

Volcano Plot for Adverse Event Analysis in R

Kamlesh Patel, AVP

Rang Technologies Inc, Piscataway, New Jersey

Kamlesh.patel1@rangtech.com

Agenda

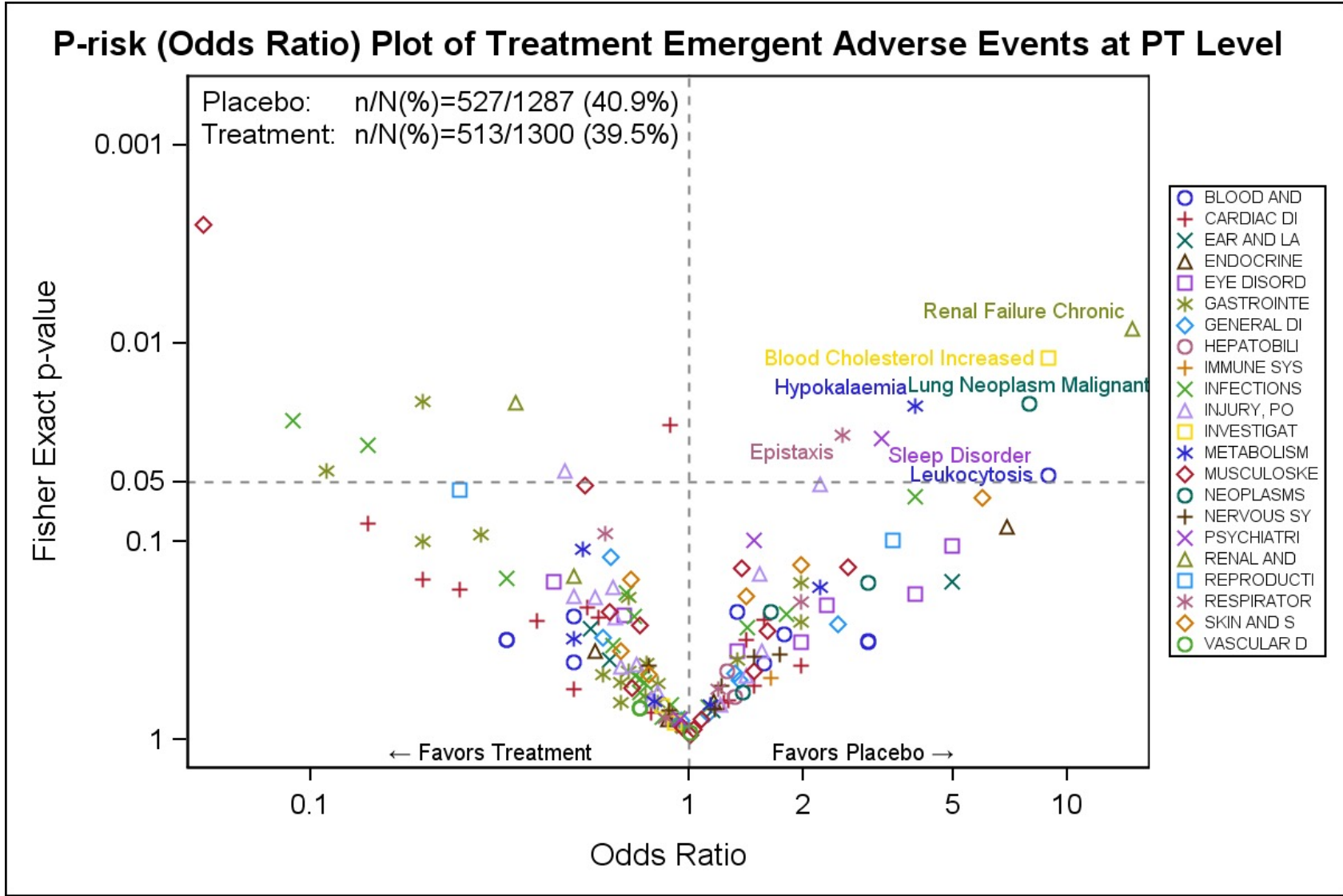
- Introduction
- Background Concepts
- Steps
 - Step 1: Setting Up Environment
 - Step 2: Data Preparation
 - Step 3: Calculate Freq
 - Step 4: Calculate Odds Ratio And P Value
 - Step 5: Pre-graph Processing For Volcano Plot
 - Step 6: Graphics Of Volcano Plot
- Summary
- Q&A

Introduction

- Safety Analysis - new types graphs
- Various types of safety analysis
 - Volcano Plot
- R trend increasing !!!

Background Concepts

- Volcano Plot
 - One type of scatter plot
 - Identify changes in one group/set compared to another group
 - X – Fold Change (E.g. Log of fold change b/w two group/conditions)
 - Fold Change - Relative Risk/Risk Difference/Odds Ratio etc
 - Y - Level of Significance (E.g. Negative Log10 of P Value)



Sample Image from https://blogs.sas.com/content/graphicallyspeaking/files/2016/05/Volcano_RR.png



STEP 1: SETTING UP ENVIROMENT

- Load libraries
- Load datasets using read.csv

```
#####  
# Step: 1 - Get Data into R  
#####  
library("tidyverse")  
library("shiny")  
#Import ADAE  
adae <- read.csv(file = "D:/sample_adae1.csv",  
                 #fileEncoding="UTF-8-BOM",  
                 blank.lines.skip = TRUE,  
                 skipNul = TRUE,  
                 header = TRUE)  
  
#Import ADSL  
adsl <- read.csv(file = "D:/sample_adsl1.csv",  
                 #fileEncoding="UTF-8-BOM",  
                 blank.lines.skip = TRUE,  
                 skipNul = TRUE,  
                 header = TRUE)
```

STEP 2: DATA PREPERATION

- Take ADAE dataset
- Select only variables which are needed
- Unique AEDECODs per subject (counting 1 preferred term per subject)

```
#####  
# Step: 2 - Data Preparation  
#####  
#Lower variable names for better handling  
names(adae) <- tolower(names(adae))  
  
#Select few variables | Select one AEDECOD out  
of multiple per subject  
adae2x <- adae %>%  
  select(usubjid, aedecod, aebodsys,  
trt01p)%>%  
  unique() #To get one AEDECOD per subject for  
counting subject numbers
```

STEP 3: CALCULATE FREQ

- Calculate the frequency of AEDECODS
- Merge both dataset.
- Goal - to get one column with # of event count and # of no event count per each AEDECOD.

```
#####  
# Step: 3 - Calculate Freq  
#####  
#Get Freqs of AEDECOD  
#AE with events  
frq1 <- data.frame(xtabs( ~ aeDecod + trt01p, data  
= adae2 ))  
#Total subjects from ADSL  
frq2 <-data.frame (xtabs( ~ trt01p, data = adsl ))  
  
#Merged  
frq3 <- merge(frq1, frq2, by="trt01p")  
  
frq4 <- frq3 %>%  
  mutate(no = Freq.y - Freq.x, #total - # with event  
         yes = Freq.x,         # all with AE  
         trt01p=if_else(trt01p == "Study Drug A",  
"druga", "active")) %>%  
  arrange(aeDecod, trt01p) %>%  
  select(c(trt01p, aeDecod, yes, no))
```

STEP 4: CALCULATE ODDS RATIO AND P VALUE



- Create 2x2 contingency table for event and no-event for both groups.
- Challenge - we have hundreds of AEDECODs.
- Executing fishers test for each AEDECOD.

```
#####  
# Step: 4 - Calculate Odds Ratio and P Value  
#####  
dfx_all <- setNames(data.frame(matrix(ncol = 5, nrow  
= 0)),  
  c("aedecod", "pval", "ci_1", "ci_2", "or"))  
  
for (p in unique( freq4$aedecod) ) {  
  df1<- subset(freq4, aedecod == p, select = ( c(-  
aedecod, -trt01p) ) )  
  stat1<- fisher.test(df1)  
  
  dfx <- data.frame(aedecod=p,  
                    pval=stat1[[1]],  
                    ci_1=stat1[[2]][1],  
                    ci_2=stat1[[2]][2],  
                    or = stat1[[3]][1],  
                    row.names = NULL)  
  
  #print(p)  
  dfx_all <- rbind (dfx_all,dfx)  
}
```

STEP 5: PRE-GRAPH PROCESSING FOR VOLCANO PLOT

- Take negative log of p value for y axis as below. This ensures data are arranged as needed in volcano plot.

- `logpval = -log10(pval)`

- For x-axis, we convert folds change (odds ratio) into log base 2 by taking log.

- `logfold = log2(or)`

- Finally, we want to display label of AEDECODs which are falling in category 1 and 2 i.e. significant p value with significant folds positive or negative.

```
#####  
# Step: 5 - Pre-Graph Processing for Volcano Plot  
#####  
#-----  
vplot1 <- dfx_all %>%  
  mutate( logpval = -log10(pval),  
          logfold = log2(or) ,  
          aelable = ifelse(  
            (logfold > 0.6 | logfold < -0.6) &  
            (logpval > 1.5),  
            aedecod, "" ),  
          pnt_col = ifelse(  
            (logfold > 0.6) & (logpval > 1.5),  
            "A", ifelse(  
              (logfold < -0.6) & (logpval > 1.5),  
              "B", "C" ))  
          ) %>%  
  filter( is.finite (logfold) & is.finite  
          (logpval) )
```

STEP 6: GRAPHICS OF VOLCANO PLOT



- We are all set to create volcano plot using ggplot2.
- We can adjust as per our need.
- We can create interactive RShiny App as well.

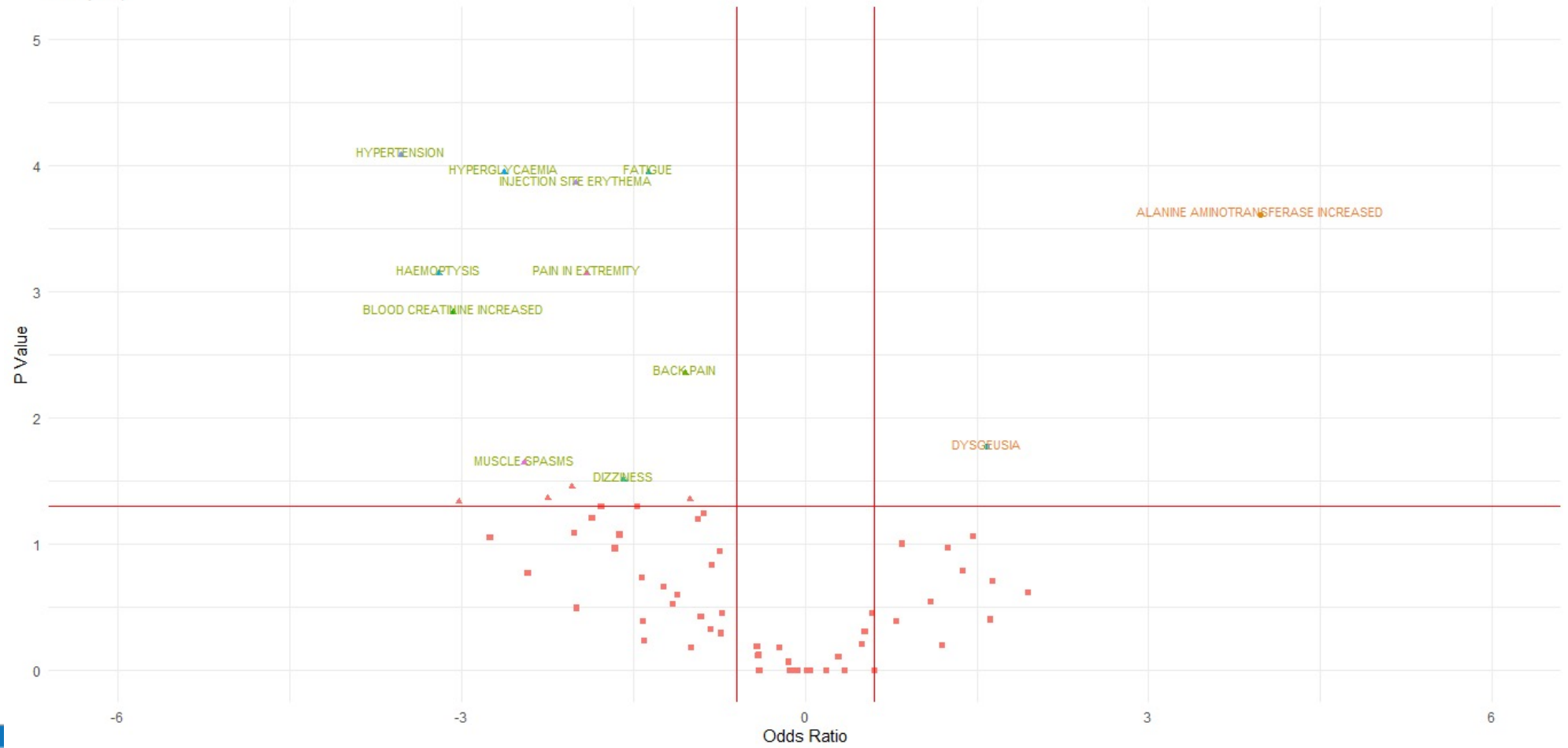


STEP 6: GRAPHICS OF VOLCANO PLOT

```
#####  
# Step: 6 - Graphics of Volcano Plot  
#####  
#Plotting  
ggplot(vplot1,  
       aes(x=logfold, y=logpval, color=pnt_col, shape=pnt_col, label =  
aedecond )) +  
  geom_point(aes(colour =aelable)) +  
  labs(title = "Volcano Plot", subtitle = "Safety Pop",  
       x="Odds Ratio",  
       y="P Value") +  
  theme_minimal() +  
  geom_vline(xintercept=c(-0.6, 0.6), col="red") +  
  geom_hline(yintercept=-log10(0.05), col="red") +  
  coord_cartesian(xlim = c(-6,6) ,  
                  ylim = c(0,5)) +  
  geom_text(aes(label = aelable), size = 3,vjust = 0.14) +  
  labs(shape="Treatment Group") +  
  guides(color = FALSE) +  
  theme(legend.position = "none")
```



Volcano Plot
Safety Pop



Summary

- Volcano plot is nice and effective way of explaining adverse event analysis
- R – Strong for graphics
- Easy to get around for custom new plots in R
- RShiny can make interactive!!!

Thanks....

Questions?