Animated Multi-dimensional Scatter Plot Visualization for Longitudinal Clinical Trial Data Reporting and Exploring Using SAS Statistical Graphic Procedures

Jianfei Jiang, Tao Shu, Eli Lilly and Company, Indianapolis, IN, USA

ABSTRACT

Clinical trial data are often in longitudinal format with repeated measurements, when the aim is to analyze the changes occurring over time. One of the biggest challenges is how to summarize and present large clinical trial data in a simple and easy-to-understand way to give business intelligence, and sponsors of trials insightful information in order to make important decisions. Graphic visualizations promote effective communication of complex data to a varied audience in a more readily understandable format. The effectiveness of many temporal visualizations, such as line and bar, however, is compromised when dealing large clinical data. In this paper, we demonstrate how to generate animated scatter plot to depict the impact of different treatments on the endpoint measurements and/or clinical response across multiple time periods, and to explore relationships amongst variables. Using visual attributes such as color, size and shape, we can add up to 6 dimensions to a scatter plot. In order to create smooth animation between the scheduled time points, linear data interpolation is adopted. A sequence of images is generated by a macro loop using SAS PROC SGPANEL procedure, and the animated GIF is created using ODS PRINTER destination. Two examples, a rain-drop scatter plot and a bubble scatter plot, are presented in this paper based on a synthetic example dataset, to illustrate the visualization techniques for multidimensional data reporting and exploring. Note that no clinical meaningfulness could be interpreted from the synthetic data used.

INTRODUCTION

Scatter plot is one of the most commonly used 2-dimensional coordinate system to examine the relation between two numeric variables X and Y. More advanced forms, such as multi-panel scatter plot, allows us to split the observations according to a categorical variable to explore the correlation in a comparative way. Additional categorical/numerical variables could be included by modifying marker attributes (such as color, shape, size) in the plot. Using SAS Statistical Graphic (SG) procedures, we are able to present multi-dimensional visualization on 2D surface. In order to display key variables simultaneously in a comparative way over time, we use SAS 9.4 system options and ODS printer to create animation, which allow us to view the clinical data dynamically and interactively.

EXAMPLE 1. FIVE-DIMENSIONAL ANIMATED RAINDROP VISUALIZATION USING SGPANEL PROCEDURE

In this example, using a synthetic dataset, we first demonstrated how to use a 4-deminstional multi-panel scatter plot (**Figure 1**) to explore the relationships among clinical response (color of marker), post baseline disease activity score (Y-axis) and baseline disease activity score (X-axis) in different treatment arms (panels) at the end of clinical trial using SGPANEL procedure (**Example Code 1**) [1,2,3]. As shown in **Figure 1**, there is no improvement in disease activity score at the end of trial compared to the baseline value for patients in group A. Only a small portion of patients with low baseline disease activity achieved good clinical outcome. In contrast,

in group B and C, the majority patients responded to the treatments well as indicated by the decreasing of post baseline disease activity, regardless of the baseline disease status. In all treatment groups, the clinical outcome is highly correlated with the change of post baseline disease activity score from baseline, indicating this particular measurement is a good predicator of the clinical outcome.

Example Code 1. SAS9.4 Code for generating raindrop animation in figure 2:

```
%MACRO animation (avisitn);
%DO i = 200 %TO &avisitn. %BY 2;
        DATA synth0&i.;
            SET final;
            WHERE avisitn=&i.;
        RUN;
PROC SORT data=synth0&i.; by decending trtn colgrp;run;
PROC SGPANEL data = synth0&i.  noautolegend;
        FORMAT trtn trtnam. colgrp grpnam. rem remstatus.;
        PANELBY trtn/columns=3 NOVARNAME layout=columnlattice;
        STYLEATTRS DATACONTRASTCOLORS = (red green)
            DATASYMBOLS = (circlefilled)
            DATALINEPATTERNS = (solid);
            COLAXIS LABEL = "Baseline Disease Activity Index"
            LABELATTRS = (size=12pt weight=bold) valueattrs = (size=10 ) values = (0 to 60 by 20);
    ROWAXIS LABEL = "Post Baseline Disease Activity Index"
            LABELATTRS = (size=12pt weight=bold) grid values = (0 to 60 by 20) valueattrs = (size=10 );

        SCATTER x=bDAI y=AVAL / group=rem markerattrs=(symbol=circle size=10px) ;
            KEYLEGEND / ACROSS = 5 POSITION = BOTTOM  NOBORDER
                            TITLE='Clinical Response';
        lineparm x=0 y=0 slope=1    / lineattrs=(color=black thickness=0.1)
                                    LEGENDLABEL = 'Baseline';
        lineparm x=0 y=0 slope=0.5 / lineattrs=(color=pink thickness=0.1)
                                    LEGENDLABEL = '50% Improvement';
        lineparm x=0 y=0 slope=0.25/ lineattrs=(color=orange thickness=0.1)
                                    LEGENDLABEL = '75% Improvement';
        lineparm x=0 y=0 slope=0.1 / lineattrs=(color=green thickness=0.1)
                                    LEGENDLABEL = '90% Improvement' ;
    RUN; %END;
    %MEND ANIMATION;

    options papersize=('9 in', '7.5 in') printerpath=gif animation=start animduration=0.03 animloop=yes
    noanimoverlay;
    ods printer file='C:\figure2.gif';
    ods graphics / width=9in height=7.5in imagefmt=GIF;
    %animation (700);
    options printerpath=gif animation=stop;
    ods printer close;
    ods graphics off;
```
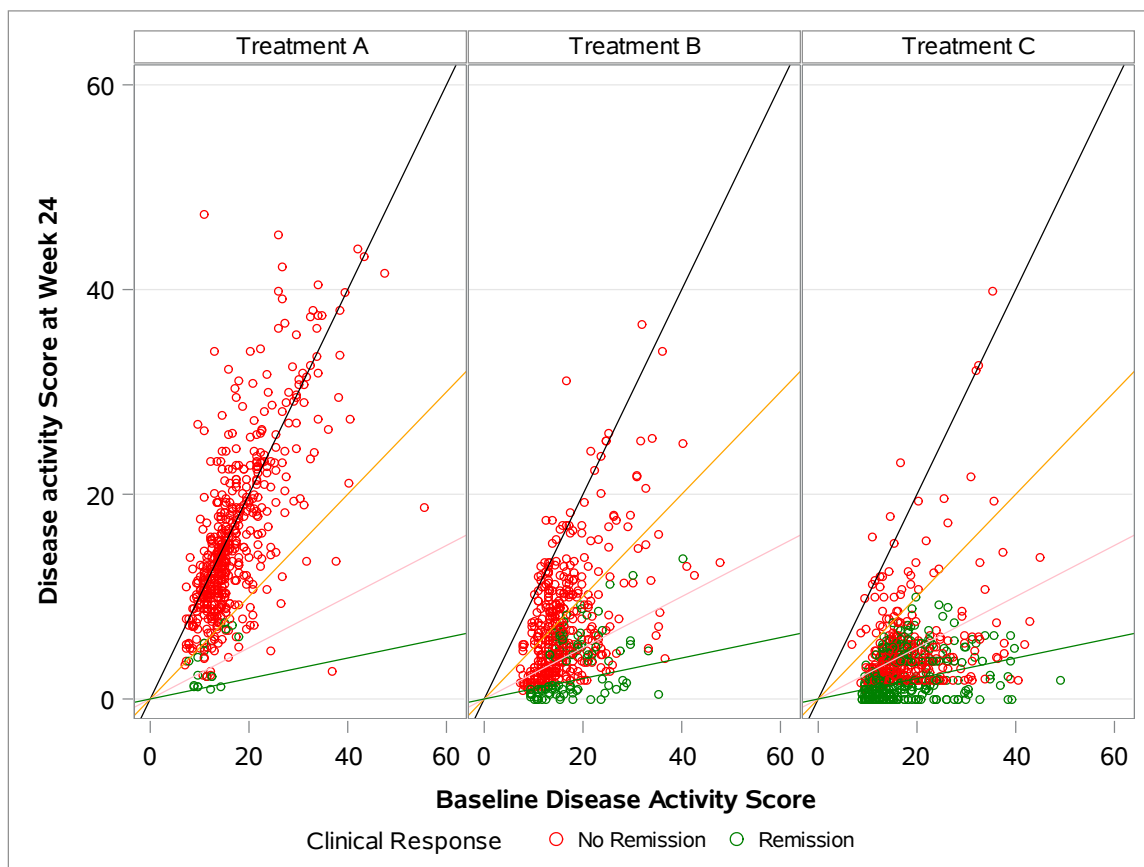
Figure 1. Four-dimensional scatter Plot analysis of post-baseline disease activity score (Y-axis) against baseline disease activity score (X-axis) with clinical response (green, remission; red, no remission) after 24-week treatment with drug A, B and C. Note that no dynamic development of disease activity score or clinical response could be shown in the snap shot**.**

Often, there is a need to present comparisons of longitudinal clinical data among treatment arms not only at the endpoint, but also at each time point the measurement occurs. In **Figure 2**, we used time as a dimension by making an animated raindrop plot for other variables over time. In order to create a smooth animation between two time points, a linear data interpolation was used to insert 50 additional records between two scheduled visits. A sequence of images was generated by a macro loop using SAS PROC SGPANEL procedure, and the animated GIF is created using ODS PRINTER destination (**Example Code 1**). The resulting GIF animation was converted into MP4 files, and embedded in PDF file (**Figure 2**). As shown in the animation, presenting changes of disease activity and clinical response over time provided a more sophisticated perspective. The change of disease activity was much more robust in group C compared to group B in the early stage of trial, and a clinical response to treatment C was seen relatively quickly compared to treatment B. In most case, the good clinical outcomes in both group B and C could be maintained over the trial course. In addition, change of disease activity was highly correlated with the clinical outcome in all three groups. In short, the presented multi-dimensional animated raindrop was very helpful in data reporting or exploration.
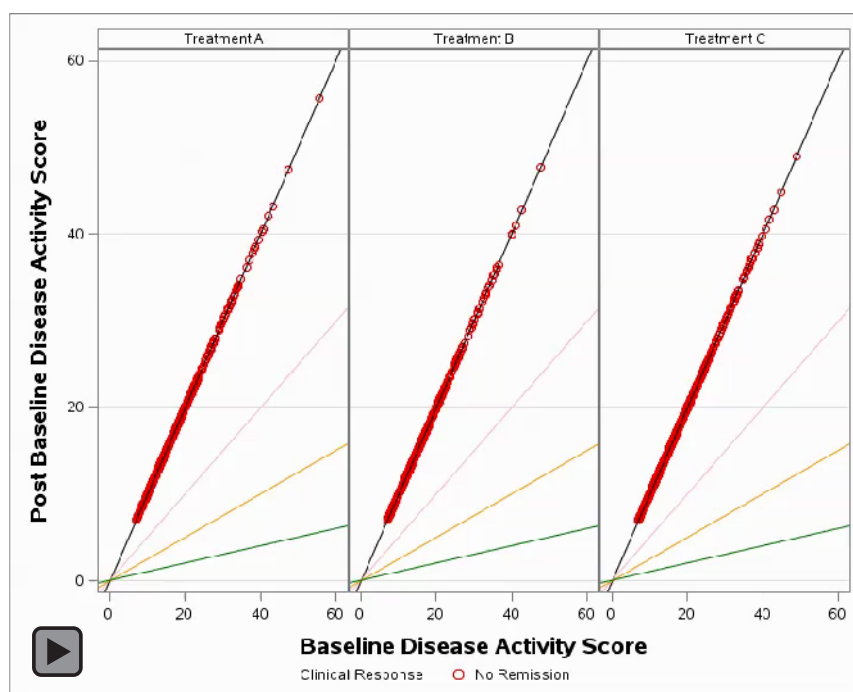
Figure 2. Five-dimensional animated scatter plot analysis of post-baseline disease activity score (DAS) (Y-axis) against baseline DAS (X-axis) with clinical response (green, remission; red, no remission) over a 24-week treatment period (Placebo, comparator, or experimental therapeutic drug). Note that the disease activity score and clinical response could be dynamically visualized for all three treatment arms simultaneously over time.

EXAMPLE 2. ANIMATED SIX-DIMENSIONAL BUBBLE PLOT USING SGPANEL PROCEDURE

Most real world clinical data have many more dimensions. Utilizing other marker attributes, such as marker size and shape, we could add more dimensions in visualization. In this example, we attempted to plot a 6-dimensional animated bubble plot to uncover the relationship among key measurements (three continuous variables, a categorical treatment variable) and their association with a categorical clinical response variable over time (**Example Code 2**). As demonstrated in a static 5-dimensional bubble plot (**Figure 3**) at the 24-week endpoint, the baseline disease activity was plot on X-axis against a primary efficacy measurement change from baseline on Y-axis. The size of bubble indicated a continuous biomarker variable. The observations were split into three panels according to the treatment arms (A, B and C) and the colors of marker were used to indicate the clinical response (red, no clinical response; blue, low disease activity, LDA; green, remission). From the multi-dimensional bubble plot, we learned that patients in group A responded to treatment moderately, regardless of the baseline disease activity score, while in group B and C, patients with low disease activity score responded much better to the treatments compared to patient with high disease activity score. The clinical response was correlated to the primary measure change from baseline in all three arms, but not correlated to the biomarker variable. Patients in group B and C responded to the treatments in a similar pattern. Finally, treatment C seemed more effective than B as revealed by the better clinical response (green and blue).

Example Code 2. SAS 9.4 code for creating animated bubble plot in Figure 4:

```
%MACRO draw(color=);
%DO i=&vmin.*100  %TO &vmax.*100 ;
DATA _null_;
     SET draw6;
WHERE visit=&i ;
     CALL SYMPUT ('week',trim(avisit));
 RUN;

 PROC SGPANEL DATA=draw6 ;
     WHERE visit=&i;
     PANELBY trtpn/ spacing=5 novarname columns=3;
     BUBBLE x=phyga y=chg2 size=esr2/ BRADIUSMAX=20 BRADIUSMIN=2 COLORMODEL=(&color)

     COLORRESPONSE=oc
              DATASKIN= MATTE  fillattrs=(transparency=0.4)  ;
              KEYLEGEND ;
                      LABEL chg2="Primary Measure Change From Baseline";
                      LABEL phyga="Baseline Disease Activity Score";
                      TITLE "&week";
                      ROWAXIS max=80 min=0 labelattrs=(Weight=bold);
                      COLAXIS  min=20 labelattrs=(Weight=bold);
                      FORMAT trtpn trt.;
RUN;
%END;
%MEND draw;

options nobyline papersize =( '9 in ', '7.5 in ') printerpath = gif animation = start  animduration =0.033
animloop = no noanimoverlay ;
ods printer file ="C:\figure4.gif " ;
ods graphics / width =9 in height =7.5 in imagefmt =gif ;
%draw(color=red blue green);
options printerpath =gif animation = stop ;
ods printer close ;
```

In order to explore or report the data in multivariate fashion over time, we created a six-dimension animated bubble plot in a similar way as we presented in example 1 (**Example Code 2**). In **figure 4**, we could simultaneously view the three key events along with the clinical response in three different treatment panels for individual patient over the trial course. The time-to-response for three different therapeutic strategies was displayed side by side for comparison. Clearly, the time-to-response was relatively short in group C.
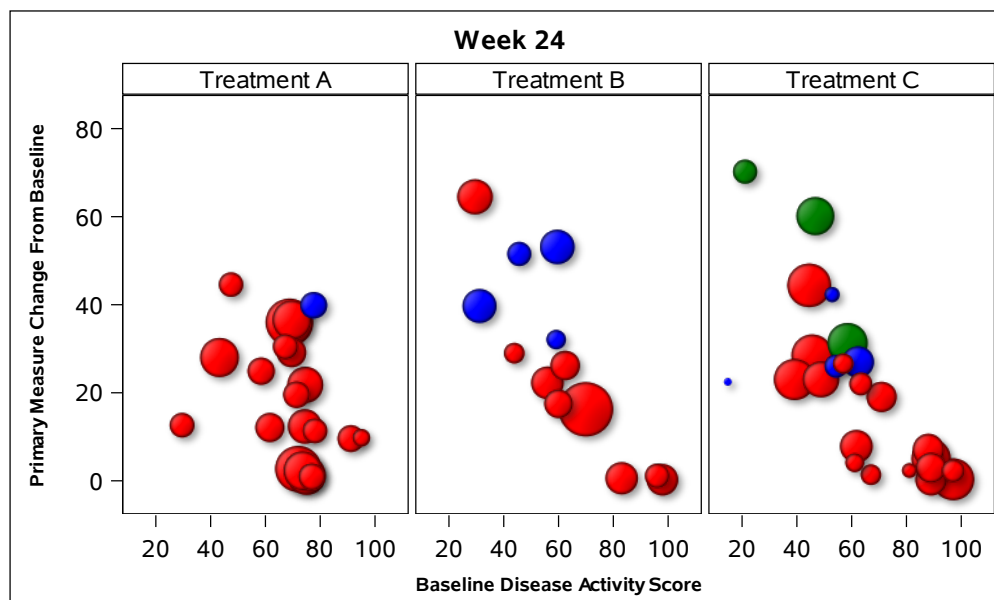
Figure 3. Five-dimensional scatter plot analysis of primary endpoint change from baseline (Yaxis) against baseline disease activity score (X-axis) with clinical response (red, no clinical response; blue, low disease activity, LDA; green, remission) and a biomarker measurement (bubble size) after a 24-week treatment (Treatment A, B or C) period. Note that no dynamic development of the presented variables could be visualized in the snap shot.
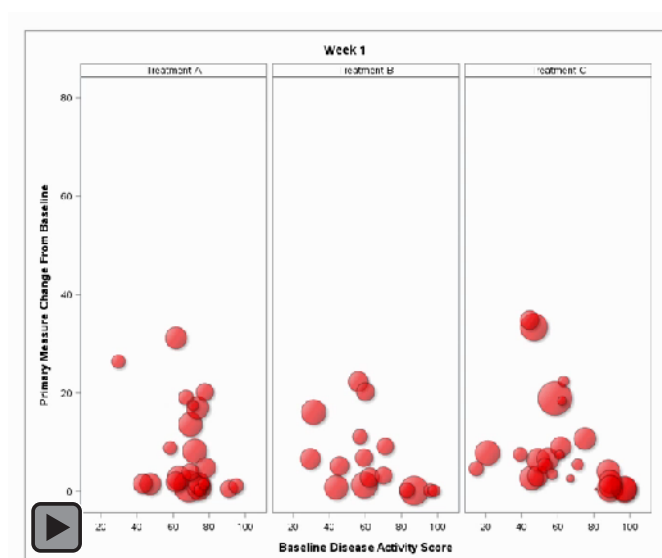

Figure 4. Six-dimensional animated scatter plot analysis of primary endpoint change from baseline (Y-axis) against baseline disease activity score (X-axis) with clinical response (red, no clinical response; blue, low disease activity, LDA; green, remission) and a biomarker measurement (bubble size) over a 24-week treatment period (Placebo, comparator, or experimental therapeutic drug). Note that the correlations among multiple variables could be dynamically visualized over time.

CONCLUSION

Using a synthetic dataset, we have demonstrated how to use SAS 9.4 SG procedure to create multi-dimensions animated visualization for reporting or exploring longitudinal clinical data. Clearly, the graphical procedures presented in this paper help to increase the interpretability of multi-dimensional clinical data and assist in communicating the trial result.

REFERENCES

1. http://support.sas.com/documentation/94/index.html
2. SAS® 9.4 Language Reference: Concepts, Sixth Edition
3. Sanjay Matange. Animation using SGPLOT. Graphically Speaking. https://blogs.sas.com/content/graphicallyspeaking/2013/05/23/animation-using-sgplot/ May 23, 2013.

ACKNOWLEDGMENTS

We are thankful to our colleague Andrew McCarthy, Kriss Harris, and Emily Seem for providing advice and expertise that greatly assisted this project.

RECOMMENDED READING

Sanjay Matange. *SAS® Press Clinical Graphs Using SAS.*

CONTACT INFORMATION

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

Jian Jianfei
Eli Lilly and Company
Email: jiang_jianfei@lilly.com

Tao Shu
Eli Lilly and Company
Email: shu_tao@lilly.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.