A Plot and a Table per Page Times Hundreds in a Single PDF File

Daniel J. Leprince and Elizabeth Li

DIEM Computing Services, Inc., Fremont, CA

Abstract

In the Windows environment, it is relatively easy to integrate a figure and a table in a single spreadsheet or a page in a word-processor. When hundreds of these pages need to be generated and stored in a format that cannot be altered by the end-user then the SAS System and the Adobe Acrobat Distiller are more appropriate tools. This paper will present the SAS code of a specific application to produce a postscript file containing multiple pages of a plot and a table per page. The Adobe Acrobat Distiller is used to transform the postscript file into a Portable Document File (PDF).

In this paper, macros and several SAS/Graph features are presented (note, annotate dataset, and gplot, gslide, greplay and gprint procedures). Emphasis is put on the use of hardware fonts which can dramatically reduce the overall size of a document.

Key words: Annotate, note, gplot, gslide, greplay, and gprint.

Introduction

In clinical trials, it is not unusual to have multiple figures per subject per visit. When the additional information is too complex or heterogeneous to be incorporated in a figure, it is assembled in an accompanying table. In a Window environment, a spreadsheet is sufficient to produce the required information for a few subjects. The task becomes overwhelming when tens or hundreds of figures and tables are required. Despite the use of spreadsheet macros, there is still a lot of data manipulation for each subject and visit combinations which increased the chance of errors.

One solution, is to use the SAS system to generate plots and tables. Then use the greplay procedure to concatenate each figure and its accompanying table on the same page, for each subject and visit combination. The output file is a high resolution postscript (.ps) file that can be easily "distilled" into a Portable Document Format file (.pdf) with the Adobe Acrobat Distiller.

The final result is a single document that can be read but not modified by the Adobe Acrobat Reader.

Methodology

We will assume that the data has been successfully validated and that the critical variables have been incorporated into the analysis dataset.

DO LOOP

First, create a "do loop" dataset to identify each unique subject by visit (period) combinations and the total number of these combinations.

```
proc sort data=pk
        out=doloop(keep=subjid period f_dose)
        nodupkey;
    by subjid period;
    run;
```

*Then, create a concatenated macro of period * subjid and final dose to be resolved inside the do loop section. The subject and period are fixed length variables. Since the final dose can be missing, it is concatenated at the end of the string "comb". The macro variable ncombo contains the total number of combinations and hence the total number of pages of the pdf file.*

```
data _null_;
    set doloop end=last;
    comb=trim(left(period))||subjid||trim(left(f_dose));
    call symput('combo'||left(_n_),trim(comb));
    if last then call symput('ncombo',left(_n_));
    run;
```

*Start the macro doit that will allow a do loop through each of the subjects * visit combinations.*

```
%macro doit;

%do i=1 %to &ncombo;
```

*If it is the first subject * visit combination, start a new file, if not, append all the following figures to the file previously created. This macro variable value is defined under the goptions gsfmode.*

```
    %if &i=1 %then  %do;  %let mode=replace;  %end;
    %else              %do;  %let mode=append;  %end;
```

SUBSETTING MACROS
Extract the value of subject, period and final dose from the macro variable combo for a given iteration of the do loop. Macro variables period and subjects are used to subset the SAS datasets.

```
%let period=%substr(&&combo&i,1,1);
%let subjid=%substr(&&combo&i,2,7);
%let f_dose=%substr(&&combo&i,9);
```

Redefine a derived period (visit) for a given value of the macro period.

```
%if &period=1 %then %do; %let tperiod=DAY 1;
%end;
%else %if &period=2 %then %do; %let tperiod=DAY
14;     %end;
…..
```

Display current macro variable values for debugging purposes.

```
%put ### combo&i=&&combo&i period=&period;
%put ### subjid=&subjid ncombo&i=&ncombo
          f_dose=&f_dose;
```

DESCRIPTIVE STATISTICS
Calculate descriptive statistics (min max median) that will be incorporated as notes in the plot.

```
proc means data=pk noprint;
    where subjid="&subjid" & period=&period;
    var valuen;
    output out=stats min=min max=max
           median=median;
     run;

data _null_;
    set stats;
    call symput('Cmin',trim(left(put(min,8.))));
    call symput('Cmax',trim(left(put(max,8.))));
    call symput('Median',trim(left(put(median,8.))));
    run;
```

Display the newly calculated macro value in the log.

```
%put ########## Cmin=&Cmin;
%put ########## Cmax=&Cmax;
%put ########## Median=&Median;
```

RANGE OF RESPONSE AXIS
When the variability of the response variable can be estimated, it is time to define a restricted set of yaxis ranges to ease the comparison between plots. The

ranges defined in the data _null_ were matching 80%, 17%, 3%, and the outliers of the population.

```
data _null_;
    if &cmax < 1800              then max=1800;
    else if 1800 <= &Cmax <= 3500 then max=3500;
    else if 3500 < &Cmax  <= 6000  then max=6000;
    else                              max=&Cmax;
    call symput('max1',compress(put(max,best8.)));
    run;
```

Display current macro values.

```
%put ### combo&i=&&combo&i period&i=&period;
%put ### subjid=&subjid ncombo&i=&ncombo;
%put ### max1=&max1;
```

ANNOTATE
Create an annoplot dataset to show actual values of the yaxis for each data point. Please note that the hardware font is first enclosed into two double quotes and then one single quote for device driver PS1200. Otherwise, the SAS system will complain that the font can't be found and will substitute a software font.

```
data anno;
    length text $5 style $12;
    retain function 'label' style '"Helvetica"' xsys ysys
        '2' hsys '1' size 3.9;
    set pk(keep=subjid period act_tim valuen
        smplhour);
    where subjid="&subjid" & period=&period;
```

Since the first two data points are very close to one another and that their values can overlap on the figure, one is set above the data point (position='2') and the following one below the data point (position='8').

```
    if smplhour~=0.5 then position='2';
      else                 position='8';
    x=act_tim;
    y=valuen;
    text=trim(left(put(valuen,5.)));
    output;
    run;
```

YAXIS
Defining the yaxis label. Rotate the label by 90 degrees. Offset the start of the plot by 3 percent to the right of the yaxis to prevent the line from starting on the axis.

```
axis1 label=(a=90 j=l font='Helvetica' h=2.5pct
"TESTOSTERONE, SERUM (TOTAL) NG/DL")
```

```
    width=12
    order=(%gaxis(ymin=0,ymax=&max1))
    major=(width=12)
    minor=(number=1 width=12)
    value=(height=2.5pct)
    length=60pct
    offset=(3pct)
    ;
```

XAXIS
Defining xaxis label. Purposefully extend the xaxis by an extra 4 hours to leave an unobstructed space to insert the gplot "notes" on descriptive statistics.

```
axis2 label=(justify=center f='Helvetica' h=2.5pct
"TIME ELAPSED IN HOURS")
    width=12
     order=(0 to 32 by 4)
     major=(width=12 height=1.5)
     minor=(number=1 width=12)
     value=(height=2.5pct)
     offset=(3pct)
     length=80pct ;
```

Use a dot symbol and connect the points with a dotted line.

```
symbol1 color=black interpol=join value=dot
        height=2 pct line=2 width=16;
```

TITLE & FOOTNOTE PER PAGE
Create a gslide for titles and footnotes associated with each unique page.   Use device to create a high resolution postscript file and use hardware font.

```
goptions rotate=portrait ftext='Helvetica' htext=2.5
nodisplay device=ps1200 gunit=pct;

proc gslide;
    title1   j=c h=1.60  f='Helvetica' 'Serum Total
Testosterone Levels in Males';
    title2   j=c h=1.60  f='Helvetica' "Study &protocol.
Subject &i, Period=&tperiod, Final Dose: &f_dose
unit";
```

Define program, location, running date and page X of Y values.

```
 footnote j=l h=1.0   f='Courier' "%str(&pgmid..sas
Date: &sysdate)"
        j=r "Page &i of &ncombo";
     run;
     quit;
```

GPLOT AND NOTE
Plotting the 24 hours PK figure. In order to prevent the vertically compressed appearance, there is a need to define the vertical spread of the figure to 5 inches.

```
goptions vsize=5.0 in;
```

To avoid repeating the gslide titles in the gplot, we need to set them to missing.

```
title; footnote;
```

Subset the PK data to the proper subject * period combination, define the yaxis and xaxis, set the vertical reference line at zero. Insert notes in the upper right part of the figure for descriptive statistics.

```
proc gplot data=pk;
     where subjid="&subjid" & period=&period;
      plot valuen*act_tim / vaxis=axis1  haxis=axis2
                            vref=(0)      lvref=3
                            anno=anno   frame;
 note move=(80.5,81) h=2.5pct  "Cmin =&cmin";
 note move=(80.5,77) h=2.5pct  "Cmax=&cmax";
 note move=(80.5,73) h=2.5pct  "Cmedian=&median";
 format valuen 5.;
 run;  quit;
```

GPRINT
Create the accompanying table for the figure as a temporary file that will be inserted as a figure by the gprint procedure..

```
filename tabletmp "&rtfdir.temphorm.lst";
proc printto new file=tabletmp;
     run;

title1 "OTHER HORMONES MONITORED
BESIDES SERUM TOTAL TESTOSTERONE (T)";
 title2 "Subject: &i, Period: &tperiod";
   footnote1  "%justify(%str(   (1)  STime=>
Scheduled time))";
   footnote2  "%justify(%str(   (2)  DTime=> Postdose
Draw Time))";
   footnote3  "%justify(%str(   (3)  BT   =>
Bioavailable T (ng/dL)))";
    …..
```

Produce the table with proc report by subsetting to the right subject and period combination.

```
  proc report data=sasdata.hormone nowd spacing=1
       split='*' missing headline headskip ls=70 ps=26;
       where subjid="&subjid" & period=&period;
       column smplhour act_tim  bt dht dhtt est estt fsh
```

```
              lh shb;
define smplhour / display "STime*(HH)(1)"    right
       width=8;
define act_tim  / display "DTime (2)*(HH:%%)"
       center width=9  format=8.2;
define bt    / display "BT*(3)"     right  width=4;
……
run;

proc printto;   run;
```

Define 60 columns and 30 rows for the output to be incorporated into gprint

```
goptions hpos=60 vpos=30 ftext='Courier' htext=2.8;
```

Prevent a repeat of titles & footnotes in the table.

```
title; footnote;
```

Incorporate the table into a figure.

```
proc gprint fileref=tabletmp
            ctext=black;
      run;  quit;
```

GREPLAY: A PAGE AT A TIME
Assemble the titles and footnotes (gslide) with the plot (gplot) and the table (gprint) into a postscript file. Use a high resolution postscript device (ps1200).

```
filename gsasfile "&rtfdir.&pgmid..ps";
goptions reset=all htext=3 rotate=portrait
        device=ps1200 ftext='Helvetica'
        colors=(black) cback=white gunit=pct
        gsfmode=&mode gsfname=gsasfile
        vsize=9.25 in hsize=7.0 in horigin=.95 in
        vorigin=.78 in display;
```

Replay the title slide, the PK plot, and the table.   The greplay code define the size and location of each unit and their order of appearance.

```
proc greplay igout=gseg nofs tc=cat1x2;
      tdef lab1x2
      1/ ulx=0 uly=100urx=100 ury=100
         llx=0 lly=  0   lrx=100 lry=  0
      2/ ulx=0 uly=92   urx=100 ury=92
         llx=0 lly=46    lrx=100 lry=46
      3/ ulx=5 uly=44   urx=100 ury=44
         llx=5 lly= 0    lrx=100 lry= 0;
      template lab1x2;
      treplay 1:gslide 2:gplot 3:gprint;
      run;
      quit;
```

CLEANING UP
If this program is run in interactive mode and hundreds of plots are created, it is recommended to clean up the catalog containing the figures after each iteration.

```
proc greplay igout=gseg nofs;
      delete _all_;
       run; quit;
```

End the "do loop".

```
%end;

%mend doit;
%doit;
```

Once the postscript file is generated, it is time to "distill" it into a pdf format with Adobe Acrobat Distiller. Set the Settings / Job Options / General / Default Page Size / width to 612.0 and the height to 792 points.   You then can drop the file in the distiller or double click on your postscript file to have it automatically translated by the distiller.   PDF files are about one third the size of a postscript file.

General Considerations
The code presented here illustrates the minimal elements required to succeed in creating multiple pages of a plot and a table per page.   It is assumed that extensive validation of the data to be plotted has been performed. Graphic displays are data driven, factual and unforgiving for outliers, missing values, and data entry errors.

SAS also allows the capture of numerous subject characteristics or descriptive statistics by SAS macro variables to be incorporated in the gplot (with the notes), titles and footnotes.

Although, the use of several "%put" statements may appear excessive, it will prove invaluable for debugging the program or isolating problem data. Especially when considering that we are turning off the following options (mprint mlogic symbolgen) during production. Even then, the log still exceeded 3MB in size in this case.

If hardware fonts are not defined, SAS will default to a software font requiring more storage space. For a few figures, the size of the output file may not matter much. The current program was generating more than

700 pages of output. The size of the postscript file was reduced by more than than 18 times (92.9MB to 4.5MB) when hardware fonts were used instead of software fonts. This dramatic reduction in file size was also accompanied with a similar reduction in computer time to run the program. The use of hardware fonts also provides an extra benefit, it allows the use of the find function in the PDF document unlike graphics containing SAS software fonts.

To obtain a list of available hardware fonts for a given SAS provided device, execute the following code:

```
Proc gdevice c=sashelp.devices nofs;
    list DeviceDriverName; * ex. list ps1200;
     run; quit;
```

Additional macros used by the program.

The macro justify pads a string with trailing blanks to the length of a linesize option to create a left justified title or footnote. Use the STR macro to preserve blanks in a string

```
%macro justify(string);
   %let nblanks = %eval(%sysfunc(getoption(linesize))
- %length(&string));
   &string%sysfunc(repeat(%str( ),&nblanks - 1))
%mend justify;
```

The macro gaxis is used to determine the most economical format for the start, stop and by values of the axis order.

```
%gaxis(YMIN=Minimum_value_of _variable,
        YMAX=Maximum_value_of_variable,
     NVALUES=Maximum_number_of_axis_values);

%macro gaxis(ymin=,ymax=,nvalues=11);
   %local i j by min max start stop;

   %do i = -3 %to 9;
     %do j = 1 %to 3;
       %let by  = %sysevalf(%scan(1 2
            5,&j)*10**&i);
       %let min=%sysevalf(&ymin/&by,floor);
       %let max=%sysevalf(&ymax/&by,ceil);
       %if %eval(&max - &min + 1) <=
          &nvalues %then %do;
         %let start=%sysevalf(&by*&min);
         %let stop=%sysevalf(&by*&max);

              &start to &stop by &by
```

```
         %let i = 9;
         %let j = 3;
       %end;
     %end;
   %end;
%mend gaxis;
```

Reference

SAS Institute Inc., SAS/GRAPH Software®. Software: Reference, Version 8, Cary, NC: SAS Institute Inc., 1999. 1240pp.

Acknowledgments

To George Black, a senior SAS programmer consultant, who graciously introduced me to his justify and gaxis macros.

Information About Authors

Daniel J. Leprince and Elizabeth Li are employees at DIEM Computing Services, Inc., a SAS Quality Partner™, providing statistical consulting, SAS program development, and data management services for biotechnology and pharmaceutical companies.

Daniel has performed data analysis in biotechnology and pharmaceutical companies for more than six years. He has participated in two submissions to the Food and Drug Administration (FDA) and has been presenting papers at regional meetings.

Elizabeth is a SAS certified professional™. She has been using SAS since 1987 and has presented several papers in user groups at the regional level.

Contact information:
38851 Litchfield Circle, Fremont, CA 94536
Phone: (510) 744-1304, Fax: (510) 744-1412
Email: dleprince@diemcomputing.com
Email: elizabethli@diemcomputing.com
            www.diemcomputing.com

Trade Mark

SAS, SAS Certified Professional, and SAS Quality Partner are registered trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration.

Serum Total Testosterone Levels in Males
Study A 99-99-99 Subject 1, Period=DAY 32, Final Dose: 80 unit

Cmin =248
Cmax =1705
Cmedian=441
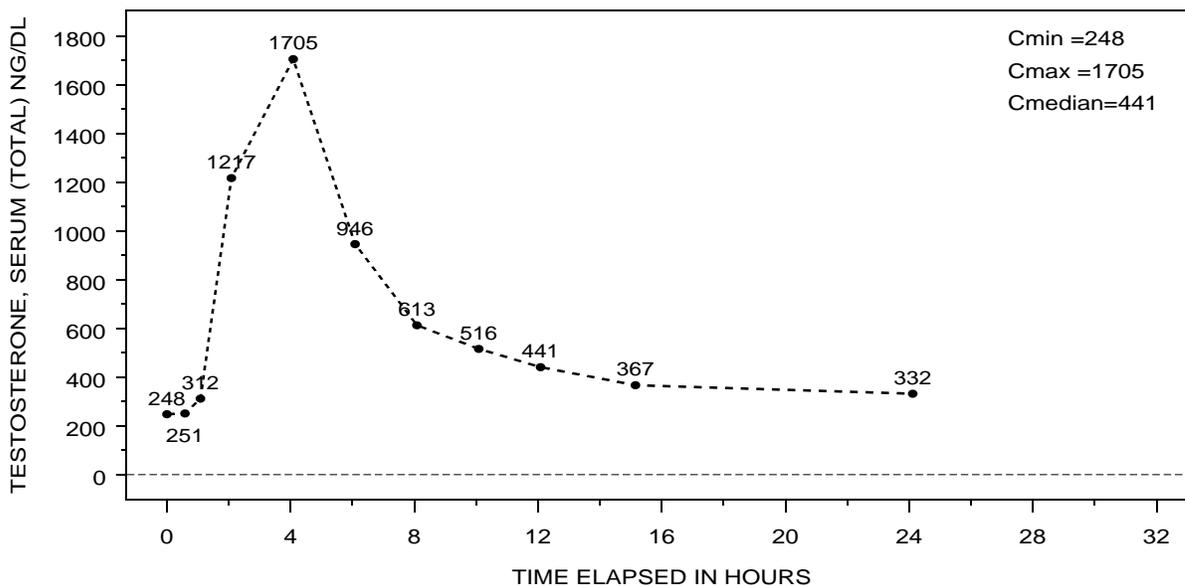
```
OTHER HORMONES MONITORED BESIDES SERUM TOTAL TESTOSTERONE (T)
                 Subject: 1, Period: DAY 32

   STime DTime (2)    BT   DHT    DHTT      E2    E2T    FSH     LH   SHBG
  (HH)(1)  (HH:%%)    (3)   (4)    (5)     (6)    (7)    (8)    (9)   (10)
 -------------------------------------------------------------------------
       0    0.00     150    51  0.2056    2.3  0.0093  0.44   0.16     20
       2    2.08     950   120  0.0986    3.1  0.0025                  22
       4    4.08    1376   244  0.1431    6.9  0.0040                  21
       6    6.08     723   144  0.1522    4.8  0.0051                  20
      24   24.12     217    98  0.2952    2.8  0.0084                  21
```

```
(1)  STime=> Scheduled time
(2)  DTime=> Postdose Draw Time
(3)  BT   => Bioavailable T (ng/dL)
(4)  DHT  => Dihydrotestosterone (DHT), Serum (ng/dL)
(5)  DHTT => DHT/T
(6)  E2   => Estradiol, Serum (ng/dL)
(7)  E2T  => E2/T
(8)  FSH  => Follicle Stimulating Hormone (FSH-ICMA) (miu/mL)
(9)  LH   => Luteinizing Hormone (LH-ICMA) (miu/mL)
(10) SHBG => Sex Hormone Binding Globulin (nmol/L)
```