

# Supplementary Appendix: R-Code and Additonal Analyses

## Contents

<b>Introduction</b>	<b>1</b>
<b>Data Preparation</b>	<b>1</b>
Loading Packages . . . . .	1
Loading Dataset . . . . .	2
Recoding the variables . . . . .	2
<b>Descriptive Analysis</b>	<b>3</b>
Quantitative Variables . . . . .	3
Categorical Variables . . . . .	3
<b>Associaton between steroid therapy and outcome</b>	<b>4</b>
Age over 60 years, Steroid Therapy and Outcome . . . . .	4
Presence of Infection, Steroid Therapy and Outcome . . . . .	5
Severity, Steroid Therapy and Outcome . . . . .	6
Diabetes, Steroid Therapy and Outcome . . . . .	7
Composite Plot . . . . .	8
Association between Early Steroid Therapy, Severity and Outcome . . . . .	9
<b>Exact Logistic Regression</b>	<b>10</b>
<b>Additional Analyses</b>	<b>10</b>
Exploration of FiO2 . . . . .	10
Inititaion of steroid therapy and outcome . . . . .	13

## 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 statsitlcal 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(gtsummary)
library(ggplot2)
library(psych)
```

```

library(gridExtra)
library(tidyverse)
library(plotly)
library(elrm)
library(ggribes)
library(ggpubr)
library(reshape2)
library(lubridate)
library(wesanderson)
library(magrittr)

```

## 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("Excel Steroids 24.07.20 (2).xlsx")

```

## 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, "Appropriate", "Inappropriate"))
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$Infection_cat == "sterile", "Sterile", "Infected")
df$Infection_cat <- factor(df$Infection_cat)
df$sex <- ifelse(df$sex == "F", "Female", "Male")
df$sex <- factor(df$sex)
df$dm <- factor(df$dm)
df$htn <- ifelse(df$htn == 1, "Hypertensive", "Non-Hypertensive")
df$htn <- factor(df$htn)
df$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$ards_cat <- as.factor(df$ards_cat)
df$early_high <- factor(ifelse(df$steroid_cat == "Early High", "Early High", "Others"))
df$pfr <- ifelse(df$pf_d1 > 150, "PFR >150", "PFR \u2264 150")
```

## 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	26	50.54	15.09	50.00	36.75	60.00
saturation_admission	Survived	25	80.88	14.51	85.00	76.00	89.00
pf_d1	Survived	26	159.46	53.53	154.50	117.25	197.75
fio2_storm	Survived	25	0.48	0.18	0.40	0.39	0.50
lympho_1	Survived	24	976.54	463.70	848.00	587.50	1334.50
steroid_total	Survived	26	1250.38	961.92	1030.00	490.00	1780.00
steroid_duration	Survived	26	5.12	1.93	5.00	5.00	6.00
steroid_first_day	Survived	26	413.27	324.14	500.00	90.00	500.00
age	Died	19	65.47	12.84	66.00	61.00	69.50
saturation_admission	Died	19	73.74	14.42	80.00	65.00	85.00
pf_d1	Died	19	139.53	56.42	130.00	103.50	164.00
fio2_storm	Died	18	0.58	0.17	0.55	0.50	0.68
lympho_1	Died	17	753.88	379.64	635.00	531.00	924.00
steroid_total	Died	19	1322.89	908.86	1240.00	670.00	1900.00
steroid_duration	Died	19	5.05	3.84	5.00	3.50	5.50
steroid_first_day	Died	19	383.68	345.57	500.00	60.00	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).

```
table2 <- df %>% select("age_60_above", "sex", "dm", "htn",
  "ards_cat", "Infection_cat", "comorbidity", "multimorbidity",
  "steroid_cat", "app_steroid", "outcome")
tab <- gtsummary::tbl_summary(table2, by = outcome, label = list(age_60_above ~ "Age (years)",
  sex ~ "Gender", dm ~ "Diabetes",
  htn ~ "Hypertension",
  ards_cat ~ "ARDS Category",
  Infection_cat ~ "Infection Status",
  comorbidity ~ "Comorbidity",
  multimorbidity ~ "Multimorbidity",
  steroid_cat ~ "Steriod Therapy",
```

```
as_kable(tab)
```

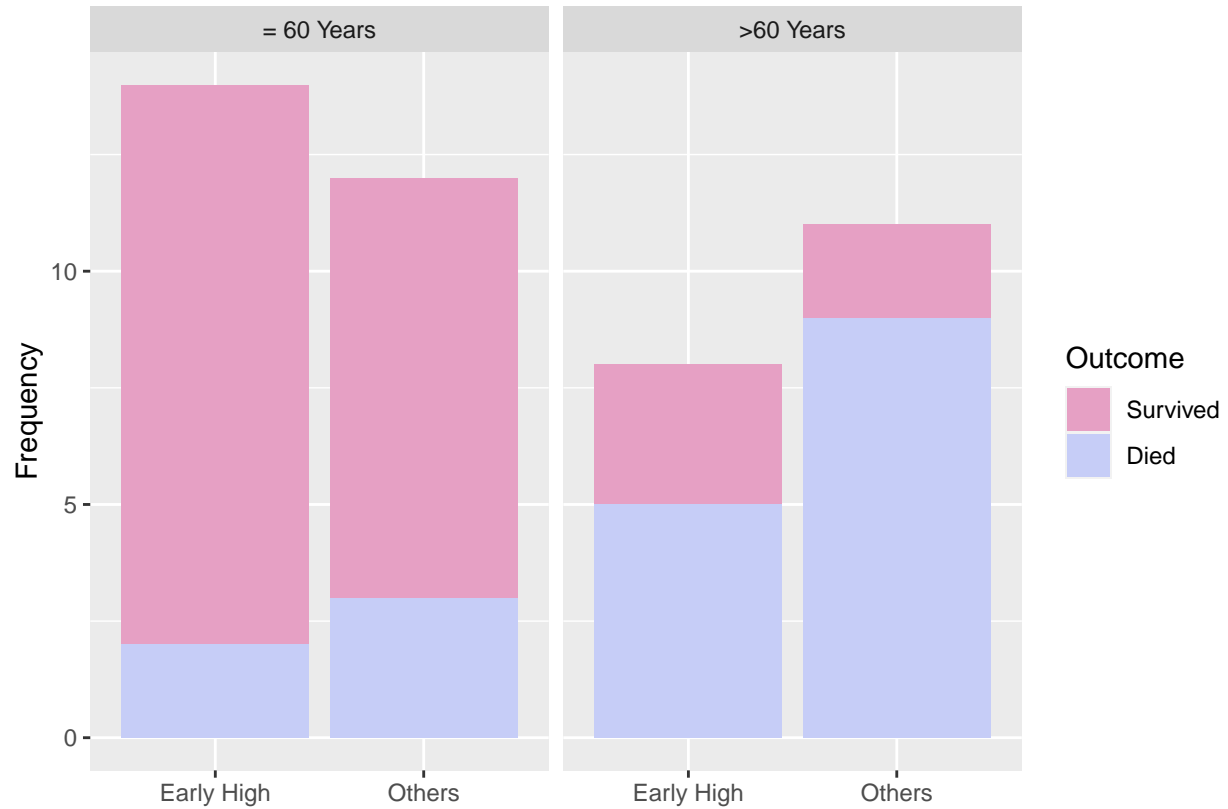
```
app_steroid ~ "Appropriatness of Ther  
outcome ~ "Outcome")
```

Characteristic	Survived, N = 26	Died, N = 19
Age (years)		
60 Years	21 (81%)	5 (26%)
>60 Years	5 (19%)	14 (74%)
Gender		
Female	9 (35%)	2 (11%)
Male	17 (65%)	17 (89%)
Diabetes		
Diabetic	8 (31%)	9 (47%)
NonDiabetic	18 (69%)	10 (53%)
Hypertension		
Hypertensive	11 (42%)	9 (47%)
Non-Hypertensive	15 (58%)	10 (53%)
ARDS Category		
1	6 (23%)	1 (5.3%)
2	15 (58%)	14 (74%)
3	5 (19%)	4 (21%)
Infection Status		
Infected	5 (19%)	8 (42%)
Sterile	21 (81%)	11 (58%)
Comorbidity		
Present	8 (31%)	9 (47%)
Absent	18 (69%)	10 (53%)
Multimorbidity		
Present	0 (0%)	0 (0%)
Absent	26 (100%)	19 (100%)
Steroid Therapy		
Early High	15 (58%)	7 (37%)
Early Low	8 (31%)	3 (16%)
Late	3 (12%)	9 (47%)
Appropriatness of Therapy		
Appropriate	22 (85%)	8 (42%)
Inappropriate	4 (15%)	11 (58%)

## Associaton between steroid therapy and outcome

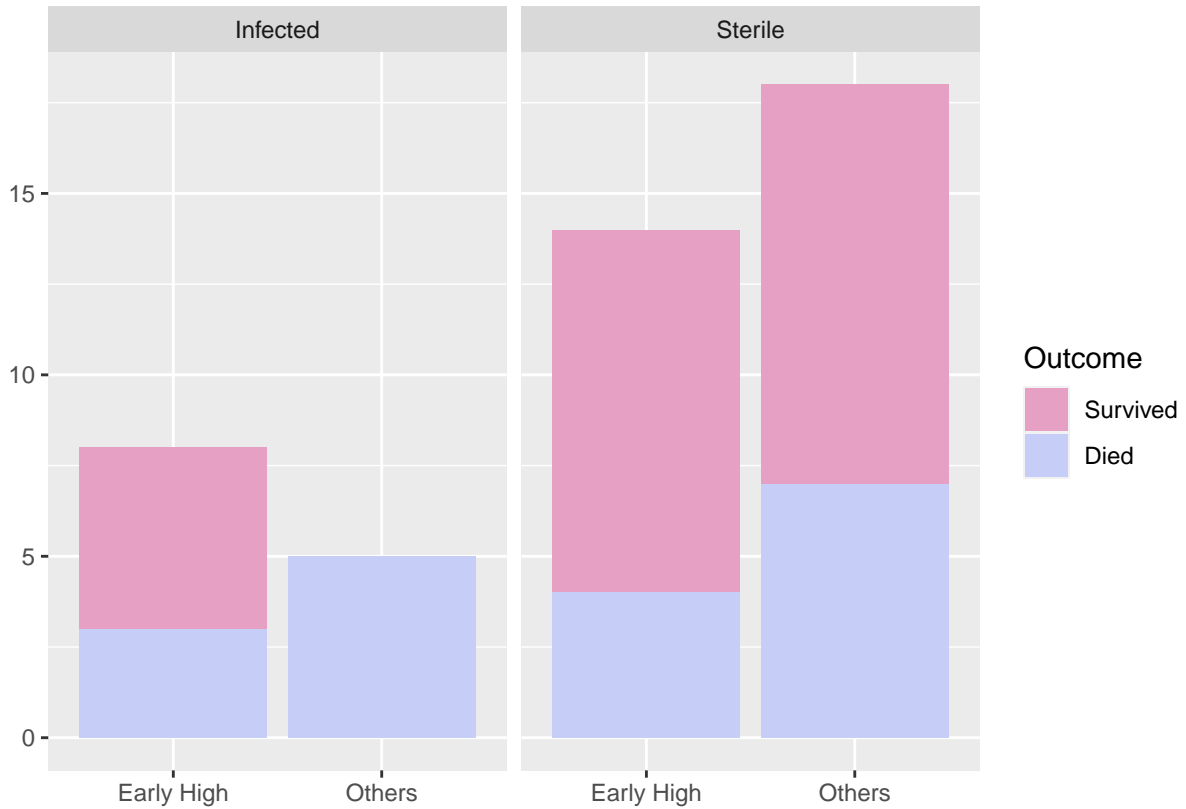
### Age over 60 years, Steroid Therapy and Outcome

The barplot below depicts the relationship between age, appropriateness of steroid therapy and outcome. The x-axis is represented by the outcome of the disease while the y-axis is respresented by the number of study participants in each group. The colour of the barplot is represented by the apropriateness of the steroid therapy (i.e. Early High Dose) while the facets represent the age group of the study participant.



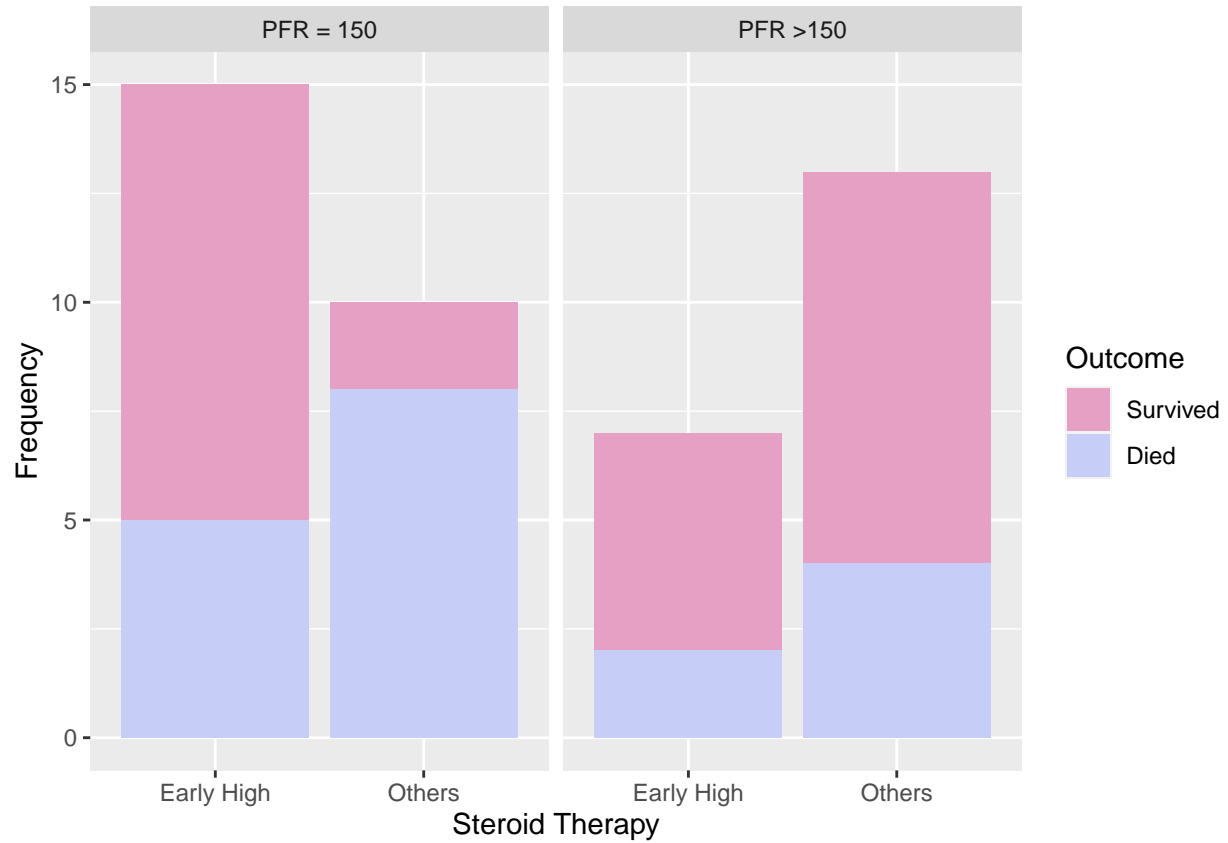
### Presence of Infection, Steroid Therapy and Outcome

The barplot below depicts the relationship between presence of infection, appropriateness of steroid therapy and outcome. The x-axis is represented by the outcome of the disease while the y-axis is represented by the number of study participants in each group. The colour of the barplot is represented by the appropriateness of the steroid therapy (i.e. Early High and Early Low) while the facets represent the status of infection of the study participant.



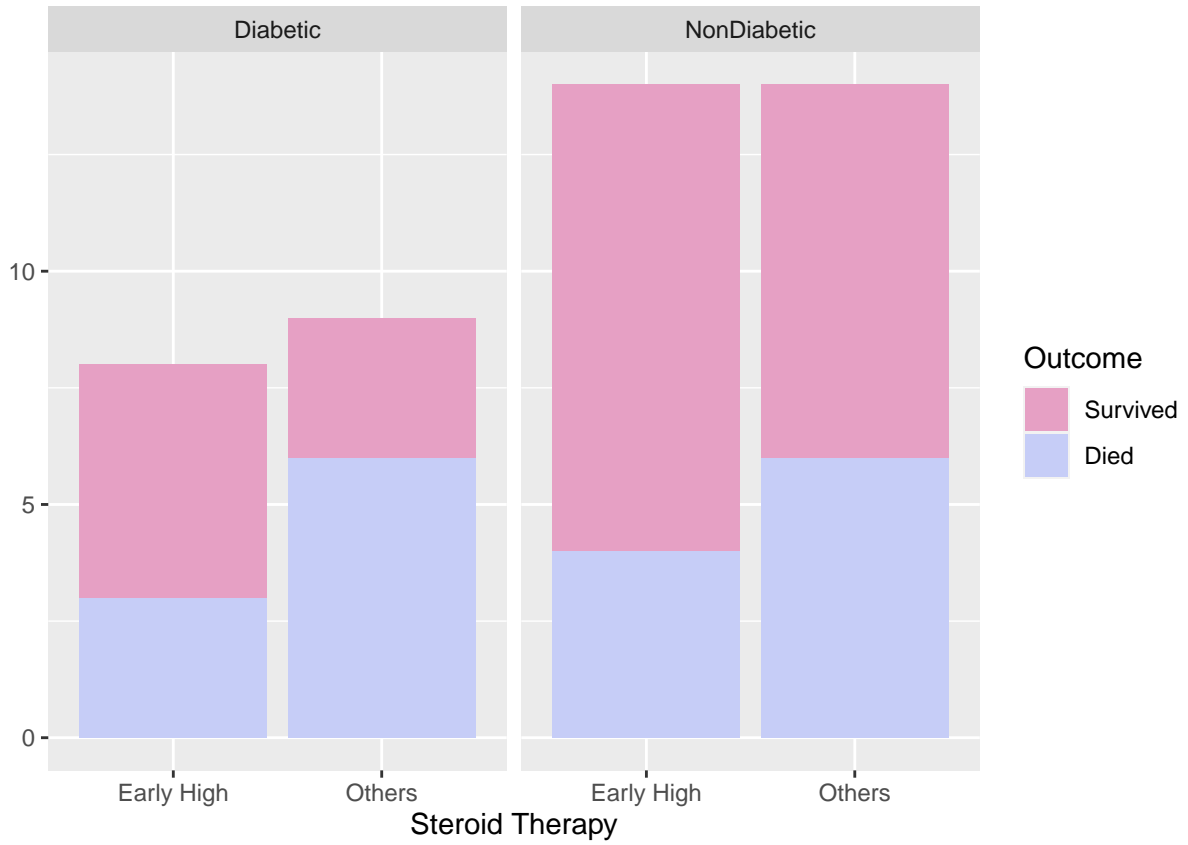
### Severity, Steroid Therapy and Outcome

The barplot below depicts the relationship between the PFR at day1  $\leq$   $\$150\text{mmHg}$  (severe) and  $\$>\$150\text{mmHg}$  (less severe), steroid therapy and outcome. The x-axis is represented by the outcome of the disease while the y-axis is represented by the number of study participants in each group. The colour of the barplot is represented by the appropriateness of the steroid therapy (i.e. Early High and Early Low) while the facets represent the status of infection of the study participant.



### Diabetes, Steroid Therapy and Outcome

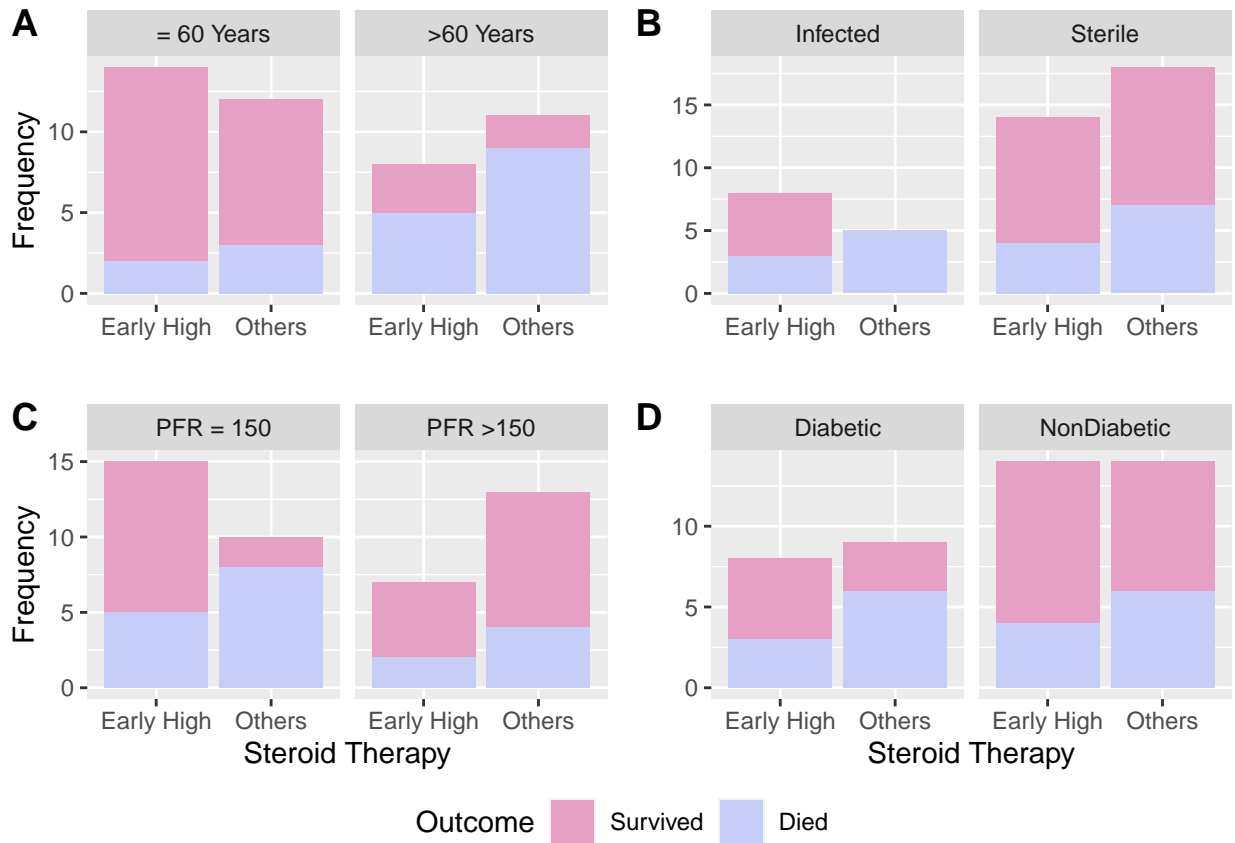
The barplot below depicts the relationship between the presence of diabetes, steroid therapy and outcome. The x-axis is represented by the outcome of the disease while the y-axis is represented by the number of study participants in each group. The colour of the barplot is represented by the appropriateness of the steroid therapy (i.e. Early High and Early Low) while the facets represent the status of diabetes of the study participant.



### Composite Plot

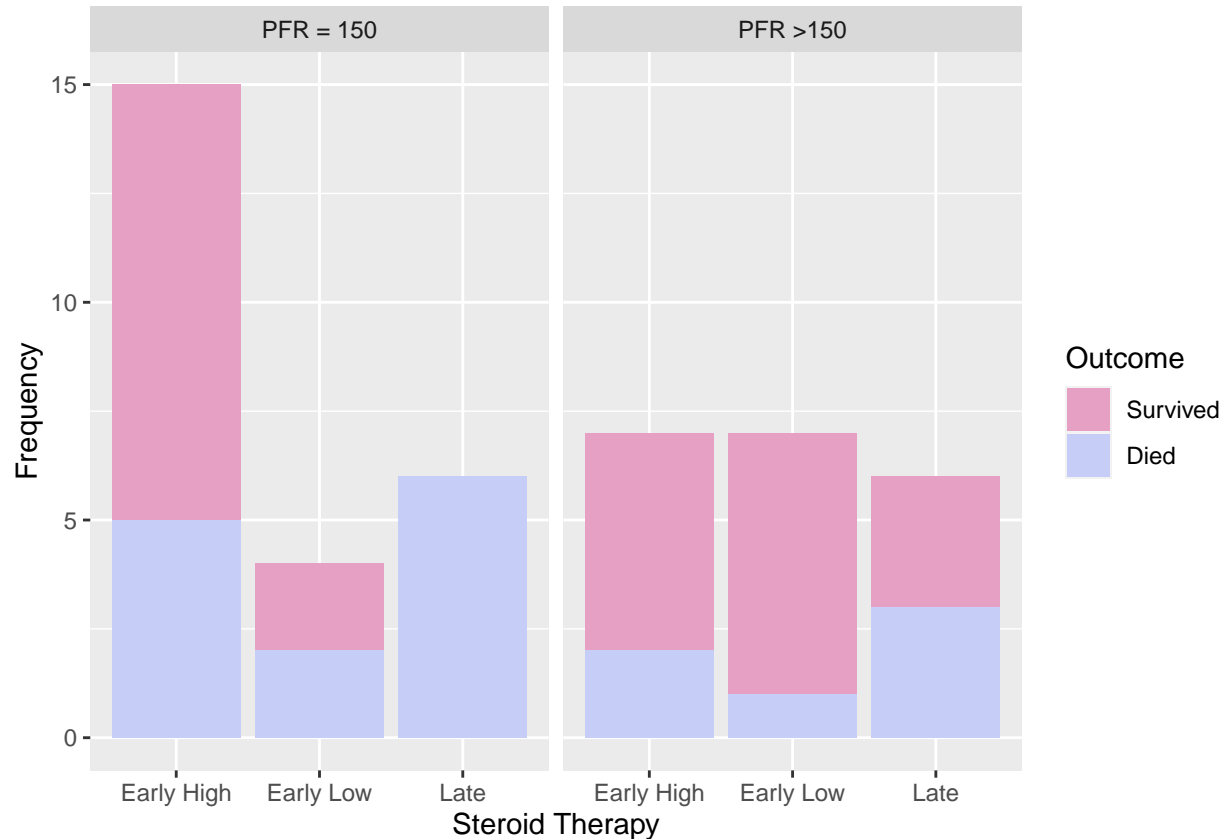
```
ggarrange(a, b, c, d, labels = c("A", "B", "C", "D"),
  common.legend = TRUE, legend = "bottom", nrow=2, ncol=2)
```





### Association between Early Steroid Therapy, Severity and Outcome

```
df %>%
  group_by(outcome, steroid_cat, pfr) %>%
  summarise(count = n()) %>% # summarise by each group
  ggplot(aes(x = steroid_cat, y = count, fill = outcome)) +
  geom_bar(stat = "identity") +
  facet_grid(~ pfr) +
  labs(x = "Steroid Therapy", y = "Frequency", fill = "Outcome") +
  scale_fill_manual(values = wes_palette(n = 2, name = "GrandBudapest2"))
```



## Exact Logistic Regression

Exact Logistic Regression was performed to demonstrate the independent effect of steroids on the outcome after adjusting for age and infection status of the study participants. This was done by using the R package `e1rm`. The `e1rm()` function in the package was used to perform the regression analysis. 50000 iterations and a burn-in of 5000 was set for the purpose of running the model. The `set.seed()` function was used for the purposes of reproducibility of the results. The summary of the model and the traceplot along with the histogram of sampled values were used for the diagnostic purposes of the model.

The following code summarises the model and plots its diagnostics. A trace plot and histogram of the sampled values of each sufficient statistic of interest is produced using the `plot.e1rm()` function and the sampled values within the burn-in period are also plotted.

## 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.

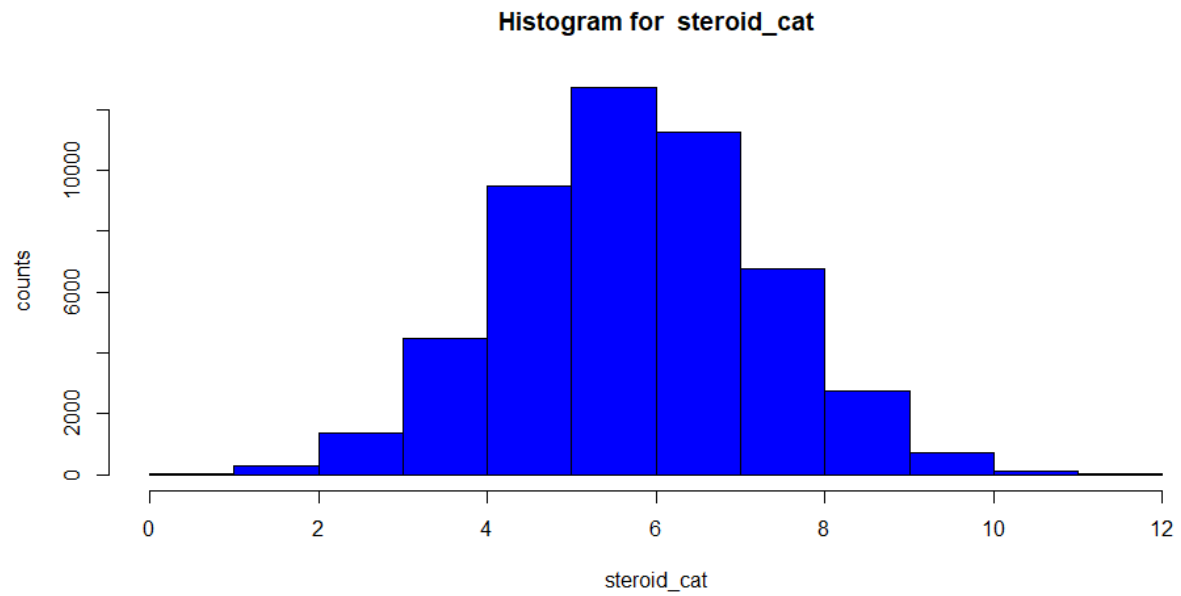
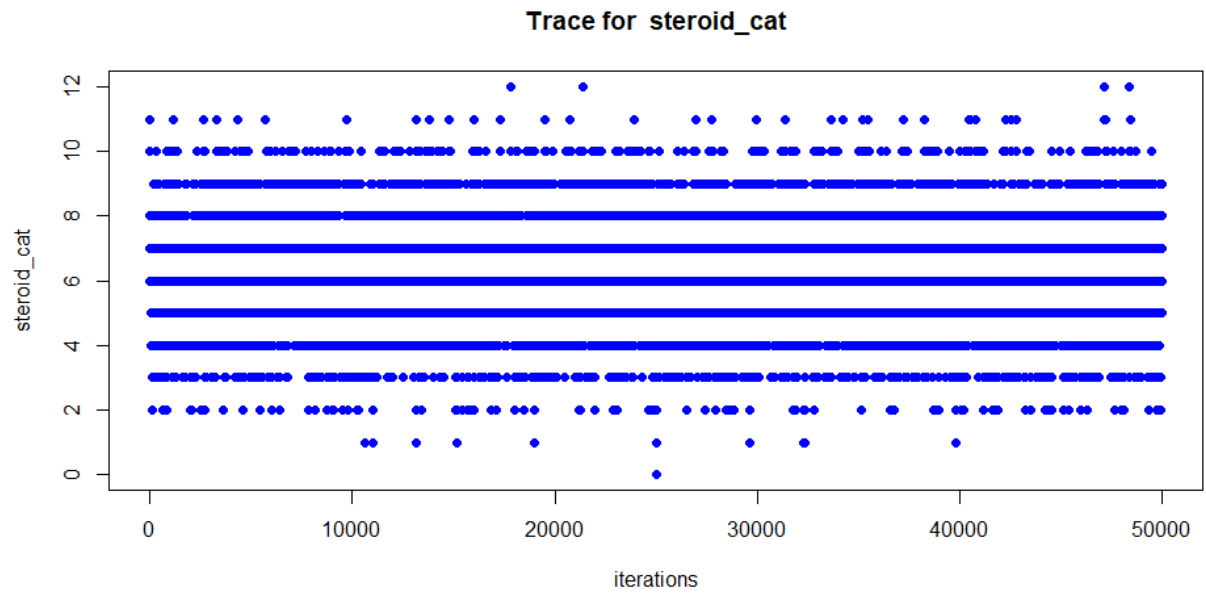
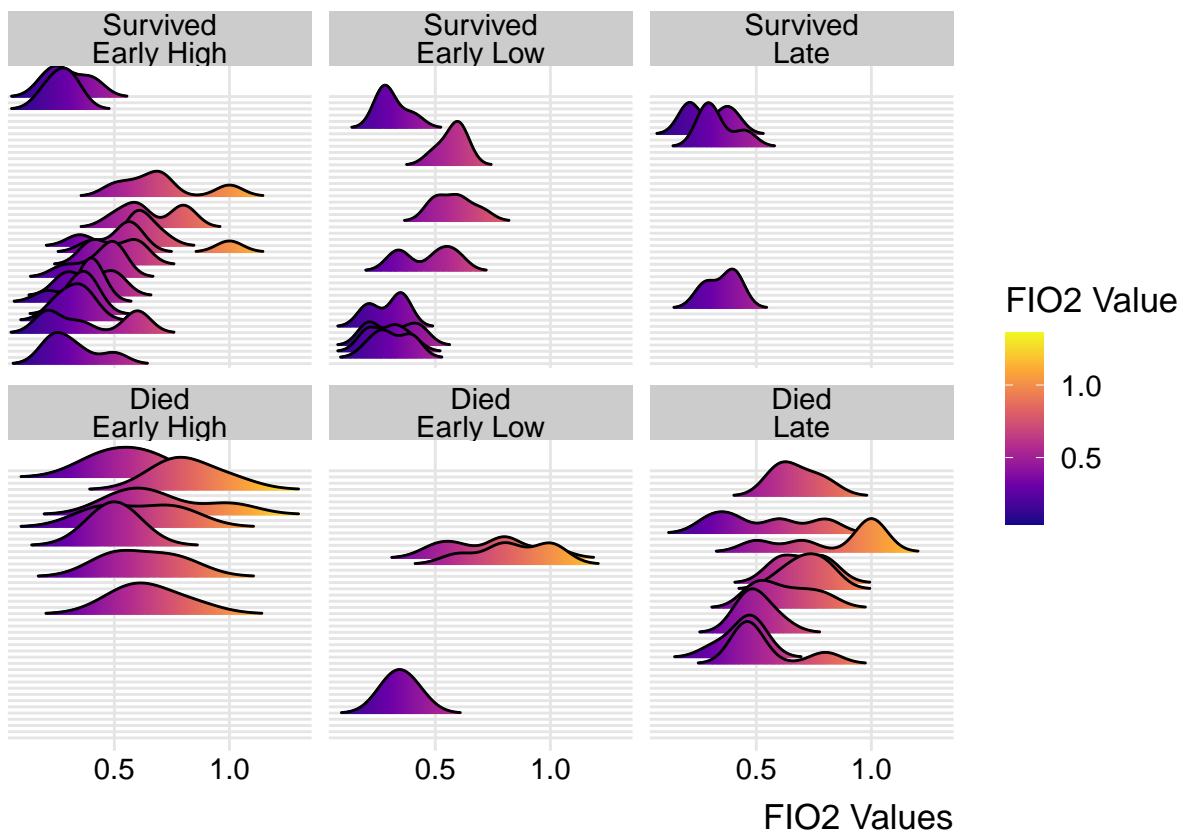


Figure 1: Trace Plot of the Exact Conditional Logistic Regression Model

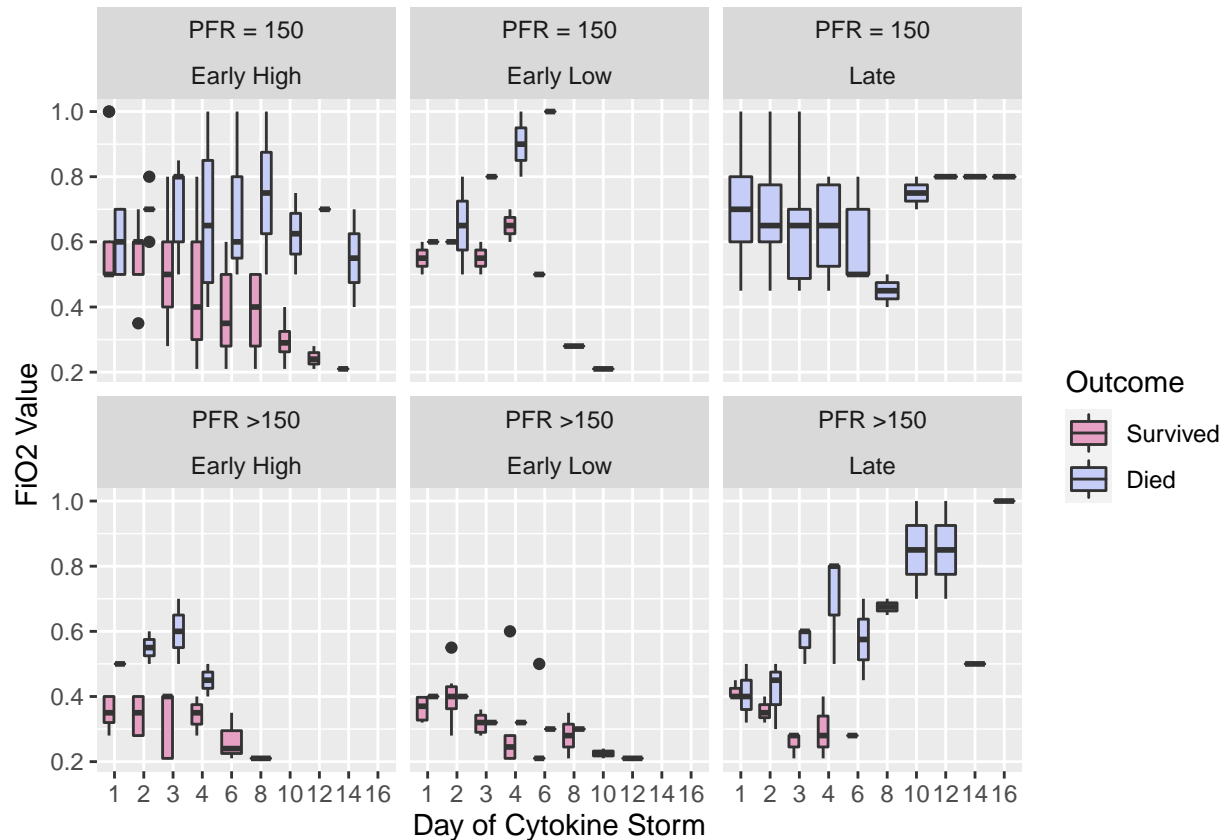
```
df %>% select (name, age, sex, dm, htn, steroid_cat, outcome, ards_cat, Infection_cat,
              date_symptom, date_cytokine_storm, date_steroid, date_ward_shift, date_discharge,
              date_death, fio2_storm, fio2_d2,fio2_d3,fio2_d4,fio2_d6) %>%
mutate(date_outcome = coalesce(date_ward_shift, 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 Value", 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~steroid_cat, nrow = 2) +
  scale_x_discrete(labels = c("1", "2", "3", "4", "6", "8", "10", "12", "14", "16"))
```



### Initiation of steroid therapy and outcome

```
df <- df[-17,]
df$date_end <- ifelse(is.na(df$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")
```

