PharmaSUG 2022 - Paper QT-102

Getting odds ratio at Preferred Term (PT) level for safety analysis in R

Kamlesh Patel, Jigar Patel, Rang Technologies, Piscataway, NJ

ABSTRACT

Safety Analysis is vital part of regulatory review. Various types of adverse event (AE) analysis are being performed in industry for reviewing safety by comparing each preferred term (PT) level statistical parameters in two treatment groups. Hence, calculating various statistical parameters at preferred term level becomes necessary. We will discuss about getting Odds Ratio (OR) for two treatments with respect to each preferred term. Calculating Odds ratio in R for multiple records together is very tricky. This can be also used in various other safety analysis like volcano plot creation. We will be presenting a simple way to calculate odds ratio for hundreds of preferred terms together in R for various safety analysis.

INTRODUCTION

Safety analysis is very important part of regulatory agency review. There are many new types of safety analysis are being performed to display safety of new drug. Some of new approaches includes new graphs (e.g., Volcano Plot) which are based on statistics derivation at preferred term level. In current industry trend of using R for many new analyses over SAS, we have selected specific problem of derivation of odds ratio (OR) at specific preferred term level in R.

Some basic understanding of statistics, R and clinical data is needed for our audience. In this paper, we have taken problem of calculation of odds ratio for each preferred term (PT) in adverse event data while comparing safety of study drug over reference drug. There are many ways to calculate regular odds ratio for whole treatment group, but there is not any promising option to get odds ratio for each preferred term (PT) in ADAE. We have taken example of creation of volcano plot.

For Volcano plot creation, we need on odds ratio and p value. We will discuss about how to get AE preferred term specific odds ratio.

CHALLENGE - NEED OF ODDS RATIO IN R

There is not any specific way to get statistics (e.g., odds ratio) at preferred term level when comparing one treatment group with another. Hence, we discuss on how to get odds ratio at each preferred term level in R in this paper. We have taken case of Volcano plot for creation as sample.

For the volcano plot creation, we need to odds ratio for each preferred term from one group over another. We can get odds ratio in R using *fisher.test* function for one group over another group.

PREREQUISITE – WHAT WE HAVE AND WHAT WE NEED

We have ADSL, ADAE datasets in CSV format. In ADAE, there are two treatment groups, and we want to calculate the odds ratio for one group over another one.

Below is the snapshot of sample ADAE –

studyid	domain	usubjid	aeseq	aespid	aeterm	aedecod	aebodsys
ABC-123	AE	ABC123-0123-1506	1	6	ARTHRALGIA- BOT...	ARTHRALGIA	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS
ABC-123	AE	ABC123-0123-1506	2	10	ARTHRALGIA IN S...	ARTHRALGIA	MUSCULOSKELETAL AND CONNECTIVE TISSUE DISORDERS
ABC-123	AE	ABC123-0123-1506	3	15	COUGH	COUGH	RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS
ABC-123	AE	ABC123-0123-1506	4	12	RED COLORED KN...	ERYTHEMA	SKIN AND SUBCUTANEOUS TISSUE DISORDERS
ABC-123	AE	ABC123-0123-1506	5	13	RED COLORED ELB...	ERYTHEMA	SKIN AND SUBCUTANEOUS TISSUE DISORDERS
ABC-123	AE	ABC123-0123-1506	8	3	SWOLLEN LYMPH ...	LYMPHADENOPAT...	BLOOD AND LYMPHATIC SYSTEM DISORDERS
ABC-123	AE	ABC123-0123-1506	9	14	SORE THROAT	OROPHARYNGEAL...	RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS
ABC-123	AE	ABC123-0123-1506	10	2	GENERALIZED RAS...	RASH GENERALISED	SKIN AND SUBCUTANEOUS TISSUE DISORDERS
ABC-123	AE	ABC123-0123-1506	11	5	RASH ON FOREAR...	RASH GENERALISED	SKIN AND SUBCUTANEOUS TISSUE DISORDERS

Display 1. Sample ADAE

STEPS OF GETTING ODDS RATIO (OR) IN R

Now, we will discuss step by step code for getting odds ratio in R.

STEP A: GETTING READY

You can import data using *read.csv* function and get datasets (ADSL, ADAE) into R environment. Then, you can move further with calculation of odds ratio.

```
##############################
# Step: A - Getting Ready
##############################
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_adsl.csv",
                 #fileEncoding="UTF-8-BOM",
                 blank.lines.skip = TRUE,
                 skipNul = TRUE,
                 header = TRUE)
```

STEP B: DATA PREPARATION

You can take ADAE dataset to select certain variables you need. You can do this task using *select()* function from dplyr. Then, we select unique AEDECODs per subject using *unique()* from dplyr. This function selects one preferred term (AEDECOD) per subject.

```
##############################
# Step: B - Data Preparation
##############################
#Select variables and get one AEDECOD out of multiple AEs per subject
adae2x <- adae %>%
  select(usubjid, aedecod, aebodsys, trt01p)%>%
  unique() #To get one AEDECOD per subject for counting subject numbers
```

STEP C: CREATE 2 X 2 CONTINGENCY LIKE DATASET

You can calculate the count of each AEDECODS by treatment groups. This number by each AEDECOD is number of subjects with event. Now, you do not have number of subjects without event for each AEDECOD. Hence, you need to get count of total number of subjects in each group. You can get count treatment groups from ADSL.

Both count datasets are merged by treatment group to get total number and number with event together. Hence, total number in each treatment group is now with number of events. Now, you can calculate the number of subjects without events by subtracting number of events from total events. This dataset is almost ready. This is not 2 X 2 contingency table, but this dataset is Each AEDECOD X 2 X 2 table.

```
##############################
# Step: C - Create 2 X 2 Contingency like dataset
##############################
#Get Count 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 D: DERIVE ODDS RATIO

You have dataset from which you can create 2x2 contingency table for each AEDECOD by sub-setting it. However, challenge is we have hundreds of AEDECODs. So, we have to create a loop for executing fishers test for each AEDECOD.

After each loop, you need to stack each outcome dataset with statistics after each loop to make final datasets.

```
##############################
# Step: D - Derive Odds Ratio
##############################
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)
}
```

OUTPUT

You can get finally dataset with all statistics you need for each AEDECOD.

aedecod	pval	ci_1	ci_2	or
ABDOMINAL DISCOMFORT	5.078930e-01	0.000000000	6.9981022	0.00000000
ABDOMINAL DISTENSION	4.320652e-01	0.033703971	Inf	Inf
ABDOMINAL INFECTION	5.078930e-01	0.000000000	6.9981022	0.00000000
ABDOMINAL MASS	4.320652e-01	0.033703971	Inf	Inf
ABDOMINAL PAIN	5.049580e-02	0.931002662	9.1574058	2.75429241
ABDOMINAL PAIN LOWER	3.196038e-01	0.316608822	210.7381188	3.98573119
ABDOMINAL PAIN UPPER	8.400868e-02	0.836355973	13.8808982	3.06574199
ABSCESS ORAL	1.000000e+00	0.000000000	51.2239208	0.00000000
ACNE	5.804990e-01	0.136443620	156.9138212	2.64274892

Display 2. Output dataset with statistics

CONCLUSION

As the trend of new analysis in safety is increasing, new statistics derivation for programmer in diverse types of data as well as performing analysis in R will become more and more essential. For adverse event analysis, calculating odds ratio in R at each preferred term level can made easy with code and it can make analysis and graphical work easy.

REFERENCES

Author name: Soetewey, Antoine. “Fisher's exact test in R: independence test for a small sample.” *Stats and R*. 2020-01-28. Available at < https://statsandr.com/blog/fisher-s-exact-test-in-r-independence-test-for-a-small-sample/>.

Stats (version 3.6.2). “fisher.test: Fisher's Exact Test for Count Data.” Accessed March 29, 2022. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/fisher.test.

ACKNOWLEDGMENTS

We would like to acknowledge help of Yuan (Ari) Yao for the her review.

CONTACT INFORMATION <HEADING 1>

Your comments and questions are valued and encouraged. Contact the author at:

Kamlesh Patel
Rang Technologies Inc
Kamlesh.patel1@rangtech.com
https://www.ranglifesciences.com/