

Visualizing Enrollment Over Time

Laura Gruetzner, BioStat Solutions, Inc., Frederick, MD

ABSTRACT

Taking a closer look at patient enrollment over time provides valuable insight and can support strategic decisions or further analysis for just about any type of clinical trial. You can often visualize early-phase studies with small sample sizes on the patient level, resulting in a compact and comprehensive summary of enrollment and events like progression, crossover, or adverse reactions. For larger trials with multiple treatment arms, area charts or simple line plots of patients per cohort over time let you keep track of the developing trial population, periods of slow enrollment, and proportions between arms.

This paper explores how to graphically inspect enrollment patterns over time for early- and late-phase clinical trials in an example-based approach. Focus is placed on different ways to visualize enrollment for dose escalation, randomized controlled two-arm, and basket (or umbrella) trials using SAS® 9.4 ODS Graphics.

INTRODUCTION

Keeping track of enrollment over time can prove important when analyzing clinical trial data. Often the study protocol contains instructions regarding randomization (e.g., treatment arm allocation ratios or randomization by specific biomarkers) and expected time to enrollment completion. Creating a visual summary of the growing patient population over time is an effective way to explore how, for example, periods of slow enrollment or randomization discrepancies between clinical sites can cause deviation from protocol defined targets. Comparability between interim analyses conducted at different points in time and the statistical power of results can be evaluated and ensured.

This paper explores ways to visualize enrollment for dose escalation, randomized controlled two-arm, and basket (or umbrella) trials using the SAS 9.4 SGPLOT procedure and Graph Template Language. SAS codes including step-by-step explanations and guidance on data preparation are provided in each section.

DOSE ESCALATION TRIALS

In dose escalation trials, increasing or decreasing dosages - depending on the presence or absence of dose limiting toxicities (DLTs) - are administered to a pre-specified number of patients per dosing cohort. With study populations commonly below 40, you can attain a comprehensive status report through a plot of enrollment and dose assignment over time on a patient-by-patient basis.

The example below is based on a classic 3+3 dose escalation design. For more details regarding this design, see Le Tourneau et al. (2009).

GENERATION OF THE PLOT WITH PROC SGPLOT

You can easily create this type of plot with PROC SGPLOT. One line is drawn for each subject with colors indicating the different cohorts. DLTs and first exposures are then scattered on top of the lines. Depending on patient counts and preference, you can either display subject IDs on the left of the respective lines or omit them to prevent crowding the plot. The two legends are placed at the top left corner and bottom of the graph for this example.

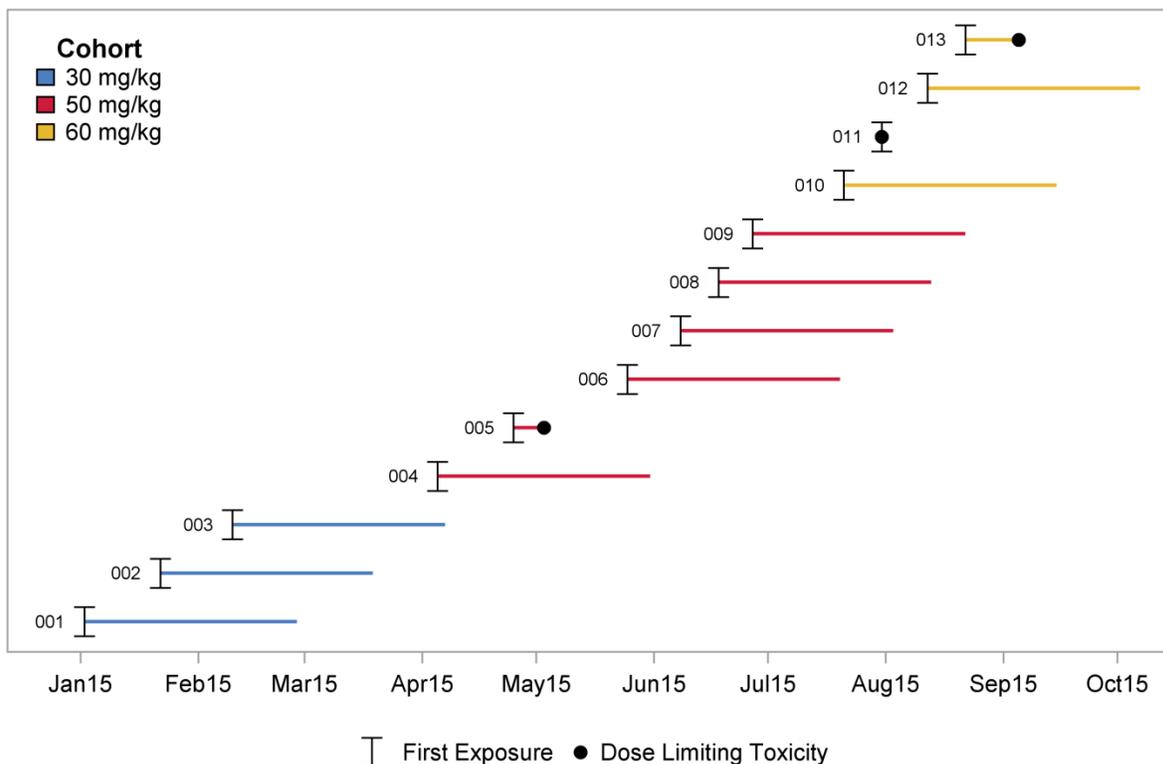


Figure 1. Plot of enrollment and dose escalation over time

THE INPUT DATA

The goal is to create a plot consisting of one line per subject indicating time frame of enrollment. To achieve this, you need to prepare an input data set capturing subject id, treatment arm, treatment start and end date. First exposures and DLTs will be drawn separately and require respective additional columns. You can easily create this data set based on the corresponding Study Data Tabulation Model (SDTM) domains DM and AE, if available. Make sure there are two observations per subject and the data set is sorted by order of recruitment and date to ensure lines appear in the correct order. In this example subject IDs are assumed to reflect order of recruitment.

	SUBJID	ARM	REF_TP	DATE	FIRST_EXP	DLT
1	001	30 mg/kg	RFSTDTC	02JAN2015	02JAN2015	.
2	001	30 mg/kg	RFENDTC	27FEB2015	.	.
3	002	30 mg/kg	RFSTDTC	22JAN2015	22JAN2015	.
4	002	30 mg/kg	RFENDTC	19MAR2015	.	.
5	003	30 mg/kg	RFSTDTC	10FEB2015	10FEB2015	.
6	003	30 mg/kg	RFENDTC	07APR2015	.	.
7	004	50 mg/kg	RFSTDTC	05APR2015	05APR2015	.
8	004	50 mg/kg	RFENDTC	31MAY2015	.	.
9	005	50 mg/kg	RFSTDTC	25APR2015	25APR2015	.
10	005	50 mg/kg	RFENDTC	03MAY2015	.	03MAY2015
11	006	50 mg/kg	RFSTDTC	25MAY2015	25MAY2015	.
12	006	50 mg/kg	RFENDTC	20JUL2015	.	.

Figure 2. The first 12 observations of the input data set for the dose escalation plot

CODE AND EXPLANATION

This code produces the dose escalation plot (Figure 1):

```

proc sgplot data=ESCDAT;
  series x=DATE y=SUBJID / group=SUBJID lineattrs=(pattern=1 thickness=2)
        name='cohorts' groupplc=ARM; 1

  scatter x=FIRST_EXP y=SUBJID / markerattrs=(color=black symbol=Ibeam
        size=16) name='f_exp' legendlabel='First Exposure'
        datalabel=SUBJID datalabelpos=left;
  scatter x=DLT y=SUBJID / markerattrs=(color=black symbol=circlefilled)
        name='DLT' legendlabel='Dose Limiting Toxicity'; 2

  xaxis display=(nolabel) values=('01JAN2015'd to '01OCT2015'd by month)
        offsetmax=.05 tickvalueformat=MONYY.;
  yaxis display=(nolabel novalues noticks); 3

  keylegend 'cohorts' / across=1 position=topleft location=inside
        type=linecolor title='Cohort'
        titleattrs=(weight=bold);
  keylegend 'f_exp' 'DLT' / position=bottom location=outside; 4
run;

```

1: The SERIES statement of *DATE* vs. *SUBJID* grouped by *SUBJID* draws a line for each subject. To construct line colors and the legend entries by dose groups instead of subjects, the GROUPLC= option is set to *ARM*. The NAME= option allows you to reference this statement later when constructing the legend.

2: The first SCATTER statement plots the variable *FIRST_EXP* and creates markers at respective time points. Data labels are set to display the subject ID with the DATALABEL= option. The variable *DLT* is plotted in a similar fashion. Legend labels are defined via the LEGENDLABEL= option.

3: You can specify the range of the x-axis as well as tick marks with the VALUES= option. The format for the x-axis is chosen to only display month and year for this example. The OFFSETMAX= option extends the axis on the upper end by 5%, creating some space for the yellow line displaying subject 012. All attributes of the y-axis are suppressed.

4: Two legends are created, one for the first SERIES statement displaying the cohorts and one for the two SCATTER statements. To ensure one legend item appears for each cohort in the first legend, you need to specify the option TYPE=LINECOLOR. The ACROSS=1 option places each legend item in a new line.

ADVANTAGES AND DRAWBACKS

The dose escalation plot over time summarizes a lot of information and is easy to create. It presents when each subject started and ended the trial while the colored cohorts and visualized DLTs allow for an easy identification of the optimal dose (in this case 50mg/kg). As the code is easily set up to run on different interim data cuts without many modifications, the plot can be used to support strategic decisions and communication throughout the trial.

In the case of very slow enrollment, you may consider x-axis breaks to avoid wasted space. For studies with many cohorts or populations exceeding 30, drawing one line per subject could result in a crowded graph. In those scenarios, you may find the area chart approach described in the last section (BASKET AND UMBRELLA TRIALS) more suitable.

RANDOMIZED CONTROLLED TWO-ARM TRIALS

In a classic randomized controlled two-arm trial, the active comparator and control arm are often randomized disproportionally, e.g., adhering to a 2:1 or 3:1 allocation ratio. After trial start, it is important

to explore whether actual groups reflect planned proportions and counts over time. If not discovered and corrected early, discrepancies can negatively impact planning, waste valuable resources, and even influence power assumptions for future analyses stated in the protocol.

The following example is based on a two-armed treatment vs. placebo trial with a 2:1 allocation ratio. A total patient population of 600 patients is expected to be recruited within 21 months of trial start assuming a gradual rate of enrollment. In this example, a 2x1 panel plot is chosen to display actual and planned patient counts in addition to the proportions of the two treatment arms. You can create this panel plot via the procedures `TEMPLATE` (by using the `LAYOUT LATTICE` statement) and `SGRENDER`.

GENERATION OF THE PLOT WITH PROC TEMPLATE AND PROC SGRENDER

Two separate panels are drawn denoting patient counts and proportions of the two treatment groups for each month. The first panel consists of planned and actual patient numbers denoted by simple line plots, while the second displays monthly proportions between treatment arms in the form of bar plots. A reference line indicates the expected percentage of placebo patients at 33%, reflecting the 2:1 allocation ratio. In the legend at the bottom of the graph, actual and planned patient counts in the first panel are distinguished by line patterns, while colored boxes identify treatment arms.

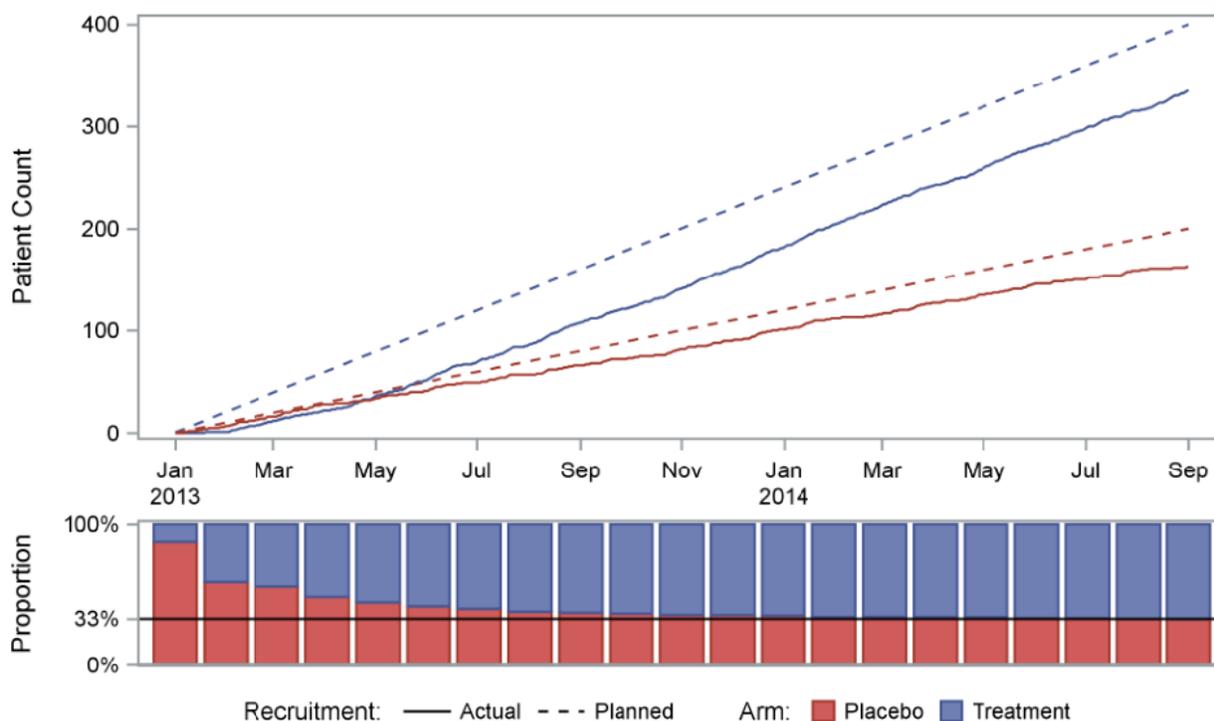


Figure 3. Planned and actual patient enrollment over time in a randomized controlled two-arm trial

THE INPUT DATA

The data set contains one row per arm and day for the planned enrollment time frame (01JAN2013 – 01SEP2014 in this example). To construct the variable `PLANNED` that reflects estimated patient counts, expected rates of enrollment were calculated based on the planned total number of patients and length of enrollment period. Observed actual patient counts are captured in the variables `ACTCOUNT`. For the bar plots, proportions of the two arms were calculated for each month (see `PERCENT`). The variable `MONTH` is only populated for the first day of each month, ensuring that x-axes of both panels align and the right number of bars is drawn.

	DATE	ARM	PLANNED	ACTCOUNT	MONTH	PERCENT
119	01MAR2013	Treatment	39.408866995	22	MAR13	76%
120	01MAR2013	Placebo	19.704433498	8	MAR13	24%
121	02MAR2013	Treatment	40.065681445	22	.	.
122	02MAR2013	Placebo	20.032840722	8	.	.
123	03MAR2013	Treatment	40.722495895	22	.	.
124	03MAR2013	Placebo	20.361247947	8	.	.

Figure 4. Excerpt of the input data set for the plot of planned and actual patient enrollment over time

CODE AND EXPLANATION

This code produces the plot of planned and actual patient enrollment over time in a randomized controlled two-arm trial (Figure 3):

```

/* create template for 2x1 panel plot */
proc template;
  define statgraph twoarms;
    beginnograph;
      layout lattice / rowweights=(.75 .25); 1

      legenditem type=line name="a" / lineattrs=(color=black pattern=1
        thickness=1.5) label="Actual";
      legenditem type=line name="b" / lineattrs=(color=black pattern=2
        thickness=1.5) label="Planned"; 2

      /* first panel */
      layout overlay / yaxisopts=(label='Patient Count')
        xaxisopts=(display=(tickvalues ticks)
          timeopts=(viewmax='01SEP2014'd));
      seriesplot x=DATE y=ACTCOUNT / group=ARM lineattrs=(pattern=1);
      seriesplot x=DATE y=PLANNED / group=ARM lineattrs=(pattern=2);
      endlayout; 3

      /* second panel */
      layout overlay / yaxisopts=(label='Proportion'
        linearopts=(tickvaluelist=(0 .33 1)))
        xaxisopts=(display=none);
      barchart category=MONTH response=PERCENT/ group=ARM name='c'
        grouporder=ascending;
      referenceline y=.33 / lineattrs=(color=black);
      endlayout; 4
    endlayout;

    /* legend */
    layout globallegend;
      discretelegend "a" "b" / title='Recruitment:'
        itemsize=(linelength=20);
      discretelegend "c" / title='Arm: ';
      endlayout; 5
    endgraph;
  end;
run;

```

```

/* call template with sgrender */
ods listing close;
ods html style=htmlblue;

proc sgrender data=TWOARMDAT template=twoarms;
run;

ods html;
ods listing; 6

```

- 1: A lattice layout is defined. The ROWWEIGHTS= option controls the height and number of panels. By default, stating two row weights results in a 2x1 layout.
- 2: The LEGENDITEM statement allows to custom create the legend entries for the actual and planned patient counts. The TYPE= option lets you pick the type of legend (here 'line'), while the NAME= option is used to reference the legend item later in the LAYOUT GLOBALLEGEND statement. Attributes like line pattern and color are set via the LINEATTRS= option.
- 3: For the first panel, the SERIESPLOT statement is used to draw lines for the variables *PLANNED* and *ACTCOUNT* grouped by *ARM*. Patterns are set to match the legend items defined above. You can control axis appearances with the XAXISOPTS= and YAXISOPTS= options. Note that the options TIMEOPTS= and VIEWMAX= are set to cut off the x-axis at Sep2014. This prevents the addition of the next tick mark (Nov2014) and ensures correct alignment of the bar plots with the x-axis.
- 4: Bars for each level of the variable *MONTH* are drawn in the second panel via the BARCHART statement grouped by *ARM*. The GROUPORDER= option is set to 'ascending', forcing Placebo to appear at the bottom of each bar. A reference line at y=33% is created with the REFERENCELINE statement. The LINEAROPTS= option of the YAXISOPTS= option allows you to assign custom tick values. All attributes of the x-axis are suppressed with the DISPLAY=NONE option.
- 5: The legend is defined by the LAYOUT GLOBALLEGEND statement. Custom defined legend items distinguishing planned and actual counts are displayed first ("a" and "b"). The ITEMSIZE= option allows you to control the lengths of lines displayed in the legend. To create colored boxes for each treatment arm, the BARCHART statement from the second panel is referenced ("c"). All items are arranged in a row at the bottom of the plot by default.
- 6: The actual plot is created using PROC SGRENDER by referencing the input data, the above defined template 'twoarms', and the style 'htmlblue' that was chosen for this example. For example code on how to assign custom colors to the treatment arms, see the APPENDIX.

ADVANTAGES AND DRAWBACKS

By providing information about planned vs. actual enrollment in combination with patient proportions in the two treatment arms, this plot creates an in-depth summary of the trial status at different points in time. It is easy to create and interpret, making it an excellent tool for strategic planning and communication. Grouping instead of stacking the bars in the second panel allows you to easily adjust the plot to fit three or four treatment arms.

Depending on the data, you may find that the x-axes of both panels are not automatically aligned correctly. Additional months might get added to the axis of the first panel by default but not to the second, resulting in a mismatch of bars and respective months on the displayed x-axis. This can be prevented by specifying the range of values on the x-axis in the first panel as mentioned under 3.

BASKET AND UMBRELLA TRIALS

In the last few years basket and umbrella studies have emerged as a key trial design in oncology. In basket designs, patients with the same genetic mutation are grouped into treatment arms (or 'baskets') based on cancer type, whereas in umbrella trials, patients with the type of cancer (e.g., breast cancer) are grouped by genetic mutation. As these types of trials often research rare tumor types and mutations,

sample sizes in each subgroup may be small and enrollment slow. For more details regarding these designs, see Menis et al. (2014).

The following example is based on a 10-basket design with up to 30 patients expected to enroll per basket. In this scenario, you can keep track of recruitment by creating area chart plots of enrollment over time grouped by basket.

GENERATION OF THE PLOT WITH PROC TEMPLATE AND PROC SGRENDER

For this example of a trial with 10 treatment groups, two area charts indicating total patient counts and percent of enrolled subjects grouped by basket are merged into a 2x1 panel plot, utilizing the procedures TEMPLATE and SGRENDER. The plot provides information on population development within each basket, new baskets that are defined after the start of the study (e.g., basket B4 in Figure 5 below), enrollment stops (occurring around 07/2011 in this example), and the number and sizes of baskets at times of different interim analyses. Y-axes are displayed on both sides of the graph to ease reading, while the shared x-axis is moved in between the two panels to save space. Reference lines denoting times of interim analyses (labeled "Int. An.") create an additional layer of information while a legend above the panels describes the different baskets.

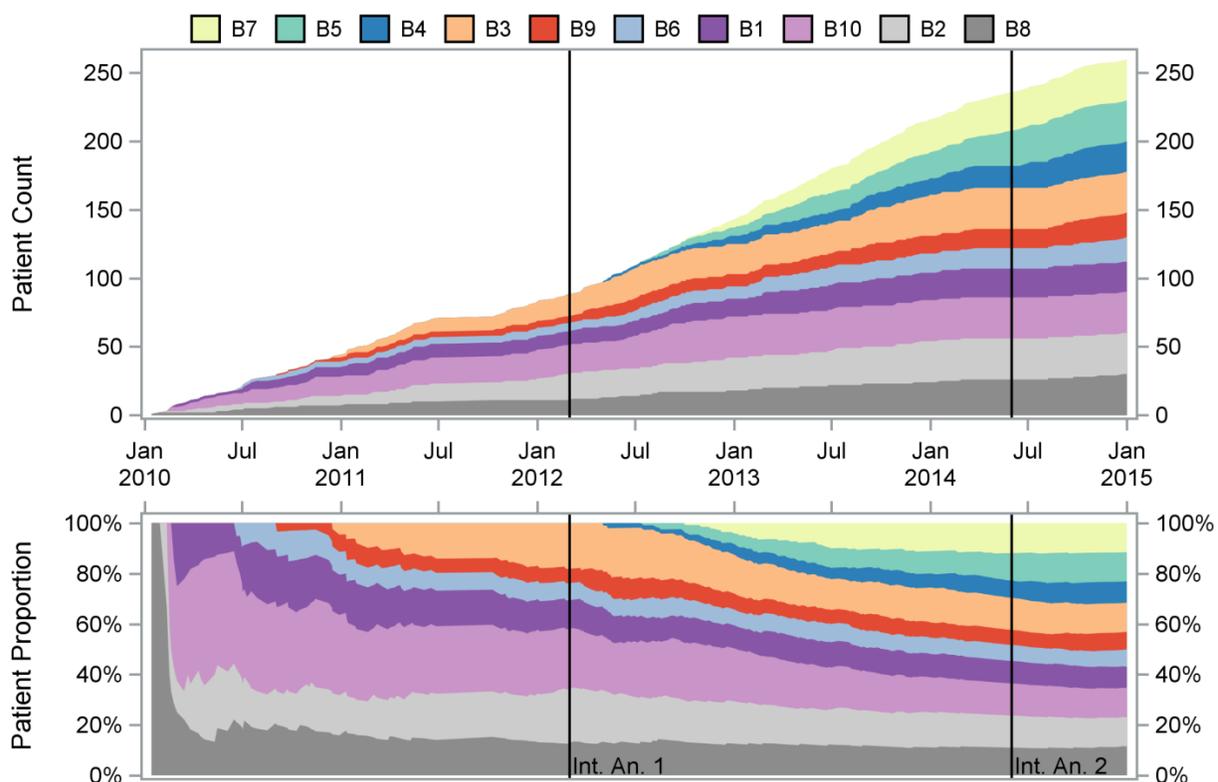


Figure 5. Patient enrollment over time by basket

THE INPUT DATA

The goal is to create cumulative band plots with one band per basket, indicating patient counts for the first panel and percentages for the second. The source data needs to be transformed so that there is one observation for each basket and enrollment date. You need to sort it by date and preferred order of baskets which is saved in a separate column (variable ORDER in the excerpt below). Depending on the data, a different basket order will produce the most informative plot. In this example, baskets are ordered

by descending date of first enrolled subject; the basket that was first populated with patients is last in order (B8 in this example) and the basket that was last populated is first (B7 in this example).

To create the overlapping band plots, you need to calculate cumulative patient counts and percent according to the chosen order of baskets. To achieve this, first count the number of patients enrolled in each basket at each enrollment date (variable *COUNT* in the data excerpt below) and the total number of patients enrolled at each date (variable *TOTAL*). Then, calculate cumulative counts and percentages, starting with the last basket in your prespecified order (variables *CUMCOUNT* and *CUMPERCENT*).

Preparing the data this way allows you to create grouped band plots by basket. The values *CUMCOUNT* and *CUMPERCENT* for the data point 26JAN2012 are identical for baskets B7, B5, B4 and B3. With each band for a specific basket overlaying the previous, when plotting the variables *CUMCOUNT* and *CUMPERCENT*, B3 will be the only basket visible, overlaying B7, B5 and B4 completely. The band for B9 is plotted next (*CUMCOUNT*=73), covering most of the chart for B3 (*CUMCOUNT*=89) so that only the part equivalent to the difference of 4 patients will be visible.

	ENRLDT	BASKET	ORDER	COUNT	TOTAL	CUMCOUNT	CUMPERCENT
881	26JAN2012	B7	1	0	89	89	100%
882	26JAN2012	B5	2	0	89	89	100%
883	26JAN2012	B4	3	0	89	89	100%
884	26JAN2012	B3	4	16	89	89	100%
885	26JAN2012	B9	5	5	89	73	82%
886	26JAN2012	B6	6	6	89	68	76%
887	26JAN2012	B1	7	10	89	62	70%
888	26JAN2012	B10	8	21	89	52	58%
889	26JAN2012	B2	9	19	89	31	35%
890	26JAN2012	B8	10	12	89	12	13%
891	31JAN2012	B7	1	0	90	90	100%
892	31JAN2012	B5	2	0	90	90	100%
893	31JAN2012	B4	3	1	90	90	100%
894	31JAN2012	B3	4	16	90	89	99%

Figure 6. Excerpt of the input data set for the area chart of enrollment by basket

CODE AND EXPLANATION

This code produces the plot of patient enrollment over time by basket (Figure 5):

```

/* create template for 2x1 panel plot */
proc template;
  define statgraph basket;
    beginingraph;
      layout lattice / rowweights=(.60 .40); 1
    /* first panel */
    layout overlay / yaxisopts=(label='Patient Count')
                    y2axisopts=(display=(tickvalues ticks))
                    xaxisopts=(display=(tickvalues ticks));
    bandplot x=ENRLDT limitlower=0 limitupper=CUMCOUNT /
             group=BASKET name='a';
    bandplot x=ENRLDT limitlower=0 limitupper= CUMCOUNT / yaxis=y2
             group=BASKET;
    referenceline x='01MAR2012'd / lineattrs=(color=black);
    referenceline x='01JUN2014'd / lineattrs=(color=black);
  endlayout; 2

```

```

/* second panel */
layout overlay / x2axisopts=(display=(ticks)) yaxisopts=(label=
    'Patient Proportion' linearopts=(thresholdmax=1));

    bandplot x=ENRLDT limitlower=0 limitupper=CUMPERCENT/ xaxis=x2
        group=BASKET;
    bandplot x=ENRLDT limitlower=0 limitupper=CUMPERCENT / xaxis=x2
        yaxis=y2 group=BASKET;
    referenceline x='01MAR2012'd / xaxis=x2 lineattrs=(color=black)
        curvelabel="Int. An. 1" curvelabelposition=min;
    referenceline x='01JUN2014'd / xaxis=x2 lineattrs=(color=black)
        curvelabel="Int. An. 2" curvelabelposition=min;
endlayout; 3

/* legend */
sidebar;
    layout overlay;
        discretelegend 'a' / across=10 border=false title='Basket'
            itemsize=(fillheight=12);
    endlayout;
endsidebar; 4
endlayout;
endgraph;
end;
run;

/* call template with sgrender */
ods listing close;
ods html style=basket;

proc sgrender data=BASKDAT template=basket;
run; 5

ods html;
ods listing;

```

1: A lattice layout is defined. The ROWWEIGHTS= option controls the height and number of panels. By default, stating two row weights results in a 2x1 layout.

2: For the first panel, a band plot is drawn via the BANDPLOT statement with *ENRLDT* on the x-axis and between 0 and the variable *CUMCOUNT* on the y-axis. By using *BASKET* as a group variable, band plots for each basket are overlaid following the order of observations in the data set. You can create a y-axis on the right side by a second BANDPLOT statement with the option YAXIS=Y2. Axis attributes are assigned in the X/Y/Y2AXISOPTS= option in the LAYOUT OVERLAY statement. Reference lines at times of interim analyses are drawn by stating the respective dates in the same format as the variable *ENRLDT* (in this case Date9).

3: Similar to the first panel, overlaying band plots are created grouped by *BASKET*, this time using *CUMPERCENT* as the upper limit. To display tick marks at the top of the plot, you can specify the option XAXIS=X2 and suppress label and tick values through the X2AXISOPTS= option. Again, the BANDPLOT statement is repeated with the additional option YAXIS=Y2 to create a y-axis on the right. Reference lines are labeled and placed through the options CURVELABEL= and CURVELABELPOSITION=.

4: To build the legend, a sidebar is defined with the SIDEBAR statement. The DISCRETELEGEND statement referencing the labeled plot in the first panel draws the legend at the top. The ACROSS= option allows you to arrange the groups in one row, while the size of the colored squares is adjusted with the ITEMSIZE= option. By default, baskets are shown in the same order as the variable *BASKET* in the data set. You can change this via the SORTORDER= option.

5: The actual plot is created using the SGRENDER procedure by referencing the input data and newly defined template. Color scheme and fonts are provided by the customized style 'basket' for this example (see APPENDIX for reference code). You can use pre-existing styles, e.g., 'htmlblue', instead.

ADVANTAGES AND DRAWBACKS

Especially for basket or umbrella design trials that cover a lot of different groups, an area chart of enrollment provides a comprehensive summary of the developing study population over time. The chart clearly visualizes periods of slow enrollment and addition of treatment groups. The combination of total patient counts and group proportions make this graph an excellent tool for communication and planning.

While area charts are great for showing changes in the total population, patient counts within each group are harder to track. Compared to the previous two examples and depending on the number of groups, data preparation is significantly more time consuming and you may find it hard to choose colorblind-friendly palettes.

CONCLUSION

This paper presents three different approaches to visualize enrollment over time. All result in a comprehensive status report and create valuable tools for strategic planning and communication between investigators, sponsors, and investors. Depending on study design and size of the patient population you may find that line, bar, or area plots of patient counts and group proportions offer the best tool to visualize enrollment. Often, you can enhance plots by including dates of adverse events, interim analyses, or planned proportions between cohorts.

REFERENCES

Le Tourneau C, Lee JJ, and Siu LL. Dose Escalation Methods in Phase I Cancer Clinical Trials. *JNCI Journal of the National Cancer Institute*. 2009; 101(10): 708–720. Accessed February 15, 2017. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2684552/>.

Menis J, Hasan B and Besse B. New clinical research strategies in thoracic oncology: clinical trial design, adaptive, basket and umbrella trials, new end-points and new evaluations of response. *European Respiratory Review*. 2014; 23: 367-378. Accessed February 28, 2017. <http://err.ersjournals.com/content/23/133/367>

CONTACT INFORMATION

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

Name: Laura Gruetzner
Enterprise: BioStat Solutions, Inc., Frederick, MD
E-mail: lgr@BioStatSolutions.com / ltgruetzner@gmail.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.

APPENDIX

This code creates the template used for the plot of patient enrollment over time by basket (Figure 5):

```
ODS PATH work.templat(update) sasuser.templat(read)
      sashelp.tmplmst(read); 1

proc template;
  define style styles.basket;
    parent=styles.htmlblue; 2

    class GraphData1 / color=#edf8b1;
    class GraphData2 / color=#7fcdbb;
    class GraphData3 / color=#2c7fb8;
    class GraphData4 / color=#fdbb84;
    class GraphData5 / color=#e34a33;
    class GraphData6 / color=#9ebcda;
    class GraphData7 / color=#8856a7;
    class GraphData8 / color=#c994c7;
    class GraphData9 / color=#cccccc;
    class GraphData10 / color=#8c8c8c; 3
  end;
run;
```

1: This statement will cause the template below to be written to the WORK library where read/write access is ensured.

2: A new style named 'basket' is defined. The SAS predefined style 'htmlblue' is chosen as a parent style, meaning the new template inherits all style attributes. This way, only attributes for style elements that you want to differ from the style htmlblue need to be specified.

3: As this template is supposed to set colors for 10 different groups (baskets in Figure 5), ten style elements are defined using the CLASS statement. The style element GRAPHDATA1 refers to the first group value (Basket 7 in Figure 5), the element GRAPHDATA2 to the second, etc. You can now assign colors via the COLOR= option.