CONDENSING HIGH VOLUMES OF STATISTICAL INFORMATION INTO POWERFUL GRAPHS

Anik Chatterjee,
Principal Programmer
Novartis Healthcare Pvt. Ltd.

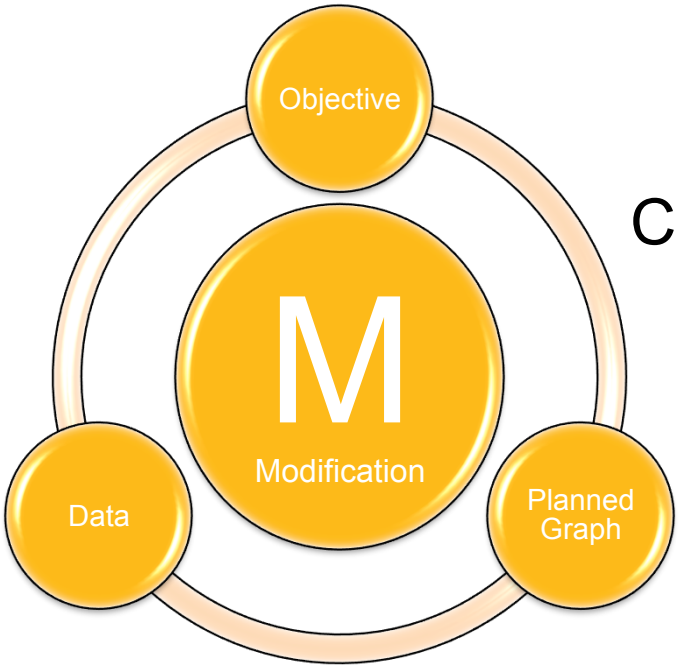
Jimmy Fernandes,
Senior Principal Programmer
Novartis Pharma AG

NOVARTIS

Agenda :

- Introduction
- Motivation
- Case Studies and Discussion

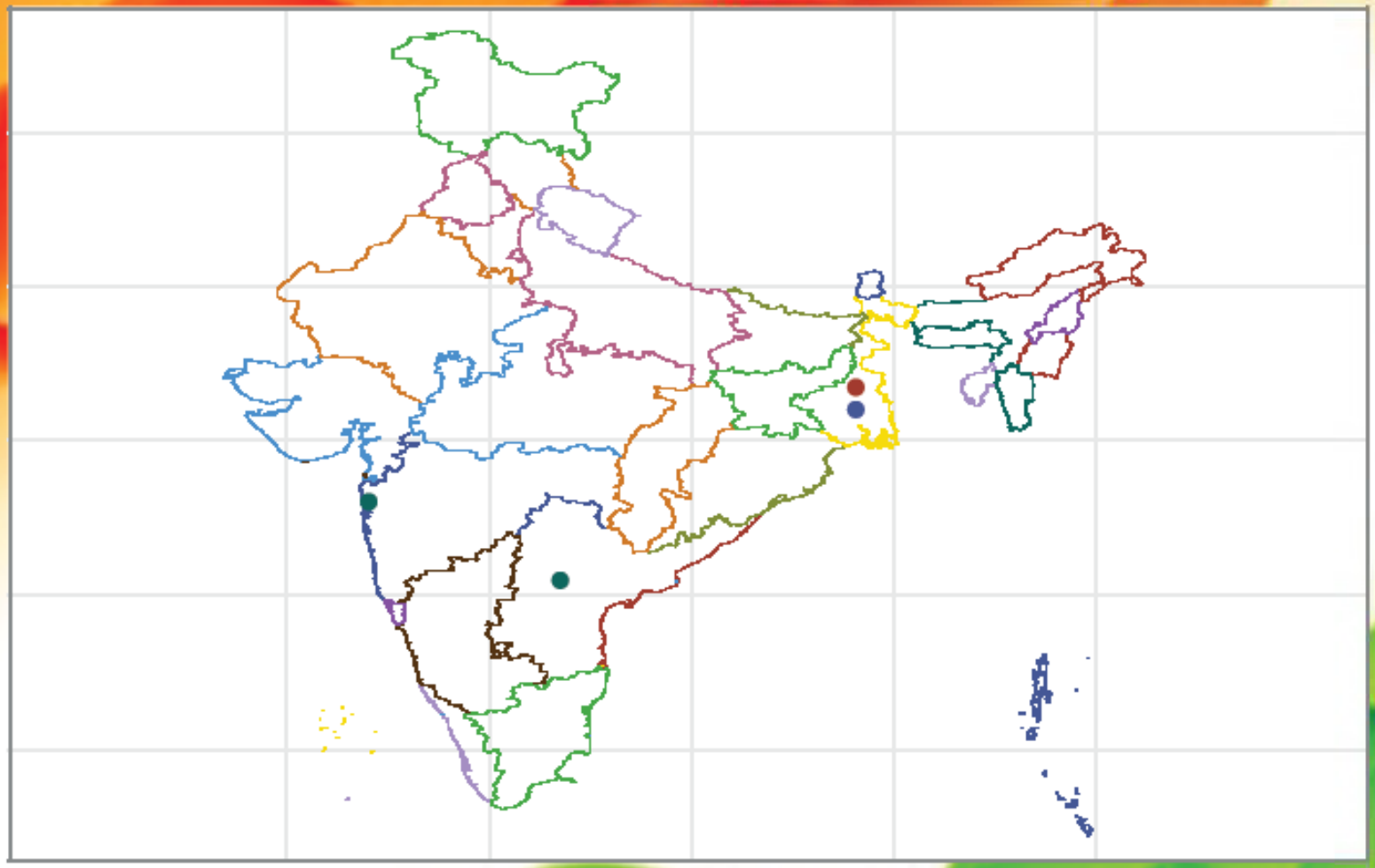
Forest Plot



Cumulative distribution plot

NOVARTIS

INTRODUCTION - Anik Chatterjee



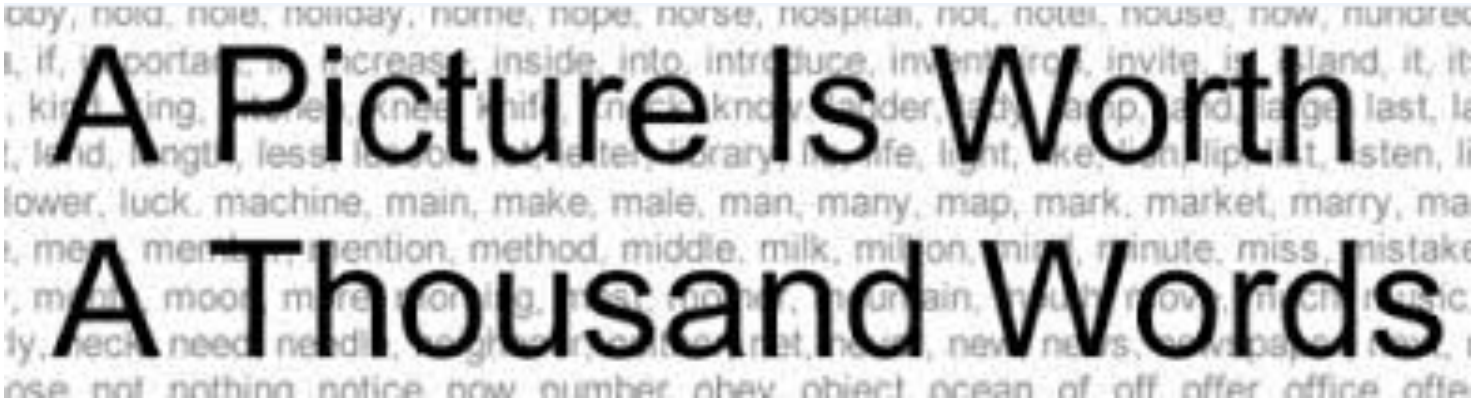
Hobbies

Introduction

```
proc template;
define statgraph Graph;
begingraph;

   /* Define the marker symbols */
   symbolimage name=edu image="/XXXX//education.jpg" / voffset=0.5;
   symbolimage name=mar image="/XXXX//marriage.jpg" /  voffset=0.5;
   symbolimage name=job image="/XXXX//job.jpg" / voffset=0.5;
    symbolimage name=son image="/XXXX//family.gif" / voffset=0.5;

   /* Create legend entries for the markers */
   legenditem type=marker name="MS-Stat" /
     markerattrs=(symbol=edu size=50px) label="";
   legenditem type=marker name="Marriage" /
     markerattrs=(symbol=mar size=70px) label="";
   legenditem type=marker name="Job" /
     markerattrs=(symbol=job size=60px) label="";
   legenditem type=marker name="Son" /
     markerattrs=(symbol=son size=70px) label="";


    /* DEFINING ATTRIBUTE MAP FOR STATUS MARKERS*/
    discreteattrmap name="wbmap";
      value "MS-Stat" / markerattrs=(symbol=edu color=blue size=55px/*symbol=circlefilled*/);
      value "Marriage" / markerattrs=(symbol=mar color=maroon size=80px/*symbol=circlefilled*/);
                                        value "Job" / markerattrs=(symbol=job color=darkgreen
size=55px/*symbol=circlefilled*/);
                                        value "Son" / markerattrs=(symbol=son color=black
size=57px/*symbol=circlefilled*/);
    enddiscreteattrmap;
    discreteattrvar attrvar=eve var=eve attrmap="wbmap";
                                        entrytitle halign=center 'INTRODUCTION - Anik Chatterjee';
  layout lattice / rowdatarange=data columndatarange=data
                rows=2 columns=2 rowgutter=0 columngutter=0 rowweights=(0.75 0.25) columnweights=(0.83 0.17);
    layout overlay / xaxisopts=(display=(line) griddisplay=on) yaxisopts=(display=(line) griddisplay=on);
      scatterplot x=x y=y / group=id name='scatter' clusterwidth=0.5 markerattrs=(size=1);
      scatterplot x=x1 y=y1 / group=cat name='scatter2' clusterwidth=0.5 markerattrs=(size=8 symbol=circlefilled);
    endlayout;

    layout overlay;
      entry _id='' halign=center '' / valign=center;
    endlayout;

    layout overlay / xaxisopts=(display=(ticks tickvalues line) tickvalueattrs=(size=8 style=italic weight=normal)
                linearopts=(viewmin=2004.0 viewmax=2009.0 tickvaluesequence=(start=2004.0 end=2009.0
increment=1.0)))
                yaxisopts=(display=(line) griddisplay=off linearopts=(viewmin=1 viewmax=1.24));
      scatterplot x=dt y=y2 / datalabel=eve group=eve name='scatter3';
      seriesplot x=dt y=y2 / name='series' connectorder=xaxis;
    endlayout;

    layout overlay;
                                        entry _id='' halign=center '' / valign=center;
    endlayout;

/*BACKGROUND IMAGE*/
                                                drawimage "/XXXX//India.jpg" /
        width=100 widthunit=percent layer=back drawspace=graphpercent scale=fitheight;

                                                drawtext textattrs=(style=italic size=8pt) "Hobbies"/
                                        anchor=top border=true width=15 widthunit=percent
                                        xspace=layoutpercent yspace=layoutpercent
                                        x=93 y=98 justify=center ;
                                                drawimage "/XXXX//run.jpg" /
        x=93 y=85 drawspace=layoutpercent width=14 height=14;
                                                drawimage "/XXXX//driving.jpg" /
        x=93 y=68 drawspace=layoutpercent width=14 height=14;
                                                drawimage "/XXXX//Imusic.jpg" /
        x=93 y=51 drawspace=layoutpercent width=14 height=14;
                                                drawimage "/XXXX//travel.jpg" /
        x=93 y=34 drawspace=layoutpercent width=14 height=14;

                                                drawimage "/XXXX//we.jpg" /
        x=93 y=16 drawspace=layoutpercent width=16 height=18;

  endlayout;
endgraph;
end;
```

Motivation

Motivation



NOVARTIS

Motivation

Let us take an example through the series of pictures

Case Study

Forest Plot

Case Study - 1

Forest Plot

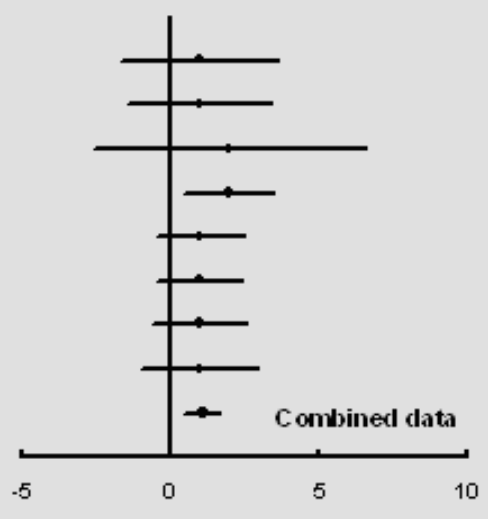
Objective:

- To compare treatment differences between two treatment arms across multiple subgroups over different visits.

Planned Graph :

A forest plot over multiple subgroups where difference between means are presented.

Subgroups are presented in Y-axis.



NOVARTIS

Case Study - 1

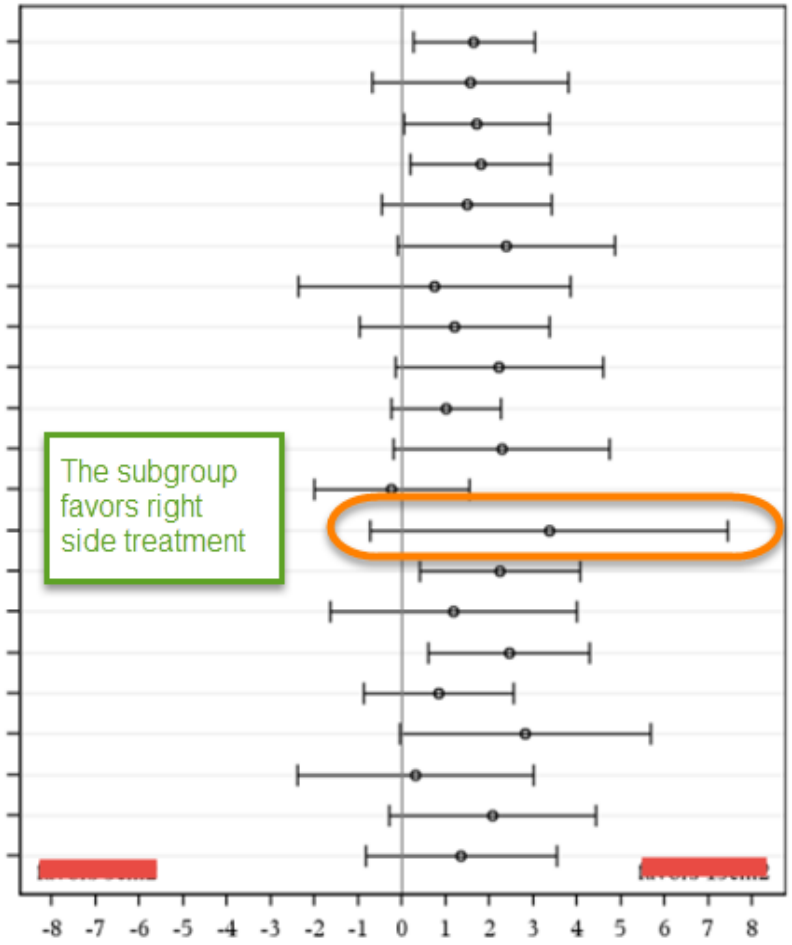
Forest Plot

More Insights about data:

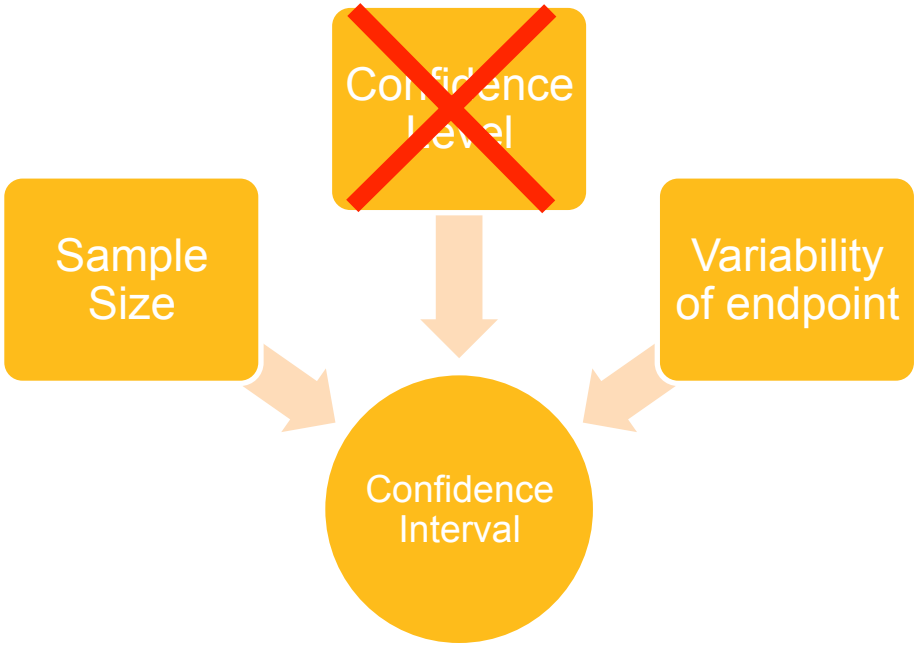
- Combined data sample size is > 600.
- More than 20 subgroups.
- Smallest subgroup size is 20 for a particular visit.
- 4 visits – baseline, week 8, week 16 and week 24

Case Study - 1

Forest Plot



What Impacts the Width of CI?



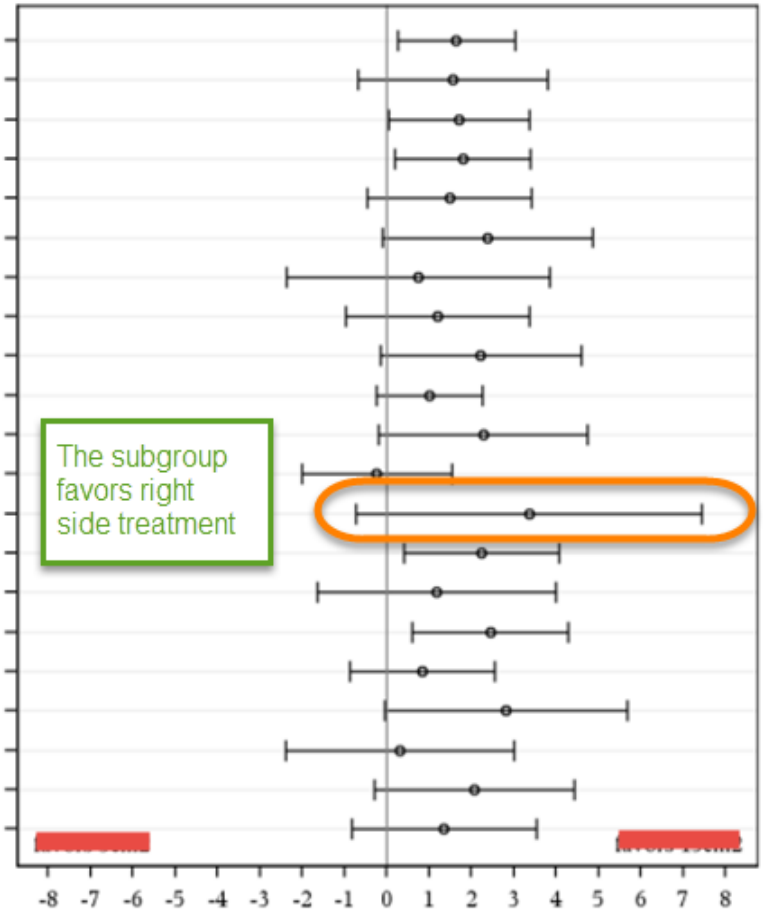
NOVARTIS

Case Study - 1

Forest Plot

- Does the subgroup have a large enough sample size?
- Is that particular subgroup consistent across all visits?

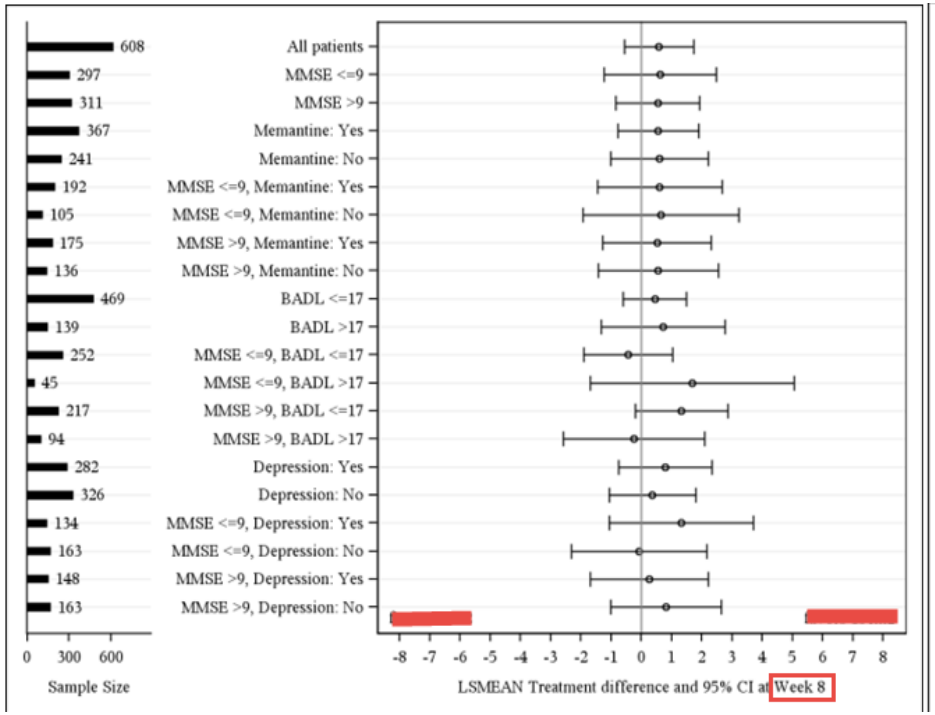
How we can make things better?

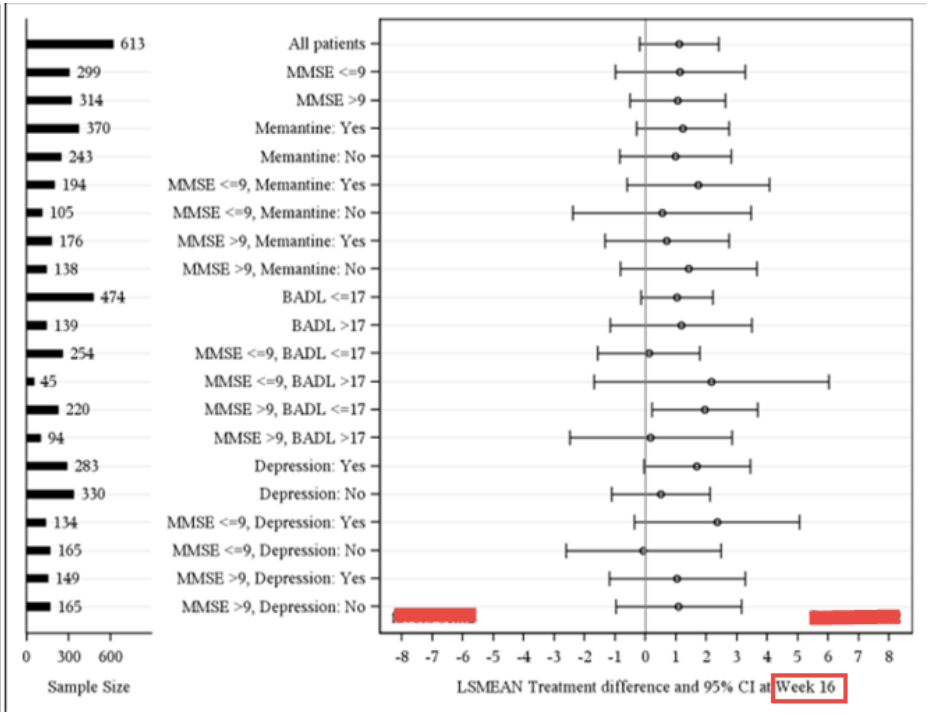


NOVARTIS

Case Study - 1

Forest Plot

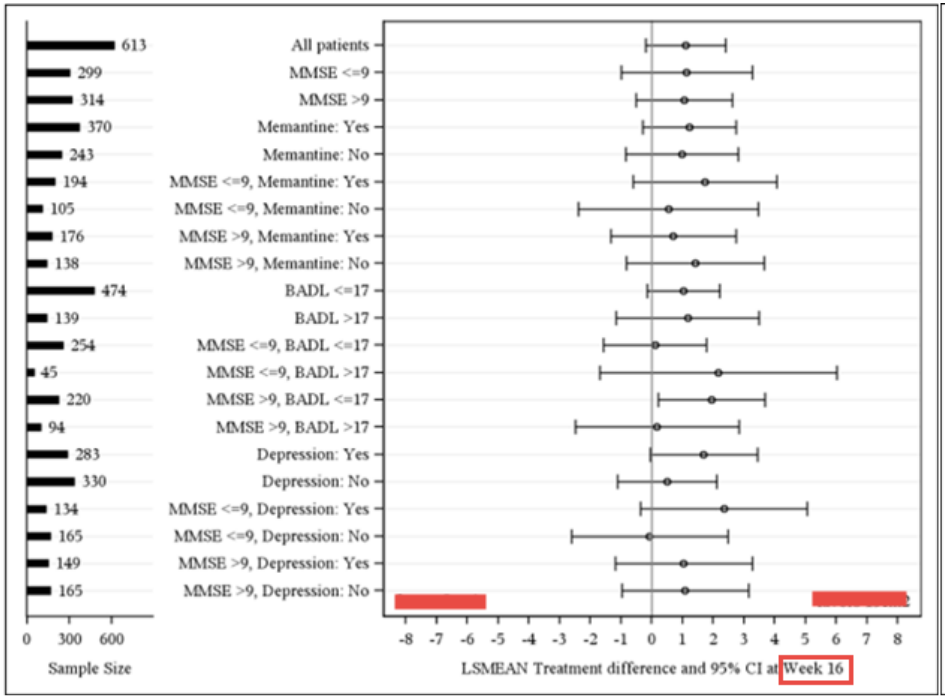


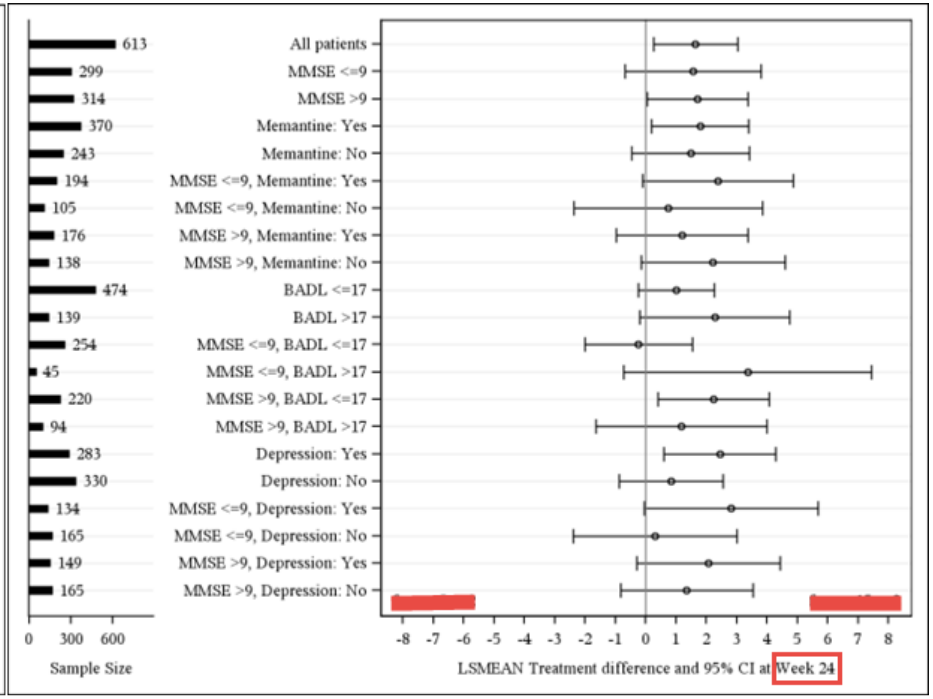


NOVARTIS

Case Study - 1

Forest Plot





NOVARTIS

Case Study - 1

Forest Plot

Summary:

Powerful plot to

- Help understand treatment effect across all subgroups simultaneously
- Compare the TE and CI across visits
- Understand if the CI may be influenced by the sample size

Execution:

In the example provided three different visits are presented which will be automatically detected by the macro during execution. Also, this graph can easily be adapted for odds ratio instead of LS Means and CI where reference line could be printed at 1 instead of 0.

The plot has two panels:

Left side panel : States the total subgroup sample size and a bar diagram for each subgroup. This illustrates different sample sizes across all subgroups.

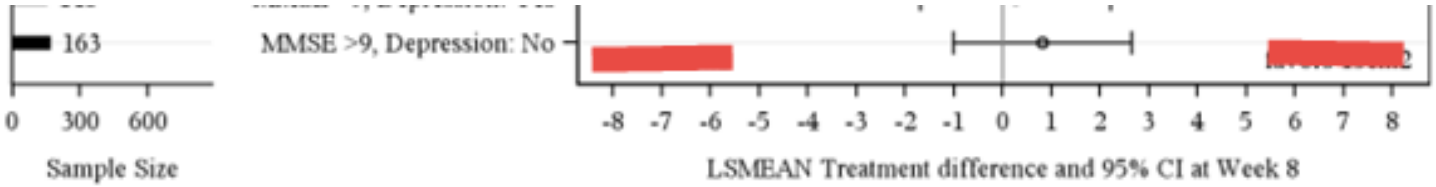
Right side panel : Forest plot comparing the estimated LS mean treatment differences and the 95% confidence intervals across all subgroups

Forest plot.txt

Case Study - 1

Forest Plot

```
xaxisopts=(griddisplay=off
           %if &lblx ne " " %then %do;
             label=(%unquote(&lblx.))
           %end;
           %else %do;
             label=("LSMEAN Treatment difference and 95% CI at %scan(&lvalue1,&lpc,|)")
           %end;
```




```
entry halign=left "&txt1" / valign=bottom textattrs=(size=8);
entry halign=right "&txt2" / valign=bottom textattrs=(size=8);
```

```
      labelattrs=(size=8) tickvalueattrs=(size=8) linearopts=(viewmin=-&xmax viewmax=&xmax
      tickvaluesequence=(start=-%SYSEVALF(&xmax-0.5) end=%SYSEVALF(&xmax-0.5) increment=1.0)))
```


NOVARTIS

Case Study - 1

Forest Plot

```
/*all distinct visits*/
proc sql noprint;
  select distinct &vis. format=8., count(distinct &vis.) into: lvalue separated by "|", :lcnt
    from forest order by &vis.;
  select distinct &vis. format=&visfmt.. into : lvalue1  separated by "|" from forest order by &vis.;

/*SELECTING AXIS RANGE FOR X AXIS AND Y AXIS*/

  select max(&item) into: ymax from forest;
  select max(abs(&up)), max(abs(&low)) into : mlw, : mup from forest;
/*SELECTING AXIS RANGE FOR BAR PLOT*/
  select ceil(max(&ss)/100)*100 into : x1mx from forest;
quit;

%let xmax = %SYSEVALF(%sysfunc(ceil(%sysfunc(max(&mup,&mlw))))+.5);

/*EXECUTING MACRO SG FOR ALL VISIT VALUE*/
%do lpc=1 %to &lcnt;
  %sg(val = %scan(&lvalue,&lpc,|));
%end;
```

Calling Proc Template and sgrender in loop for all visit consecutively

```
SYMBOLGEN:  Macro variable LCNT resolves to       3
MLOGIC(FORESTPLOT):  %DO loop beginning; index variable LPC; start value is 1; stop value is 3; by value is 1.
MLOGIC(SG):  Beginning execution.
SYMBOLGEN:  Macro variable LVALUE resolves to 3|4|5
SYMBOLGEN:  Macro variable LPC resolves to 1
MLOGIC(SG):  Parameter VAL has value 3
MLOGIC(FORESTPLOT):  %DO loop index variable LPC is now 2; loop will iterate again.
MLOGIC(SG):  Beginning execution.
SYMBOLGEN:  Macro variable LVALUE resolves to 3|4|5
SYMBOLGEN:  Macro variable LPC resolves to 2
MLOGIC(SG):  Parameter VAL has value 4
```

Case Study

Cumulative distribution plot

Case Study - 2

Cumulative distribution plot

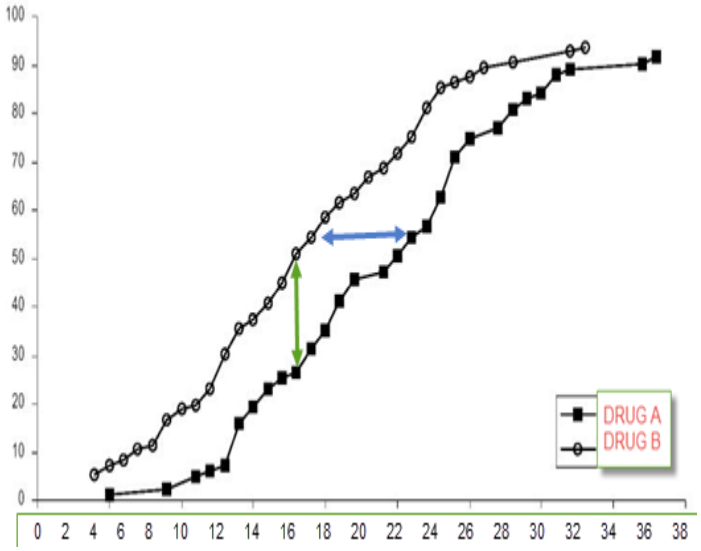
Objective:

- To compare the cumulative percentage of subjects with values below (or above) respective values on the X-axis for a particular endpoint (*change from baseline in this example*) for two different treatment groups

Planned Graph :

A cumulative plot with both the treatment groups under consideration.

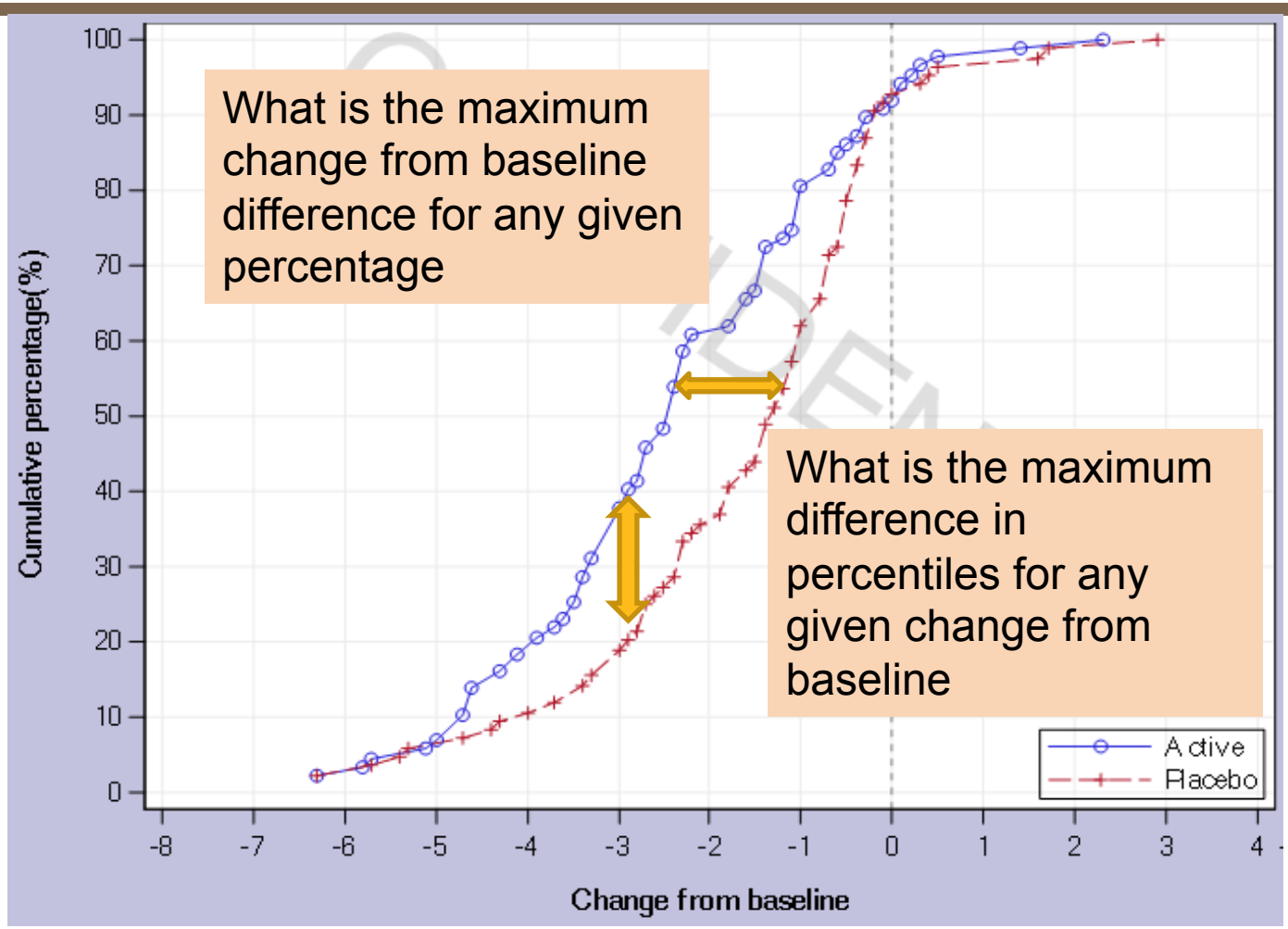
X-axis => Change from baseline
Y-axis => Cumulative Percentage(%)



NOVARTIS

Case Study - 2

Cumulative distribution plot



ISSUE: We may not have matching data points in both the treatment arms to compute the difference.

NOVARTIS

Case Study - 2

Cumulative distribution plot

How To Compare Exactly One to One

1. Create matching records for unobserved change from baseline.
2. Create matching records for unobserved cumulative percentage.
3. Interpolate values for unobserved matching data points using LOESS.

Case Study - 2

Cumulative distribution plot

How To Compare Exactly One to One

1. Create matching records for unobserved change from baseline.
2. Create matching records for unobserved cumulative percentage.
3. Interpolate values for unobserved matching data points using LOESS.

Case Study - 2

Cumulative distribution plot

Creating matching data points for unobserved change from baseline

Original Data

	treat	change	CUM_PCT
1	Active	-6.3	2.2988505747
2	Placebo	-6.3	2.380952381
3	Active	-5.8	3.4482758621
4	Active	-5.7	4.5977011494
5	Placebo	-5.7	3.5714285714
6	Placebo	-5.4	4.7619047619
7	Placebo	-5.3	5.9523809524
8	Active	-5.1	5.7471264368
9	Active	-5	6.8965517241
10	Active	-4.7	10.344827586
11	Placebo	-4.7	7.1428571429
12	Active	-4.6	13.793103448
13	Placebo	-4.4	8.3333333333
14	Active	-4.3	16.091954023
15	Placebo	-4.3	9.5238095238
16	Active	-4.1	18.390804598
17	Placebo	-4	10.714285714
18	Active	-3.9	20.689655172

With Matching Data Points for Unobserved Changes

	treat	change	CUM_PCT
1	Active	-6.3	2.2988505747
2	Active	-5.8	3.4482758621
3	Active	-5.7	4.5977011494
4	Active	-5.4	.
5	Active	-5.3	.
6	Active	-5.1	5.7471264368
7	Active	-5	6.8965517241
8	Active	-4.7	10.344827586
9	Active	-4.6	13.793103448
10	Active	-4.4	.
11	Active	-4.3	16.091954023
12	Active	-4.1	18.390804598
13	Active	-4	.
14	Active	-3.9	20.689655172
15	Active	-3.7	21.83908046
16	Active	-3.6	22.988505747
17	Active	-3.5	25.287356322
18	Active	-3.4	28.735632184

Case Study - 2

Cumulative distribution plot

How To Compare Exactly One to One

1. Create matching records for unobserved change from baseline.
2. Create matching records for unobserved cumulative percentage.
3. Interpolate values for unobserved matching data points using LOESS.

Case Study - 2

Cumulative distribution plot

Creating matching data points for unobserved cumulative percentage

Original Data

treat	change	CUM_PCT
Active	-6.3	2.2988505747
Placebo	-6.3	2.380952381
Active	-5.8	3.4482758621
Placebo	-5.7	3.5714285714
Active	-5.7	4.5977011494
Placebo	-5.4	4.7619047619
Active	-5.1	5.7471264368
Placebo	-5.3	5.9523809524
Active	-5	6.8965517241
Placebo	-4.7	7.1428571429
Placebo	-4.4	8.3333333333
Placebo	-4.3	9.5238095238
Active	-4.7	10.344827586
Placebo	-4	10.714285714
Placebo	-3.7	11.904761905
Active	-4.6	13.793103448
Placebo	-3.4	14.285714286
Placebo	-3.3	15.476190476

With Matching Data Points For Unobserved Cumulative Percentages

	treat	change	Cumulative Percent
1	Active	-6.3	2.2988505747
2	Placebo	.	2.2988505747
3	Active	.	2.380952381
4	Placebo	-6.3	2.380952381
5	Active	-5.8	3.4482758621
6	Placebo	.	3.4482758621
7	Active	.	3.5714285714
8	Placebo	-5.7	3.5714285714
9	Active	-5.7	4.5977011494
10	Placebo	.	4.5977011494
11	Active	.	4.7619047619
12	Placebo	-5.4	4.7619047619
13	Active	-5.1	5.7471264368
14	Placebo	.	5.7471264368
15	Active	.	5.9523809524
16	Placebo	-5.3	5.9523809524
17	Active	-5	6.8965517241

Case Study - 2

Cumulative distribution plot

How To Compare Exactly One to One

1. Create matching records for unobserved change from baseline.
2. Create matching records for unobserved cumulative percentage.
3. Interpolate values for unobserved matching data points using LOESS.

Case Study - 2

Cumulative distribution plot

Interpolate matching values using LOESS for cumulative percentages

```
proc loess data=out1;
   model cum_pct = change/smooth=0.05;
   score data=PredPoints1;
   ods Output ScoreResults=ScoreOut1;
   by treat;
run;
```

treat	change	CUM_PCT	p_CUM_PCT
Active	-6.3	2.2988505747	2.2988505747
Active	-5.8	3.4482758621	3.4482758621
Active	-5.7	4.5977011494	4.5977011494
Active	-5.4	.	5.1724137931
Active	-5.3	.	5.3639846743
Active	-5.1	5.7471264368	5.7471264368
Active	-5	6.8965517241	6.8965517241
Active	-4.7	10.344827586	10.344827586
Active	-4.6	13.793103448	13.793103448
Active	-4.4	.	15.325670498

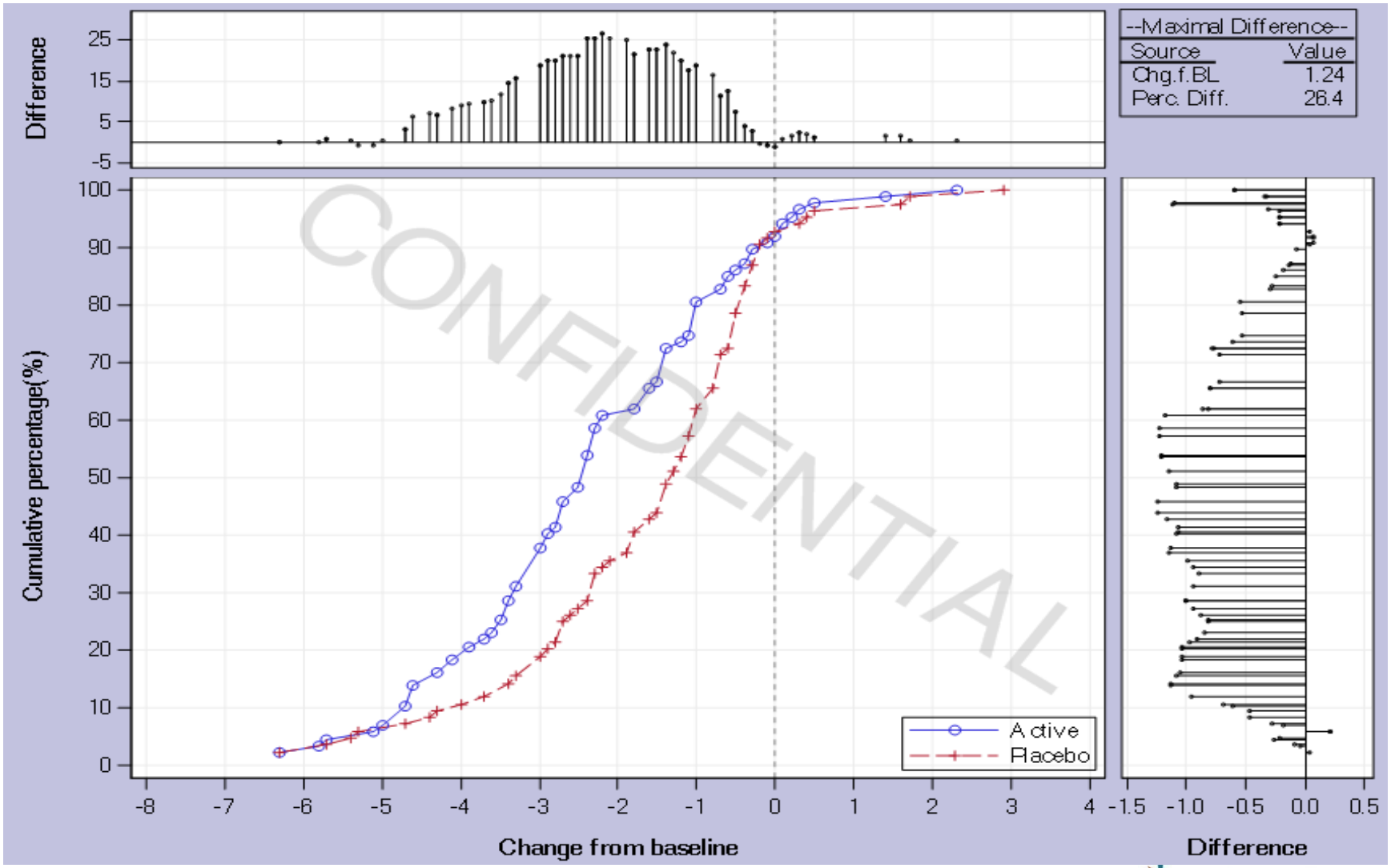
Interpolate matching values using LOESS for cumulative change

```
proc loess data=out1;
   model change = cum_pct/smooth=0.05;
   score data=PredPoints2;
   ods Output ScoreResults=ScoreOut2;
   by treat;
run;
```

treat	CUM_PCT	change	p_change
Active	2.2988505747	-6.3	-6.3
Active	2.380952381	.	-6.264285714
Active	3.4482758621	-5.8	-5.8
Active	3.5714285714	.	-5.789285714
Active	4.5977011494	-5.7	-5.7
Active	4.7619047619	.	-5.614285714
Active	5.7471264368	-5.1	-5.1
Active	5.9523809524	.	-5.082142857
Active	6.8965517241	-5	-5
Active	7.1428571429	.	-4.978571429
Active	8.3333333333	.	-4.875
Active	9.5238095238	.	-4.771428571
Active	10.344827586	-4.7	-4.7

Case Study - 2

Cumulative distribution plot



NOVARTIS

Case Study - 2

Cumulative distribution plot

Summary

The plot helps to provide an overview of :

- Distribution of the differences (both horizontal and vertical)
- The difference in percentiles for any given change from baseline (and vice versa)

TRT comparison.txt

Suggestion and Discussion

Anik Chatterjee
Principal Programmer
Novartis, India

Contact
anik.chatterjee@novartis.com
anikchatterjee@gmail.com

Jimmy Fernandes
Senior Principal Programmer
Novartis Pharma AG, Switzerland

Contact
jimmy.fernandes@novartis.com

NOVARTIS