SAS Programs Validation In A GCP Regulated Environment

Maria-Theresa S Pangan
Ischemia Research and Education Foundation (IREF), San Francisco, CA

ABSTRACT

"*If it is not documented, it didn't happen*" is a validation motto followed in Good Clinical Practices (GCP) regulated environments. GCP requires that in clinical research, any software application developed with the intention of manipulating, transforming, analyzing, and/ or reporting data either in table or graphical form should be validated and adequately documented. These validations and their documentation are subject to audit and inspection under GCP principles. This paper illustrates how the validation and documentation processes are performed on SAS programs used in analyzing data collected in clinical trials.

INTRODUCTION

SAS software has become a leading data analysis tool used by Pharmaceutical and Clinical Research Organizations. As thousands of lines of SAS code is written to give meaning to collected data, it has become imperative to investigate how this code manipulates data in order to answer specific research questions. Documentation produced by the validation activity provides evidence that the program does what it is supposed to do. Validation is an added benefit to programmers and statisticians because it allows them to gain better understanding of the data and structure their code more efficiently. IREF began to practice SAS Programs Validation not only because FDA requires it but because it provides some degree of assurance that results of statistical analysis are accurate and reliable whether those results are intended for journal publication or for submission to FDA. The whole concept of SAS validation revolves around the 3 vital questions:

1. What is the program's intended purpose?
2. What does the program actually do?
3. Are 1 and 2 the same?

In order to provide evidence that the above questions are satisfied, the SAS code being validated is analyzed and documented as follows:

- Graphical representation of data flow
- Identify sources of requirements
- State requirements being satisfied
- Identify input test data set used
- Identify output produced
- Identify test program used
- State expected results
- State actual results
- Conclusion
- Attachments (e.g. SAS log)

EXAMPLE

A SAS program that groups heart rate data will be used as an example. The original SAS program is shown below and the data flow is shown in Figure 1.

SAS Code and Data Flow

```
/* SAS Program for WUSS98 paper --
validation program example */
/* Program Name : example.sas */
/* Purpose: Process heart rate to
create data set for low and high heart
rate */
/* January 25 1999 */
/* Ischemia Research, San Francisco */

data example lowgrp highgrp;
  input pat_id $ 1-4 gender $ 6 hrt;
  * categorize heart rate into low and
high;
```

```
  if hrt ne . then do;
    if hrt =<90 then output lowgrp;
     else output highgrp;
  end;
  * Output observations into 2 datasets
based on group;


  cards;
0001 F  90
0002 M 100
0003 M  55
0004 F  95
0005 F 101
0006 M  89
0007 F  95
0008 M  89
0009 F  72
0010 M  92
0011 M .
run;

title 'Observations in the low group
dataset';
proc print data=lowgrp; run;

title 'Observations in the high group
dataset';
proc print data=highgrp; run;
```

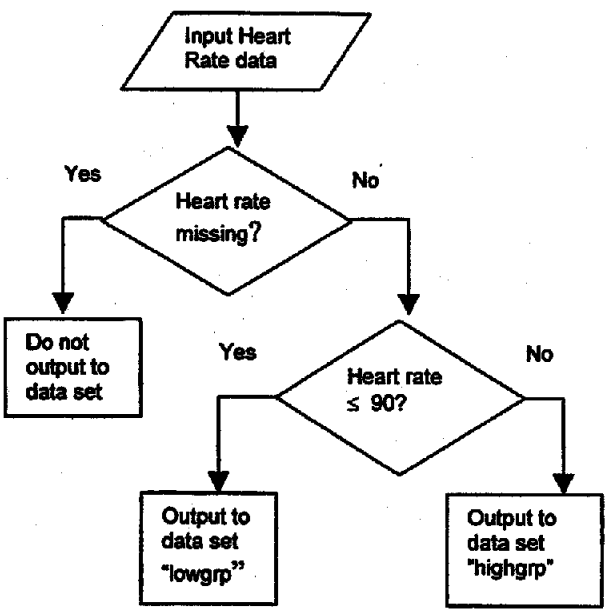

Figure 1. Data flow of example.sas

Sources of Requirements

The section in the study protocol or data analysis procedures where the requirement was derived should be referenced. Any communication between the sponsor and the organization analyzing the data can be referenced here as source of requirement.

Example:

> **Page 2 of Protocol USA-1001 dated 01Jan1999, entitled *Assessing the Efficacy of Drug X in Lowering High Blood Pressure* contains the analyis plan for measuring the effect of Drug X in patients with high blood pressure. The plan calls for analyzing the characteristics of subjects with low heart rate and high heart rate.**

Statement of Requirements

Group observations based on the value of heart rate (hrt). If heart rate is less than or equal to 90 then classify heart rate into the "low" group. If heart rate is greater than 90 then classify heart rate into the "high" group. Output the two groups into separate data sets.

Test Data Set

The actual data will be used to test the requirements.

Outputs Produced

Outputs produced will be data sets named "lowgrp" and "highgrp". A Proc Print of the two data sets will also be produced.

SAS Test Program

The SAS test program used will be test.sas. This program will examine the original data before the 2 data sets lowgrp and highgrp are created. The code is shown below:

```
/* SAS Program for WUSS98 paper --
validation program example */
/* Program Name : test.sas */
/* Purpose: This program will test
results of example.sas */
/* January 25 1999 */
/* Ischemia Research, San Francisco */
```

```
data example ;
  input pat_id $ 1-4 gender $ 6 hrt;
  length grp $4;
  * categorize heart rate into low and
high;
  if hrt ne . then do;
    if hrt =<90 then grp='low';
     else grp='high';
  end;
  cards;
0001 F  90
0002 M 100
0003 M  55
0004 F  95
0005 F 101
0006 M  89
0007 F  95
0008 M  89
0009 F  72
0010 M  92
0011 M  .
run;


title 'Observations in the Low Group';
proc print;
  where grp='low';
run;

title 'Observations in the High Group';
proc print;
  where grp='high';
run;
```

Expected Results

Heart rate data should be classified into "low" group and "high" group. Heart rate less than equal to 90 should be classified "low" and heart rates greater than 90 should be classified "high".

Actual Results

The Proc Print of the 2 data sets created by executing example.sas is shown below:

Observations in the low group dataset

OBS	PAT_ID	GENDER	HRT
1	0001	F	90
2	0003	M	55
3	0006	M	89
4	0008	M	89
5	0009	F	72

Observations in the high group dataset

OBS	PAT_ID	GENDER	HRT
1	0002	M	100
2	0004	F	95
3	0005	F	101
4	0007	F	95
5	0010	M	92

The output of Proc Print in the test program shows the classification of heart rate data into "low" group and "high" group.

Observations in the Low Group

OBS	PAT_ID	GENDER	HRT	GRP
1	0001	F	90	low
3	0003	M	55	low
6	0006	M	89	low
8	0008	M	89	low
9	0009	F	72	low

Observations in the High Group

OBS	PAT_ID	GENDER	HRT	GRP
2	0002	M	100	high
4	0004	F	95	high
5	0005	F	101	high
7	0007	F	95	high
10	0010	M	92	high

Conclusion:

Comparison of the actual results obtained by executing the SAS program example.sas and the test program test.sas shows that the logic used in example.sas is processing the to meet the requirement correctly.

SUMMARY

The above process illustrates how a SAS program is documented for FDA reviewers or auditors. A well documented SAS program helps speed up the validation process because it helps the validator identify the critical code that needs to be examined more closely. Consequently, a well documented SAS program validation aids FDA reviewers in understanding the data thus helping to speed up the drug approval process.

REFERENCES

Code of Federal Regulations, Food and Drugs, Title 21, Part 11, "Electronic Recordds, Electronic Signatures: Final Rule" (U.S. Government Printing Office, Washington, DC).

Food and Drug Administration (FDA) Techinical Report, Software Development Activities, Draft, Office of Regulatory Affairs, October 1996.

Teri Stokes, "Installing GCP Systems at Investigator Sites," Applied Clinical Trials, January 1991, 46-50.

ACKNOWLEDGEMENTS

The author would like to thank Nick Paszty, IREF's Manager of Application Development, for his valuable suggestions and editing assistance with this paper.

AUTHOR CONTACT:

Maria-Theresa Pangan
Computer Validation Specialist
Ischemia Research and Education Foundation (IREF)
250 Executive Park Blvd., Ste 3400
San Francisco, CA 94134
Phone: (415) 715 2337
Email: tsp@iref.org