

## SCSUG-2018

### Standardisation of clinical laboratory data using SAS

Laxmi Samhitha Bontha.

#### Abstract:

For regulatory bodies to better understand clinical data, data should be in a standard format. To achieve this the collected data must be cleaned and combined from various sources. Standardization of data improves decision making, productivity and quality. Laboratory data is one of the largest and relatively complex in clinical trial data. Laboratory data is collected both on case report forms or by using electronic data capture, obtained from different sources i.e. central laboratories or local laboratories and obtained in different formats like SAS® datasets, delimited text, Excel® and CSV files. These cause huge discrepancies in the laboratory data. Standardization of data will lead to effective analysis and generation of reports.

This paper gives an insight on how SAS can be used to convert clinical data to standard format of study data tabulation model.

#### Introduction:

The laboratory dataset is one of the most crucial datasets which indicates the drug safety and efficacy because of which laboratory dataset is present in most of the clinical trial submissions. Laboratory data is often split into subsets for quicker analysis.

- LBTESTCD and LBTEST:

LBTESTCD is the topic variable in the laboratory domain. Controlled terminology has been defined for both LBTEST and LBTESTCD. The controlled terminology codelist is extensible. For example, the test names may be present as Blood Glucose instead of Glucose, it is required by the programmer to correct these test names. Different approaches can be followed based on the data complexity and quantity. If there are fewer laboratory tests, conditional statements can be used. Proc format and pre-defined macros are also widely used.

TESTNM	LBTEST
Blood Glucose	Glucose
NK cells	Natural Killer cells
Alk Phos	Alkaline Phosphatase

```
if upcase(testnm)="BLOOD GLUCOSE" then lbtest="Glucose";  
else if upcase(testnm)="NK CELLS" then lbtest="Natural Killer Cells";  
else if upcase(testnm)="ALK PHOS" then lbtest="Alkaline Phosphatase";
```

- LBDTC:

LBDC is timing variable and it must follow ISO8601 format. Usually in raw data, date and time are in two separate variables. In ISO8601 format date and time are separated by letter "T". Date has to be in yymmdd10 format.

LBDT	LBTM
29-Mar-18	20:45
30-Mar-18	20:45
31-Mar-18	21:00
28-Mar-18	22:00
29-Mar-18	23:00

```
if lbdtd ne "" and lbtm ne "" then
lbdtc=strip(put(input(lbdtd,date11.),yymmdd10.))||"T"||
strip(put(input(lbtm,time5.),time5.));
```

- LBBFL

Baseline value is the last non-missing value on or before treatment start date. Unscheduled visits are not considered while determining baseline value.

Lab.sas7bdat

USUBJID	LBTESTCD	VISITNUM	VISIT	LBSTRESN	LBDC	RFSTDTC
ABC123-001	RBC	1	SCREENING	3.4	2018-09-01	2018-09-10
ABC123-001	RBC	2	VISIT 1	3.3	2018-09-10	2018-09-10
ABC123-001	RBC	3	VISIT 2	5.7	2018-09-20	2018-09-10
ABC123-001	RBC	4	FOLLOW-UP	4.3	2018-09-22	2018-09-10
ABC123-002	RBC	1	SCREENING	4.3	2018-09-10	2018-09-12
ABC123-002	RBC	99.1	UNSCHEDULED VISIT	5.7	2018-09-11	2018-09-12
ABC123-002	RBC	2	VISIT 1	.	2018-09-12	2018-09-12
ABC123-002	RBC	3	VISIT 2	8.5	2018-09-23	2018-09-12

```

proc sort data=lab out=baserecs;
by usubjid lbtestcd visitnum lbdtc;
where lbstresc ne "" and int(visitnum) ne 99;
run;

proc sort data=lab ;
by usubjid lbtestcd visitnum lbdtc;
run;

data baserecs1(keep=usubjid lbtestcd visitnum lbbfl);
set baserecs(where =( input(lbdtc, yymmdd10.) le input(rfstcd, yymmdd10.)));
by usubjid lbtestcd visitnum lbdtc;
if last.lbtestcd;
lbbfl="Y";
run;

data labs;
merge lab baserecs1;
by usubjid lbtestcd visitnum;
run;

```

- LBSTRESC

International system of units or SI units are the most widely used standard units. If units must be converted from g/dL to g/L, then the following code can be used.

```

if lborresu="g/dL" and lbstresu="g/L" then
lbstresc=put(input(lborres,best.)*10,best.);

```

- LBSEQ

Sequence is derived once all the other variables for laboratory domain have been obtained. For this the laboratory domain is sorted by the key variables for lb domain i.e. Studyid, usubjid, lbtestcd, lbspec, visitnum, lbtptref and lbtptnum and then sequence is assigned.

```
proc sort data=lb;  
by STUDYID USUBJID LBTESTCD LBSPEC VISITNUM LBTPREF LBTPNUM;  
run;
```

---

```
data lb;  
set lb;  
by STUDYID USUBJID LBTESTCD LBSPEC VISITNUM LBTPREF LBTPNUM;  
if first.usubjid then lbseq=1;  
else lbseq=lbseq+1;  
run;
```

### **Conclusion:**

This paper gives a generalized idea on how to standardize laboratory data. However, lab data greatly varies and programmer has to code accordingly.

### **References:**

1. Dealing with Lab Data – Stacking the Deck in Your Favor Jennifer Fulton, Westat, Houston, TX, Paper 182-2010, SAS Global Forum 2010.
2. Base SAS 9.4 Procedures Guide, Second Edition, SAS Institute, Cary, NC.
3. Global Checklist to QC SDTM Lab Data, Murali Marneni, Sekhar Badam, PPD, LLC, Morrisville, NC PPD. PharmaSUG 2018 – Paper DS-13.

### **Acknowledgements:**

I would like to thank Girish Kankipati, Vasanth Karri, Prithiviraj Loganathan and Suryakiran Potharaju for their encouragement. I would also like to thank my family and friends for their support.

### **Contact Information:**

Your comments and questions are valued and encouraged.

Contact the author at:

Laxmi Samhitha Bontha,  
Lamar University,  
bls26595@gmail.com