Make it personal: Upgrade your figures by adding individual patient data to the common figure types

Yuliia Bahatska, inVentiv Health Clinical, Berlin, Germany; Christiane Ahlers, Bayer AG, Wuppertal, Germany;

ABSTRACT

Figures are believed to be the most flexible way of data presentation in the world of clinical trials. Indeed, with the abilities of the GTL and graphic procedures in SAS$^{®}$ today it is only the power of the statistician's imagination that defines the complexity of the figures in the analysis. In the clinical study reports it is usually referred to summary statistics tables for the general overview and more detailed information about a certain patient can be found in the listings. Figures can present both of these. In this paper we would like to share with you our experience of creation of the figures that contain both the combination of summary statistics and individual patient data. The main idea is to enable the reviewer to trace how the values for a single patient changed throughout the study and at the same time put them into context with the derived statistics or other patients' values and also to identify outliers. This kind of data presentation provides all clinical functions and in the end all readers of the document with a deeper and a more detailed overview of the clinical trial results.

INTRODUCTION

Figures are a powerful tool to present clinical study data and offer flexible solutions to various kinds of medical questions. Indeed, with the abilities of the GTL and graphic procedures in SAS today it is only the power of the statistician's imagination that defines the complexity of the figures in the analysis. In CSRs summary statistics are presented to give an overview of data distribution in the study population. In case individual patients with certain characteristics are of special interest, patient level data is presented in listings. Figures can present both of these. In this paper we would like to share with you our experience of creation of the figures that contain both the combination of summary statistics and individual patient data. The main idea is to enable the reviewer to trace how the values for a single patient changed throughout the study and at the same time put them into context with the derived statistics or other patients' values and also to identify outliers. This kind of data presentation provides all clinical functions and in the end all readers of the document with a deeper and a more detailed overview of the clinical trial results.

ENHANCED MOSAIC PLOTS

For one of the projects we needed to show how the distribution of patients across several mutually exclusive categories and two mutually exclusive subcategories changed over time. For each patient and visit one category value as well as one subcategory value was available. Values were collected on an ordinal scale and based on non-dichotomous data for the categories and dichotomous data for the subcategories. A commonly used approach to provide an overview about such data is a frequency table, as below, showing the number of patients in each category and visit:

		Treatment (N=x)
Visit x	n	x (100.0%)
	Category 1	xx (xx.x%)
	Subcategory 1	xx (xx.x%)
	Subcategory 2	xx (xx.x%)

Table 1: Frequency table by visit, category and subcategory

Also transition tables are often used to show how the patients changed categories over time:

		Baseline	
		Category 1	Category 2
Visit x	Category 1	xx (xx.x%)	xx (xx.x%)
	Category 2	xx (xx.x%)	xx (xx.x%)

Table 2: Shifts from baseline

A listing showing the information on each patient will help to complete the overview:

Subject Number	Visit	Category	Subcategory
Subject 1	Visit 1	<Category value>	<Subcategory value>
	Visit 2	<Category value>	<Subcategory value>
Etc.			

Table 3: Listing of individual data

We decided to combine all of above in one figure which would:
- show the distribution of patients across the categories and during different visits
- indicate the proportion of patients within each subcategory at a certain visit
- show the patients who switch between visits and categories

For our case a mosaic plot based on two ordinal-scaled variables was chosen to present the data. For each visit a separate mosaic-like part was created as below. On the x-axis the four visits are displayed and the y-axis shows the percentage of patients in each category and visit.
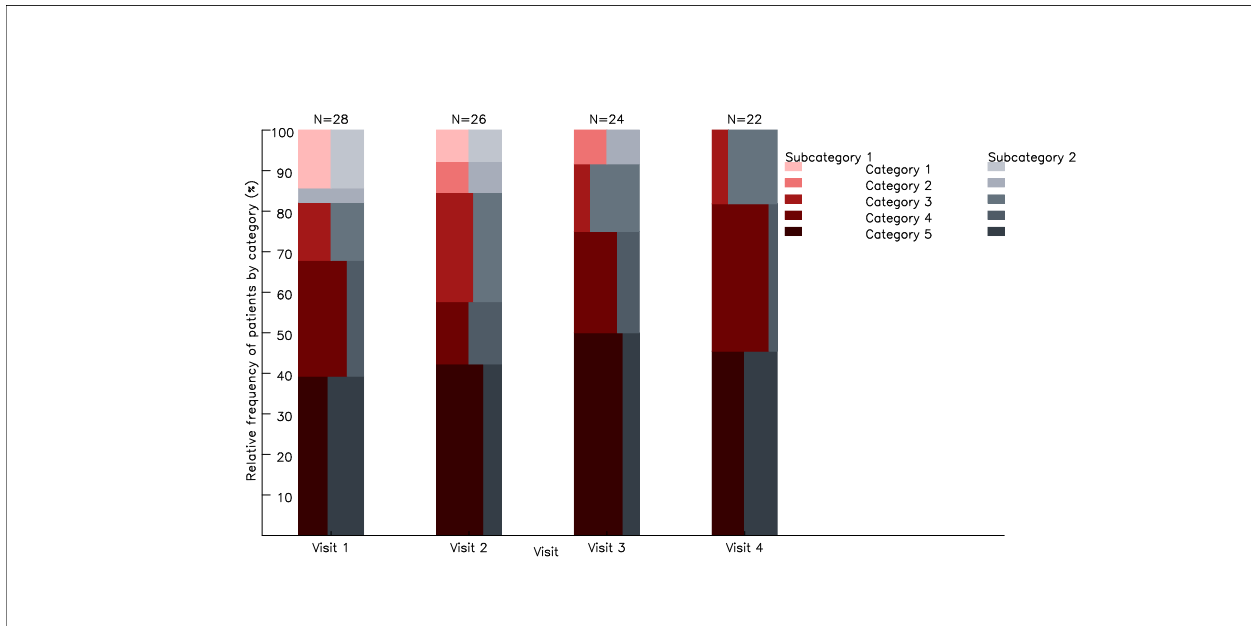

Figure 1 Mosaic plot, first version

After that it was decided to enhance the figure and add a line for each patient who switched from one category to another between two consecutive visits. The lower the category displayed in the mosaic plot, the 'worse' it is, i.e. the 'best' category is displayed on the top. For patients switching to a 'better' category a green line is drawn and for those who switched to a 'worse' category a red line. Proportion of patients switching into a better category are bordered by horizontal green lines, those switching to a worse

2

category are bordered by horizontal red lines. The N on the top of each bar indicates the overall number of patients who had values for the visit. This number is used as the denominator for the percentages. The resulting figure is presented below.
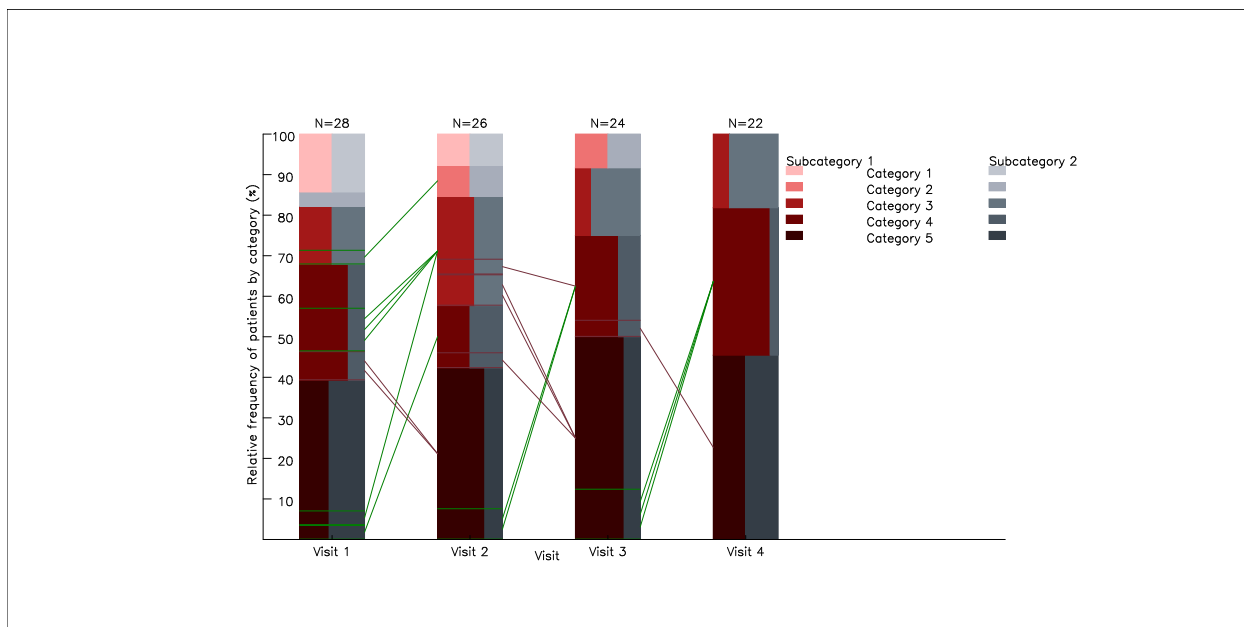

Figure 2 Mosaic plot final version

Although this figure looks quite complicated it only consists of three key elements: lines, rectangles and labels. To create this figure we use Annotate macros %line, %bar and %label. First of all, we calculate all the frequencies for the categories and subcategories by visit and with quite simple manipulations create the data set that we use for bars. In this paper we will not concentrate on how we got this data set as anyone familiar with the FREQ procedure and the RETAIN statement can easily create it. The data set FOR_BARS looks as below:

CATEGORY	SUBCATEGORY	VISIT	X_START	Y_START	X_END	Y_END
Category 1	Subcategory 1	Visit 1	xx	Xx	xx	xx
Etc.						

Data set 1: FOR_BARS

Once this data set is created we compile the Annotate macros with %annomac and then apply %bar macro to create bars themselves:

```
%annomac;
DATA BARS;
      %dclanno;
      SET FOR_BARS;
      XSYS='2';
      YSYS='2';
      %bar(x1=X_START,y1=Y_START,x2=X_END,y2=Y_END,bartyp=0,pattern=1);
RUN;
```

After execution of this code we only need to populate the variable COLOR so that each bar is coloured based on category and subcategory. Similarly we calculate all required variables to draw the connecting

lines in the data set FOR_LINES. This data set contains one record per switcher and additional records for horizontal lines to indicate the borders:

CATEGORY	SUBCAT EGORY	VISIT	X_START	Y_START	X_END	Y_END	SWITCHE R_TYPE
Category 1	Subcat egory 1	Visit 1	xx	Xx	Xx	xx	Better
Etc.							

Data set 2: FOR_LINES

Next step is to create a line for each switcher and for the borders. This is done based on the variable SWITCHER_TYPE:

```
DATA LINES;
     %dclanno;
     SET FOR_LINES;
     XSYS='2';
     YSYS='2';
     IF SWITCHER_TYPE="Better" THEN DO;
          %line(X_START,Y_START,X_END,Y_END,GREEN,1,0.3);
     END;
     ELSE IF SWITCHER_TYPE="Worse" THEN DO;
          %line(X_START,Y_START,X_END,Y_END,RED,1,0.3);
     END;
RUN;
```

Using Annotate macros it is even possible to create the legend and the axes with the same logic as above. The final step is to set all the pieces together into one data set ANNO and to create the figure using the GANNO, the GSLIDE and the GREPLAY procedures:

```
FILENAME FIGURE "your file";

GOPTIONS RESET=all FTEXT=simplex HTEXT=1.5 BORDER HPOS=102 VPOS=60
HSIZE=100 VSIZE=50 GSFMODE =REPLACE GSFLEN=8092 DEVICE=emf RESET=global
CTEXT=black GSFNAME=FIGURE;

PROC GANNO ANNOTATE=ANNO GOUT=tempcat NAME='ganno';
RUN;
QUIT;

PROC GSLIDE GOUT=tempcat NAME='gslide';
RUN;
QUIT;

PROC GREPLAY NOFS TC=tempcat IGOUT=tempcat GOUT=work;
     TDEF bild
     1/LLX=0      LLY=0
       ULX=0      ULY=100
       URX=100    URY=100
       LRX=100    LRY=0

     2/LLX=15     LLY=15
       ULX=15     ULY=80
       URX=80     URY=80
       LRX=80     LRY=15
```

```
        ;
        TEMPLATE bild;
        TREPLAY 1:gslide 2:ganno;
        DELETE _ALL_;
   QUIT;
```

With the data presented in this fashion, the reviewer gains an immediate overview of the categories and subcategories over time. First of all, trends over time regarding categories can be identified based on the height of the coloured bars. Furthermore, it is possible to investigate for each category the proportion of patients within subcategories which is shown by the gray part of the respective category rectangle. In addition, the category switching patterns are also of interest and can be readily seen. This figure shows for each patient who changes categories between two visits, direction (by colour of the connecting line), the base and new category. This figure has proven to be of high value for our project. Nevertheless, it may not be appropriate for all situations where trends over time for categories and subcategories shall be presented. In case of a large sample, a high number of categories or visits this figure might get too crowded to provide a good overview. To compare different groups (e.g. treatment arms or subgroups) an option would be to provide this display per group. Here, the number of patients included is limited. In case a larger database is used, it might be difficult to indicate individual switching behaviours. For these situations, we suggest to present the proportion of switching patients with triangles. The left side of the triangle represents the proportion of switchers into a certain subsequent category. The right vertical identifies the subsequent category.

ENHANCED SCATTER PLOTS

For the other project we needed to compare the distributions of values between two visits and three treatment groups and to investigate individual changes. Usually for this one would need a table with descriptive statistics:

Treatment group	Visit	n	Mean	SD	Min	Median	Max
<Treatment group> (N=x)	Visit 1	xx	xx.x	xx.x	xx.x	xx.x	xx.x
	Visit 2	xx	xx.x	xx.x	xx.x	xx.x	xx.x
<Treatment group> (N=x)	Visit 1	xx	xx.x	xx.x	xx.x	xx.x	xx.x
	Visit 2	xx	xx.x	xx.x	xx.x	xx.x	xx.x
<Treatment group> (N=x)	Visit 1	xx	xx.x	xx.x	xx.x	xx.x	xx.x
	Visit 2	xx	xx.x	xx.x	xx.x	xx.x	xx.x
Etc.							

Table 4: Descriptive statistics table by treatment and visit

and a listing as below:

Subject Number	Treatment Group	Value at Visit 1	Value at Visit 2
Subject 1	Group 1	xx	xx
Subject 2	Group 2	xx	xx
Etc.			

Table 5: Listing of individual data

To gain a better visualisation of the results we decided to create a figure. Scatter plots are the simplest kind of figures so the first version was not very informative. As you can see below it is hard to see
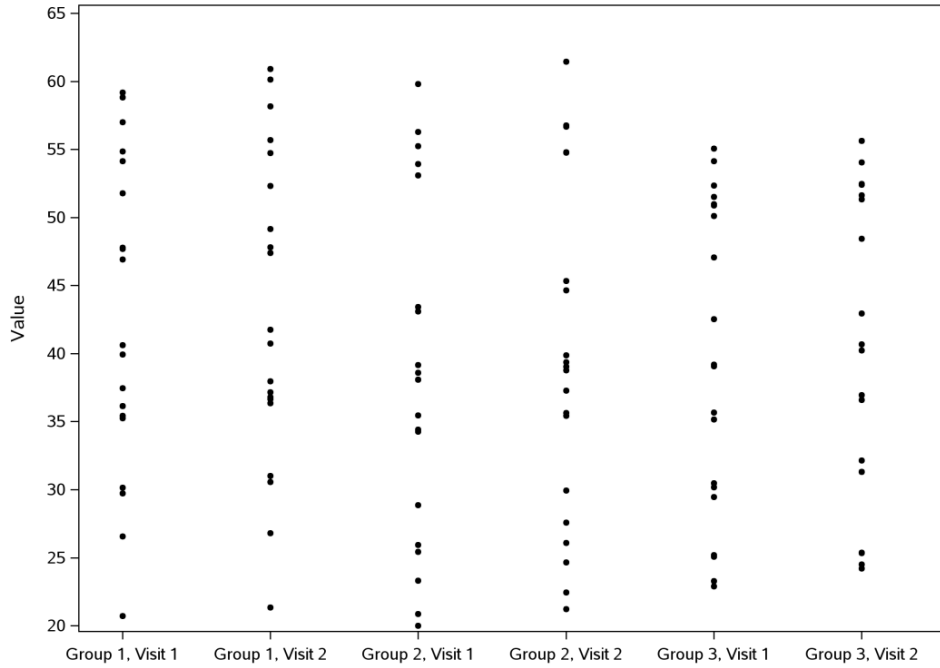
whether the value remained the same or changed:

Figure 3: Scatter plot, first version

Then we decided to connect the values for each patient to make the change easier to observe and we also added green lines indicating how the quartiles changed from visit 1 to visit 2, with red lines for the mean and blue ones for the median:
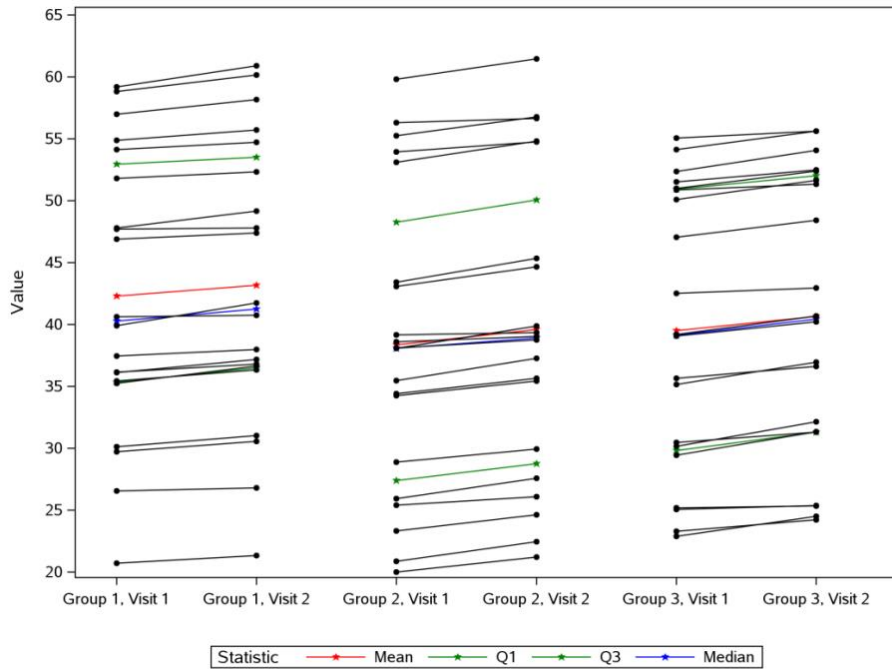

Figure 4: Scatter plot, final version

The original data set PATIENT has the following structure:

USUBJID Unique Subject Identifier	GROUP Treatment Group	VISIT Visit	AVAL Value
1	Group 1	Visit 1	xx
1	Group 1	Visit 2	xx
2	Group 2	Visit 1	xx
Etc.			

Data set 3: PATIENT

For our convenience we also add a variable XVAR that combines treatment and visit, i.e. for Treatment 1 and Visit 1 XVAR="Group1, Visit 1" and so on, and two supplementary formats:

```
PROC FORMAT;
     VALUE GROUP
           1='Group 1, Visit 1'
           2='Group 1, Visit 2'
           3='Group 2, Visit 1'
           4='Group 2, Visit 2'
           5='Group 3, Visit 1'
           6='Group 3, Visit 2';

     VALUE STAT
           1='Mean'
           2='Q1'
           3='Q3'
           4='Median';
RUN;
```

Then we create a data set with the descriptive statistics and make a set with the original data set:

USUBJID Unique Subject Identifier	GROUP Treatment Group	VISIT Visit	XVAR Supplementary Variable Format= GROUP.	AVAL Value	STAT_ NUMBER Statistic Number Format= STAT.	STAT Statistic	INDEX
	Group 1	Visit 1	1-Group 1, Visit 1		1-Mean	Xx	1
	Group 1	Visit 1	1-Group 1, Visit 1		2-Q1	Xx	2
	Group 1	Visit 1	1-Group 1, Visit 1		3-Q3	Xx	3

	Group 1	Visit 1	1-Group 1, Visit 1		4-Median	Xx	4
Etc.							
1	Group 1	Visit 1	1-Group 1, Visit 1	xx			5
1	Group 1	Visit 2	2-Group 1, Visit 2	xx			5
2	Group 2	Visit 1	3-Group 2, Visit 1	xx			6
Etc.							

Data set 4: ANALYSIS

The variables XVAR and STAT_NUMBER are numeric with a format applied, and the table above shows native and formatted values separated by '-'. Also the variable INDEX is added as it will be required later. For the descriptive statistics it is equal to STAT_NUMBER and for observations with individual patient data it equals USUBJID+4. After this we define styles so that for all the patients we have the same line style, symbol and colour and for each kind of statistic we have a different colour and a filled star as a symbol:

```
%MACRO define_style;
    PROC SQL NOPRINT;
        SELECT COUNT(DISTINCT USUBJID) INTO :subjnum FROM PATIENT;
    QUIT;

    PROC TEMPLATE;
        DEFINE STYLE Styles.REPORT;
        parent = styles.journal;
        STYLE GraphDataDefault FROM GraphDataDefault /
        markersize = 4px MARKERSYMBOL = 'circle';
        STYLE GraphData1 FROM GraphData1 / CONTRASTCOLOR = red
        MARKERSYMBOL = 'starfilled';
        STYLE GraphData2 FROM GraphData1 / CONTRASTCOLOR = green
        MARKERSYMBOL = 'starfilled';
        STYLE GraphData3 FROM GraphData1 / CONTRASTCOLOR = green
        MARKERSYMBOL = 'starfilled';
        STYLE GraphData4 FROM GraphData1 / CONTRASTCOLOR = blue
        MARKERSYMBOL = 'starfilled';
        %DO i = 5 %TO %eval(&subjnum+4);
            STYLE GraphData&i FROM GraphData1 /
            MARKERSYMBOL ='circlefilled' CONTRASTCOLOR = black;
        %END;
        END;
    RUN;
%MEND;
```

To create a separate series plot and scatter plot for each treatment group we split the variable STAT into three separate variables STAT1, STAT2, STAT3, each populated for the respective treatment group. Also

a separate series plot and scatter plot are created for the individual patient data. The final step is to create a template for the figure and to run the PROC SGRENDER:

```
PROC TEMPLATE;
    DEFINE STATGRAPH Tmpl;
        BEGINGRAPH;
            LAYOUT LATTICE / ROWS=1 ROWWEIGHTS=(1) COLUMNS=1
            COLUMNWEIGHTS=(1) COLUMNGUTTER=0;
            CELL;
            LAYOUT OVERLAY / XAXISOPTS=(LABEL=" " offsetmax=0.1
            offsetmin=0.05 TYPE=linear
            linearopts=(viewmin=1 viewmax=6 tickvaluelist=(1 2 3 4 5
            6) tickvalueFormat=group. tickvaluefitpolicy=rotate))

            YAXISOPTS=(LABEL="Value" TYPE=linear
            linearopts=(viewmin=20 viewmax=65 tickvaluelist=(20 25
            30 35 40 45 50 55 60 65) tickvalueFormat=2.0));

            SERIESPLOT X=XVAR Y=STAT1 / NAME="series1"
            GROUP=STAT_NUMBER INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SCATTERPLOT X=XVAR Y=STAT1 / NAME="scatter1"
            GROUP=STAT_NUMBER INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SERIESPLOT X=XVAR Y=STAT2 / NAME="series2"
            GROUP=STAT_NUMBER INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SCATTERPLOT X=XVAR Y=STAT2 / NAME="scatter2"
            GROUP=STAT_NUMBER INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SERIESPLOT X=XVAR Y=STAT3 / NAME="series3"
            GROUP=STAT_NUMBER INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SCATTERPLOT X=XVAR Y=STAT3 / NAME="scatter3"
            GROUP=STAT_NUMBER INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SERIESPLOT X=XVAR Y=AVAL / NAME="series4" GROUP=USUBJID
            INDEX=INDEX INCLUDEMISSINGGROUP=false;

            SCATTERPLOT X=XVAR Y=AVAL / NAME="scatter4"
            GROUP=USUBJID INDEX=INDEX INCLUDEMISSINGGROUP=false;

            DiscreteLegend "series1" "scatter1" / TITLE="Statistic"
            MERGE=true;

            ENDLAYOUT;
            ENDCELL;
            ENDLAYOUT;
        ENDGRAPH;
    END;
RUN;

ODS LISTING GPATH="your output path" IMAGE_DPI=300 STYLE=REPORT;
    ODS GRAPHICS / RESET IMAGEFMT  = PNG
                    BORDER    = OFF
                    HEIGHT    = 600px
                    WIDTH     = 800px
```

9

```
                        IMAGENAME = "lineplot_stat";

    PROC SGRENDER TEMPLATE=tmpl DATA=ANALYSIS;
    RUN;
    ODS GRAPHICS off;
```

This example code illustrates the idea of creating such a figure. Nevertheless, it can be modified and extended according to the user's needs.

With the data presented like this, the reviewer can easily identify the distribution of data by group and visit. Additionally the change over time can be seen overall as well as for every individual patient. Interpretation of overall changes needs to be done carefully as changes of means are presented in this case rather than means of changes which may be different in case patients have missing values at either of the visits. In situations with informative missings, one might address this topic by plotting in addition means only for those patients with non-missing values at both visits and connect only those means.

This display is intended to present the change over time in values in case only a limited number of patients is included in the analysis. Within a small database this display is very useful to identify distributions and time courses for multiple groups. With larger samples the figure might not be useful due to the high number of overlaying lines.

ENHANCED BOX PLOTS

In one of our projects we needed to analyze the number of patients who met the laboratory criteria of Hy's law by visit. We used the definition from ADaM examples provided by CDISC [1]. To satisfy the criteria the patient required to have total bilirubin elevations to be greater than 1.5 times ULN in addition to transaminase (alanine transaminase [ALT] or aspartate transaminase [AST]) elevations of greater than 1.5 times ULN. We wanted to present the data that would combine the information from the following frequency tables:

Parameter	Visit	Number (%) of subjects with high results
ALT	Visit 1	xx / xx (xx.x%)
	Visit 2	xx / xx (xx.x%)
Etc.		

Table 6: Frequency table for subjects with high results

,

Visit	Number (%) of subjects who satisfy Hy's law criteria
Visit 1	xx / xx (xx.x%)
Visit 2	xx / xx (xx.x%)
Etc.	

Table 7: Frequency table for subjects who satisfy Hy's law criteria

and a listing showing the information on each patient with enzyme elevations to complete the overview:

Subject Number	Visit	ALT/ULN	AST/ULN	Bilirubin/ULN	Hy's law criteria met?
Subject 1	Visit 1	x.x	x.x	x.x	Y/N

	Visit 2	x.x	x.x	x.x	Y/N
Etc.					

Table 8: Listing of individual data
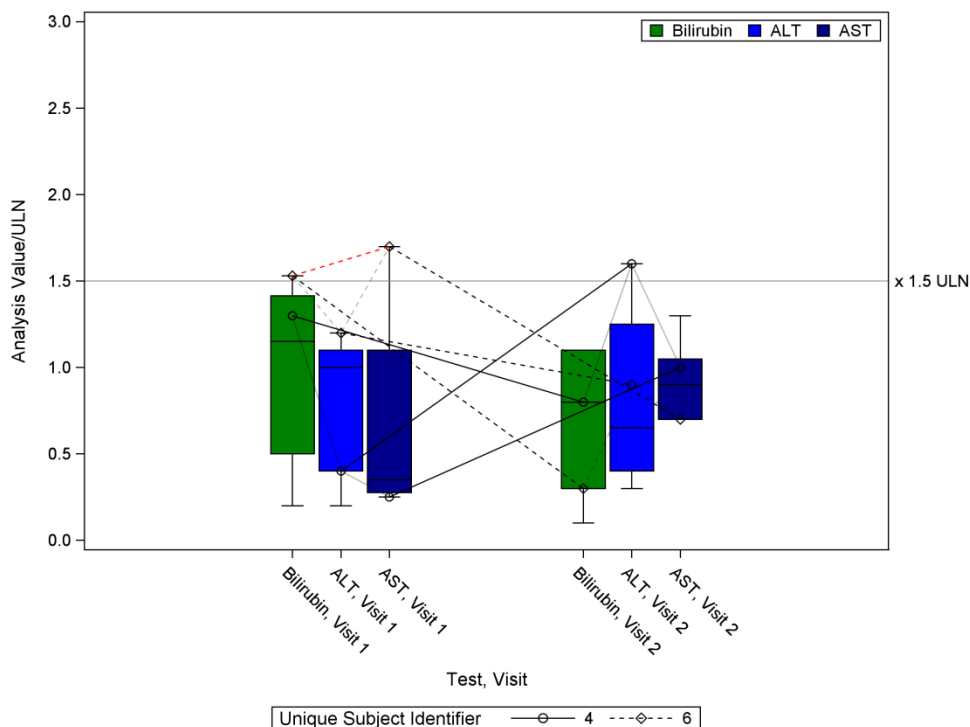
The result we came up with looks like below:

Figure 5: Box Plot

Here we analyse the values relative to upper limit of normal for three laboratory test parameters. Boxes show 25th to 75th percentile, horizontal line is the median and the whiskers show minimum and maximum values. For those patients who have any of these ratios greater or equal to 1.5 we create a scatter plot connected with the lines to provide a detailed overview of the values and their changes over time. The values for each such patient are connected with separate lines between visits and between tests. In case the Hy's law criteria which are defined as above are met at one of the visits, the line connecting the increased values is red. Although the plot looks rather complicated it is nothing more than a combination of box plots, series plots and scatter plots. In total there are three box plots: one for each laboratory test, three scatter plots: one for each laboratory test and six series plots: three connecting values between tests, two connecting values between visits and one for those values that satisfy the criteria. The original data set contains four variables: USUBJID, PARAM, VISIT and RATIO. Similarly to the previous example we create variable XVAR that combines PARAM and VISIT.

USUBJID Unique Subject Identifier	VISIT Visit	PARAM Paramer	XVAR Supplementary Variable Format=XVAR.	RATIO Value
1	Visit 1	Bilirubin	1-Bilirubin, Visit 1	xx
1	Visit 2	Bilirubin	2-Bilirubin, Visit 2	xx
Etc.				

Data set 5: TOTAL

For each of the elements specified above a set of variables is created based on RATIO and XVAR, i.e. XVAR1 and RATIO1 are only populated for bilirubin values, XVAR2 and RATIO2 for alanine transaminase, XVAR3 and RATIO3 for aspartate transaminase, XVAR4 and RATIO4 are only populated for bilirubin for those patients who have at least one of the values of RATIO greater or equal to 1.5 etc., XVAR7 and RATIO7 are used for Visit 1 and XVAR8 and RATIO8 for Visit 2 and finally, XVAR9 and RATIO9 are only populated for those values that fit the criteria in order to create red lines. This code will only work in SAS 9.4 as it provides an opportunity to have x-axis linear for box plots which allows us to nicely match box plots with all the other elements. The code below concludes our derivations:

```
PROC TEMPLATE;
      DEFINE STATGRAPH Tmpl;
      BEGINGRAPH;
           LAYOUT LATTICE / ROWS=1 COLUMNS=1 COLUMNWEIGHTS=(1)
           COLUMNGUTTER=0;
           CELL;
           LAYOUT OVERLAY / XAXISOPTS=(LABEL="Test, Visit" TYPE=linear
           linearopts=(viewmin=0 viewmax=10 tickvaluelist=(1 2 3 7 8 9 )
           tickvalueFormat=XVAR. tickvaluefitpolicy=ROTATE ))
           YAXISOPTS=(LABEL="Analysis Value/ULN" TYPE=linear
           linearopts=(viewmin=0 viewmax=3 tickvaluelist=(0 0.5 1 1.5 2
           2.5 3 ) tickvalueFormat=3.1));

           BOXPLOT X=XVAR1 Y=RATIO1 / NAME="boxplot1" DISPLAY=(CAPS FILL
           MEDIAN) EXTREME=FALSE FILLATTRS=GRAPHDATA1
           BOXWIDTH=0.15;

           BOXPLOT X=XVAR2 Y=RATIO2 / NAME="boxplot2" DISPLAY=(CAPS FILL
           MEDIAN ) EXTREME=FALSE FILLATTRS=GRAPHDATA2
           BOXWIDTH=0.15;

           BOXPLOT X=XVAR3 Y=RATIO3 / NAME="boxplot3" DISPLAY=(CAPS FILL
           MEDIAN ) EXTREME=FALSE FILLATTRS=GRAPHDATA3
           BOXWIDTH=0.15;

           SERIESPLOT X=XVAR4 Y=RATIO4 / NAME="series1" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false;

           SCATTERPLOT X=XVAR4 Y=RATIO4 / NAME="scatter1" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false;

           SERIESPLOT X=XVAR5 Y=RATIO5 / NAME="series2" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false;

           SCATTERPLOT X=XVAR5 Y=RATIO5 / NAME="scatter2" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false;

           SERIESPLOT X=XVAR6 Y=RATIO6 / NAME="series3" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false;

           SCATTERPLOT X=XVAR6 Y=RATIO6 / NAME="scatter3" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false;

           SERIESPLOT X=XVAR7 Y=RATIO7 / name="series4" GROUP=USUBJID
           INCLUDEMISSINGGROUP=false datatransparency=0.7;
```

```
                SERIESPLOT X=XVAR8 Y=RATIO8 / NAME="series5" GROUP=USUBJID
                INCLUDEMISSINGGROUP=false datatransparency=0.7;

                SERIESPLOT X=XVAR9 Y=RATIO9 / NAME="series6" GROUP=USUBJID
                INCLUDEMISSINGGROUP=false LINEATTRS=(color=red);

                REFERENCELINE Y=1.5 / NAME="referenceline1" CURVELABEL="x 1.5
                ULN" datatransparency=0.3;

                MERGEDLEGEND "series1" "scatter1" / TITLE="Unique Subject
                Identifier";
        ENDLAYOUT;
        ENDCELL;
        ENDLAYOUT;
    ENDGRAPH;
    END;
    RUN;

    PROC TEMPLATE;
        DEFINE STYLE Styles.Report;
        parent = styles.journal;
        STYLE graphdata1 FROM graphdata1 /
            COLOR = green markersymbol = "Circle" linestyle = 1;
        STYLE graphdata2 FROM graphdata1 /
            COLOR = blue markersymbol = "Diamond" linestyle = 2;
        STYLE graphdata2 FROM graphdata1 /
            COLOR = darkblue markersymbol = "Diamond" linestyle = 3;

    RUN;

    ODS LISTING GPATH="your output path" IMAGE_DPI=300 STYLE=REPORT;
    ODS GRAPHICS / RESET OUTPUTFMT = PNG
        BORDER = OFF HEIGHT = 600px WIDTH = 800px IMAGENAME = "box_stat";

    PROC SGRENDER TEMPLATE=tmpl DATA=TOTAL;
    RUN;

    ODS GRAPHICS off;
```

This graphical display provides summary statistics as well as individual patient data for different laboratory tests. Subjects with values larger than 1.5 times ULN are identified and highlighted in red in case they meet the defined criteria combination. In addition the development over time is indicated for these patients of interest. With this a variety of analyses is covered and presented within one overview. In situations with different criteria (e.g. different thresholds between laboratory parameters) this approach can be taken over by adjusting the reference line, the algorithm to identify patients for who series plots are drawn and those line which indicate that the patient fulfills the criteria.

ENHANCED MEANDEVIATION PLOTS

In another case we would like to share with our readers we were interested in the change of values during the study. The study drug was supposed to increase the certain laboratory parameter and we wanted to combine the information from the following outputs:

Visit	n	Mean	SD	Min	Median	Max
Visit 1	xx	xx.x	xx.x	xx.x	xx.x	xx.x

Visit 2	xx	xx.x	xx.x	xx.x	xx.x	xx.x

Table 9: Descriptive statistics table by visit

,

	Number (%) of subjects who reached target value
Visit 1	xx / xx (x.x%)
Visit 2	xx / xx (x.x%)
Etc.	

Table 10: Frequency table for subjects who reached the target value by visit

and a listing as below:

Subject	Visit	Values
Subject 1	Visit 1	xx
	Visit 2	xx
Etc.		

Table 11: Listing of individual data

Our suggestion is presented below. Here meandeviation plot is combined with line plot. For each of the patients with at least one value less than 9 an individual set of lines is created to trace the values until the patient reaches the stable value of 9, i.e. the values remain high or until the study end whatever is first. Individual patients with values below the threshold can be identified based on different line types. Subject identifiers are given in the legend. The number of patients at each visit is indicated at the bottom.
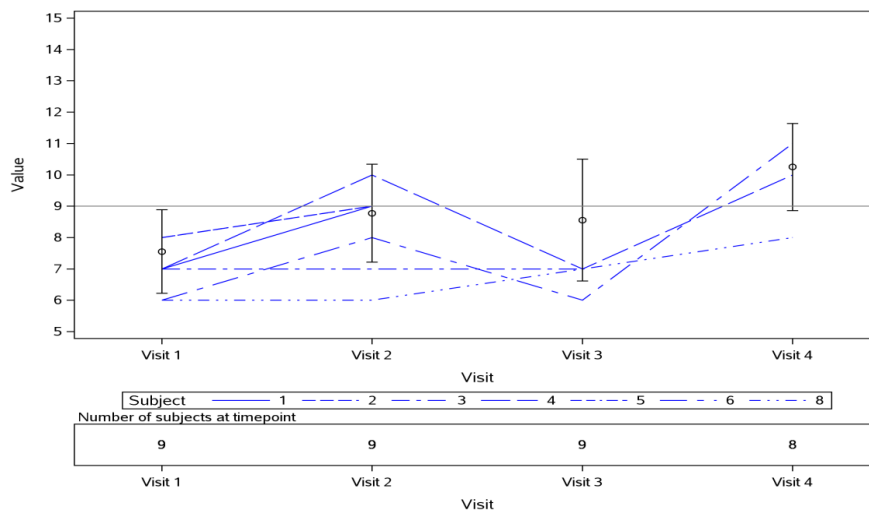

Figure 6: Meandeviation Plot

The original data set contains variables USUBJID, VISIT and AVAL. We make a subset of all the patients who have at least one value less than 9 and make a set with the data set with descriptive statistics. Also we calculate number of patients present at each of the visits and store it in the variable COUNT. The resulting data set is called TOTAL:

USUBJID Unique Subject Identifier	VISIT Visit Format=VISIT.	AVAL Value	MEAN Mean	ERRORLOWER Mean-SD	ERRORUPPER Mean+SD	COUNT Number of Subjects
1	Visit 1	xx				xx
Etc.						
	Visit 1		xx	xx	xx	xx
Etc.						

Data set 6: TOTAL

And then the final step is to define the template:

```
PROC TEMPLATE;
     DEFINE STATGRAPH Tmpl;
     BEGINGRAPH;
          LAYOUT LATTICE / ROWS=2 ROWWEIGHTS=(0.8 0.2) COLUMNS=1
          COLUMNWEIGHTS=(1) COLUMNGUTTER=0;
          CELL;
          LAYOUT OVERLAY / XAXISOPTS=(LABEL="Visit" TYPE=linear
          offsetmax=0.1 offsetmin =0.1 linearopts=(viewmin=1 viewmax=4
          tickvaluelist=(1 2 3 4) tickvalueFormat=Visit.
          tickvaluefitpolicy=rotate)) YAXISOPTS=(LABEL="Value"
          TYPE=linear
          linearopts=(viewmin=5 viewmax=15 tickvaluelist=(5 6 7 8 9 10
          11 12 13 14 15 ) tickvalueFormat=best12.));

          SERIESPLOT X=VISIT Y=AVAL / GROUP=USUBJID NAME="series1"
          INCLUDEMISSINGGROUP=false LINEATTRS=(color=blue );

          SCATTERPLOT X=VISIT Y=MEAN / NAME="scatter1"
          YERRORLOWER=ERRORLOWER YERRORUPPER=ERRORUPPER;

          REFERENCELINE Y = 9;
          DiscreteLegend "series1" / title="Subject";
          ENDLAYOUT;
          ENDCELL;

          CELL;
          CELLHEADER;
          ENTRY halign=left "Number of subjects at timepoint" /;
          ENDCELLHEADER;
          LAYOUT OVERLAY / XAXISOPTS=(LABEL="Visit" TYPE=linear
          offsetmax=0.1 offsetmin =0.1 linearopts=(viewmin=1 viewmax=4
          tickvaluelist=(1 2 3 4) tickvalueFormat=Visit.
          tickvaluefitpolicy=rotate));

          BLOCKPLOT X=VISIT BLOCK=COUNT / NAME="block" DISPLAY=(values)
          REPEATEDVALUES=true LABELPOSITION=left VALUEHALIGN=start;
          ENDLAYOUT;
          ENDCELL;
          ENDLAYOUT;
     ENDGRAPH;
     END;
```

```
RUN;

ODS LISTING GPATH="your output path" IMAGE_DPI=300;
     ODS GRAPHICS / RESET IMAGEFMT = PNG
                    BORDER = OFF
                    HEIGHT = 600px
                    WIDTH = 800px
                    IMAGENAME = "meansd_stat";

PROC SGRENDER TEMPLATE=tmpl DATA=TOTAL;
RUN;
ODS GRAPHICS off;
```

This figure provides information about the development of the laboratory parameter over time for the complete population as well as for individual patients. In this example evenly spaced visits are given. If this is not the case the visit schedule should be visualized on the x-axis (e.g. by presenting the planned study day of the visit) to avoid misinterpretations .The definition of the subgroup to be presented as line plots needs to be clarified carefully together with the medical functions. There may be situations where the development of a large proportion or even of all patients is of interest. In other cases this type of figure may be used to detect few extreme cases. These decisions depend on the context, the data and the study design.

ENHANCED STEP PLOTS

The last but not the least in this paper is the step plot. In our case we had an event that occurred for most of the patients. For this event multiple occurrences with different grades were expected. We wanted to create a so called swimmer plot that would show both the individual patient data and the picture in general and would contain the information from the following tables and listing:

	Treatment (N=x)
Number (%) of subjects who had at least one event	xx (x.x%)
Maximum grade	xx (x.x%)
Grade 1	xx (x.x%)
Grade 2	xx (x.x%)
Etc.	

Table 12: Analysis of events

,

	Treatment (N=x)
Number of events per subject	
1	xx (x.x%)
2	xx (x.x%)
Etc.	

Table 13: Number of events per subject

,

	Number (%) of subjects
Treatment duration category (days)	

16

1-20	xx (x.x%)
Etc.	

Table 14: Frequency table for treatment duration

,

Subject Number	Relative day of event onset	Relative day of event end	Grade
Subject 1	1	3	2
	6	7	3
	10	18	2
Etc.			

Table 15: Listing of individual data

As a result we created a swimmer plot which contained coloured bars for each patient according to the grade, duration and sequence of event episodes. Lengths of bars in days for one patient add up to the whole treatment duration for this patient. Patients are sorted by treatment duration. Patient with smallest treatment duration is the first one, patient with longest treatment duration is the last one. For each patient a row is given. Width of the row is 100/number of patients. Grades are presented per day in the row of a patient as filled bar in different shades of a colour, with darkest shade meaning most severe grade, green meaning no event.
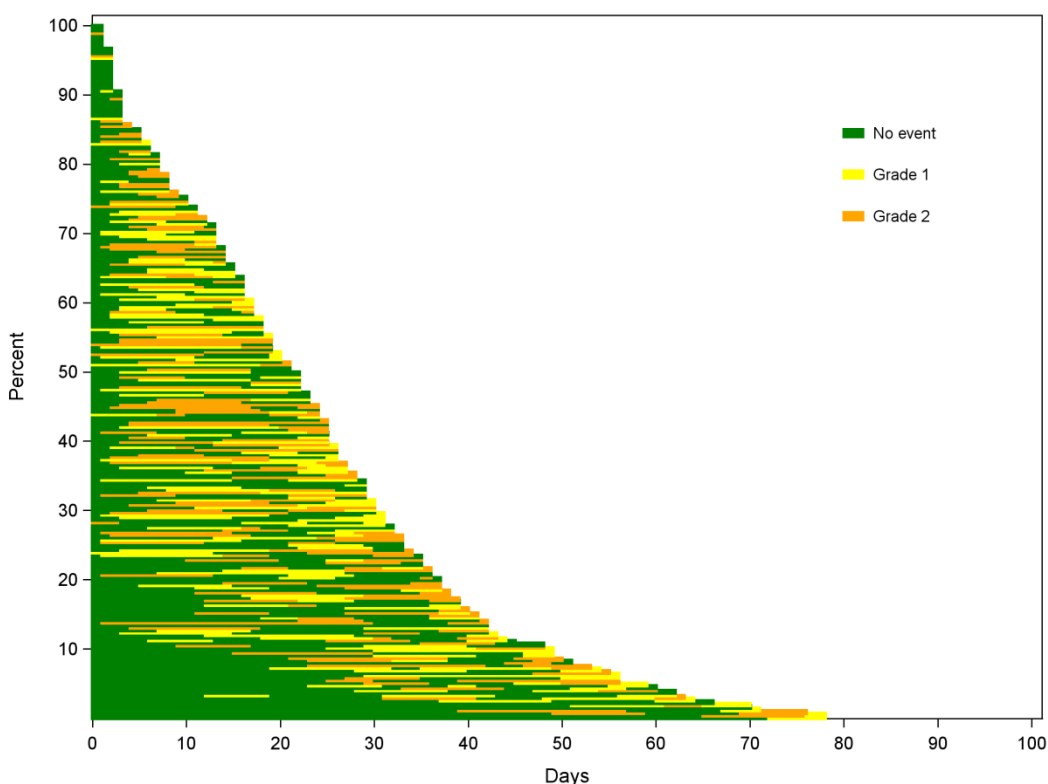

Figure 7: Step Plot

To create this figure the main idea is to prepare the data set similarly to what was done for mosaic plots. For each of the patients for each of the episodes we derive the start coordinates, width and height of the bar based on their start and end days and duration. The data set we create has the following structure:

17

USUBJID Unique Subject Identifier	GRADE Grade	X_START	Y_START	WIDTH_	HEIGHT_
Subject 1	Grade 1	xx	xx	xx	xx
Etc.					

Data set 7: FOR_BARS

In this case we decided to create such figure using Annotate macros and GTL. This approach requires SAS 9.4. We apply the Annotate macros similarly to how it was done for mosaic plots for rectangles and legend:

```
%sganno;
DATA ANNO end =_last_;
     LENGTH COLOR $200;
     SET FOR_BARS;
     IF GRADE=0 THEN COLOR="GREEN";
     IF GRADE=1 THEN COLOR="YELLOW";
     IF GRADE=2 THEN COLOR="ORANGE";
     %sgrectangle(x1=X_START,y1=Y_START,fillcolor=COLOR,linecolor=COLOR,
          display="all",drawspace="datavalue",anchor="bottomleft",
          height=HEIGHT_,width=WIDTH_);
     IF _last_ THEN DO;
        %sgtext(x1=83, y1=83, LABEL="No event", WIDTH=10,
        DRAWSPACE="datavalue", ANCHOR="bottomleft",TEXTSIZE=8)
        %sgtext(x1=83, y1=77, LABEL="Grade 1", WIDTH=10,
        DRAWSPACE="datavalue",ANCHOR="bottomleft")
        %sgtext(x1=83, y1=71, LABEL="Grade 2", WIDTH=10,
        DRAWSPACE="datavalue",ANCHOR="bottomleft")
        %sgrectangle(x1=80,y1=84,fillcolor="GREEN", linecolor="GREEN",
        display="all",drawspace="datavalue",ANCHOR="bottomleft",height=1,
        width=2);
        %sgrectangle(x1=80,y1=78,fillcolor="YELLOW", linecolor="YELLOW",
        display="all",drawspace="datavalue",ANCHOR="bottomleft",height=1,
        width=2);
        %sgrectangle(x1=80,y1=72,fillcolor="ORANGE",
        linecolor="ORANGE",display="all",drawspace="datavalue",ANCHOR="bo
        ttomleft",height=1,width=2);
     end;

RUN;
```

In order to create the figure with all the attributes in place we also need to define some dummy plot so for this data set DUMMY is used:

X	Y
0	0

Data set 8: DUMMY

```
PROC TEMPLATE;
     DEFINE STATGRAPH Tmpl;
     BEGINGRAPH;
          LAYOUT LATTICE / ROWS=1 COLUMNS=1 COLUMNWEIGHTS=(1)
          COLUMNGUTTER=0;
          CELL;
```

```
                LAYOUT OVERLAY / XAXISOPTS =(LABEL="Days" TYPE=linear
                offsetmin=0
                linearopts=(viewmin=0 viewmax=100 tickvaluelist=(0 10 20 30 40
                50 60 70 80 90 100) tickvaluefitpolicy=ROTATE ))
                YAXISOPTS =(LABEL="Percent" TYPE=linear offsetmin=0
                linearopts=(viewmin=0 viewmax=100 tickvaluelist=(0 10 20 30 40
                50 60 70 80 90 100)));

                SCATTERPLOT X=X  Y=Y / NAME="dummy";
                ANNOTATE;

                ENDLAYOUT;
                ENDCELL;
                ENDLAYOUT;
           ENDGRAPH;
           END;
      RUN;

      ODS LISTING GPATH="your output path" IMAGE_DPI=300 ;
      ODS GRAPHICS / RESET OUTPUTFMT = PNG
           BORDER = OFF HEIGHT = 600px WIDTH = 800px IMAGENAME = "step_plot";

      PROC SGRENDER TEMPLATE=tmpl DATA=DUMMY SGANNO=ANNO;
      RUN;
      ODS GRAPHICS off;
```

The step plot provides an intuitive overview of the events in the patients overall as well as in individual patients. Such an approach is extremely valuable to identify a seasonal pattern or a relationship between event occurrence or intensity and discontinuation. The users should keep in mind that when the number of patients increases it might become harder to interpret such figure. Such step plot can further enhanced to present Kaplan Meier analysis. The example is suggested by J. A. Dubin, H. Muller and J. Wang in their paper [2].

CONCLUSION

The combination of summary statistics and individual patient data within graphical presentations enables the reader to not only gain information about the population as a whole but provides the possibility to trace individual subjects. With this a more complete picture of the data is provided which leads to a better understanding of the outcome of interest, identification of patterns within individual patients or certain subsets of patients. Nevertheless, it is fundamental to decide for a certain graphical tool based on the medical question, study design properties and the available data.
It may be difficult in a larger sample to provide information for all or a large proportion of patients without crowding the display. In such cases it may be considered to only provide values for individual patients in a smaller subgroup. Common scenarios are that groups with a certain baseline characteristic are of special interest or patients with extreme values during the course of the study. In this paper we shared our ideas how to present the data in more flexible and informative ways. We hope that we were able to provide a fresh look at the common figure types and encourage our readers to think out of the box and search for the new and creative ways of data visualisation.

REFERENCES

[1] "Analysis Data Model (ADaM) Examples in Commonly Used Statistical Analysis Methods". http://www.cdisc.org/adam. The current version is downloadable from web page and available to CDISC members and non-members (access date 30Mar2017).

[2] Dubin J. A., Muller H. and Wang J. "Event history graphs for censored survival data"
 Statist. Med. 2001; 20:2951–2964 (DOI: 10.1002/sim.888).

CONTACT INFORMATION

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

Name: Yuliia Bahatska
Enterprise: Inventiv Health Clinical
Email: yuliia.bahatska@inventivhealth.com

Name: Christiane Ahlers
Enterprise: Bayer AG
Email: christiane.ahlers@bayer.com