

Volcano Plot: Demystifying Why and How DV14

Jyothi Nelakurthi
Chandana Sudini

What is Volcano Plot?

- Two dimensional scatter plot
- P-values on Y-axis
- Metrics like Relative risk, Odds ratio or Hazard ratio on X-axis.

Why Volcano Plot?

- Displays in Spatially Constrained manner.
- Graphically Presents Measure of strength versus significance of a statistical signal.
- Easily identifies large magnitude changes in voluminous data
- Can be used in exploratory analysis to emphasize important differences between the arms.

Popular use

- Originally Used in Genomics micro array experiments.
- To identify meaningful changes in thousands of genes as replicate data points and present large fold changes.
- Multiplicity adjustments.
- Display appropriate estimands and adjusted P-values to reduce false positives.

Volcano Plot in Safety Evaluation

- Adverse event profiles in safety reports.
- Summarizing laboratory abnormalities counts in defined time intervals.
- Concomitant medications which are likely to impact treatment outcome.

How to: Volcano Plot?

Overview of the data considered to create the plot-

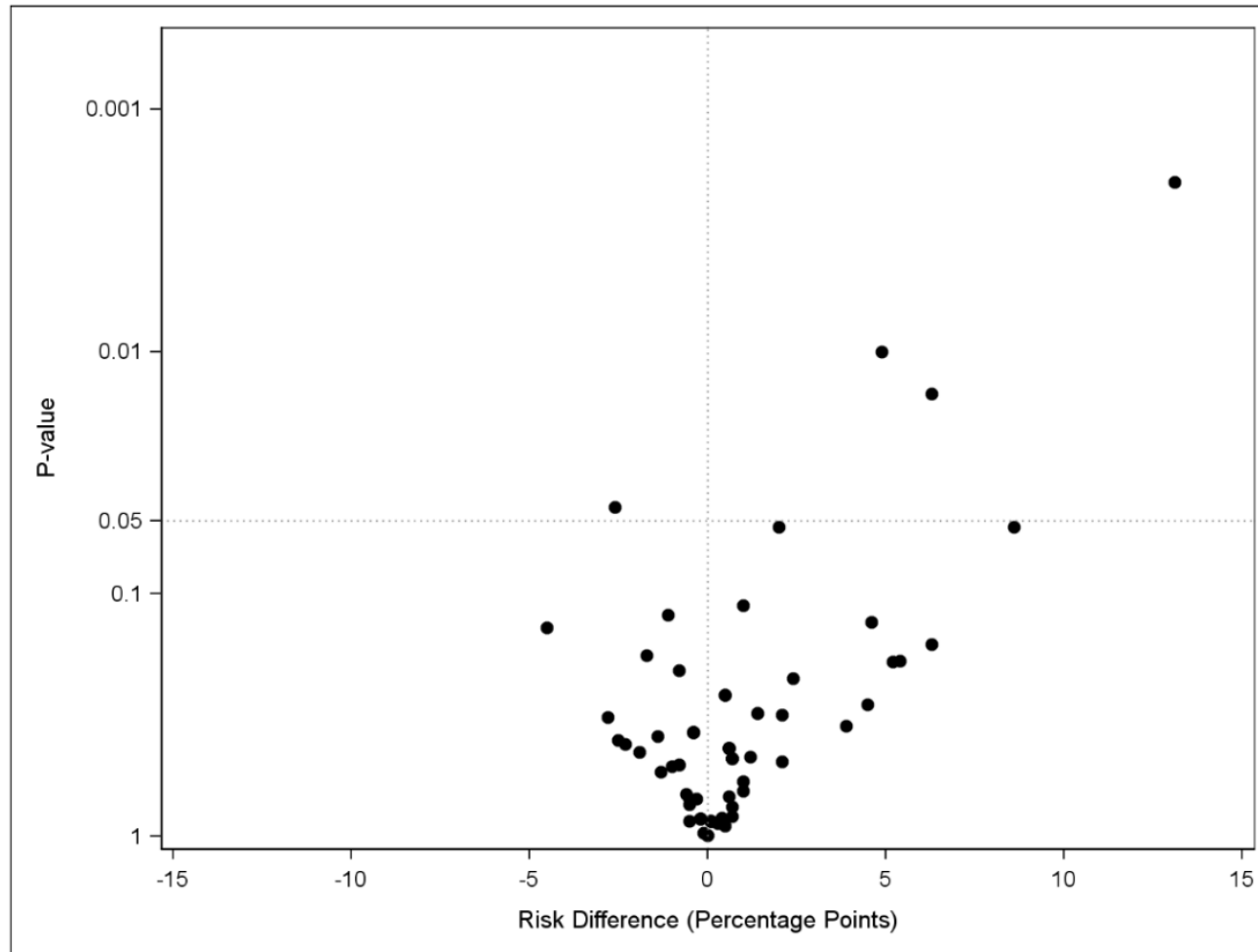
- Adverse events data from late stage clinical trial.
- 2 treatment arms
- Classified AE's using Standard Medical Dictionary.
- Used PROC SGPLOT with statements like SCATTER, KEYLEGEND, INSET ,TEXT and BUBBLE etc.

Input data used:

Body System or Organ Class	Preferred Term	Risk Difference	P-Value
Blood and lymphatic system disorders	Anemia	13.1	0.002
Infections and infestations	Cellulitis	-2.6	0.044
Nervous system disorders	Dizziness	-2.8	0.322
Gastrointestinal disorders	Abdominal pain	-4.5	0.138
Renal and urinary disorders	Acute kidney injury	6.3	0.015
Respiratory, thoracic and mediastinal disorders	Cough	0.7	0.829
Musculoskeletal and connective tissue disorders.....	Flank pain...	-1.7...	0.18....

- Risk difference is calculated as the difference between the incidence proportion of an event in Drug A and the incidence proportion of the same outcome in the Drug B.
- P-values are calculated with respect to derived risk difference.

Basic Plot

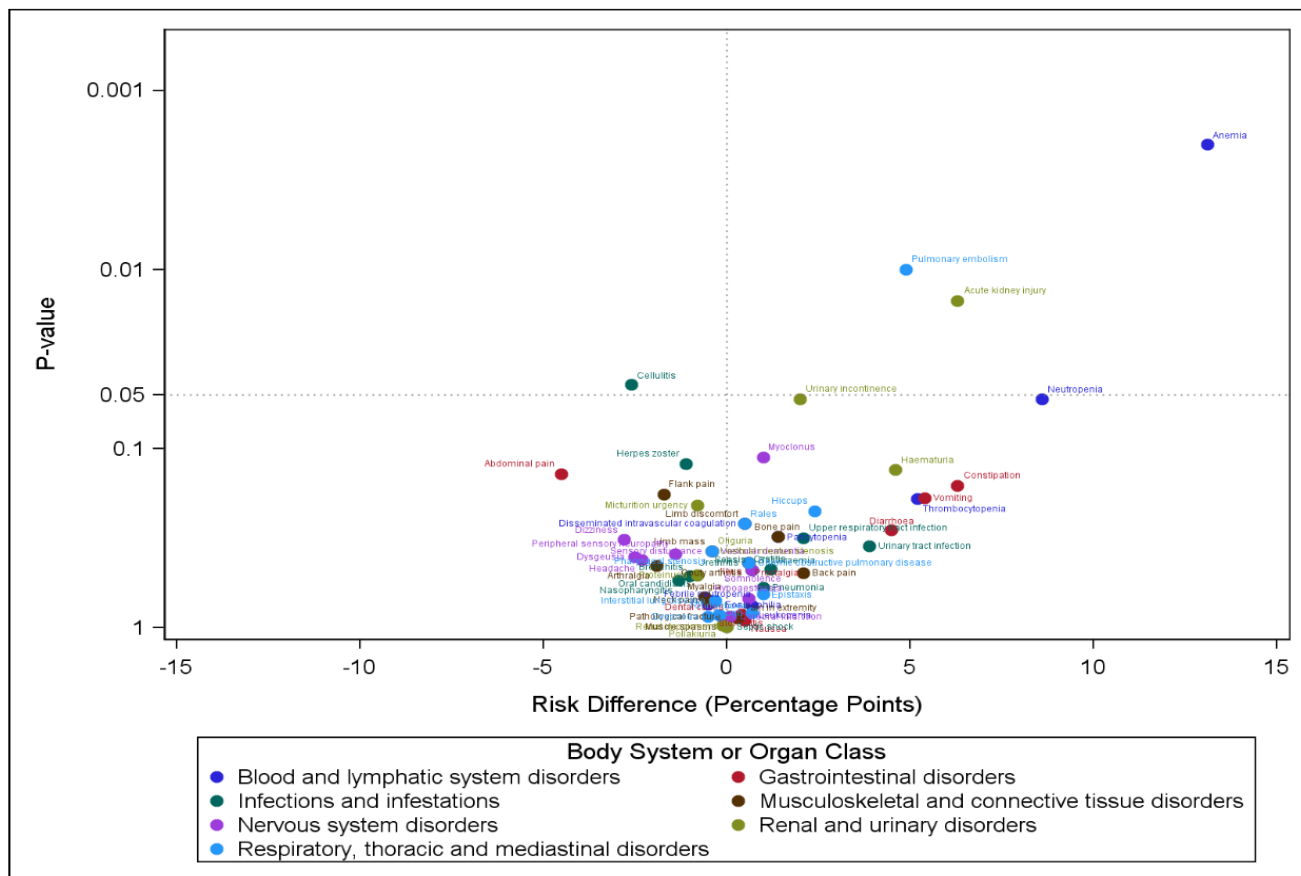


- Plot risk difference on X-axis for each SOC's preferred terms
- Plot log transformed P-values on Y-axis
- Dashed line shows $p=0.05$
- Filled circles represent an AE preferred term

Code

```
proc sgplot data=vol1v3;  
    label Risk Difference='Risk Difference (Percentage Points)';  
    label P-value='P-value';  
    scatter x=Risk y=P-value /markerattrs= (symbol=circlefilled color=black);  
    refline 0/axis=x lineattrs=(pattern=dot);  
    refline 0.05/axis=y lineattrs=(pattern=dot);  
    yaxis reverse type=log values= (1.0 0.1 0.05 0.01 0.001) offsetmin=0.1;  
    xaxis values= (-15 -10 -5 0 5 10 15);  
  
run;
```

Customized Plot 1

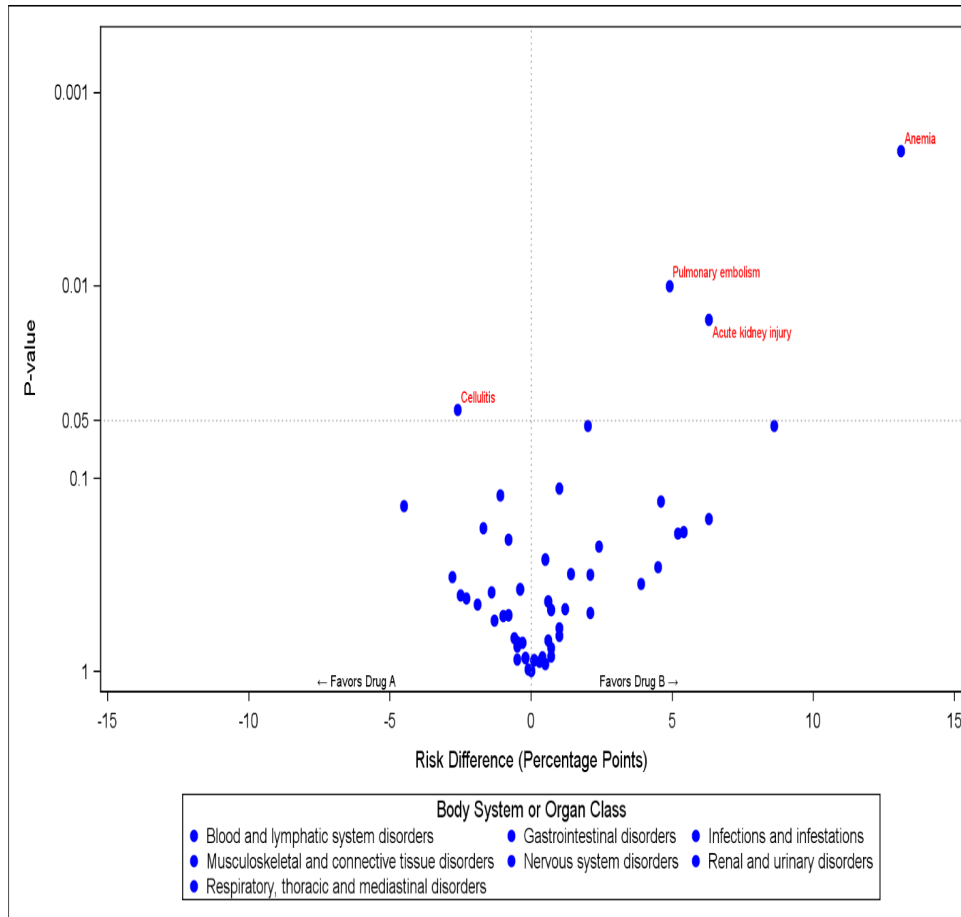


Displaying Preferred terms that belong to each SOC in same color

Code for customized plot 1

```
proc sgplot data=vol1v32;  
    label risk ='Risk Difference (Percentage Points)';  
    label pval='P-value';  
    scatter x=risk y=pval /GROUP=aebodsys DATALABEL=aedecod  
    markerattrs= (symbol=circlefilled);  
    refline 0/axis=x lineattrs=(pattern=dot);  
    refline 0.05/axis=y lineattrs=(pattern=dot);  
    yaxis reverse type=log values= (1.0 0.1 0.05 0.01 0.001) offsetmin=0.1;  
    xaxis values= (-15 -10 -5 0 5 10 15);  
run;
```

Customized plot 2



Displays labels for AE terms with p-values below 0.05 only.

Data points with low P-value appear towards top of the Plot.

Shows Favorability of an AE towards a drug.

Code for customized plot 2

```
proc format;
  value $fav
    "A" = "(*ESC*) {Unicode '2190'x} Favors Drug A"
    "B" = "Favors Drug B (*ESC*) {Unicode '2192'x}";
run;

data vol1v32;
  set vol1v3 end=last;
  if pval < 0.05 then label=aedecod;
  output;
  if last then do;
    ylbl=1.0; xlbl=-5; text="A"; output;
    ylbl=1.0; xlbl=5; text="B"; output;
  end;
run;
```

Code for customized plot 2

```
ods graphics / width=10in height=6in maxlegendarea=100;
proc sgplot data=vol1v32;
  format text $fav.;
  label risk='Risk Difference (Percentage Point)';
  label pval='P-value';
  scatter x=risk y=pval /group=aebodsys datalabel=label
  markerattrs= (symbol=circlefilled color=blue) datalabelattrs=(color=red);
  refline 0/axis=x lineattrs=(pattern=dot);
  refline 0.05/axis=y lineattrs=(pattern=dot);
  text x=xlbl y=ylbl text=text / position=bottom contributeoffsets=(ymax);
  yaxis reverse type=log values= (1.0 0.1 0.05 0.01 0.001) offsetmin=0.1;
  xaxis values= (-15 -10 -5 0 5 10 15);
run;
```

Bubble Plot

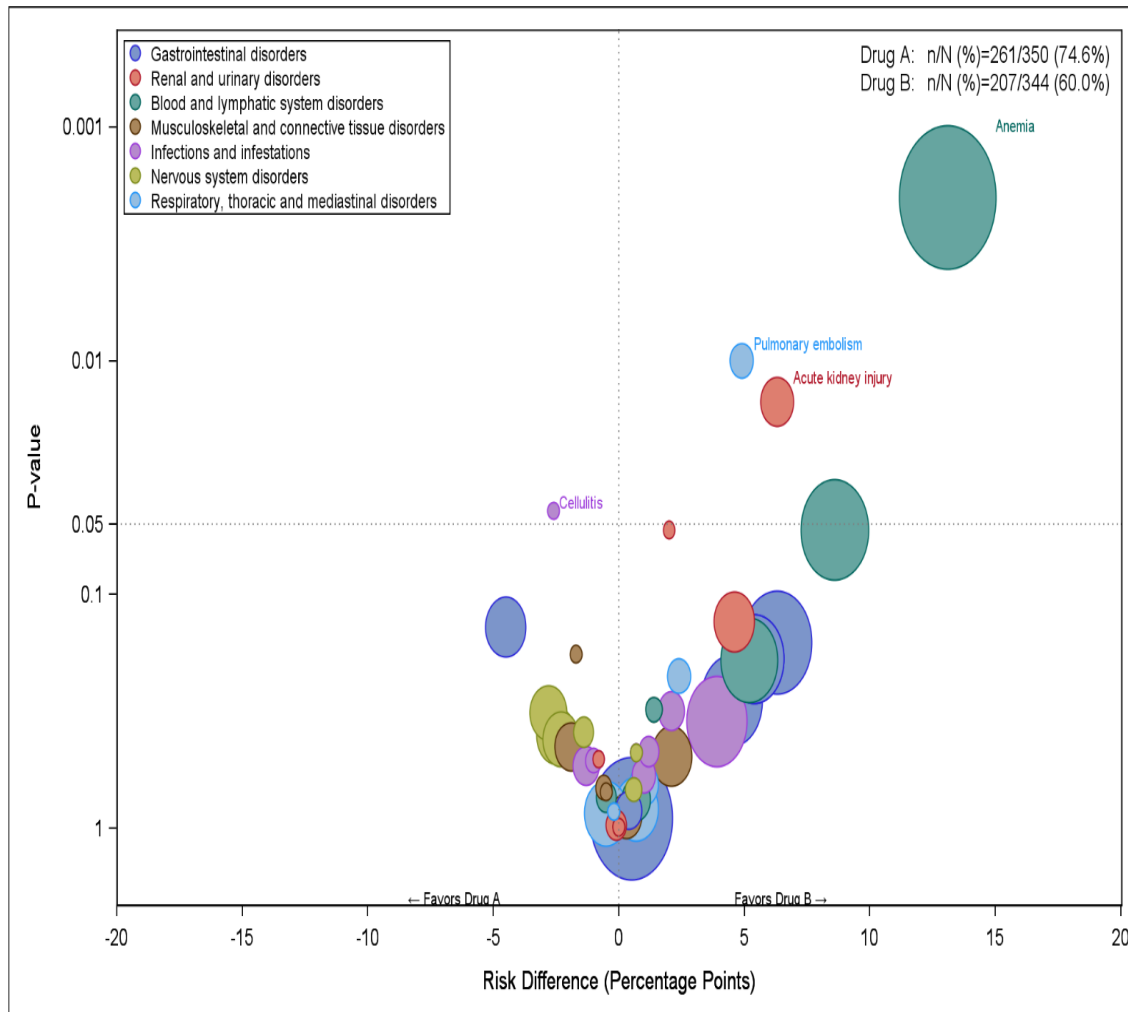
- Each bubble represents an adverse event.
- Bubble size indicates the total number of adverse events that occur in both treatments combined.
- Radius of any bubble is proportional to the square root of the event frequency.

Input data used for Bubble plot

INPUT DATA

Body System or Organ Class	Preferred Term	Risk Difference	P-Value	Bubble Size
Blood and lymphatic system disorders	Anemia	13.1	0.002	70
Infections and infestations	Cellulitis	-2.6	0.044	2
Nervous system disorders	Dizziness	-2.8	0.322	11
Gastrointestinal disorders	Abdominal pain	-4.5	0.138	13
Renal and urinary disorders	Acute kidney injury	6.3	0.015	9
Respiratory, thoracic and mediastinal disorders	Cough	0.7	0.829	15
Musculoskeletal and connective tissue disorders.....	Flank pain...	-1.7...	0.18....	2....

Bubble Plot



- Bubbles to the right indicates a higher risk for subjects in drug A
- Bubbles to the left indicates a higher risk for subjects in drug B
- Anemia is the most frequently occurring AE favoring drug B

Code for bubble plot

```
proc format;  
value $fav  
"A" = "(*ESC*) {Unicode '2190'x} Favors Drug A"  
"B" = "Favors Drug B (*ESC*) {Unicode '2192'x}";  
run;
```

```
%let N1= 261/350 (74.6%);  
%let N2= 207/344 (60.0%);
```

```
data bubble1;  
set bubble1 end=last;  
    if pval < 0.05 then label=aedecod;  
    output;  
    if last then do;  
        ylbl=1.8; xlbl=-5; text="A"; output;  
        ylbl=1.8; xlbl=8; text="B"; output;  
    end;  
run;
```

Code for bubble plot

```
proc sgplot data=bubble1;
  format text $fav.;
  label risk='Risk Difference (Percentage Points)';
  label npval='P-value';
  bubble x=risk_disp y=npval size=size / group=aebodsys datalabel=label
  name='a' bradiusmin=0 bradiusmax=39;
  refline 0/axis=x lineattrs=(pattern=dot);
  refline 0.05/axis=y lineattrs=(pattern=dot);
  inset ("Drug A:" ="n/N (%) =&N1"
  "Drug B:" ="n/N (%) =&N2") / noborder position=topright;
  text x=xlbl y=ylbl text=text / position=bottom contributeoffsets=(ymax);
  yaxis reverse type=log values= (1.0 0.1 0.05 0.01 0.001);
  xaxis values= (-20 -15 -10 -5 0 5 10 15 20) offsetmin=0 offsetmax=0;
  keylegend 'a' / across=1 location=inside valueattrs=(size=8);
run;
```

Summary

- Safety is always a primary focus in drug development process.
- Usually safety information is reported in tables which are voluminous and difficult to interpret.
- Graphical outputs can be beneficial in presenting concise summaries to communicate important implications.
- Volcano plot is one such tool used to easily identify differences in large datasets and display number of data points in the spatially constrained manner.
- This paper discussed the significance and interpretation of volcano plot and construction of graphic elements using SGPLOT procedure statements by providing an example from a late stage clinical trial.

References

- Richard C Zink, Russell D Wolfinger and Geoffrey Mann. Summarizing the incidence of adverse events using volcano plots and time intervals. Clinical Trials 2013; 10:398-406
- Mehrotra DV, Adewale AJ. Flagging clinical adverse experiences: Reducing false discoveries without materially compromising power for detecting true signals. Stat Med 2012; 31: 1918–30.
- Sanjay Matange.
<https://blogs.sas.com/content/graphicallyspeaking/2016/05/23/ctspedia-clinical-graphs-volcano-plot/#prettyPhoto>

THANK YOU !!!!!

Any Questions?