PhUSE EU Connect 2018
Paper DV11

Adverse Event Data – gRaphically

Lina Rajput, Cytel Statistical Software & Services Pvt. Ltd, Mumbai, India
Prajakta Chitale, Cytel Statistical Software & Services Pvt. Ltd, Pune, India

ABSTRACT
Adverse event reporting is an important aspect of clinical trials. In this paper, we will be discussing the visualization of adverse event data through Heat Maps and Volcano plots using R software.
A heat map is a tool that can be used to color code and present large amount of data. The intensity of the color can be used to highlight the significance/ pattern in the data. A volcano plot is a type of scatter-plot which can be used to quickly identify changes in large data sets. In this paper, we will go through the various functions available in R to enhance the quality of output.

KEY WORDS
Adverse event, heat map, volcano plot

INTRODUCTION
In clinical trial studies, the investigators are responsible for recording a subject's adverse experiences (AEs) at each clinical office visit or assessment. An AE is defined as: any noxious, pathological or unintended change in anatomical, physiological or metabolic functions as indicated by physical signs, symptoms and/or laboratory changes occurring in any phase of the clinical study whether associated with the study drug, or active comparator or placebo, and whether or not such change is considered drug related. The adverse events are also graded as per their Severity (‘Mild’, ‘Moderate’ and ‘Severe’) and Seriousness (‘Not Serious’, ‘Serious’). This categorization helps in conducting further analysis on the AE data.

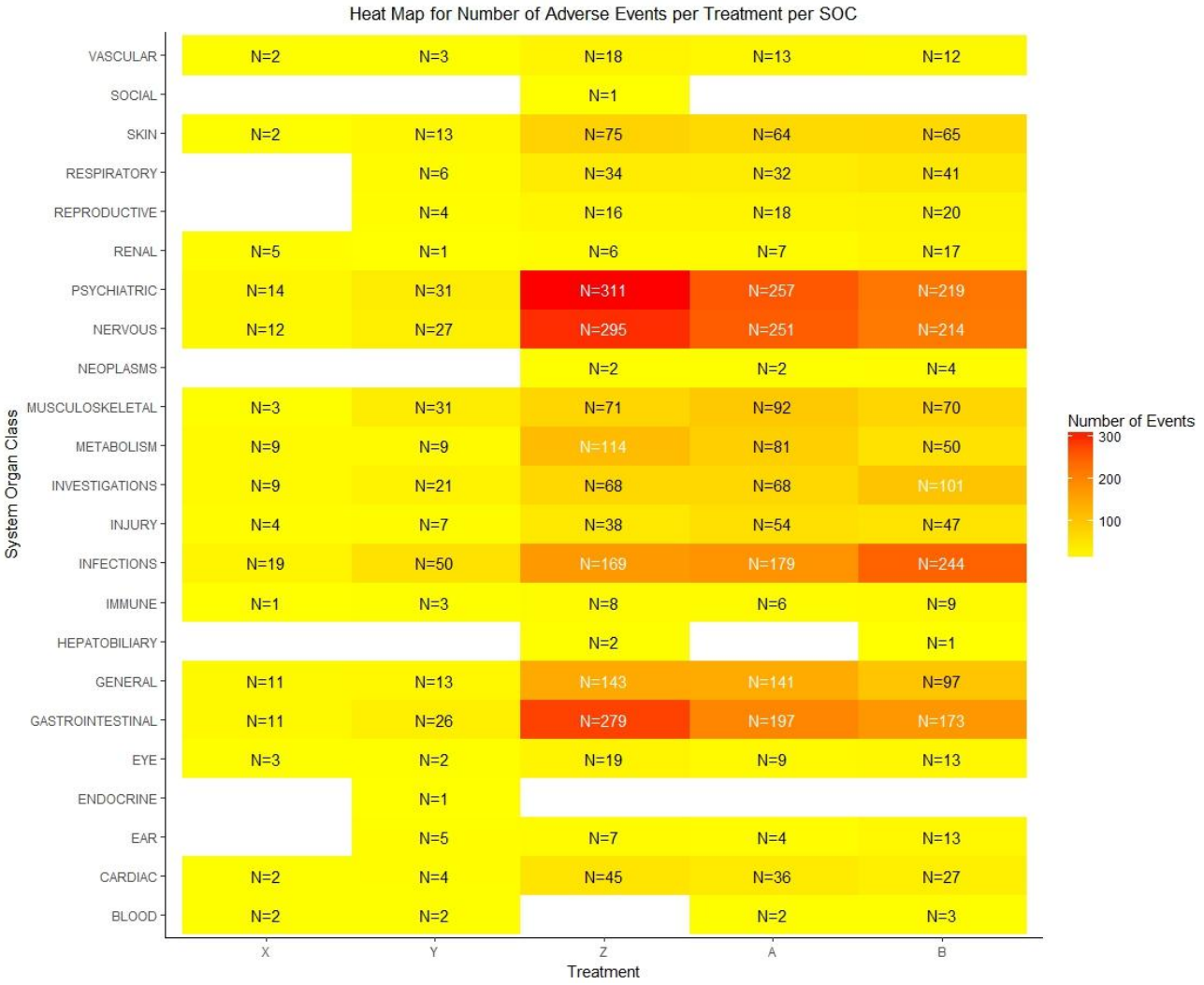
PURPOSE
Recording the adverse events in any trial generates a lot of data due to the amount of information that has to be collected, stored, analyzed and presented. The basic information that has to be analyzed and presented is based on the number of adverse events recorded as per treatments, severity, causality, seriousness etc. The approach usually followed is to summarize and present this data in tabular format. Along with this, adverse event incidence analyses are a critical component for describing the safety profile of any new intervention. The results typically are presented in lengthy summary tables.
We will describe two space-saving graphical summarization methods using R language that could be used for better interpretation of the safety profile.

HEAT MAPS
Heat maps are charts with values being represented by colors. The intensity of the color is an indication of the number i.e. higher the value more intense will be the color. Here, we will have a look at how to plot heat maps to graphically summarize the number of events recorded in a clinical trial. The data we will use is from a trial conducted for a mental health disorder. There are 5 treatments with varying drug dose amount. The AEs have been entered using a standard medical dictionary.

HEAT MAP FOR NUMBER OF ADVERSE EVENTS PER TREATMENT PER SYSTEM ORGAN CLASS (SOC):
A heat map can be generated to visualize the number of adverse events recorded per SOC per treatment. We will plot the treatments on x-axis and the SOCs on y-axis. Each tile of the heat map will be colored as per the number of events recorded for that particular combination. The color of the tile will vary from yellow to red, with low counts shown in yellow and high counts in red. The tiles with white color denote the absence of events for a particular combination of SOC and treatment. The number of events will be printed in either black or white as per a predefined threshold; counts below or equal to the threshold will be printed in black, remaining will be in white.

DATA SNIPPET 1

TRTA	SOC	OBS
Y	NERVOUS	27
X	NERVOUS	12
Z	PSYCHIATRIC	311
A	PSYCHIATRIC	257
B	PSYCHIATRIC	219
Y	PSYCHIATRIC	31
X	PSYCHIATRIC	14
B	RENAL	17
A	RENAL	7



In this plot, we are able to see the number of events reported for every SOC for all 5 arms. As seen in the legend, tiles with count less than 100 are yellow with varying intensity as per the actual count for that tile. For e.g. if we check the number of adverse events in the gastrointestinal system for treatments X and Y, the count is 11 and 26 respectively. The color of the tile for Y is deeper than the tile for X.

The color starts changing from yellow to orange and then to red as the number of events increases. This is observed in the "Psychiatric" SOC counts, with 31 events Y has a yellow tile, 257 events lead to treatment A having an orange tile and the treatment Z has a red tile due to the 311 events reported. Here the threshold for number of events has been set to 100. We can quickly identify the SOCs with number of events greater than 100 as they are printed in white. We can observe white tiles for SOCs with no events for a particular treatment like Reproductive system and treatment X.

It might be of interest to further look in to the Psychiatric and Nervous System disorders since this is a trial for a mental health disorder.

The code to generate the above plot is as follows-

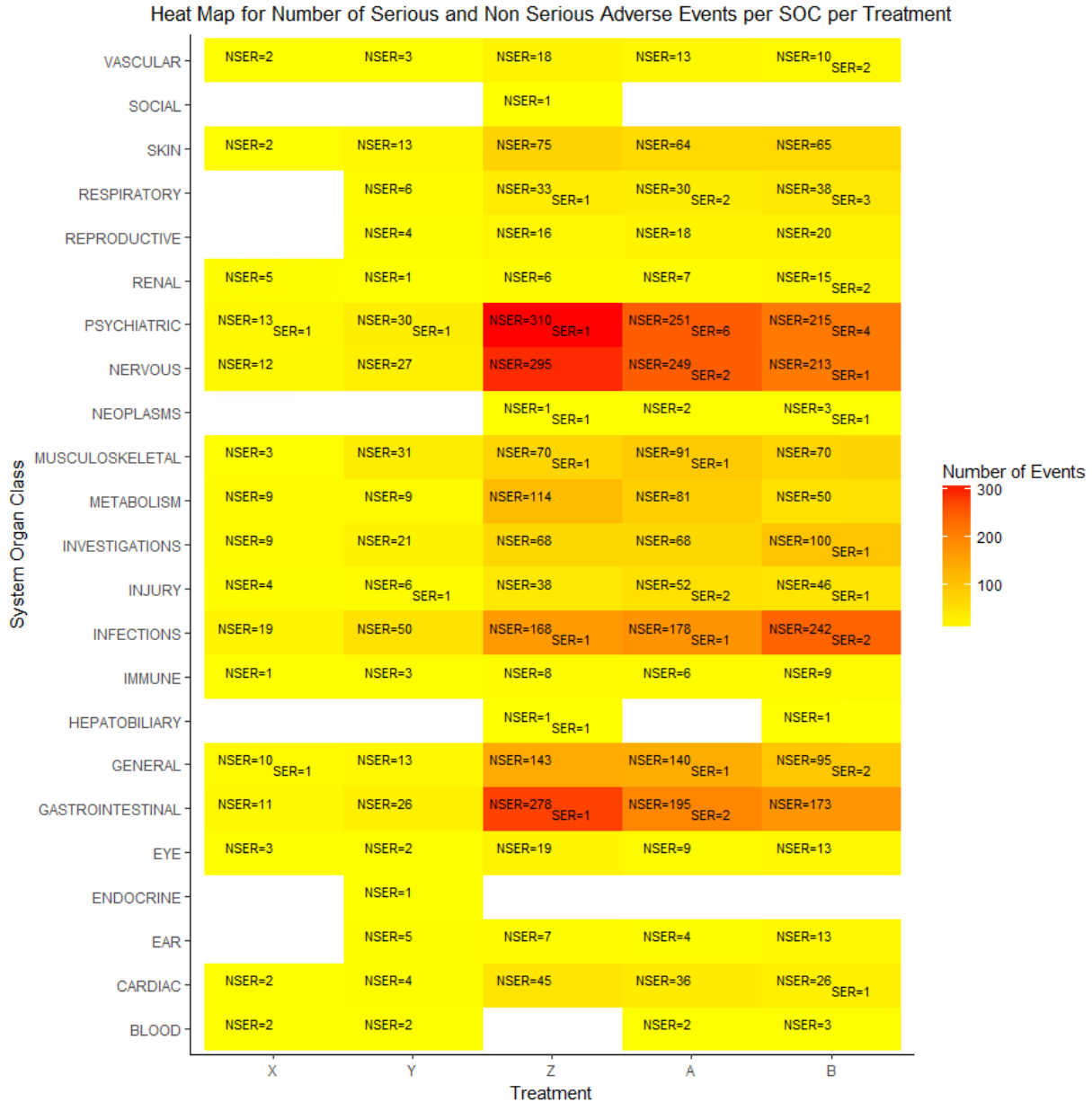
```
tb1 <-aggregate(USUBJID~TRTA+SOC, data=aeord, FUN=function(x) c(ln=length(x)))
head(tb1)
names(tb1)[names(tb1)=="USUBJID"]<- "OBS"
p1 <-  ggplot(tb1, (aes(x = TRTA, y = SOC, fill=OBS))) + geom_raster() +
       geom_text(aes(label = paste0("N=",round(OBS, 1)), color = OBS < 100),show.legend = F)+
       scale_fill_gradient(low="yellow",high="red",na.value="white")+
       scale_color_manual(values = c("FALSE" = "white", "TRUE" = "black")) +
       scale_x_discrete("Treatment") + scale_y_discrete("System Organ Class") +
       theme_classic()+ggtitle("Heatmap for Number of Adverse Events per Treatment per SOC")+
       theme(plot.title = element_text(hjust = 0.5)) + labs(fill="Number of Events")

p1
```

HEAT MAP FOR NUMBER OF SERIOUS AND NON SERIOUS ADVERSE EVENTS PER SOC PER TREATMENT

The adverse events are classified as either 'Non Serious' or 'Serious'. We can plot the number of serious and non serious events occurring per SOC per treatment as shown in the heat map below. In this plot, both the counts will be displayed in the same tile and the color of the tile will be decided by the maximum of the number of events if both events are present.

DATA SNIPPET 2

TRTA	SOC	AESER	OBS
B	PSYCHIATRIC	Y	4
A	PSYCHIATRIC	Y	6
X	PSYCHIATRIC	N	13
Y	PSYCHIATRIC	N	30
B	PSYCHIATRIC	N	215
A	PSYCHIATRIC	N	251
Z	PSYCHIATRIC	N	310

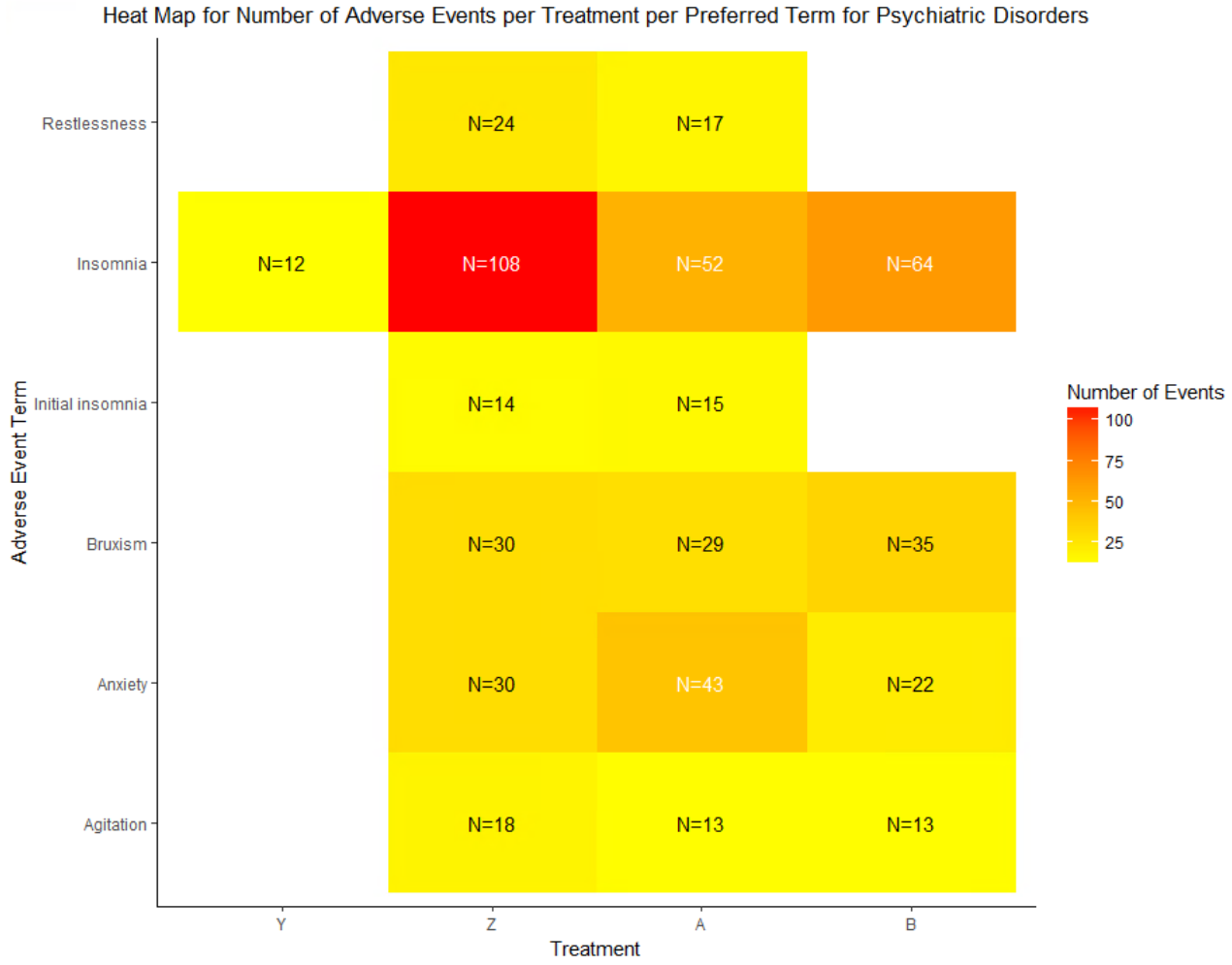


We can observe that counts for both types of events are displayed in the same tile whenever applicable. At a glance, we can conclude that the number of non serious events is far more than the number of serious events. The serious events occurring in treatment B are spread out more as compared to the other treatments. The color of the tile is as per the greater number of events when both types are reported. This is observed in the "Psychiatric" tile for arm Z. Out of the 311 events shown in the first heat map, there are 310 non serious events and only 1 serious event. Hence the color of the tile is red.

```
tb2  <- aggregate(USUBJID~TRTA+SOC+AESER, data=aeord, FUN=function(x) c(ln=length(x)))
names(tb2)[names(tb2)=="USUBJID"]<- "OBS"
tb3 <-tb2[order(tb2$TRTA, tb2$SOC, tb2$OBS),]
p2 <- ggplot(tb3, aes(TRTA, SOC, fill = OBS)) + geom_raster() +
       geom_text(data=tb3[tb3$AESER=="N",],aes(label = paste0("NSER=",round(OBS, 1))),
       show.legend = F,size=2.9, vjust=0, hjust=1) +
       geom_text(data=tb3[tb3$AESER=="Y",],aes(label = paste0("SER=",round(OBS, 1))),
       show.legend = F,size=2.9, vjust=1, hjust=0)+
       scale_fill_gradient(low="yellow",high="red",na.value="white")+
       scale_x_discrete("Treatment") + scale_y_discrete(name = "System Organ Class") +
       theme_classic()+
       ggtitle("Heatmap for Number of Serious and Non Serious Adverse Events per SOC per
       Treatment") + theme(plot.title = element_text(hjust = 0.5)) + labs(fill="Number of
       Events")
p2
```

HEAT MAP FOR NUMBER OF ADVERSE EVENTS PER TREATMENT PER PREFERRED TERM (PT) FOR A SINGLE SOC

As mentioned earlier, we can analyze the psychiatric disorders further by generating a heat map of the events reported under this SOC for all treatments. We will generate a heat map of all PTs with number of reported events greater than 10. Data is similar to data snippet 1.

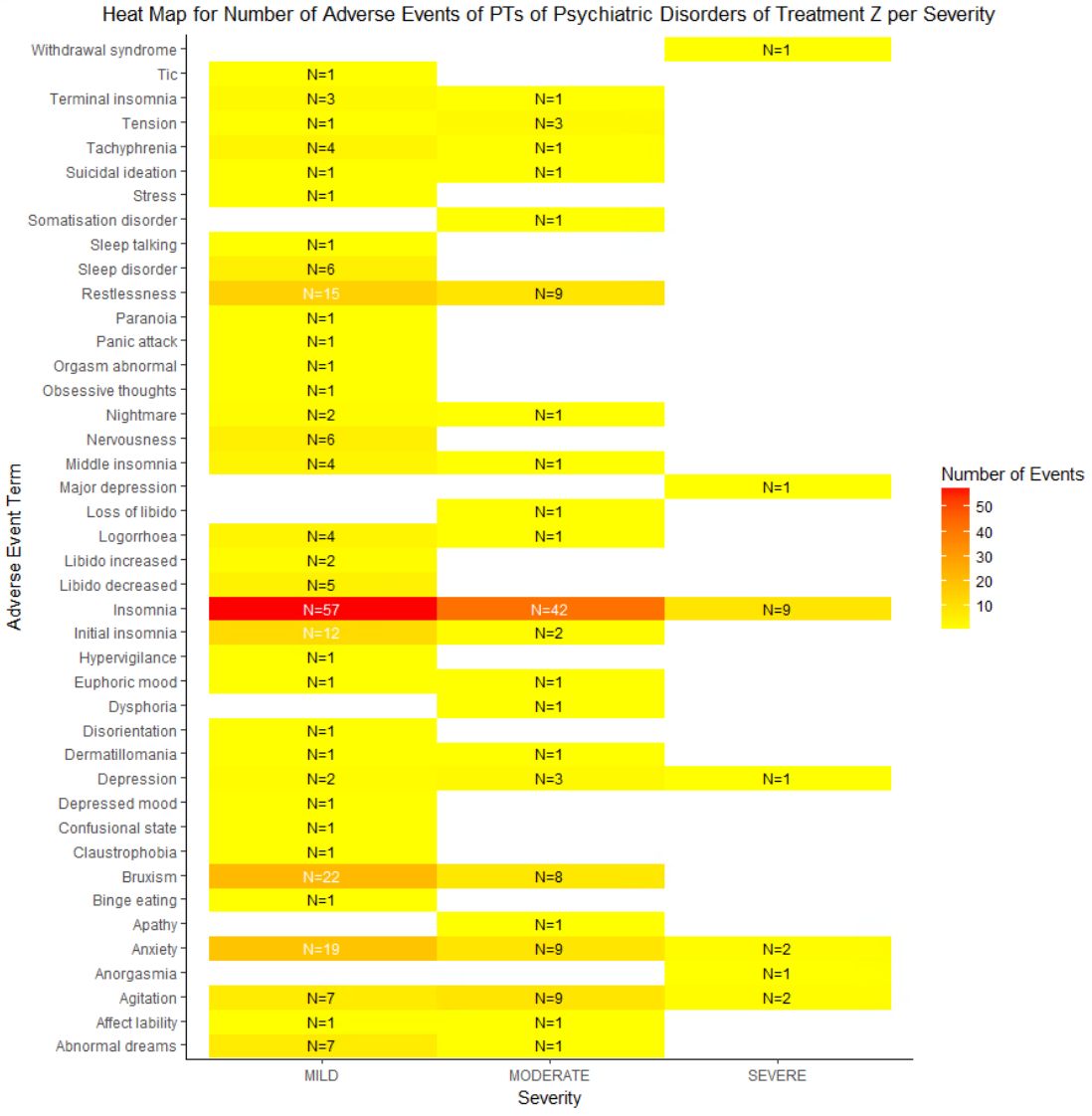


We can see that there are only 6 PTs with number of events greater than 10, the highest being Insomnia reporting 108 events for treatment Z. We can also see that treatment X has no PTs with number of events greater than 10 and treatment Y has only 1. Also, the threshold has been set to 40 and hence tiles with counts greater than 30 have the text in white.

```
sub <- subset(aeord,aeord$SOC=="PSYCHIATRIC")
tb4 <-aggregate(USUBJID~TRTA+AEDECOD, data=sub, FUN=function(x) c(ln=length(x)))
names(tb4)[names(tb4)=="USUBJID"]<- "OBS"
p3 <-  ggplot(tb4[tb4$OBS>10,], (aes(x = TRTA, y = AEDECOD, fill = OBS))) + geom_raster() +
       geom_text(aes(label = paste0("N=",round(OBS, 1)), color = OBS < 40), show.legend =
       F,size=4)+ scale_fill_gradient(low="yellow",high="red",na.value="white")+
       scale_color_manual(values = c("FALSE" = "white", "TRUE" = "black")) +
       scale_x_discrete("Treatment") + scale_y_discrete(name = "Adverse Event Term") +
       theme_classic()+
       ggtitle("Heat Map for Number of Adverse Events per Treatment per Preferred Term for
       Psychiatric Disorders") + theme(plot.title = element_text(hjust = 0.5)) +
       labs(fill="Number of Events")
p3
```

HEAT MAP FOR NUMBER OF ADVERSE EVENTS OF PTS OF A SINGLE SOC OF A SINGLE TREATMENT PER SEVERITY

We can check the severity of all the PTs for a single SOC for a single treatment. Let us analyze the Psychiatric Disorder PTs reported for treatment Z. We will plot the number of events reported for PTs against their severity which will have the value as Mild, Moderate or Severe. Data is similar to data snippet 1.

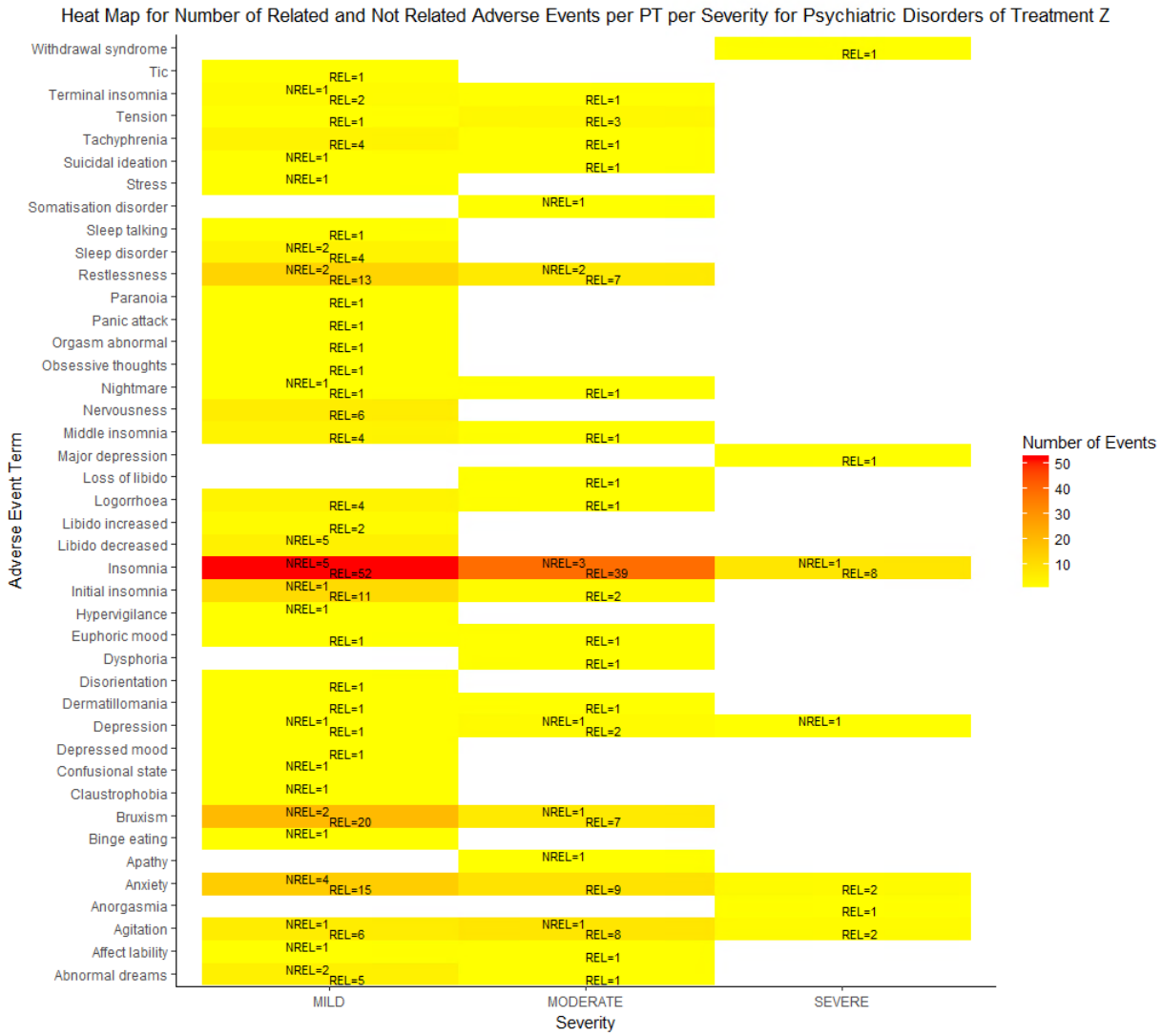


We observe that majority events are mild in nature with the highest being Insomnia having 57 events. If we check the tiles for Insomnia, we will see the total for all severity grades is 108, as observed in the previously plotted heat map. There are very few severe events as compared to the other two grades. The threshold has been set to 10 for this plot.

```
sub1 <- subset(aeord,aeord$SOC=="PSYCHIATRIC" & TRTA=="Z")
tb5  <- aggregate(USUBJID~AEDECOD+AESEV, data=sub1, FUN=function(x) c(ln=length(x)))
names(tb5)[names(tb5)=="USUBJID"]<- "OBS"
p4 <-  ggplot(tb5, aes(AESEV, AEDECOD, fill = OBS)) + geom_raster() +
       geom_text(aes(label=paste0("N=",round(OBS,1)),color=OBS<10), show.legend = F,size=3)+
       scale_fill_gradient(low="yellow",high="red",na.value="white")+
       scale_color_manual(values = c("FALSE" = "white", "TRUE" = "black")) +
       scale_x_discrete("Severity") + scale_y_discrete(name = "Adverse Event Term") +
       theme_classic()+ ggtitle("Heat Map for Number of Adverse Events of PTs of Psychiatric
       Disorders of Treatment Z per Severity") + theme(plot.title = element_text(hjust =
       0.5)) + labs(fill="Number of Events")
p4
```

HEAT MAP FOR NUMBER OF RELATED AND NOT RELATED ADVERSE EVENTS PER PT PER SEVERITY FOR A SINGLE TREATMENT

We can check the causality of the events reported in the above plot along with their severity. Let us once again consider the psychiatry PTs reported for treatment Z. We will plot the number of events reported for PTs with their causality which will have values as 'Related' or 'Not Related' against severity. Data is similar to data snippet 2.



We can see in the above plot that majority of the events are related to the study drug. We can see that for insomnia with moderate severity, 39 out of the 42 events have been categorized as related to the treatment. Also, out of the 9 severe insomnia events, 8 are related to treatment Z. Thus we can say that the insomnia observed in a subject might be drug induced.

```
tb6  <- aggregate(USUBJID~AEDECOD+AESEV+AEREL, data=sub1, FUN=function(x) c(ln=length(x)))
names(tb6)[names(tb6)=="USUBJID"]<- "OBS"
tb6 <-tb6[order(tb6$AEDECOD,tb6$AESEV, tb6$OBS),]
p5 <-  ggplot(tb6, aes(AESEV, AEDECOD, fill = OBS)) + geom_raster() +
       geom_text(data=tb6[tb6$AEREL=="NOT RELATED",],aes(label=paste0("NREL=",round(OBS,1))),
       show.legend = F,size=2.8, vjust=0, hjust=1) +
       geom_text(data=tb6[tb6$AEREL=="RELATED",],aes(label = paste0("REL=",round(OBS, 1))),
       show.legend = F,size=2.8, vjust=1, hjust=0)+
       scale_fill_gradient(low="yellow",high="red",na.value="white")+
       scale_x_discrete("Severity") + scale_y_discrete(name = "Adverse Event Term") +
       theme_classic()+
       ggtitle("Heat Map for Number of Related and Not Related Adverse Events per PT per
       Severity for Psychiatric Disorders of Treatment Z") + theme(plot.title =
       element_text(hjust = 0.5)) + labs(fill="Number of Events")
p5
```

VOLCANO PLOTS

A volcano plot is a type of scatter-plot that is used to quickly identify changes in large data sets composed of replicate data. It plots significance versus fold-change on the y and x axes, respectively. It is constructed by plotting the negative log of the p value on the y axis (usually base 10). This results in data points with low p values (highly significant) appearing toward the top of the plot. The x axis is the log of the fold change between the two conditions. The log of the fold change is used so that changes in both directions appear equidistant from the center. In the clinical domain, a volcano plot is used to view Risk Difference (RD), Relative Risk or Odds Ratio of AE occurrence between the treatment groups by preferred term.

VOLCANO PLOT OF TREATMENT EMERGENT ADVERSE EVENT AT PT LEVEL

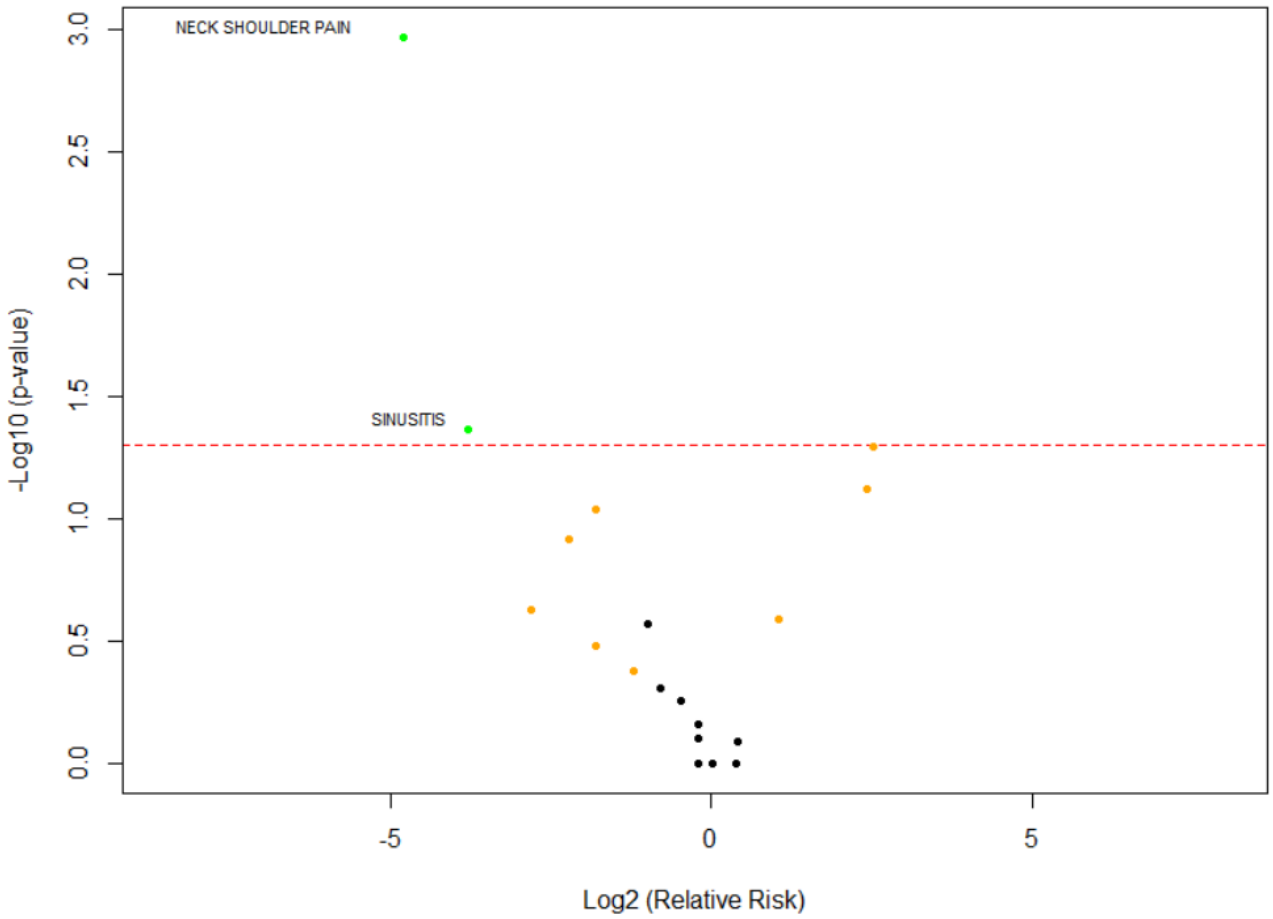
Volcano plots are also generated on the same data which is used for heat maps. Here, we have plotted negative log10-transformed p-value calculated using Fisher exact test on Y-axis and Log2-transformed relative risk on x-axis.

DATA SNIPPET 3

AETERM	rr	log2rr	p	Total	term
NECK SHOULDER PAIN	0.0359734	-4.79692576	0.001080449	5	
SINUSITIS	0.0719468	-3.79692576	0.043228800	3	
DECREASED APPETITE	5.7557437	2.52500234	0.050540193	41	DECREASED APPETITE
DRY MOUTH	5.3240629	2.41252761	0.075784395	38	DRY MOUTH
UPPER RESPIRATORY INFECTION	0.2877872	-1.79692576	0.092355370	9	
COLD SYMPTOMS	0.2158404	-2.21196326	0.121613984	5	
DRYNESS IN MOUTH	0.2158404	-2.21196326	0.121613984	5	
WEIGHT GAIN	0.2158404	-2.21196326	0.121613984	5	
HYPERHIDROSIS	0.1438936	-2.79692576	0.235878161	2	
INCREASED IRRITABILITY	0.1438936	-2.79692576	0.235878161	2	
INCREASED THIRST	0.1438936	-2.79692576	0.235878161	2	
LOWER BACK PAIN	0.1438936	-2.79692576	0.235878161	2	
NUMBNESS (BOTH HANDS)	0.1438936	-2.79692576	0.235878161	2	

PhUSE EU Connect 2018

Volcano Plots of Treatment Emergent Adverse Events at PT Level



The dashed red line on y-axis shows where p = 0.05 with points above the line having p < 0.05 and points below the line having p > 0.05. Note that smaller the p-value larger is the number on y-axis. For this graph, we have labeled the AE terms with p-values < 0.05 as they are significant.

The R code is as follows-

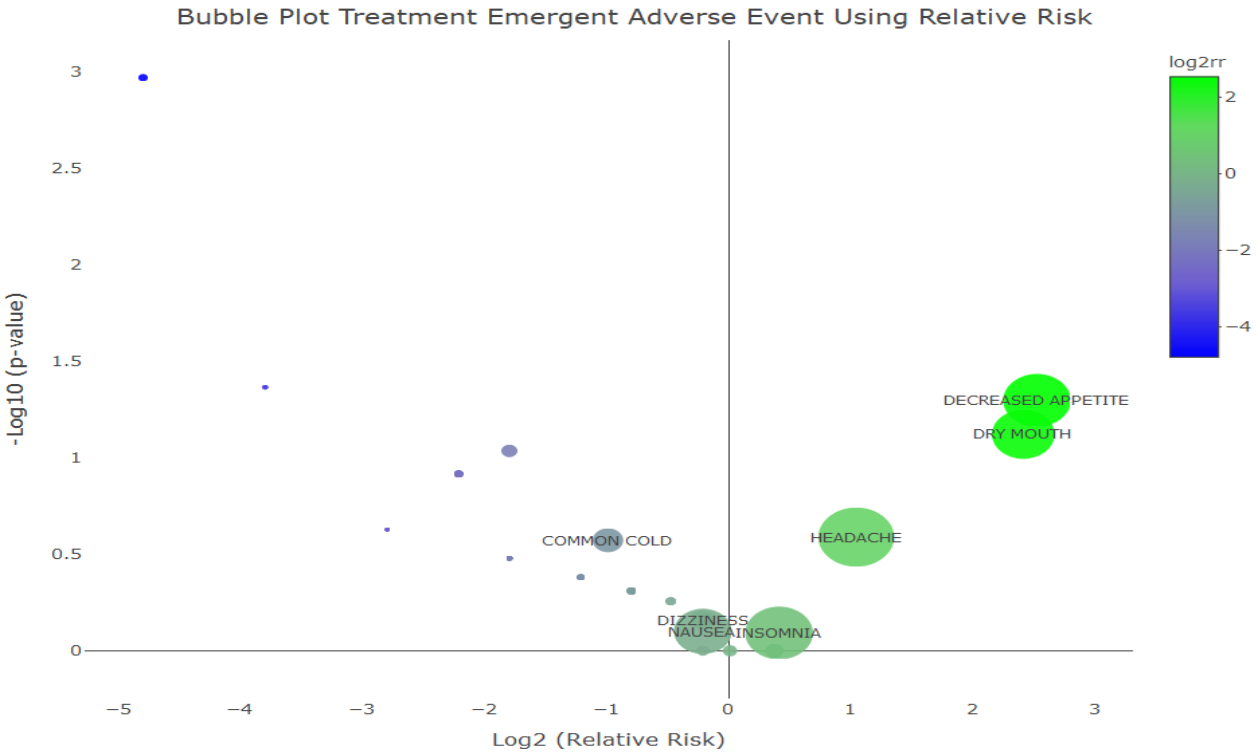
```
#Make a basic volcano plot
with(total, plot(log2rr, -log10(p), pch=20, main="Volcano Plots of Treatment Emergent Adverse
Events at PT Level", xlim=c(-8.5,8), abline(h = 1.3, col="red", lty=2),  xlab="Log2 (Relative
Risk)", ylab="-Log10 (p-value)"))

#Add colored points: red if padj<0.05, orange of log2rr>1, green if both)
with(subset(total, p<.05 ), points(log2rr, -log10(p), pch=20, col="red"))
with(subset(total, abs(log2rr)>1), points(log2rr, -log10(p), pch=20, col="orange"))
with(subset(total, p<.05 & abs(log2rr)>1), points(log2rr, -log10(p), pch=20, col="green"))

#Label points with the textxy function from the calibrate plot
library(calibrate)
with(subset(total, p<.05 & abs(log2rr)>1), textxy(log2rr, -log10(p), labs=AETERM, cex=.7))
```

BUBBLE PLOT OF -LOG10 (P-VALUE) BY RELATIVE RISK SIZED BY COUNTS

The x-axis represents the log2 relative risk and y-axis represents the –log10 p-values computed using the fisher exact test. Each bubble represents an adverse event, with bubble size indicative of the total number of adverse events that occur for both treatments combined. Specifically, the bubble area is proportional to the total number of events.

DATA SNIPPET 4

AETERM	rr	log2rr	p	Total	term
NAUSEA	0.8633615	-0.21196326	0.793537653	35	NAUSEA
INSOMNIA	1.3310157	0.41252761	0.809275857	41	INSOMNIA
HEADACHE	2.0624748	1.04437650	0.258095025	46	HEADACHE
DRY MOUTH	5.3240629	2.41252761	0.075784395	38	DRY MOUTH



As shown above, Headache is the most commonly occurring adverse event during this trial. Log2 (Relative Risk) of 0 represents no difference in risk between the 2 treatment groups, while bubbles to the right indicate a higher risk for subjects in treatment group Y (i.e. – If Log2 (RR) = 1, that is the same as an RR=2), and bubbles to the left indicate a higher risk for subjects in treatment group Z. Color, green and blue helps to emphasize adverse events that are more common for treatment Y or Z, respectively. The size of the bubble represents the total number of occurrences of the AE of interest. Names of events with total number of occurrences greater than 10 are displayed.

```
library(plotly)
p <-plot_ly(total, x = ~ log2rr, y = ~ -log10(p), text = ~ AETERM, type = 'scatter', mode =
'markers', color=~ log2rr, colors=c("blue","green") , marker = list(size = ~ Total, opacity =
0.9)) %>%
layout(title = 'Bubble Plot Treatment Emergent Adverse Event of Significance versus Relative
Risk', xaxis = list(showgrid = FALSE,  title = 'Log2 (Relative Risk)'), yaxis = list(showgrid
= FALSE, title = '-Log10 (p-value)')) %>%
add_annotations(x = total$log2rr, y = total$p1, text = total$term, xref = "x", yref = "y",
                showarrow = FALSE, font = list(size = 10))
p
```

LIMITATIONS

Heat maps are a basic tool of visualizing any data. We may not be able to perform complex analysis using heat maps.
In volcano plots, treatments are compared in a pair wise fashion resulting in only two treatments being compared at any given instance.

CONCLUSION

The graphical displays could be given more consideration in the analysis of safety data arising from clinical trials. Tabular summaries are voluminous and might be difficult to interpret however graphical summaries display information in a clear and concise manner. Graphical displays also help us identify any areas of interest or concern in the data and hence are useful for gaining insight and planning further analyses.
Volcano plots are space-saving tools that emphasize important differences between the adverse event profiles of two treatment arms.
Heat maps are useful to understand the spread of the adverse event data. They also help in identifying any peculiarities in the data on which further analysis could be conducted.
R Shiny applications can be developed to generate various types of plots described above for validation purposes.

REFERENCES

- https://www.nature.com/articles/srep24955
- https://ggplot2.tidyverse.org/reference/index.html
- https://plot.ly/r/reference/#scatter
- https://community.plot.ly/

CONTACT INFORMATION

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

Lina Rajput
Cytel Statistical Software & Services Pvt. Ltd. (Subsidiary of Cytel Inc.)
Reliable Plaza, 3rd Floor, Unit No. 301, Plot No. K-10, Kalwa Industrial Area, District Thane, Airoli, Navi Mumbai – 400 708, India
lina.rajput@cytel.com | www.cytel.com

Prajakta Chitale
Cytel Statistical Software & Services Pvt. Ltd.
S.No. 150, Lohia Jain IT Park, A Wing, 6th Floor, Paud Road,
Kothrud, Pune - 411 038
prajakta.chitale@cytel.com | www.cytel.com