Using the ODS GRAPHICS® to Create Patients Enrollment Graphs

Yiqian Jiang, Janssen Research & Development, Shanghai, China

ABSTRACT

Clinical research study enrollment data is widely used for investigators to monitor enrollment progress as it consists of subject enrollment dates. Meeting study enrollment targets is critical to ensuring there will be enough data and endpoints to ensure the predefined statistical power of the study. A Patients Enrollment plot is an effective way of graphically representing enrollment progress over time. Information like actual enrollment and projection plan can give a precise reading of study enrollment. This presentation will demonstrate how to use the ODS GRAPHICS to create graphs that display cumulative enrollment.

INTRODUCTION

Patient enrollment is critical to the success of a clinical trial, and information regarding initial plan, projection plan and enrollment progress can provide real-time data regarding enrollment to assist trial investigators, sponsors, and research staff in making timely decisions about recruiting policies and organizational decisions. In this paper, we will focus on the process of creating graphs that are effective in disseminating this information.

DESCRIPTION OF ENROLLMENT DATA USED FOR DATA VISUALIZATION

- Initial Plan: an initial projection with an estimated enrollment rate normally from site experience.

- Projection Plan: an adjusted projection with an estimated enrollment rate based on study enrollment status.

- Study Enrollment: actual study enrollment status.

APPLICACATION DESIGN

- Data visualization design: an overlaid graph to compare initial plan, projection plan, and study enrollment by month via a graphical visualization and tabulated numbers.

- Software selection: SAS ® is a validated and convenient analysis tool widely used in clinical trial data analysis. ODS graphics providing new procedures and features which enable SAS® programmers to create complex graphs quickly.

- SAS ® Code Design is showed in Figure 1. Unchanged and seldom changed input data sets are included into build-in coding section.
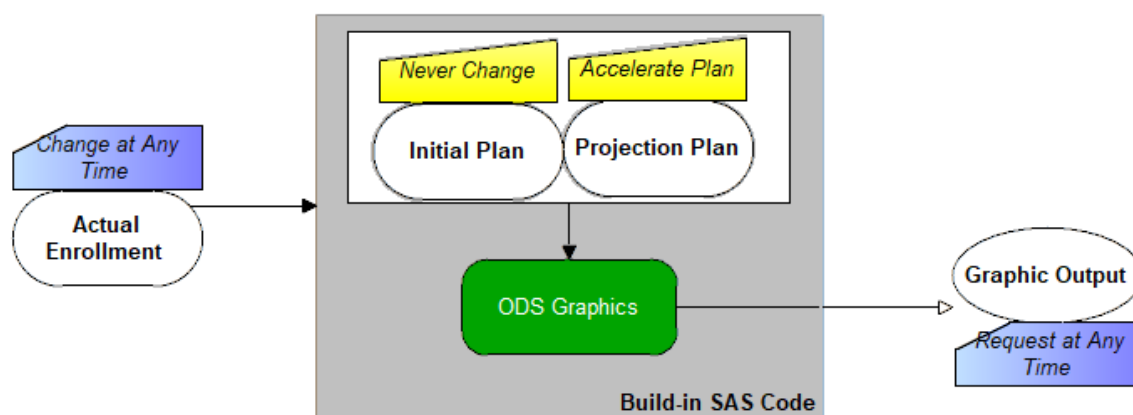

Figure 1 Code Design

DESCRIPTION OF DATA USED IN THIS PAPER

The data used in the following sections are dummy/scrambled data which have no real meaning.

- Initial plan and projection data are both estimated by month. For example, by end of 31Jul2017, initial plan is to enroll 23 patients. Projection plan is to enroll 8.18 patients. This data set is built in within the SAS code as the possibility to change is minimal.

	End Date	Initial Plan	Projection Plan
1	2017/4/30	2	.
2	2017/5/31	5	.
3	2017/6/30	9	3.69
4	2017/7/31	23	8.18
5	2017/8/31	48	17.108
6	2017/9/30	61	29.248
7	2017/10/31	78	32.908
8	2017/11/30	90	49.548
9	2017/12/31	105	60.548
10	2018/1/31	129	78.548
11	2018/2/28	132	98.468
12	2018/3/31	146	109.468
13	2018/4/30	158	120.828
14	2018/5/31	.	132.188
15	2018/6/30	.	153.548
16	2018/7/31	.	160.948

Figure 2 Initial Data and Projection Data

- Study enrollment data consist of a list of subjects along with the date each subject enrolled in the study. This data set is the input data set which might be received from a vendor extracted file or directly from a database in various formats e.g. SAS ®, EXCEL, TXT, etc.

	ID	EnrollmentDate
1	2309	04/02/2017
2	2402	04/15/2017
3	2503	04/23/2017
4	2505	05/06/2017
5	2407	05/15/2017
6	2608	05/18/2017
7	2611	06/02/2017
8	2508	06/12/2017
9	2507	07/03/2017
10	2406	07/12/2017

Figure 3 Enrollment Data (first ten patients)

PROGRAMMING STEPS IN BUILD IN SAS CODE

The programming steps that will be used to create the graph are:

1. Build in a list of all months with planned number of patient and projected number of patients.

2. Map the study enrollment data set to the timeslot of build-in by-month data and count the cumulative enrollment number during that period.

3. Graph the curves.

STEP 1: CREATE A LIST OF ALL MONTHS WITH PLANNED NUMBER OF PATIENT AND PROJECTED NUMBER OF PATIENT

Below is the code to create the list of all months. Then create a macro variable TOTAL to obtain the maximal projection number; LSTDAY to obtain the last date of current month; LSTMONTH to obtain the current month.

```
DATA plan;
  LABEL dat = "End Date" Planned="Initial Plan" Projected ="Projection Plan";
```

```
  INPUT dat $1-10 planned 13-15 projected 17-30;
  FORMAT date date9.;
  date=MDY(INPUT(SCAN(dat,2,'/'),best.),INPUT(SCAN(dat,3,'/'),best.),INPUT(SCAN(dat,1,
  '/'),best.));
  IF N(projected) THEN projected=INT(projected);
  CARDS;
2017/4/30    2
2017/5/31    5
2017/6/30    9    3.69
2017/7/31    23   8.18
2017/8/31    48   17.108
2017/9/30    61   29.248
2017/10/31   78   32.908
2017/11/30   90   49.548
2017/12/31   105  60.548
2018/1/31    129  78.548
2018/2/28    132  98.468
2018/3/31    146  109.468
2018/4/30    158  120.828
2018/5/31         132.188
2018/6/30         153.548
2018/7/31         160.948
  ;
RUN;


PROC SQL NOPRINT;
  SELECT strip(put(max(planned),best.)) INTO :total FROM plan;
QUIT;
%PUT &total;

DATA _NULL_;
  FORMAT lstday date9.;
  month=MONTH('31JUL2018'd);
  year=YEAR('31JUL2018'd);
  IF month=12 THEN lstday=MDY(1,1,year+1)-1;
  ELSE lstday=MDY(month+1,1,year)-1;
  CALL SYMPUT("lstday",put(lstday,date9. -l));
  CALL SYMPUT("lstmonth",substr(put(lstday,date9. -l),3));
RUN;
%put &lstday &lstmonth;
```

STEP 2: MERGE THE LIST OF ALL MONTHS WITH THE ORIGINAL DATA SET AND PREPARE DATA SET FOR GRAPHING

The code below will create a blank data set containing every month within a time range from an initial date to LSTDAY.

```
DATA blank;
  DO a=2017 TO 2020;
    DO b=1 TO 12;
      DO c=1;
        FORMAT endt stdt date mmddyy10.;
        endt=MDY(b,c,a)-1;
            date=endt;
        stdt=MDY(month(endt),1,year(endt));
        IF  '30APR2017'd<=endt<= "&lstday"d then OUTPUT;
      END;
    END;
  END;
  DROP a b c;
RUN;
```

The code below will merge the list of all months we created in the previous step with the original enrollment data to create a data set which contains every month with the number of subjects who enrolled in and before that month. Then include number of projected and planned subjects on each month to the output dataset.

```
PROC SQL;
  CREATE TABLE Enroll AS
  SELECT a.*,planned,projected
  FROM (SELECT DISTINCT date,stdt,endt,count(id) AS actual,
                        'Projection' AS group1,'Actual' AS group2
        FROM (SELECT a.*,b.*
              FROM Enrollment AS A RIGHT JOIN blank AS b
              ON EnrollmentDate <= endt)
        GROUP BY 1,2) AS A
  LEFT JOIN plan AS b ON a.endt=b.date;
QUIT;
```

The code below finally prepare data for graphing.

```
DATA final;
  SET enroll;
  date2=TRANWRD(PROPCASE(SUBSTR(PUT(date,date9.),3)),'2',' 2');
  FORMAT lstday date9.;
  lstday="&lstday"d;
  IF date gt lstday THEN actual=.;
run;

PROC SQL NOPRINT;
  SELECT cats('"',date2,'"') INTO :dates separated ' '
  FROM final;
QUIT;
```

STEP 3: GRAPH THE CURVES

The data is now ready for graphing! The program uses the ODS GRAPHICS procedure to graph the data as shown in the code below. The first part builds up the template for the graph.

BARCHART indicates the actual number of subject, two SERIESPLOT statementsS with different colors represent projected number of initial number respectively. LEGEND for projected line is typed with the current month. Projected curve is labeled with total number of projection patients. REFERENCELINE is added to mark an acceleration plan start month.

Two SCATTERPLOT statements are used to add actual and projection numbers like a tabulation table or an annotate at the bottom of the graph.

PROC SGRENDER in the end is to do the actual plot by linking the template ENROLL with FINAL data set.

```
PROC TEMPLATE;
DEFINE STATGRAPH enroll;
BEGINGRAPH;
ENTRYTITLE "Study01- XXXXX XXXXX Study";
LAYOUT lattice / pad = (top=0cm bottom=0cm) columns = 1 rows = 2 rowweights = (0.92
0.08) rowgutter=0 columndatarange = unionall rowdatarange=union;
LAYOUT
OVERLAY / cycleattrs=true border = false pad=(bottom=0.2cm)
         xaxisopts=(label="Date" type=discrete display=( ticks tickvalues)
         discreteopts=( tickvaluelist=(&dates) tickvaluefitpolicy=rotatealways
         tickvaluerotation=vertical))
         yaxisopts=( linearopts = (tickvaluelist = (0 50 100 150 200 250)
         viewmin=0 viewmax=250) label="Number of Patients" offsetmax=0.05 );
BARCHART x=date2 y=actual / name="A" legendlabel="Actual"   barwidth=0.4
         fillattrs=(color=red);
SERIESPLOT x=date2 y=projected / name="B" legendlabel="Projected (&lstmonth)"
```

```
            lineattrs= (pattern=1 color=blue) curvelabel="Total N=&total";
SERIESPLOT x=date2 y=planned / name="C" legendlabel="Inital Plan" lineattrs=(pattern=1
            color=green);
REFERENCELINE x='Jun 2017' /lineattrs=(pattern=2 color=grey) clip=yes
            curvelabel="Start of Acceleration Plan";
REFERENCELINE y=50 /lineattrs= (pattern=34 color=grey);
REFERENCELINE y=100 /lineattrs= (pattern=34 color=grey);
REFERENCELINE y=150 /lineattrs=(pattern=34 color=grey);
REFERENCELINE y=200 /lineattrs=(pattern=34 color=grey);
REFERENCELINE y=250 /lineattrs=(pattern=34 color=grey);
DISCRETELEGEND "A" "B" "C" / location=inside halign=right valign=top across=1;
ENDLAYOUT;
LAYOUT
OVERLAY / border=false walldisplay=none
            xaxisopts=(label="Date" display=none offsetmin=0.05 offsetmax=0.05
)
            yaxisopts=(TYPE=DISCRETE DISPLAY=(TICKVALUES) REVERSE= TRUE)
            x2axisopts  = (DISPLAY = NONE)
            y2axisopts  = (DISPLAY = NONE);
SCATTERPLOT X=date2
Y=group2/MARKERCHARACTERATTRS=(COLOR=GraphValueText:Color)
            MARKERCHARACTER=actual MARKERCHARACTERPOSITION=center jitter=auto;
SCATTERPLOT X=date2
Y=group1/MARKERCHARACTERATTRS=(COLOR=GraphValueText:Color)
            MARKERCHARACTER=projected MARKERCHARACTERPOSITION=center
jitter=auto;
ENDLAYOUT;
ENDLAYOUT;
ENDGRAPH;
END;

*** Do actual plots;
OPTIONS NODATE NONUMBER NOBYLINE;
ODS RESULTS OFF;
FOOTNOTE;
ODS LISTING image_dpi=300 gpath="C:\D\enroll\study01" ;
ODS GRAPHICS ON /
    imageName = "Enroll_study01_&sysdate."
    imagefmt  = png
    border    = off
    scale     = no
    reset     = index
    width     = 9.7 in
    height    = 6.7 in;
ODS ESCAPECHAR="~";
RUN;

PROC SGRENDER DATA= final TEMPLATE=enroll;
  FORMAT actual projected best3.;
RUN;

ODS LISTING;
ODS GRAPHICS OFF;
OPTIONS MISSING='.';
```

THE OUTPUT

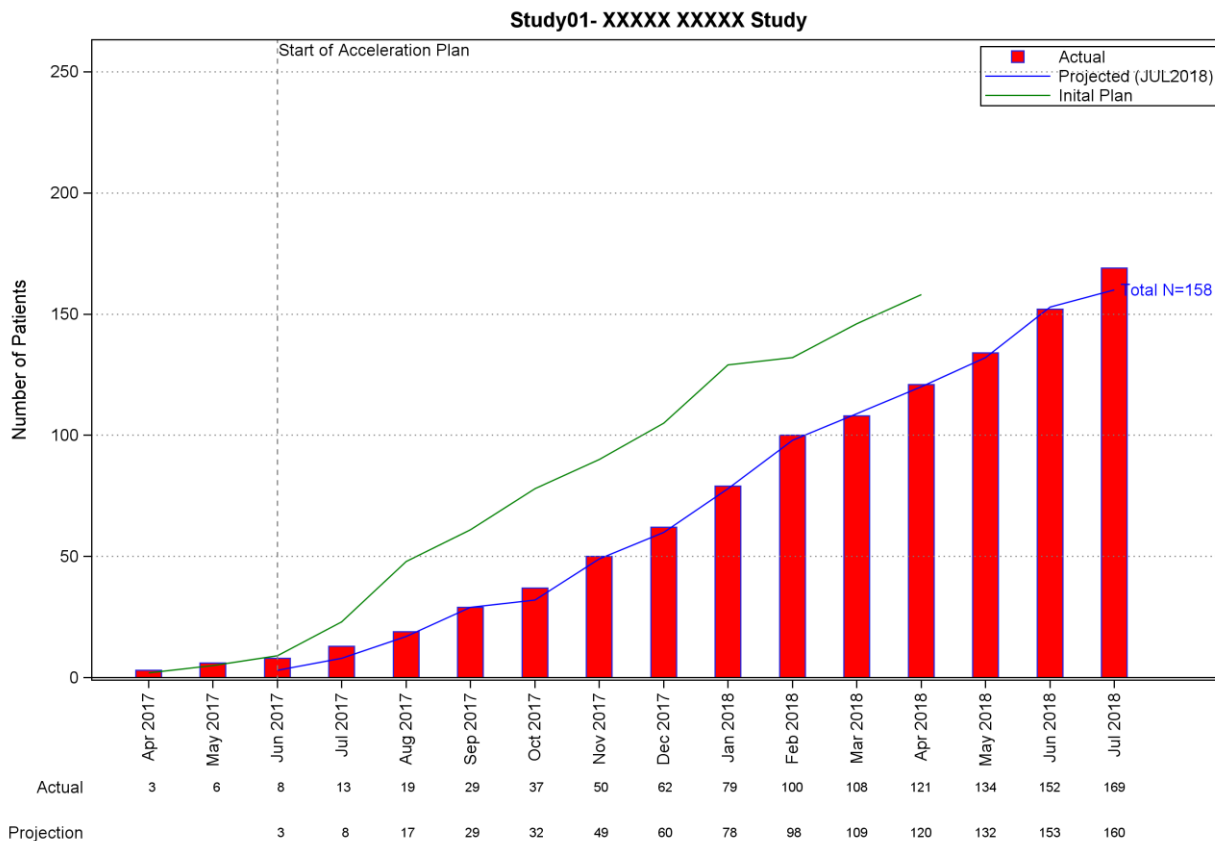The output from the codes is a PNG file containing a graph like below:

Figure 4 Sample Output

CONCLUSION

This paper provides code and step-by-step instructions on how to create a graph showing the enrollment data. We provide all the information together within one graph for clinical team to do a regular review.

We note that some of the codes could be more automatically by packaging all codes into a macro and set up some default macro parameters. Considering that an enrollment graph will have a rather long life-cycle as long as study enrollment duration, we decide to provide the code as a reference code and encourage the users to review coding carefully to make sure that they are familiar with the setting and customize everything they need to do from a programmer's perspective.

REFERENCES

Kalt, Mike and Zender, Cynthia. "Introduction to ODS Graphics for the Non-Statistician ". SAS Global Forum. 2011. Available at https://support.sas.com/resources/papers/proceedings11/294-2011.pdf

ACKNOWLEDGMENTS

The author would like to thank all of her colleagues who reviewed this paper and provided comments.

RECOMMENDED READING

- Base SAS® Procedures Guide
- SAS® SQL Procedures Guide
- SAS® ODS Graphics Guide

CONTACT INFORMATION

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

Name: Yiqian Jiang
Enterprise: Janssen Research & Development
Address: 6F, Building A, Xinyan Mansion, No. 65 Guiqing Rd
City, State ZIP: Shanghai, 200233, China
E-mail: ajiang4@its.jnj.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.