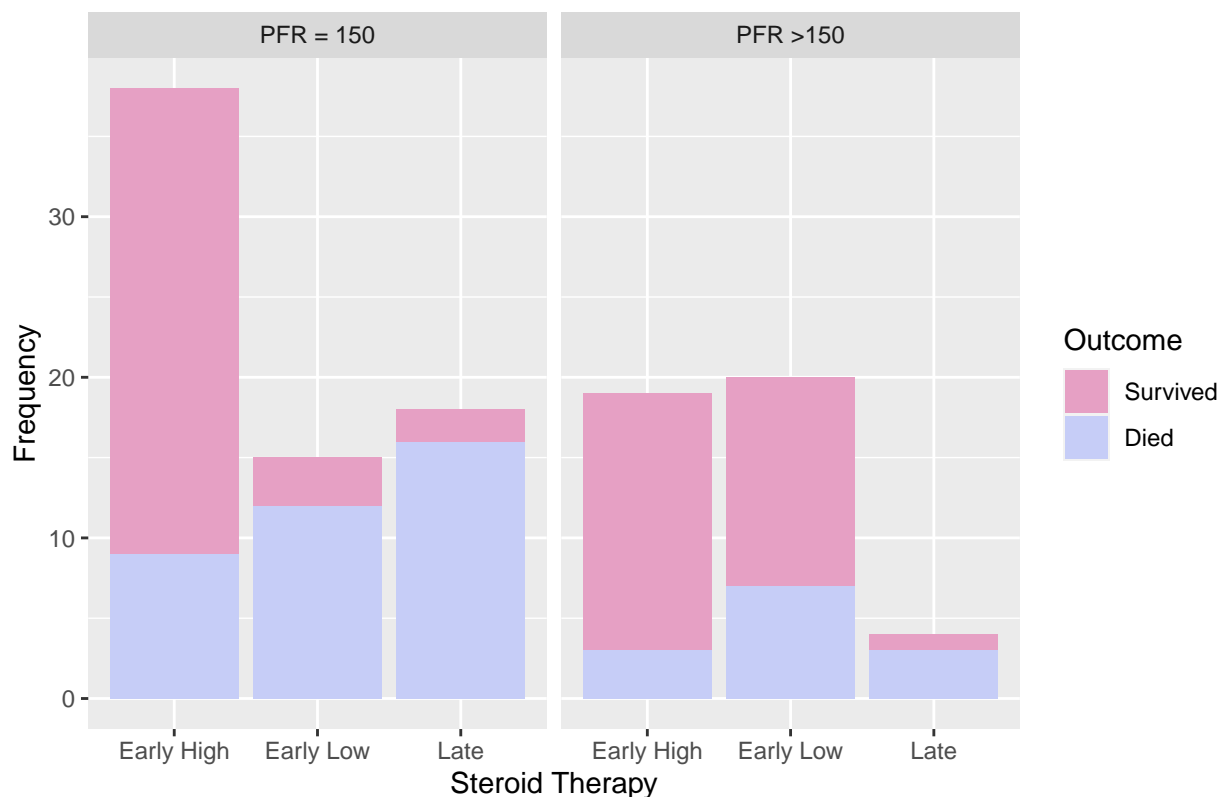
²Statistical tests performed: t-test

Association between Early Steroid Therapy, Severity and Outcome

PFR

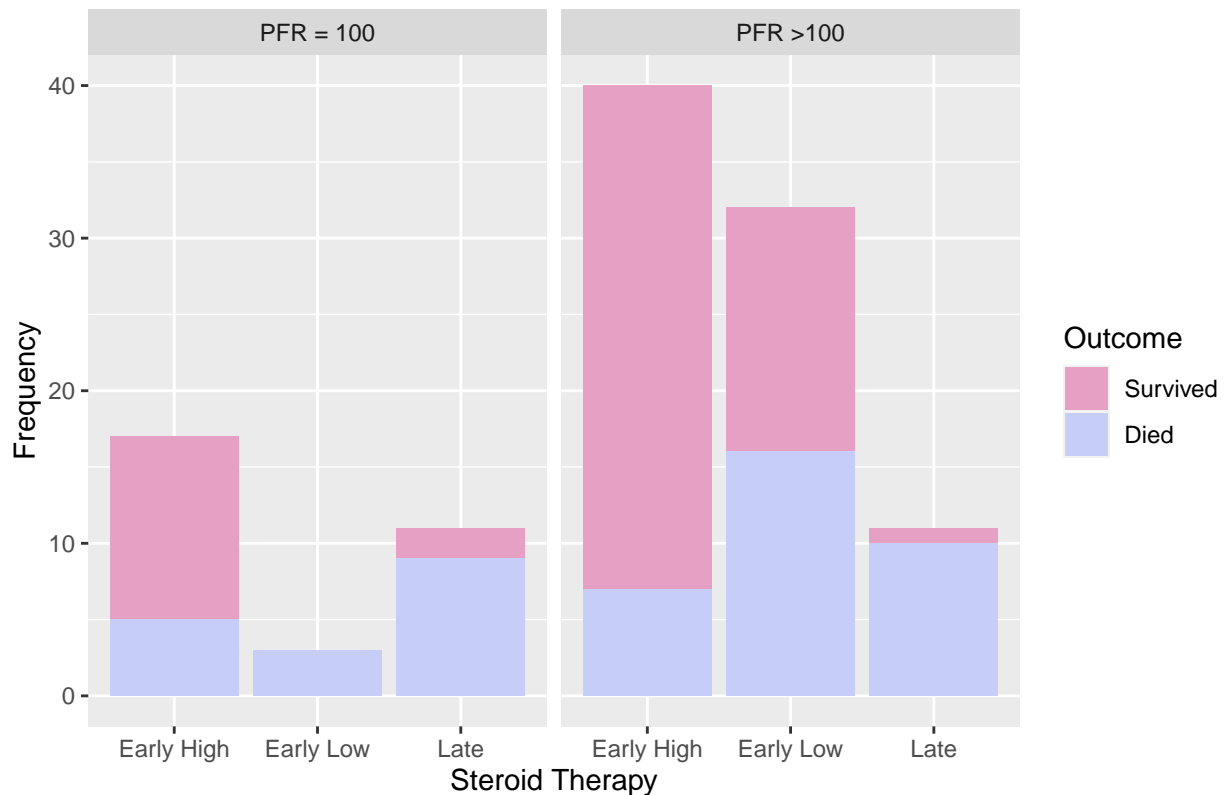
```
df %>%  
  group_by(outcome, steroid_cat, pfr_150) %>%  
  summarise(count = n()) %>% # summarise by each group  
  ggplot(aes(x = steroid_cat , y = count, fill = outcome)) +  
  geom_bar(stat = "identity") +  
  facet_grid( ~ pfr_150) +  
  labs(x = "Steroid Therapy", y = "Frequency", fill = "Outcome") +  
  scale_fill_manual(values = wes_palette(n = 2, name = "GrandBudapest2")) +  
  labs(title = "Time of start and dose of steroid therapy Vs Survival (Time/Dose)")
```

Time of start and dose of steroid therapy Vs Survival (Time/Dose)



```
df %>%  
  group_by(outcome, steroid_cat, pfr_100) %>%  
  summarise(count = n()) %>% # summarise by each group  
  ggplot(aes(x = steroid_cat , y = count, fill = outcome)) +  
  geom_bar(stat = "identity") +  
  facet_grid( ~ pfr_100) +  
  labs(x = "Steroid Therapy", y = "Frequency", fill = "Outcome") +  
  scale_fill_manual(values = wes_palette(n = 2, name = "GrandBudapest2")) +  
  labs(title = "Time of start and dose of steroid therapy Vs Survival (Time/Dose)")
```

Time of start and dose of steroid therapy Vs Survival (Time/Dose)



Cox Regression Analysis

Cox Proportional Hazards model was performed to demonstrate the independent effect of steroids on the outcome after adjusting for the effect of age. This was done by using the R packages `survival` and `survminer`. Likelihood Ratio Test and Wald Test statistics were performed for model diagnostics.

Data Preparation

As the `coxph()` function from the `survival` package needs the event to be passed as a dichotomised variable with 1 as event occurred and 0 as not, the following code prepares the data for further analysis. It also creates a new column (`date_outcome`) by combining `date_ward_shift` (favorable) and `date_death` (unfavourable) by removing NA values using `coalesce` function from the `dplyr` package. This is then used to create a time difference (in days) between the date of outcome and the date of cytokine storm. This variable is saved as `time` and appended to the `df.s` dataset.

```
library(survival) # for survival analysis
library(survminer) # for visualization
library(ggsci) # for setting the colour palette

# A new dataframe df.s object is created to perform the analyses
df.s<-as.data.frame(df)

# Event Variables
df.s$outcome <- ifelse(df.s$outcome == "Died", 1,0)
```

```

df.s$outcome<-as.numeric(df.s$outcome)
df.s$app_steroid <- factor(ifelse(df.s$app_steroid == "Suboptimal", "0", "1"))
df.s$age_60_above <- factor(ifelse(df.s$age>60, "1", "0"))
df.s$Infection_cat <-factor(ifelse(df.s$Infection_cat=="Sterile","0","1"))
df.s$pf_cat_adm <- factor(ifelse(df.s$pf_d1>150,"0","1"))

# Time Variable
### calculate the difference in time
###merge the date_ward and date_death into one new coloumn date_outcome

df.s <- df.s %>% mutate(date_outcome = coalesce(date_ward_shift, date_death))
df.s$time <- difftime(df.s$date_outcome, df.s$date_cytokine_storm,units="days")

```

Univariate Cox Model

Initially a univariate Cox regression analysis was performed for selecting the variables based on clinical and statistical significant. apply group of functions were used to perform univariate analysis across different covariates using the Surv() function from the survival package. The following code demonstrates the univariate Cox Regression analysis.

```

covariates <- c("age_60_above","Infection_cat", "pf_cat_adm","app_steroid")
univ_formulas <- sapply(covariates,
                       function(x) as.formula(paste('Surv(time, outcome)~', x)))

univ_models <- lapply( univ_formulas, function(x){coxph(x, data = df.s)})
# Extract data
univ_results <- lapply(univ_models,
                      function(x){
                        x <- summary(x)
                        p.value<-signif(x$wald["pvalue"], digits=2)
                        wald.test<-signif(x$wald["test"], digits=2)
                        beta<-signif(x$coef[1], digits=2);#coefficient beta
                        HR <-signif(x$coef[2], digits=2);#exp(beta)
                        HR.confint.lower <- signif(x$conf.int["lower .95"], 2)
                        HR.confint.upper <- signif(x$conf.int["upper .95"],2)
                        HR <- paste0(HR, " (",
                                     HR.confint.lower, "-", HR.confint.upper, ")")
                        res<-c(beta, HR, wald.test, p.value)
                        names(res)<-c("Beta Coefficient", "HR (95% CI for HR)", "Wald test",
                                     "p Value")
                        return(res)
                        #return(exp(cbind(coef(x),confint(x))))
                      })

res <- t(as.data.frame(univ_results, check.names = FALSE))
rownames(res) <- c("Age Over 60 years", "Infection Status",
                  "PFR Category","Optimal Steroids" )
kableExtra::kable(res)

```

	Beta Coefficient	HR (95% CI for HR)	Wald test	p Value
Age Over 60 years	0.88	2.4 (1.4-4.3)	9	0.0027
Infection Status	-0.65	0.52 (0.28-0.99)	4	0.046
PFR Category	-0.22	0.81 (0.42-1.5)	0.44	0.51
Optimal Steroids	-1	0.37 (0.19-0.69)	9.5	0.002

Multivariate Analysis

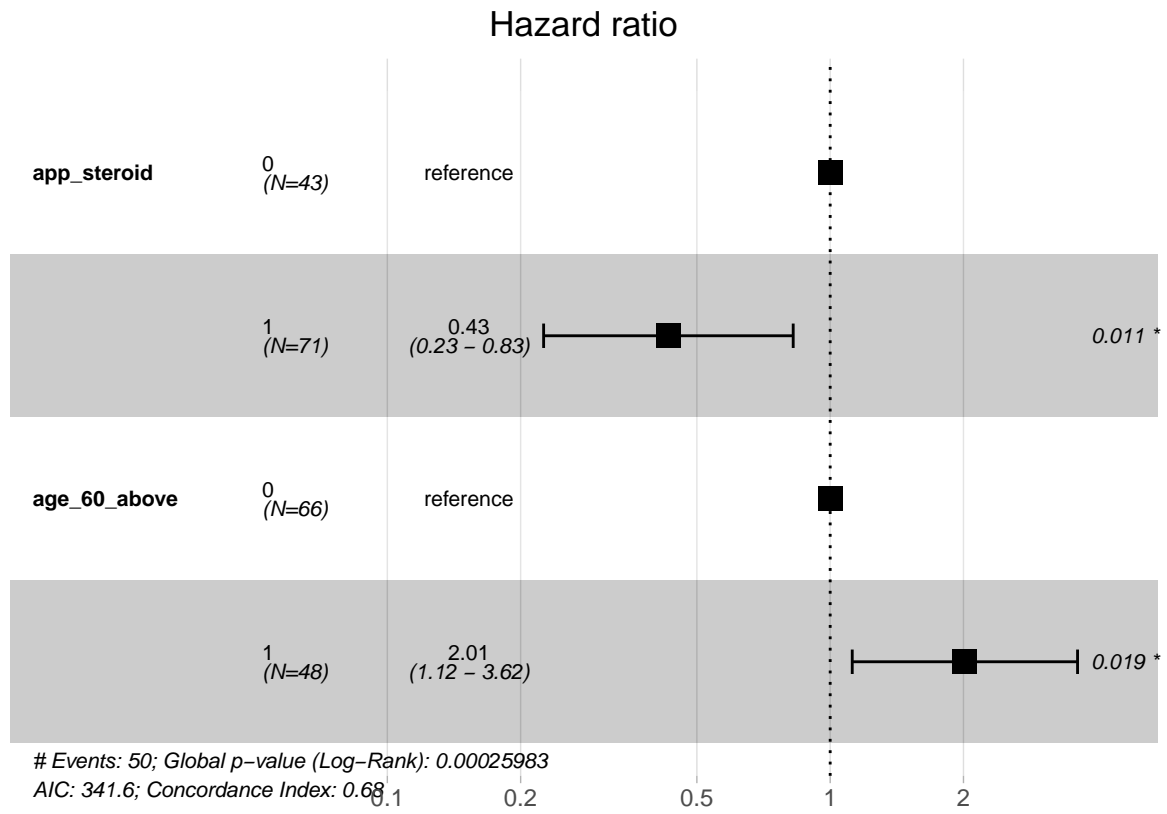
The univariate Cox Proportional Hazards model shows that **Age** and **Optimal Steroids** are statistically significant. So these variables were then included in the multivariate analysis. The following code reproduces the multivariate analysis and summarizes the results.

```
s.cox <- coxph(Surv(time, outcome) ~ app_steroid + age_60_above , data = df.s)
summary(s.cox)
```

```
## Call:
## coxph(formula = Surv(time, outcome) ~ app_steroid + age_60_above,
##       data = df.s)
##
##   n= 114, number of events= 50
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## app_steroid1 -0.8411   0.4313  0.3310 -2.541  0.0111 *
## age_60_above1  0.7005   2.0148  0.2989  2.344  0.0191 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## app_steroid1   0.4313     2.3188   0.2254   0.8251
## age_60_above1  2.0148     0.4963   1.1216   3.6194
##
## Concordance= 0.683 (se = 0.041 )
## Likelihood ratio test= 16.51 on 2 df,  p=3e-04
## Wald test               = 14.96 on 2 df,  p=6e-04
## Score (logrank) test = 16.26 on 2 df,  p=3e-04
```

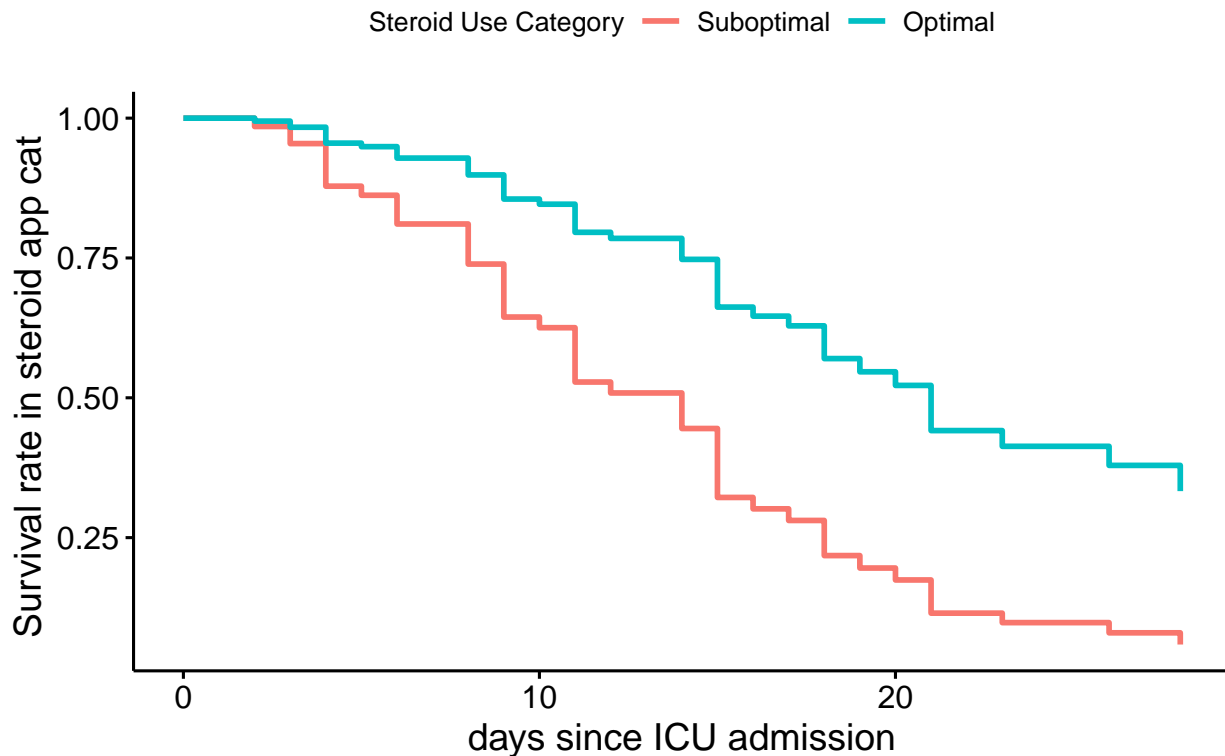
Visualization of the Cox Proportional Model

Forest Plot



Adjusted Survival Times

Survival rate



Model Diagnostics

The model diagnostics for the Cox Proportional Hazards Model can be done in three different ways:

- Testing the proportional hazards assumption.
- Examining influential observations (or outliers).
- Detecting nonlinearity in relationship between the log hazard and the covariates.

In order to check these model assumptions, **Residuals** method are used. The common residuals for the Cox model include: - **Schoenfeld** residuals to check the proportional hazards assumption - **Martingale** residual to assess nonlinearity - Deviance residual (symmetric transformation of the Martingale residuals), to examine influential observations

Testing the Proportional Hazards Assumption:

```
##           rho chisq      p
## app_steroid1 -0.129 0.832 0.3616
## age_60_above1 -0.281 3.866 0.0493
## GLOBAL           NA 4.144 0.1259
```

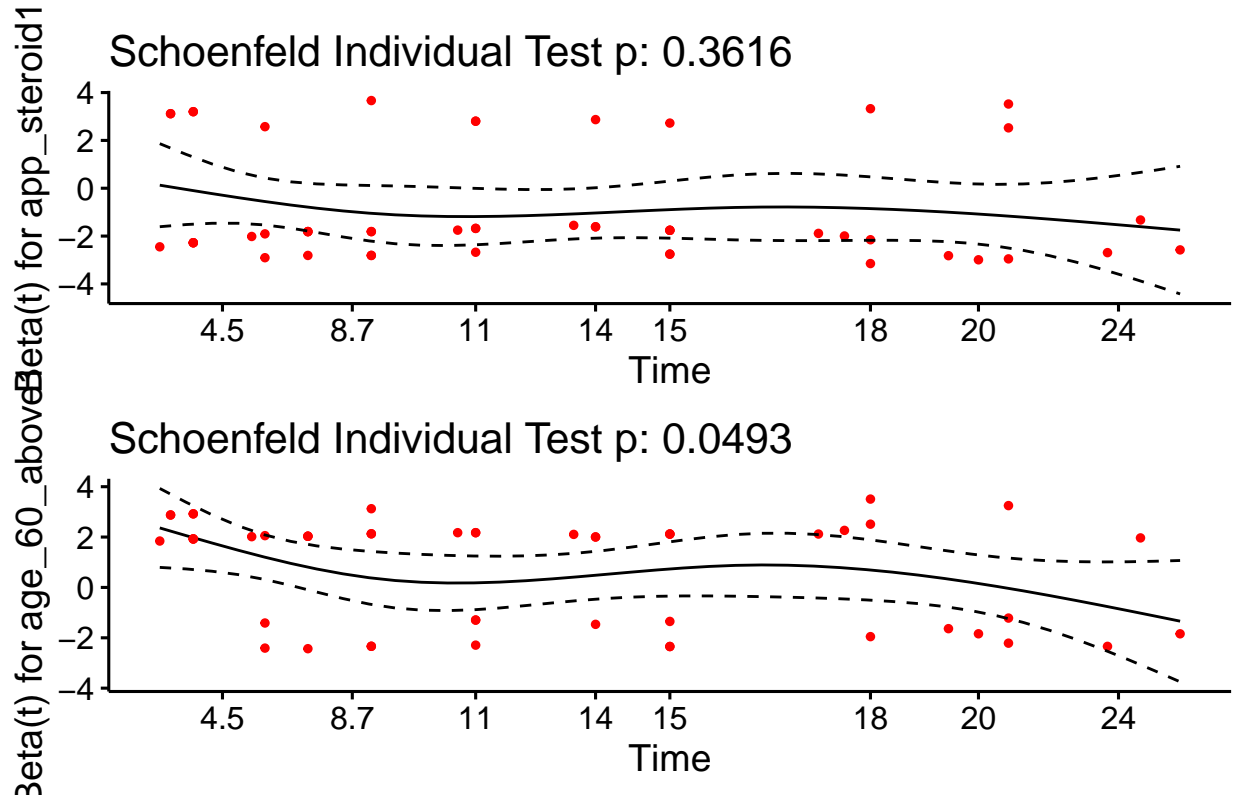
From the output above, the test is not statistically significant for each of the covariates, and the global test is also not statistically significant. Therefore, we can assume the proportional hazards.

The `ggcoxzph()` function from the `survminer` package graphically tests the Proportional Hazards Assumption by graphing the scaled Schoenfeld residuals against the transformed time. The graph below can be interpreted as the solid line is a smoothing spline fit to the plot, with the dashed lines representing a ± 2

standard-error band around the fit. As there are no systematic departures from a horizontal line, it is indicative that the assumption of proportional hazards is supported for the covariates age over 60 years and optimal steroid use.

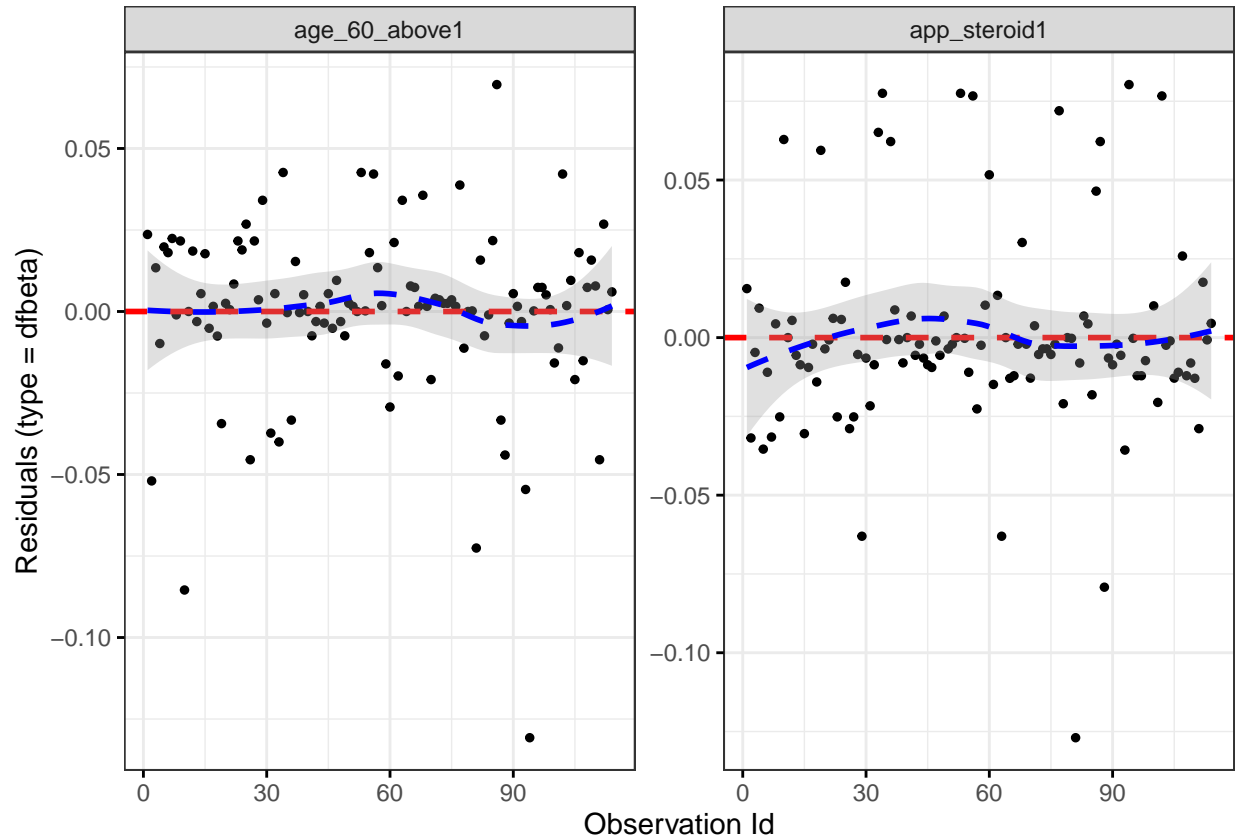
```
ggcoxzph(test.ph) ###graph shows a random departure from horizontal line.
```

Global Schoenfeld Test p: 0.1259



The function `ggcoxdiagnostics()` from the `survminer` package was used to check for influential observations.

```
ggcoxdiagnostics(s.cox, type = "dfbeta",  
                 linear.predictions = FALSE, ggtheme = theme_bw())
```



Comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is excessively influential individually, even though some of the dfbeta values for age and optimal steroid are large compared with the others.

Testing for non-linearity was not done as the variables were categorical in nature.

Additional Analyses

Exploration of FiO2

Ridge Plots

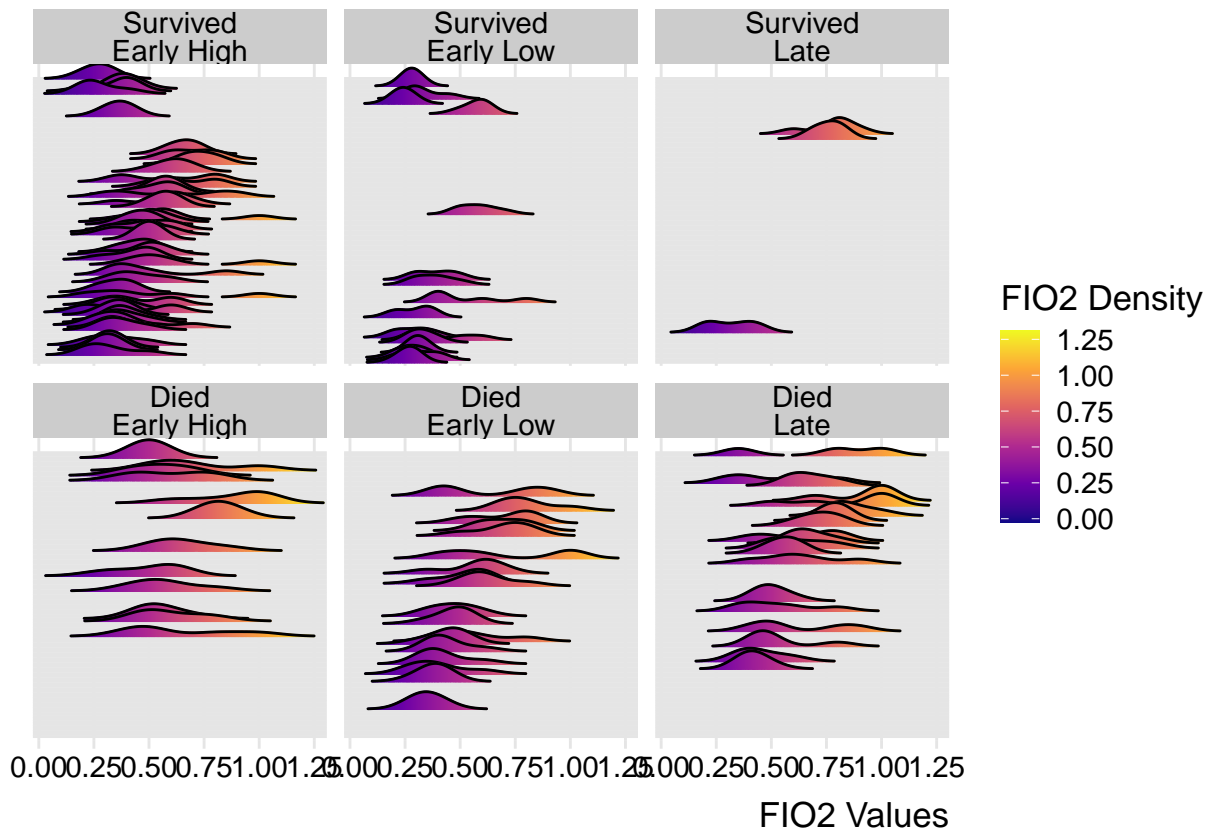
The code below creates a new column (`date_outcome`) by combining `date_ward_shift` (favorable) and `date_death` (unfavourable) by removing NA values using `coalesce` function from the `dplyr` package and plot a ridgeline plot faceted by the outcome. It is clear that the distribution of the FiO2 values vary with steroid therapy and outcome.

```
df %>% select (name, age, sex, dm, htn, steroid_cat, outcome, ards_cat, Infection_cat,
              date_symptom, date_cytokine_storm, date_steroid, date_ward_shift, date_discharge,
              date_death, fio2_storm, fio2_d2,fio2_d3,fio2_d4,fio2_d6) %>%
mutate(date_outcome = coalesce(date_ward_shift, date_death), ID = 1:dim(df)[1]) %>%
melt(., measure.vars =16:20 ,variable.name="fio2", value.name="fio2_value_trends") %>%
mutate(ID_reorder = fct_reorder(factor(ID), fio2_value_trends)) %>%
filter(fio2_value_trends < 4) %>%
ggplot() +
```

```

aes(x = fio2_value_trends, y = ID_reorder, fill = stat(x)) +
geom_density_ridges_gradient(scale = 7, rel_min_height = 0.01, gradient_lwd = 1.) +
scale_x_continuous(expand = c(0, 0)) +
scale_y_discrete(expand = expansion(mult = 0.01, add = 0.25))+
scale_fill_viridis_c(name = "FIO2 Density", option = "C") +
theme_ridges(font_size = 13, grid = TRUE) +
theme(axis.title.y = element_blank(),
      axis.ticks.y = element_blank(),
      axis.text.y = element_blank())+
xlab("FIO2 Values")+
facet_wrap(~outcome + steroid_cat, nrow = 2)

```



Boxplots

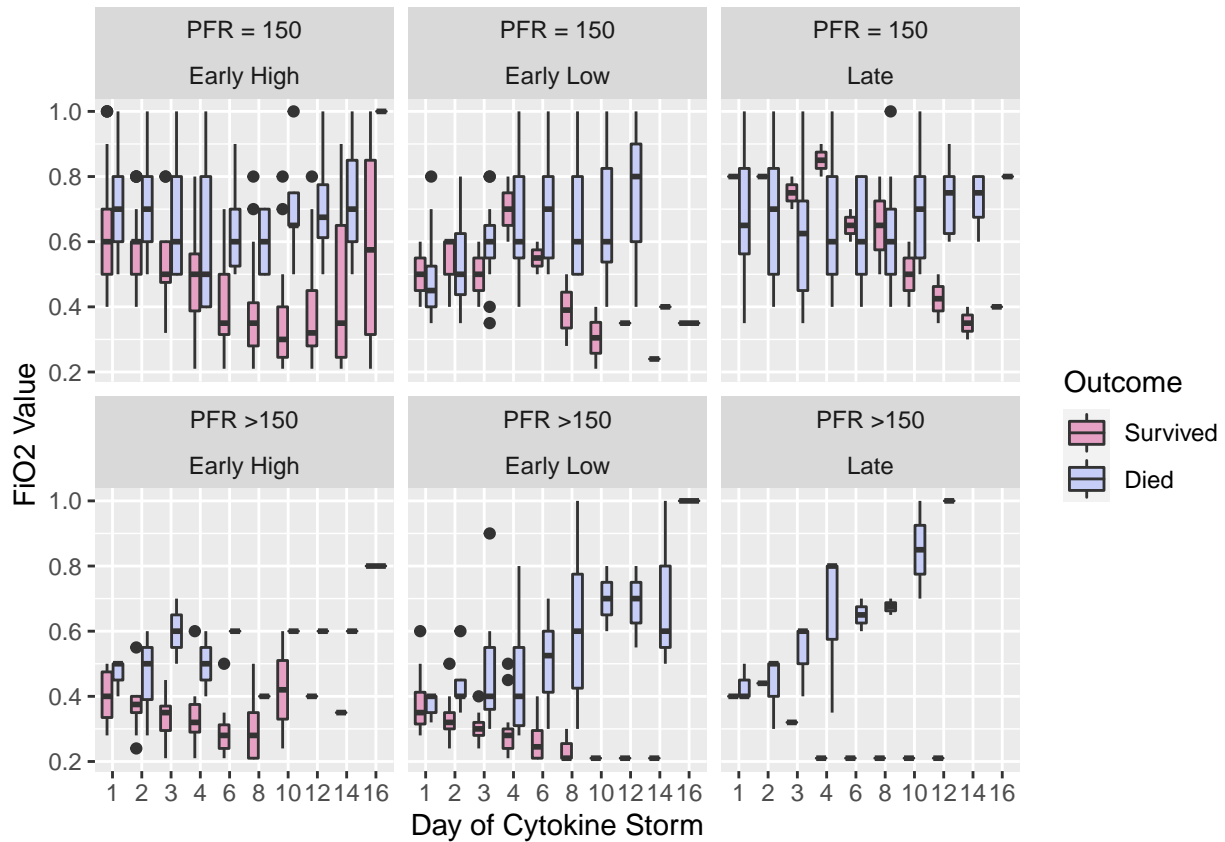
The boxplot below shows the median FiO2 of multiple days after the cytokine storm across the study population faceted by both the steroid therapy and severity. The boxplot has been grouped by the outcome of the study participants.

```

df %>% melt(, measure.vars =32:41 ,variable.name="fio2", value.name="fio2_value_trends") %>%
  filter(fio2_value_trends < 4) %>%
  ggplot() +
  aes(x = fio2, y = fio2_value_trends, fill = factor(outcome)) +
  geom_boxplot()+
  labs(x ="Day of Cytokine Storm", y = "FiO2 Value", fill = "Outcome") +
  scale_fill_manual(values = wes_palette(n = 2, name = "GrandBudapest2"))+

```

```
facet_wrap(pfr_150~steroid_cat, nrow = 2) +
scale_x_discrete(labels = c("1", "2", "3", "4", "6", "8", "10", "12", "14", "16"))
```

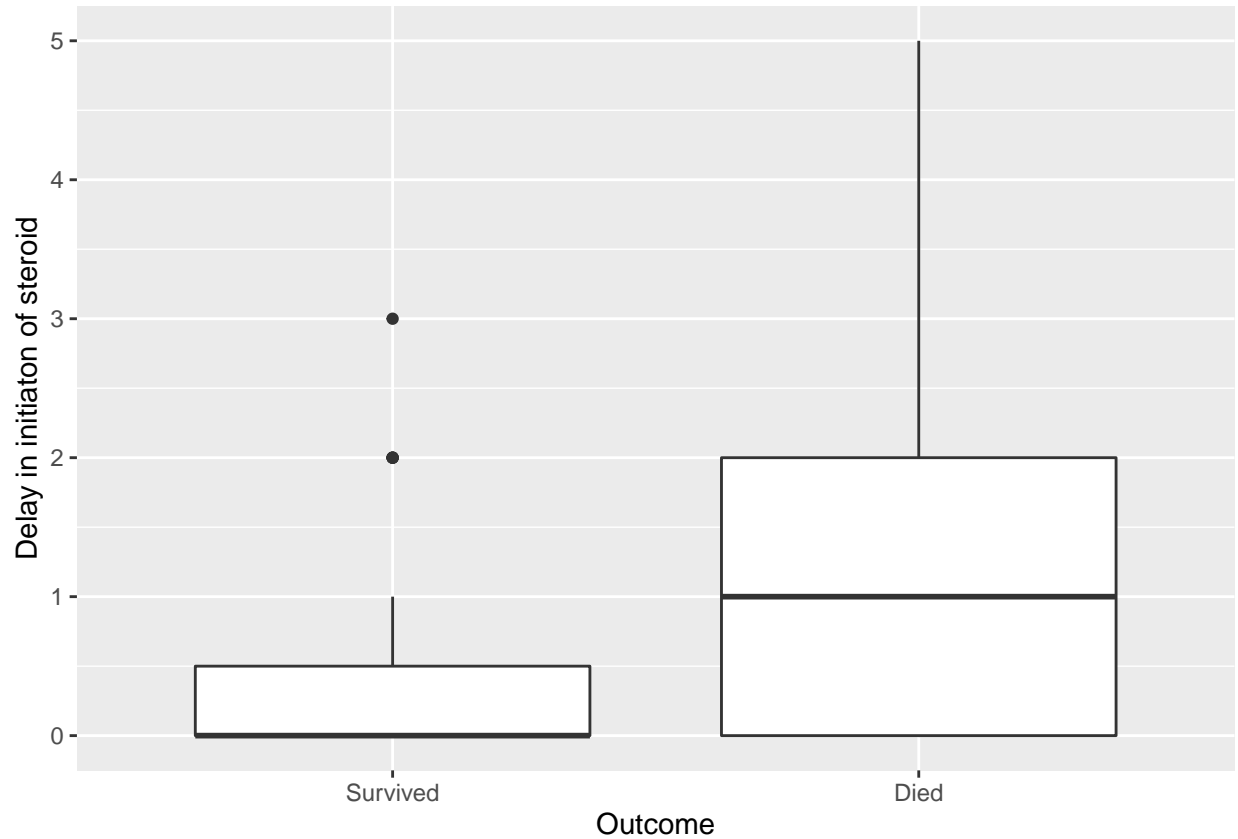


Initiation of steroid therapy after cytokine storm by outcome

The following plot shows the time difference in initiation of steroids and cytokine storm in both the outcome groups.

```
df <- df[-17,]
df$date_end <- as.Date(ifelse(is.na(df$date_death), as.character(df$date_discharge),
                             as.character(df$date_death)))

df %>%
ggplot() +
  aes(x = outcome, y = as.Date(df$date_steroid) - as.Date(df$date_cytokine_storm)) +
  geom_boxplot()+
  labs(x = "Outcome", y = "Delay in initiation of steroid")
```



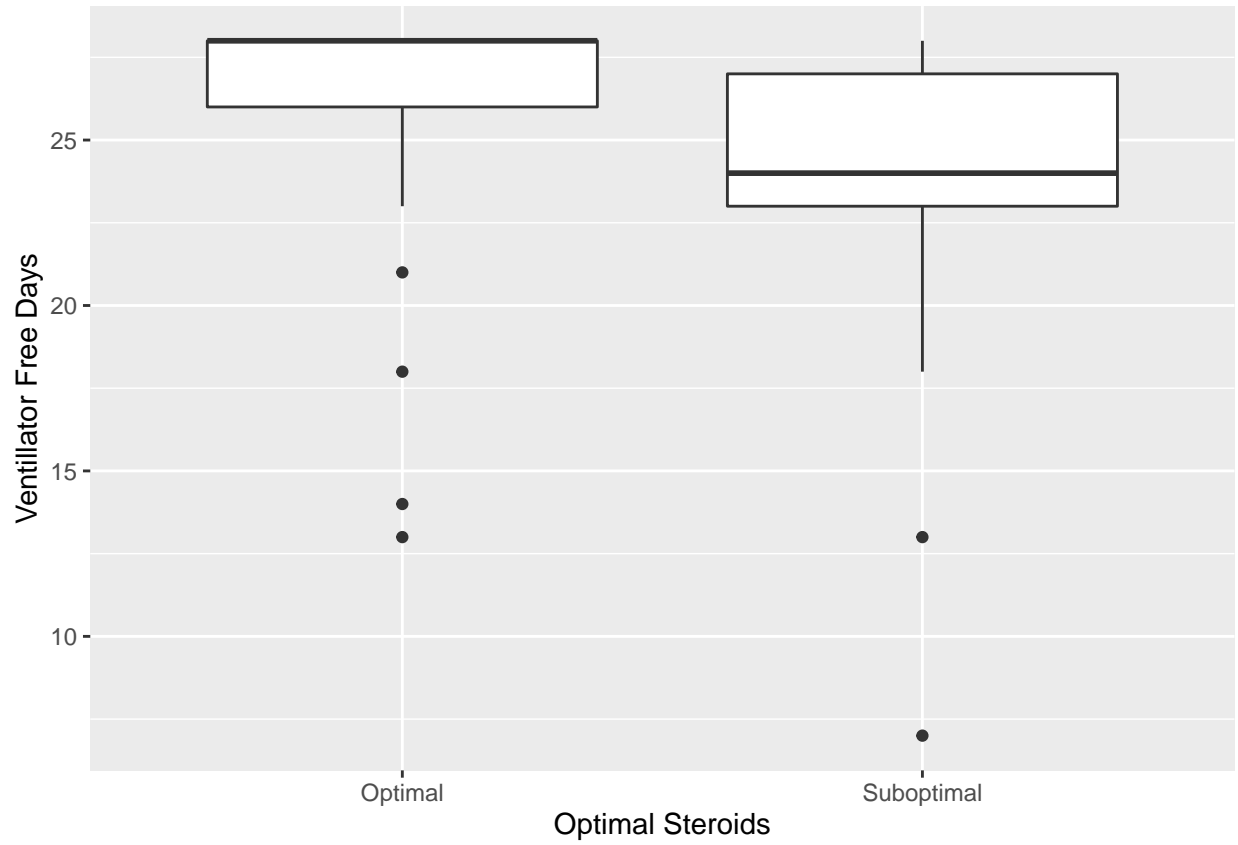
Characteristic	Survived, N = 63 ¹	Died, N = 50 ¹
Initiation of Steroids Since day of Cytokine Storm	0.00 (0.00, 0.50)	1.00 (0.00, 2.00)

¹Statistics presented: median (IQR)

Ventillator Free Days

The following plot shows the number of ventillator free days among those who had a PFR < 150 on Day 1 by the use of steroids groups. The n here is 69 i.e. those who had PFR < 150 on Day 1.

```
df %>% filter(pf_d1 < 150) %>%
ggplot() +
  aes(x = app_steroid, y = 28 - mech_ventilation_duration) +
  geom_boxplot()+
  labs(x = "Optimal Steroids", y = "Ventillator Free Days")
```

Characteristic	Optimal, N = 38 ¹	Suboptimal, N = 31 ¹
Ventillator Free Days	28.0 (26.0, 28.0)	24.0 (23.0, 27.0)
Unknown	4	0

¹Statistics presented: median (IQR)