

## Translating Statistics into Knowledge by Examples

### Using SAS Graphic Procedures

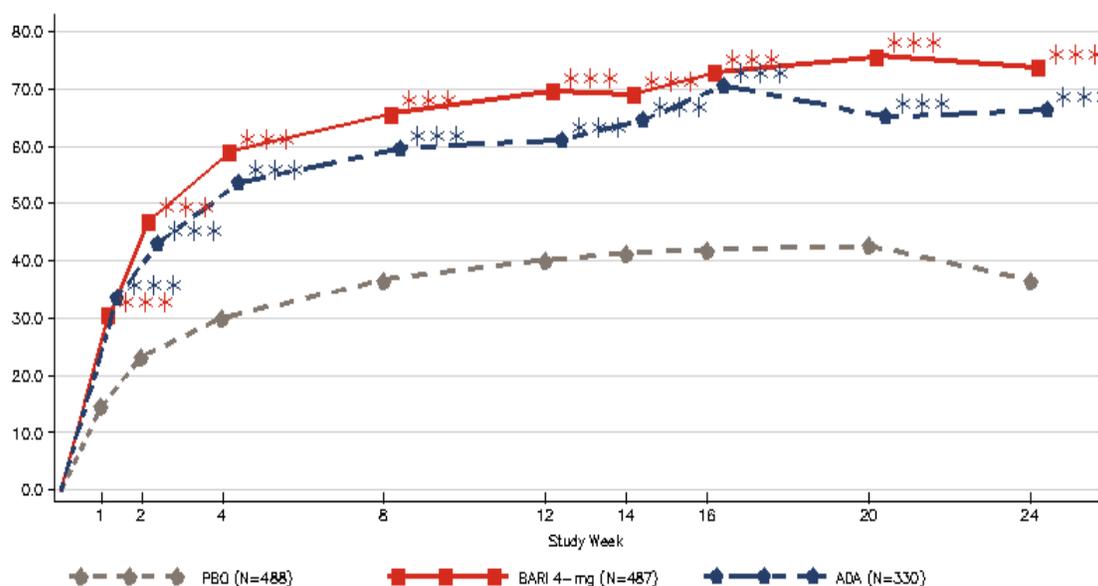
Tao Shu, Eli Lilly and Company, Indianapolis, IN, USA  
 Jianfei Jiang, Eli Lilly and Company, Indianapolis, IN, USA

#### ABSTRACT

In a clinical trial, sponsors are required to communicate safety and efficacy data clearly and efficiently to regulatory agencies and to the public. Application of report and exploratory graphic visualization is crucial in summarizing clinical trial data in ways that can facilitate the precise conveyance of the statistical analysis with compelling visual stories. Here, we employ three different examples from Lilly Rheumatoid Arthritis clinical trials to demonstrate how SAS 9.4 graphical procedures can translate statistical analysis into vivid graphical forms. In the first example, using the SGPLOT procedure, we efficiently present the  $p$ -value in line graph and bar chart graphical reports. The clinically important effect of the Lilly compound (Baricitinib), as revealed by changes in American College of Rheumatology 20% response and modified Total Sharp Score, is clearly demonstrated at multiple time points when compared to placebo. In examples 2 and 3, using the SGPANEL procedure, we generate exploratory multivariable graphics on two-dimensional surfaces, which help to identify significant factors associated with serious adverse events (Poisson regression model) or radiographic progression of structural joint damage (longitudinal linear model) in Rheumatoid Arthritis. In addition, using exploratory graphics, we are capable of looking into clinical data at both the subject and population level. In summary, use of SAS 9.4 graphical procedures results in efficient and effective data interpretation in clinical trials.

#### EXAMPLE 1

For the efficacy measures in clinical trials, there is a need to present comparisons among treatment arms not only at the end point, but also at each time point the measurement occurs. **Figure 1** compares the American College of Rheumatology 20% (ACR20) response rate of different arms in a Rheumatoid Arthritis trial at each post-baseline visit. The Y-axis represents the proportion of patients achieving a 20% improvement in the ACR criteria.



**Figure 1. Time Course of ACR20 Response Rate in a Lilly Phase III Rheumatoid Arthritis Trial.** ADA, Adalimumab; BARI 4-mg, Baricitinib 4-mg; PBO, Placebo. \*\*\*  $p < 0.001$  versus placebo.

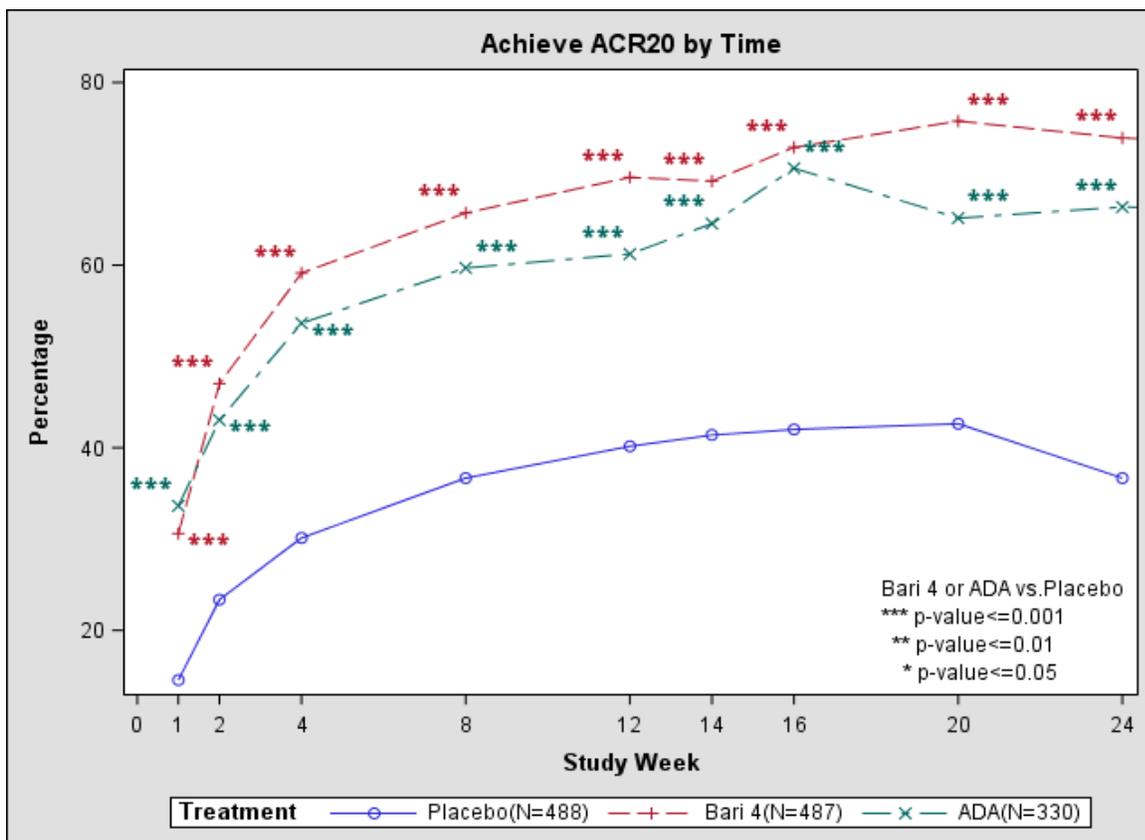
The code used to create **Figure 1**, however, was lengthy and complicated when the SAS graph annotate facility was employed (**Table 1**), and it was difficult to apply the code to different scenarios (e.g, different time points) to obtain the right format.

**Table 1. Example Code for Figure 1 Using the SAS Graph Annotate Facility.**

<pre> data anno1;   length function \$8 text \$50 color \$10;   retain xsys '2' ysys '3' color 'black' when 'a';   set lsmout (where=(trtpn="2") ) end = eof;   week = week-0.2;   week1 = left(put(week,best.));   %move (week,17);   %draw (week,15.5,black,1,1);   %label (week,14,week1,black,,1.5,,); run;  proc sql noprint; select (ceil(max(lsmean)/10))*10 into: upansd from lsmout; select max(lsmean) into: upansdj2 from lsmout; quit; %put &amp;upansd.; %if &amp;upansd. &gt; 30 %then %do;   %let stepasd=10; %end; %if &amp;upansd. &lt;= 30 and &amp;upansd. &gt; 20 and &amp;upansdj2.&gt;25   %then %do;      %let stepasd=5; %end; %if &amp;upansd. &lt;= 30 and &amp;upansd. &gt; 20 and &amp;upansdj2.&lt;25   %then %do;   %let upansd=25; %let stepasd=5; %end; %if &amp;upansd. &lt;= 20 and &amp;upansd. &gt; 10 and &amp;upansdj2.&gt;15   %then %do;   %let stepasd=5; %end; %if &amp;upansd. &lt;= 20 and &amp;upansd. &gt; 10 and &amp;upansdj2.&lt;15 % then %do;   %let upansd=15; %let stepasd=5; %end; %if &amp;upansd. &lt;= 10 %then %do;   %let stepasd=2; %end; %put &amp;stepasd.; *Create final annotation dataset for p-values; data anno2;   length function \$8 text \$50 color \$10;   retain xsys '2' ysys '2' when 'a';   set lsmanno end = eof;   if probt ^= . then do;   if probt &lt;= .05 and probt &gt; .01 then do;     if trtpn = "2" then do;       %label(week,lsmean,'*',cxd52b1e,,4,,C);     end;   if trtpn = "3" then do;     %label(week,lsmean,'*',cx263f6a,,4,,C);   end;   end;   if trtpn = "2" then do;     %label(week,lsmean,'**',cxd52b1e,,4,,C);   end;   if trtpn = "3" then do;     %label(week,lsmean,'**',cx263f6a,,4,,C);   end; end; </pre>	<pre> if probt &lt;= .01 and probt &gt; .001 then do; options orientation=landscape papersize=("11in", "8.5in") leftmargin= "1.0 in" rightmargin="1.0 in" topmargin="1.0 in" bottommargin="1.0 in" nodate nonumber; goptions reset=all device = zpng ftext=simplex ftitle=simplex cback = white ctext = black htext = 3.0 ymax=4in xmax=11in vsize=3.6in hsize=9in htitle=2 gsflen=80 gunit = pct display gsfmode = replace noborder; ods listing close; ods rtf file="&amp;_tfl.&amp;tfl_name..rtf" startpage=off style=mystyle; legend1 label=none order= ("1" "2" "3") value=(h=8 pt "PBO (N=&amp;PBO)" "BARI 4-mg (N=&amp;B4G)" "ADA (N=&amp;ADA)")   position=(bottom center outside) down=1 shape=symbol(10,3); %if &amp;tfl_name=f_acr20_nri_wk0to12 %then %do; axis1 label=(h=3 a=90 j=c "ACR20 Response Rate (% of Patients)") order = (0 to &amp;upansd. by &amp;stepasd.)   value=(h=3 ) minor=none offset=(2,3); axis2 label=(h=3 j=c " " j=c "Study Week") order = (0 to 14 by .2)   value = none   minor=none major=none; %end; %if &amp;tfl_name=f_acr20_file_1 or &amp;tfl_name=f_acr20_file_2 %then %do; axis1 label=(h=3 a=90 j=c "ACR20 Response Rate (% of Patients)") order = (0 to &amp;upansd. by &amp;stepasd.) value=(h=3 ) minor=none offset=(2,3); axis2 label = ( h = 3 j = c " " j = c "Study Week") order = (0 to 26 by .2)   value = none   minor = none major = none; data annot;   set anno1 anno2; run; ods path show; ods path(prepend) work.templat(update); proc template; define style mystyle; style usertext from usertext /   font = ("Arial")   protectspecialchars = off   outputwidth = 100%   just = c   backgroundcolor = white; end; run; %end; %if &amp;tfl_name = f_acr20_file_3 or &amp;tfl_name = f_acr20_file_4 %then %do; axis1 label = ( h = 3 a = 90 j = c "ACR20 Response Rate (% of Patients)") order = (0 to &amp;upansd. by &amp;stepasd.) value = ( h = 3 ) minor = none offset = (2,3); %end; %if &amp;tfl_name = f_acr20_file_3 or &amp;tfl_name = f_acr20_file_4 %then %do; axis2 label=(h=8 pt j=c " " j=c "Study Week") order = (0 to 54 by 0.2)   value =none   minor=none major=none; %end; symbol i=j c=cx82786f font=marker v=W l=2 h=3 w=2; symbol4 i=j c=cx82786f font=marker v=none l=2 h=3 w=2; </pre>
---	---

<pre> if probt &lt;= .001 then do;   if trtpn = "2" then do;     %label(week,lsmear,'***',cxd52b1e,,4,C);   end;   if trtpn = "3" then do;     %label(week,lsmear,'***',cx263f6a,,4,C);   end; end; end; run; </pre>	<pre> proc sort data=lsmout;   by week trtpn; run; %header(); proc gplot data = lsmout ;   plot lsmear * week = trtpn / anno=annot legend=legend1   vaxis=axis1 haxis=axis2 autovref cautohref=cxd3d3d3 noframe;   format lsmear 12.1; run; quit; </pre>
--	--

An improved figure can be easily created using the SGPLOT procedure in SAS 9.4 (<http://support.sas.com/documentation/94/index.html>) (Figure 2). Clearly, the code (see below) is a much simpler version, which can be easily modified to fit different needs.



**Figure 2. Time Course of ACR20 Response Rate in a Lilly Phase III Lilly Rheumatoid Arthritis Trial.** ADA, Adalimumab; BARI 4-mg, Baricitinib 4-mg; PBO, Placebo. \*\*\*  $p < 0.001$  versus placebo, \*\*  $p < 0.01$  versus placebo, \*  $p < 0.05$  versus placebo.

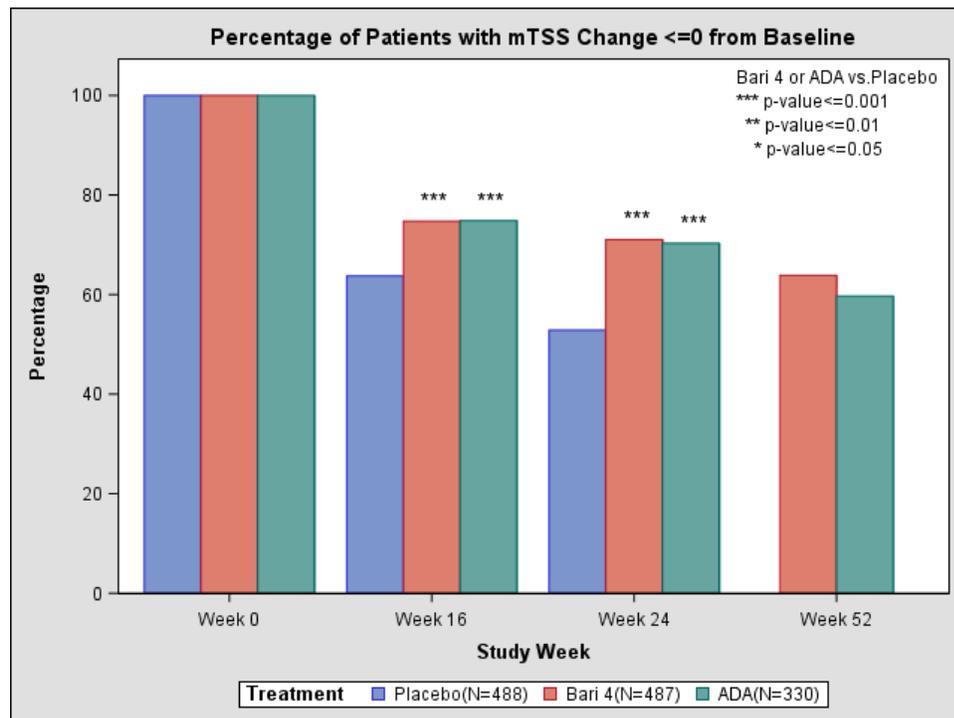
### Example SAS 9.4 code used to generate Figure 2

```

ODS HTML PATH="H:\ " (url=none) FILE="example5.html" STYLE=default ;
PROC SGPLOT DATA=t3;
  vline week / response=col1 group=trtpn datalabel=pval      DATALABELATTRS=(Family=Arial
    Size=15 Style=Italic Weight=Bold) markers;
  FORMAT trtpn $trt.;
  LABEL trtpn="Treatment";
  xaxis LABEL="Study Week" TYPE= linear VALUES=(0,1,2,4,8,12,14,16,20,24);
  yaxis LABEL="Percentage";
  INSET 'Bari 4 or ADA vs. Placebo' '*** p-value<=0.001' ' ** p-value<=0.01' ' * p-
value<=0.05' / POSITION = bottomRIGHT ;
  TITLE "Achieve ACR20 by Time";
RUN;
ODS HTML CLOSE;

```

In addition to the line graph, the SGPLOT procedure can also be used to present the p-value in a bar chart or scatter plot. **Figure 2** shows the percentage of Rheumatoid Arthritis patients with no modified Total Sharp Score (mTSS) change from baseline on erosions and narrowing of the joint space measured by X-RAY. Normally, patients with Rheumatoid Arthritis are likely to develop progressive erosion of the cartilage and bone, and sometimes even joint deformity, in their hands, wrists, and feet. Treatments that can prevent further erosion and deformity would be of great benefit to patients. In **Figure 3**, we show that both Baricitinib 4mg (Bari 4) and Adalimumab (ADA) significantly reduce the progression of joint damage in Rheumatoid Arthritis patients compared to Placebo at week 16 and week 24.



**Figure 3. Percentage of Patients with mTSS Change ≤ 0 from Baseline.** Bari 4: Baricitinib 4-mg; ADA: Adalimumab. N is the population size of the treatment arm. \*\*\*  $p < 0.001$  versus placebo.

## EXAMPLE 2

In one of the Rheumatoid Arthritis clinical trials, we attempted to develop a risk model to determine the factors impacting the development of serious adverse events (SAE). In particular, we looked at the non-rescued placebo population (355 subjects) in which no active therapy was involved. The explanatory variables in the full model included age, sex, patient disease conditions at the baseline, etc (listed in **Output 1**).

Statistical approach

Numbers of SAE occurrence were modelled by the Poisson distribution using the PROC GENMOD procedure (<http://support.sas.com/documentation/94/index.html>). The exposure time on treatment was treated as an offset in the model. The GENMOD procedure output is pictured below in **Output 1**.

### Output 1. Full Model Fit Statistics Output from the PROC GENMOD Procedure.

LR Statistics For Type3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
AGE	1	4.22	0.0400
SEX	1	0.67	0.4132
RADDURGR	3	5.28	0.1525
RF	1	9.93	0.0016
EGFRGR1	1	11.94	0.0005
SJC66	1	0.19	0.6617
TJC68	1	0.34	0.5621
HAQDI	1	6.36	0.0117
PHYGA	1	1.62	0.2031
ESR	1	0.00	0.9764
BMI	1	0.09	0.7585
HSCRP	1	3.03	0.0818

After dropping variables with high  $p$ -values ( $>0.1000$ ) by stepwise selection, we obtained the final model as shown in **Output 2**.

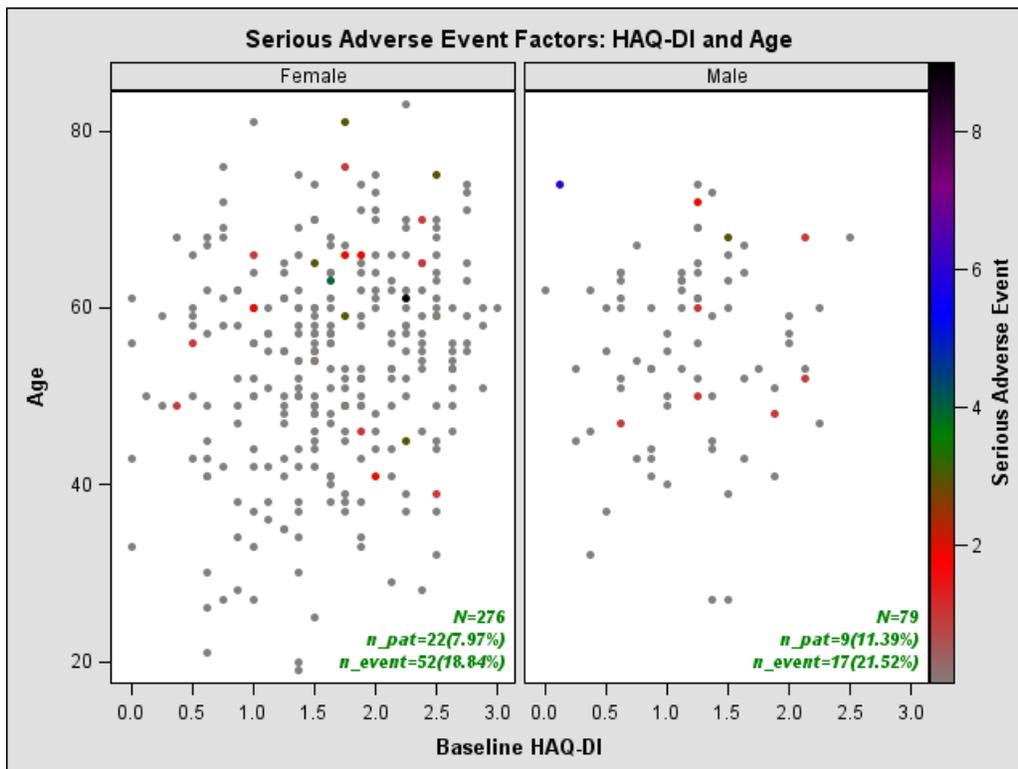
### Output 2. Reduced Model fit statistics output from SAS.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
AGE	1	4.27	0.0388
RF	1	14.71	0.0001
EGFRGR1	1	20.48	<.0001
HAQDI	1	12.12	0.0005

## SAS 9.4 Code Used to Generate Output 2.

```
PROC GENMOD DATA=adae3;  
  CLASS egfrgr1 ;  
  MODEL event=age rf egfrgr1 haqdi  
    /DIST=poisson LINK=log OFFSET=log_time type3 LRCI;  
RUN;
```

To further validate the model, we used the SGPANEL procedure to generate exploratory graphics (**Figure 4** and **Figure 5**) to examine the relationship among the significant variables and SAE. The left and right panels in the graph represent different gender groups. The Y-axis and X-axis are the significant factors associated with the dependent variable SAE in the reduced model. Each dot in the figure represents a subject. The color bar on the right side indicates the number of SAEs. The plain grey color represents no SAE. The pink represents one SAE, and the red represents two SAEs. The number of SAEs increases along the color line from red, green, blue, purple to black.



**Figure 4. Multivariate Graph Generated by PROC SGPANEL Showing the Relationship between Age, Baseline HAQ-DI and SAE.** N, population size; n\_pat, the number of the subjects having the SAE; n\_event, the number of SAEs.

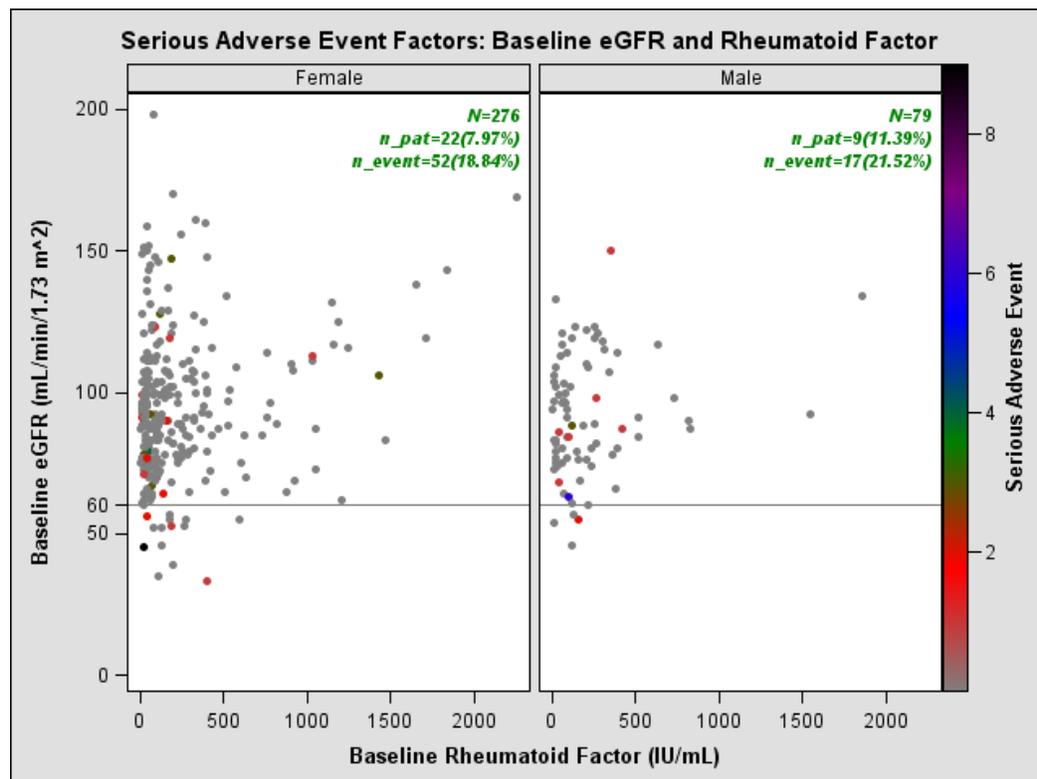
As **Figure 4** reveals, the occurrence of SAEs is slightly higher for males than females (11.39% in male versus 7.97% in female). Age and baseline Health Assessment Questionnaire-Disability Index (HAQ-DI) are two contributing factors since more colored dots are present in older patients and in patients with higher HAQ-DI.

### SAS 9.4 Code Used to Create Figure 4.

```
ODS HTML PATH ="H:\\" (URL = NONE) FILE="example6.html" STYLE=DEFAULT;
PROC SGPANEL DATA=final;
  PANELBY sex /SPACING=5 NOVARNAME COLUMNS=2;
  SCATTER X=haqdi Y=age/ COLORMODEL=(grey red green blue purple black )
COLORRESPONSE=event MARKERATTRS=(symbol=circlefilled) name="figure" ;
  INSET t sn en /POSITION=bottomright nolabel TEXTATTRS=(Color=Green Family=Arial Size=8
  Style=Italic Weight=Bold) ;
  LABEL event ="Serious Adverse Event";
  FORMAT sex $sex.;
  TITLE "Serious Adverse Event Factors: HAQ-DI and Age ";
RUN;
ODS HTML CLOSE;
```

**Figure 5** shows higher occurrence of SAEs in subjects with baseline eGFR less than 60 (mL/min/1.73 m<sup>2</sup>). There are 15 SAEs occurring in 5 of 18 subjects (27.78%) with baseline eGFR less than 60, while only 54 SAEs occurring in 26 of 337 subjects (7.72%) with baseline eGFR greater than or equal to 60. This observation is consistent with the *p*-value from the SAS output for the reduced model in **Output 2**.

In summary, the exploratory graphics provide intuitive information for the validation of the risk model that determines the significant factors associated with occurrence of SAEs.



**Figure 5. Multivariate Graph Generated by PROC SGPANEL Showing the Relationship between Baseline eGFR, Baseline Rheumatoid Factor, and SAE.** N, population size; n\_pat, the number of the subjects having the SAE; n\_event, the number of SAEs.

### EXAMPLE 3

In this efficacy analysis, we attempted to identify factors contributing to the progression of erosion and narrowing of joint space in Rheumatoid Arthritis patients. We were interested in identifying any differences among the active treatment groups in maintaining the joint structure status (no progression). The treatment arm A had 154 subjects and the treatment B had 199 subjects. Change from baseline of the modified Total Sharp Score (mTSS), which measures the structural joint damage (erosion and narrowing of joint space), was the response variable. The measurements occurred at week 12, week 24 and week 52, respectively. The longitudinal linear model was fit using the SAS MIXED procedure.

From the SAS output in **Output 3**, we see that the time (visit), erythrocyte sedimentation rate (ESR), and high sensitivity C-reactive protein (hsCRP) are the potential risks ( $p < 0.1000$ ) associated with the progression of structural joint damage. Because of these results, we re-fit the data with the longitudinal linear model with these selected variables, and excluded the insignificant variables one by one. **Output 4** is the SAS output of the final model, which includes visit, ESR, and hsCRP at baseline. No significant ( $p = 0.6642$ ) difference among these two active treatments was identified.

#### Output 3. Full Model fit statistics output from SAS 9.4 PROC MIXED Procedure.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TRT01PN	1	327	0.05	0.8179
AVISITN	1	327	4.22	0.0407
AGE	1	327	1.27	0.2612
ALCOHOL	1	327	0.02	0.8897
EGFRGR1	1	327	0.17	0.6805
RADDURGR	3	327	0.55	0.6475
cigarette	1	327	0.08	0.7728
RF	1	327	1.60	0.2072
tjc	1	327	1.87	0.1723
sjc	1	327	0.07	0.7904
haq	1	327	0.32	0.5713
DASESR	1	327	3.84	0.0509
MJS	1	327	1.12	0.2918
JNTPAIN	1	327	0.00	0.9552
PHYGA	1	327	0.05	0.8231
BASE	1	327	0.25	0.6193
ESR	1	327	12.91	0.0004
HSCRCP	1	327	4.74	0.0302
ACPA	1	327	0.19	0.6620
CRPS91S	1	327	0.01	0.9292
CDAI	1	327	0.01	0.9290
SDAI	1	327	0.01	0.9290
PATGADA	1	327	0.00	0.9967
PHYSGADA	1	327	0.21	0.6485
DAS28ESR	1	327	1.03	0.3110
TJC68	1	327	0.01	0.9170
SJC66	1	327	0.33	0.5664
HAQDI	1	327	0.84	0.3597

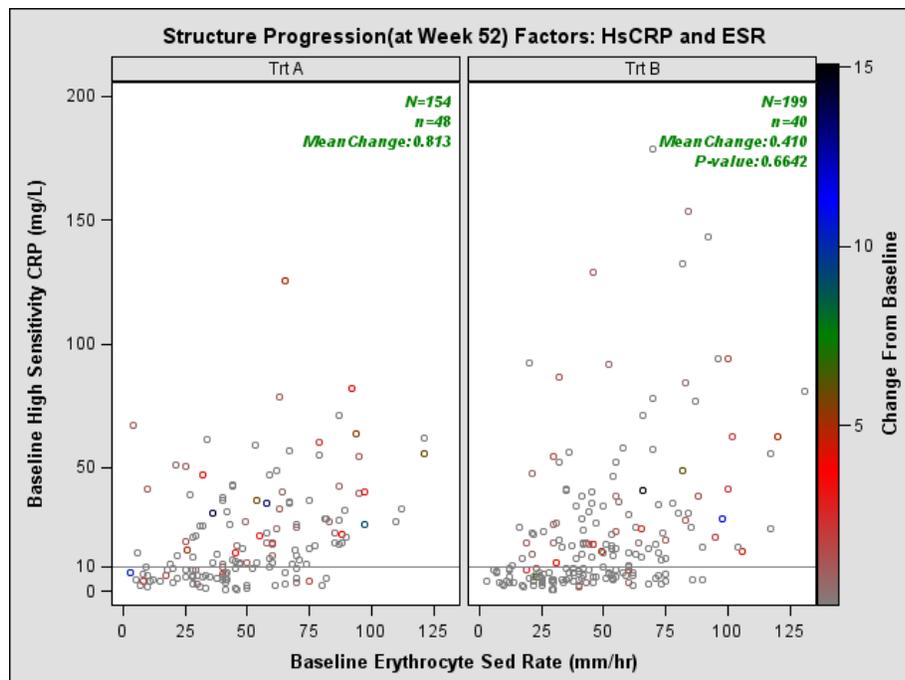
#### Output 4. Final Model fit statistics output from SAS 9.4 PROC MIXED Procedure.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TRT01PN	1	349	0.19	0.6642
AVISITN	1	349	5.86	0.0160
ESR	1	349	11.53	0.0008
HSCRCP	1	349	4.89	0.0277

#### SAS Code for the Output of the Final Model

```
PROC MIXED DATA = stata;  
  WHERE trt01pn>1;  
  CLASS usubjid trt01pn ;  
  MODEL chg = trt01pn AVISITN esr hscrp ;  
  REPEATED /type=un SUBJECT=usubjid;  
RUN;
```

Finally, we employed the SAS SG PANEL (<http://support.sas.com/documentation/94/index.html>) procedure to generate exploratory graphics to validate the final model **Figure 6** and **Figure 7**.

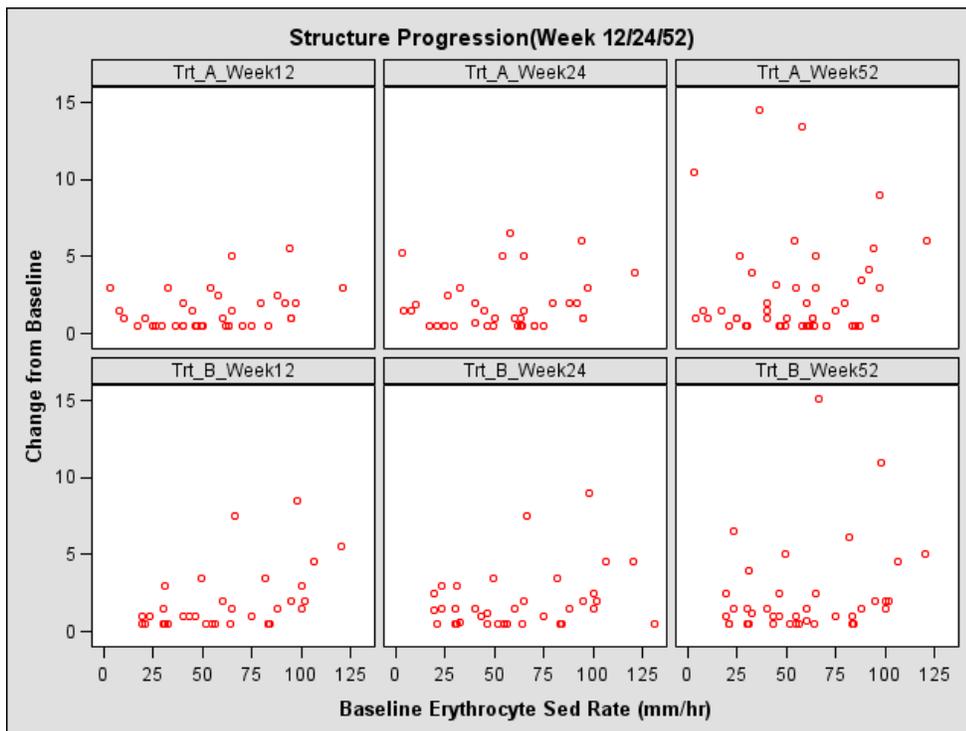


**Figure 6. Multivariate Graph Generated by the PROC SG PANEL Procedure Showing the Relationship among Baseline hSCRp, Baseline ESR, Treatment and mTSS at Week 52. N, population size; n, the number of patients having disease progression in each treatment arm.**

The mean changes from baseline, 0.813 in treatment A and 0.410 in treatment B, are not significantly different from each other ( $p = 0.6642$ ) (**Figure 6**). The colored bar on the right side indicates the change from baseline in the measure quantifying the degree of structural joint damage. Grey represents no progression, and the progression status worsens from the colors red, green, blue to black. Most of the colored dots, which indicate that the patient has progression of structural joint damage, are distributed on upper side of the hsCRP reference line ( $X=10$  mg/L). The odds of progression increase as ESR increases. The exploratory graphic visually confirms that hsCRP and ESR are risks for the progression of structural joint damage, as suggested by the SAS output from the final model.

In **Figure 7**, we demonstrate the relationship among mTSS change from baseline, baseline ESR, visit, and treatment. The red dots shown in the figure represent only those subjects with progression. In both treatments A and B, the progression becomes gradually worse over time. The mTSS changes from baseline are correlated with the baseline ESR.

In summary, using the SAS SGPANEL procedure, we effectively validated the longitudinal linear model to identify the risk factors associated with Rheumatoid Arthritis progression.



**Figure 7. Multivariate Graph Generated by PROC SGPANEL Procedure Shows the Relationship among mTSS Change from Baseline, Baseline ESR, Visit and Treatment.**

## CONCLUSION

Using three examples from Rheumatoid Arthritis clinical trials at Lilly, we have demonstrated how efficiently and effectively SAS 9.4 represents statistical analysis in report and exploratory graphics with the PROC SGPLOT and the PROC SGPANEL procedures. Clearly, the graphical procedures presented in this paper help to increase the interpretability of statistical results and assist in communicating the trial data.

## REFERENCES

1. <http://support.sas.com/documentation/94/index.html>

## ACKNOWLEDGMENTS

We are thankful to our colleague Dr. Scott Beattie, Dr. Lei Chen, Mr. Xingxing Wu, Mrs. Ming Yan, Ms. Anastasia Ilynichna Alexeeva, and Ms. Emily Seem for providing advice and expertise that greatly assisted this project.

## CONTACT INFORMATION

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

Tao Shu  
Eli Lilly and Company  
Email: [shu\\_tao@lilly.com](mailto:shu_tao@lilly.com)

Jian Jianfei  
Eli Lilly and Company  
Email: [jiang\\_jianfei@lilly.com](mailto:jiang_jianfei@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.