SESUG Paper 055-2025

Versatile Graphic Presentation of Immunogenicity Analysis Results in Vaccine Studies with a Macro that Creates Multiple Forest Plots in One Output

Hong Wang, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

Immunogenicity analysis is a crucial aspect of vaccine clinical trials. This analysis aims to evaluate the immune responses elicited by a vaccine candidate, enabling researchers and stakeholders to assess the vaccine's effectiveness in inducing protective immunity against specific pathogens. Subgroup analysis within immunogenicity studies is essential for understanding how different populations respond to a vaccine. It helps identify variations in immune responses based on demographic, clinical, or other relevant factors.

Forest plots play a vital role in summarizing and visualizing the results of immunogenicity analyses, especially in subgroup immunogenicity analyses where the forest plots of different subgroups are compared to evaluate the overall effectiveness and variability of vaccine responses across these groups.

A SAS macro has been developed to generate multiple forest plots in a single RTF output. This tool is particularly beneficial for immunogenicity subgroup analyses, as it provides a clear and concise comparison of immunogenicity results for various subgroups simultaneously, with one individual forest plot generated for each group.

This paper presents an overview of the macro's design, key features, and required input data. Additionally, it shows example outputs generated by the macro in different formats.

INTRODUCTION

Forest plots are a powerful visualization tool commonly used in immunogenicity analyses in vaccine clinical trials, as they provide a clear and concise way to present the results for multiple analysis parameters in a single visual representation. In immunogenicity subgroup analyses of the vaccine studies, the multiple forest plots shown in one output are especially useful as they facilitate the comparison of immunogenicity outcomes across different subgroups, allowing for a deeper understanding of how different factors (e.g., age, sex, race, or pre-existing conditions, etc.) may influence immunogenicity.

Given the vast need on this kind of forest plots in immunogenicity analyses of vaccine clinical trials, a SAS macro named **graph0forestplot0multiple** has been developed using SAS Graph Template Language (GTL). With help of SAS GTL and macro language, it provides diverse options for users to generate multiple forest plots in a single RTF output file with versatile layouts and formats. This multi-plot macro is effective for the commonly analyzed immunogenicity results, such as ratios between two different vaccines on geometric mean titers (GMTs) of opsonophagocytic activity (OPA), geometric mean concentrations (GMCs) of immunoglobulin G (IgG), or differences between vaccines on rates of specific immunologic responses, etc. across multiple subgroups.

MACRO DESIGN AND KEY FEATURES

To support the comparison of the immunogenicity results across multiple subgroups in a flexible way, this macro generates forest plots with versatile layouts and formats.

It can generate either two or three individual forest plots in a single RTF output at user’s choice, with one forest plot for each subgroup. This is quite useful in immunogenicity subgroup analysis, enabling a concise and efficient comparison of results across different subgroups, and aiding reviewers understand the commonness and differences among subgroups clearly.

Each individual forest plot can be further split into the top and the bottom parts if needed, with independent settings of axes, headers, and labels, etc. for each part, controlled by users. This feature fits

the need when different results in a plot are analyzed differently. For instance, if a study vaccine has both shared and unique serotypes (a serotype is a specific subgroup of a pathogen, against which a vaccine has immunogenicity), compared to the comparator, and those two kinds of serotypes are displayed with different x-axis ticks, this feature may be utilized.

Another feature of this macro is that user can choose to display the supporting data for each individual plot, which helps reviewers get more reference information. These data may include the source data used to generate the forest plots, such as OPA GMT or IgG GMC ratios. Additionally, the ranges for numbers of subjects contributing to the analysis, across all serotypes, are displayed as well, in a format of "**n1=(min, max) n2=(min, max)**", where "n1" and "n2" represent different vaccine treatment groups, and "min" and "max" denote the minimum and maximum numbers of subjects contributing to analysis across all serotypes, respectively, in each plot.

Furthermore, the vertical reference lines can be displayed at specific points on x-axis in selected plot(s) if needed, and the x-axis positions for those lines are independent in each plot per user's choice. Along with the presence of reference lines, the macro shows another functionality: the color of those serotypes that meet specific criterion can change to a different one at user's choice, e.g., if a reference line of 0.5 is specified in a given subgroup, and the lower 95% confidence interval (95% CI) of OPA GMT ratio for a specific pneumococcal serotype in this subgroup is lower than 0.5, then the color of that serotype changes to red, while the color of other serotypes in that subgroup still keeps the default black. This color difference clearly shows whether any pre-defined statistical hypothesis (e.g., superiority or non-inferiority of the study vaccine vs. the comparator) is fulfilled at specific serotypes. Multiple criteria are available to provide flexibility in exploration of the results from multiple analysis perspectives:

- if the lower limit of 95% CI of the analyzed point estimate (e.g., ratio of GMT OPA or GMC IgG, etc.) for a specific serotype is lower than or equal to the value of the reference line. This is useful to verify the hypotheses for non-inferiority of a vaccine compared to another one.
- if the upper limit of 95% CI of the analyzed point estimate for a specific serotype is higher than the value of the reference line. This can be used to verify the hypotheses for superiority of a vaccine over another.
- users can also select criteria that if the analyzed point estimate is lower than or greater than 1 so the color of specific serotypes that meet those criteria changes to another one.

Macro graph0forestplot0multiple also offers various cosmetic features: users can set different x-axis ticks in different plots, display those ticks in either linear or logarithmic scale, adjust the sizes of top/bottom parts for the individual plots as well as the total graphic area in the RTF output, sorting the serotypes in the plots in a specific order, or change the plot color etc., for an optimal display that serves the analysis well.

The basic structure of SAS GTL used in this macro can be summarized as the following:

```
proc template;
  define statgraph forest_plot_template_name;
    begingraph;
      LAYOUT LATTICE statement for numbers of individual plots (2 or 3);
        SCATTERPLOT statements to generate individual plots;
        AXISTABLE statements to display source data, if needed;
        SIDEBAR statements for column headers and specific labels;
        REFERENCELINE statements for vertical reference lines as needed;
        …
      ENDLAYOUT;
      *** same layout lattice statement as above is included for the
          bottom part of the plots if each individual forest plot is split
          into top and bottom parts per user's choice;
    endgraph;
  end;
run;
```

In general, this macro is powerful and versatile. Even though initially designed for immunogenicity subgroup analyses in vaccine studies, it yet provides a lot of flexibility for users to create forest plots for any other analyses where applicable. As long as the structure of the input data complies with the corresponding requirements specified in the “INPUT DATA” section below, this macro will generate the ideal output.

INPUT DATA

To generate multiple forest plots correctly by this macro, the input data with specific formats are needed.

The input data should contain the corresponding information of all subgroups represented in the plots, in either multiple input datasets (one for each subgroup) or a single dataset (with one subset of that dataset for each subgroup).

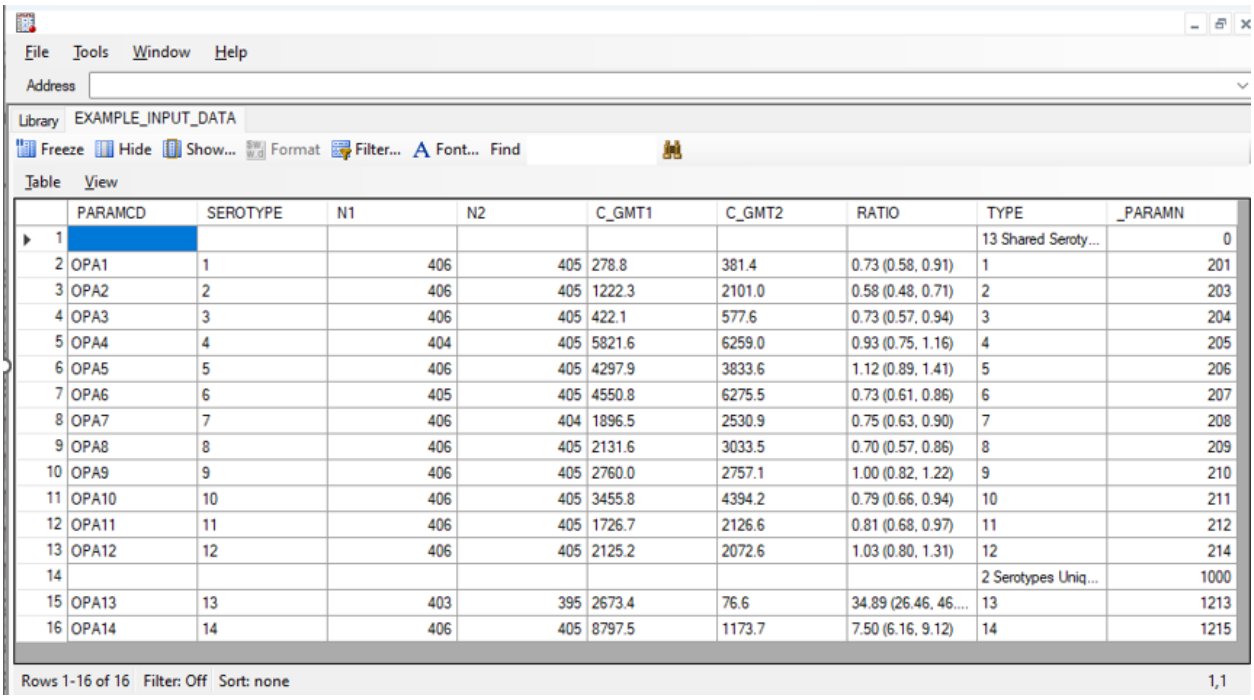
As this macro was originally created for vaccine immunogenicity analyses, a variable containing serotypes is assumed in the input data, and each serotype in that variable is displayed as a row label in the plots. Other category variables besides serotype are also acceptable, though, if needed.

A variable that contains the analysis results to be displayed in the forest plots, must be present in the input data, and those results should be in a format of “**point estimate (lower confidence interval, upper confidence interval)**”. The “point estimate” may refer to the OPA GMT or IgG GMC, etc.

Also, the two numeric variables, with the numbers of subjects contributing to the analysis, for the two vaccine treatment groups in comparison, respectively, should be available in the input data as well.

In immunogenicity analyses of vaccine clinical trials, the forest plots usually support corresponding immunogenicity tables and are usually generated after those tables, so the output data of the immunogenicity tables may be designed to include all required variables of this forest plot macro, therefore they can be directly used as input data of this macro to generate desirable forest plots in an efficient way.

The following screenshot shows an example input SAS dataset of macro graph0forestplot0multiple to generate a RTF output with three individual forest plots for OPA GMT ratios between a polyvalent pneumococcal study vaccine and its comparator, and this dataset is the input data for one of the three plots. The variable RATIO is used as the source data to create the forest plot, and its values can also be displayed in the output as supporting data if user chooses to do so. The variable SEROTYPE will be used as the row labels for the forest plots, order by variable _PARAMN. Variables N1 and N2 are used to calculate the ranges for numbers of subjects contributing to analysis in the output. The output for this input dataset (along with two other input datasets not shown in this paper but with similar structure), is shown in “OUTPUT STRUCTURE” section below as Figure 1.



	PARAMCD	SEROTYPE	N1	N2	C_GMT1	C_GMT2	RATIO	TYPE	_PARAMN
1								13 Shared Seroty...	0
2	OPA1	1	406	405	278.8	381.4	0.73 (0.58, 0.91)	1	201
3	OPA2	2	406	405	1222.3	2101.0	0.58 (0.48, 0.71)	2	203
4	OPA3	3	406	405	422.1	577.6	0.73 (0.57, 0.94)	3	204
5	OPA4	4	404	405	5821.6	6259.0	0.93 (0.75, 1.16)	4	205
6	OPA5	5	406	405	4297.9	3833.6	1.12 (0.89, 1.41)	5	206
7	OPA6	6	405	405	4550.8	6275.5	0.73 (0.61, 0.86)	6	207
8	OPA7	7	406	404	1896.5	2530.9	0.75 (0.63, 0.90)	7	208
9	OPA8	8	406	405	2131.6	3033.5	0.70 (0.57, 0.86)	8	209
10	OPA9	9	406	405	2760.0	2757.1	1.00 (0.82, 1.22)	9	210
11	OPA10	10	406	405	3455.8	4394.2	0.79 (0.66, 0.94)	10	211
12	OPA11	11	406	405	1726.7	2126.6	0.81 (0.68, 0.97)	11	212
13	OPA12	12	406	405	2125.2	2072.6	1.03 (0.80, 1.31)	12	214
14								2 Serotypes Uniq...	1000
15	OPA13	13	403	395	2673.4	76.6	34.89 (26.46, 46....	13	1213
16	OPA14	14	406	405	8797.5	1173.7	7.50 (6.16, 9.12)	14	1215

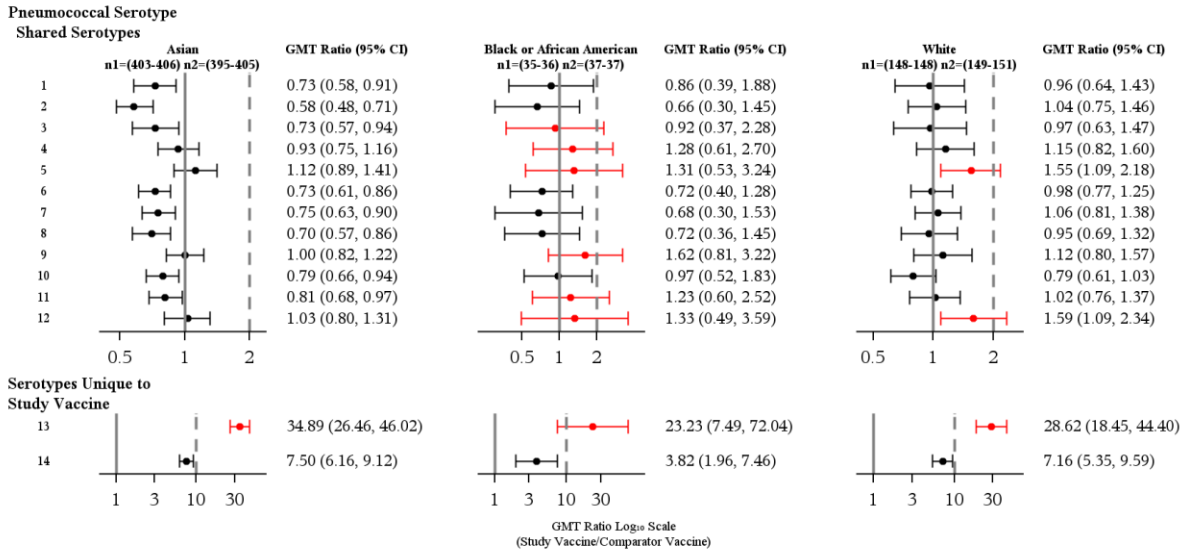
OUTPUT STRUCTURE

An example output created by macro graph0forestplot0multiple is shown below as Figure 1, for comparison of OPA GMT ratios between study and comparator vaccines on selected pneumococcal serotypes among different race groups. It has three individual forest plots for respective race groups, and each individual plot is split into top and bottom parts, with unique headers and x-axis scales. The source data for those plots as well as the numbers of subjects contributing to the analysis (n1 and n2) are also displayed. Additionally, any serotypes with upper 95% CI value greater than those of the corresponding dashed reference line, are highlighted in red, indicating the superiority of study vaccine over the comparator on those serotypes, in terms of OPA GMT.

Figure 1 also illustrates a few other features of this macro: logarithmic scale is used for x-axis to fit the wide range of input data, and the x-axis ticks and the positions of reference lines between the upper and lower plots are different to accommodate the different analysis needs, etc.

Note: all data and plots shown in this figure are from dummy data for demonstration purpose only.

Figure 1 Example Forest Plots of Post-Baseline OPA GMT Ratios by Race (Per-Protocol Population)

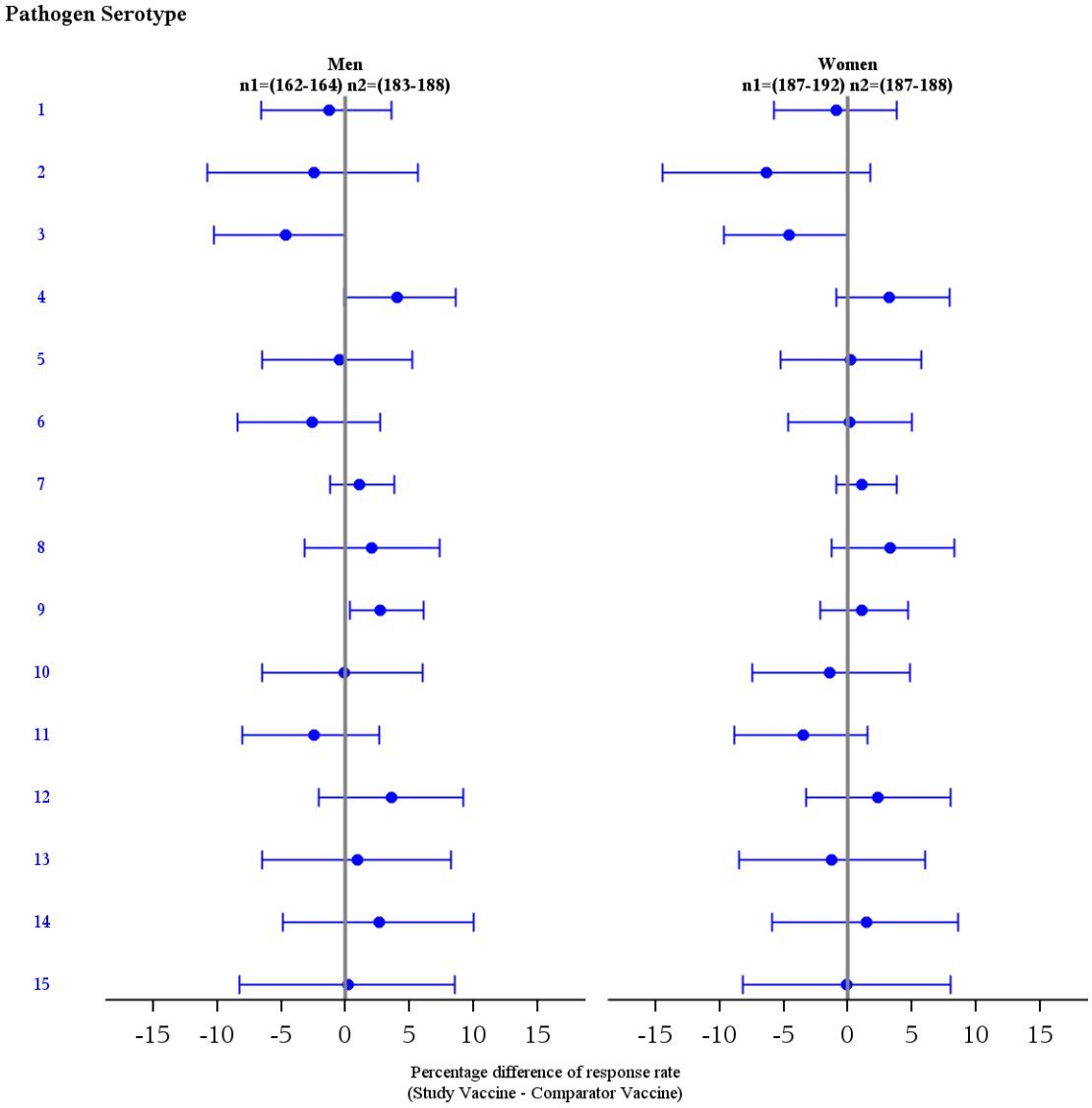


n1=Number of subjects contributing to the analysis from study vaccine across all serotypes

n2=Number of subjects contributing to the analysis from comparator vaccine across all serotypes

The following is another example output (Figure 2), created by same macro, but in a different format: two individual forest plots instead of three included in the output, source data not displayed, each individual forest plot not split into two parts, reference lines not displayed, linear instead of log x-axis scale used, and a different color of the forest plots present, etc.

Figure 2 Example Forest Plots of Post-Baseline Immune Response Rate Difference[a] by Sex (Per-Protocol Population)



[a] Immune response rate difference refers to the difference between the study vaccine and comparator groups on percentage of subjects having specific immune response.

Given the versatility of this macro, users can generate output in various formats other than those of Figures 1 and 2, to meet specific analysis needs, by adjusting different macro parameters that control number of individual forest plots to be displayed, whether individual plots need to be split into top and bottom parts, with or without source data and/or reference lines, and linear or log x-axis scale, etc.

EXAMPLES OF MACRO CALLS

The following is an example macro call, and the output of this call is shown in Figure 1:

```
%graph0forestplot0multiple(
outdir=%str(&outputpath.),
outname=example_call,
titles=Example Forest Plots of Post-Baseline OPA GMT Ratios by Race~-2n
      (Per-Protocol Population),
footers=%str(n1=Number of subjects contributing to the analysis from study
               vaccine across all serotypes~-2n
             n2=Number of subjects contributing to the analysis from
               comparator vaccine across all serotypes),
panel23=3,
ratiodiff=r,
inds1=work.example_input_data,
inds2=work.example_input_data2,
inds3=work.example_input_data3,
wheret1=%str(if serotype in ('1' '2' '3' '4' '5' '6' '7' '8'
             '9' '10' '11' '12')),
wheret2=%str(if serotype in ('1' '2' '3' '4' '5' '6' '7' '8'
             '9' '10' '11' '12')),
wheret3=%str(if serotype in ('1' '2' '3' '4' '5' '6' '7' '8'
             '9' '10' '11' '12')),
whereb1=%str(if serotype in ('13' '14')),
whereb2=%str(if serotype in ('13' '14')),
whereb3=%str(if serotype in ('13' '14')),
stype1=serotype,
stype2=serotype,
stype3=serotype,
rdci1=ratio,
rdci2=ratio,
rdci3=ratio,
smallnf1=n1,
smallnf2=n1,
smallnf3=n1,
smallns1=n2,
smallns2=n2,
smallns3=n2,
hdrtop=Pneumococcal Serotype,
hdrtop2=%str(  Shared Serotypes),
hdrbtm=%str(Serotypes Unique to),
hdrbtm2=Study Vaccine,
colhdt1=Asian,
colhdt2=Black or African American,
colhdt3=White,
tickt1=0.5 1 2,
tickt2=0.5 1 2,
tickt3=0.5 1 2,
tickb1=1 3 10 30,
tickb2=1 3 10 30,
tickb3=1 3 10 30,
reflnt1=2,
reflnt2=2,
reflnt3=2,
reflnb1=10,
reflnb2=10,
```

```
reflnb3=10,
lblxb=GMT Ratio Log~{unicode '2081'x}~{unicode '2080'x} Scale|
      (Study Vaccine/Comparator Vaccine),
sortvar=_paramn,
xtype=log,
fgcolor=black,
fgcolorx=red,
grhprtn=67 33,
bottomgraph=Y,
grhhtl=4.2,
crit=2,
datacolumn=Y,
coldata=%nrstr(GMT Ratio (95% CI)));
```

Here are several notes for a better understanding of above code:

- The value “&outputpath” of the macro parameter “outdir” specifies the path of the folder where the output RTF file of this macro will be generated.
- The macro parameter “outname” defines the actual RTF output file name.
- “~-2n” shown in the macro call indicates a soft return to change lines in titles, footnotes, and labels of the plot, where “~” is the escape character setup for RTF output, default to this macro.
- The Unicode used in x-axis label is for display of special characters like subscripted texts and the trademark, etc.
- The macro parameter “panel23” prompts user to select either two or three individual forest plots to be displayed; “bottomgraph=Y” indicates that each individual forest plot is split into two (top and bottom) parts; “datacolumn=Y” ensures the macro display the source data of the forest plots.
- The parameters “inds1-3” specify the input dataset names for the three individual plots, respectively, in the format of “library.dataset”.
- The parameters “wheret1-3” and “whereb1-3” indicate the selection criteria of the input datasets for top and bottom parts, respectively, of the three corresponding individual plots.
- The parameter “ratiodiff” indicates if ratio or difference will be displayed for specific immunogenicity analysis results. A solid vertical line is displayed at x=1 in each forest plot if the ratios of the analysis results between two treatment groups are displayed (i.e., if “ratiodiff=r” for Figure 1), while that line is displayed at x=0 for difference between treatment groups when “ratiodiff=d”.
- The variable names in input datasets for analysis results (i.e., ratio of OPA GMT between two treatment groups in Figure 1), and numbers of subjects contributing to analysis in the two treatment groups are specified in macro parameters “rdci1-3”, “smallnf1-3”, and “smallns1-3”, for the three individual forest plots, respectively.
- “xtype=log” creates base 10 log scale for x-axis in each plot, instead of linear scale, to fit the actual source data; Parameters “tickt1-3” and “tickb1-3” allow users to input the customized ticks displayed for the top and bottom parts in the three individual plots, respectively.
- Macro parameter “sortvar” indicates name of the variable the values of which are sorted to display the rows of the plots. In Figure 1, the serotypes are displayed in the ascending order of this variable.
- Macro parameters reflnt1-3 and reflnt1-3 define the positions of the reference lines displayed at x-axes, for top and the bottom parts of the three individual forest plots, respectively. If the value for one of those parameters is missing then no reference line will be displayed in that corresponding part/plot; “crit=2” in the example call indicates that if the upper limit of 95% CI of the analysis result for a specific serotype is greater than the value of the corresponding reference line, the

color of that serotype changes to red (controlled by macro parameter “fgcolorx”) in corresponding plot.

- Other parameters in the example macro call mainly focus on the texts displayed in different elements of the plots, such as titles, footnotes, column headers, and axis labels, etc.

CONCLUSION

The macro **graph0forestplot0multiple**, which generates multiple forest plots in a single RTF output, is a valuable tool for vaccine immunogenicity analyses as well as any other applicable studies. It enables users to customize their output to meet complex user requirements, providing a clear and comprehensive visual representation of data across multiple subgroups.

With sophisticated control over the formatting and layout of the output, this macro allows users to do fine-tuning on the output forest plots and address their specific needs. This capability ensures that the generated output fully satisfies user requirements, offering significant advantages in the immunogenicity analyses during vaccine development and evaluation.

REFERENCES

“Graph Template Language (GTL)” https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/grstatgraph/p0891gx3y0z8xqn1k9ijhv5xughi.htm

“SAS Macro Language” https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/mcrolref/titlepage.htm

ACKNOWLEDGMENTS

The author would like to thank Peng Wan, Shunbing Zhao, and Changhong Shi for their support and guidance in authoring this paper.

CONTACT INFORMATION <HEADING 1>

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

Hong Wang
Merck & Co., Inc., Rahway, NJ, USA
hong.wang11@merck.com