# How to build up the data flow that fits for analysis purpose of Laboratory data

Ruiying Liu, Fortrea, Beijing, China

## ABSTRACT

Laboratory analysis is a common part for clinical studies. It is important for statistical programmers to understand the data flow of Lab in clinical trial so they could generate statistical outputs more quickly and correctly.

In this article, we will focus on understanding the analysis requirements of Lab analysis and build up the workflow of data.

## UNDERSTAND LAB ANALYSIS REQUIREMENTS FROM PROTOCOL/SAP

Laboratory analysis is an indispensable part in a study of clinical trial. We could get useful information from Protocol/SAP in a specific study, for example:

**Category of Assessment:** Hematology, Chemistry, Urinalysis, Coagulation, Thyroid Function Test and Pregnancy, etc.

**Abnormal Laboratory Values:** Any abnormal finding in laboratory assessment at screening considered as clinically significant

**Clinical Laboratory Evaluations:** Central/Local laboratory will be used for all specimen collected.

**Limit of Quantification:** For laboratory values below the lower limit of quantification (LLOQ) like "< xxx" or "⩽ xxx", or above the upper limit of quantification (ULOQ) like "> xxx" or "⩾ xxx", LLOQ or ULOQ (xxx) as applicable will be imputed for purposes of calculating descriptive statistics.

**Toxicity Grade:** Hematological and biochemical values will be graded based on NCI CTCAE version 4.03. This is required for some studies.

**Requirements for Tables and Listings**: We need to summarize Clinical Laboratory values(hematology/general chemistry) values and changes from baseline for each visit for the purpose of descriptive analysis. The number and percentage of participants in each result category (normal, abnormal, clinically insignificant,

abnormal, clinically significant, and not applicable) will also be summarized by visit. We also need to summarize the values(numbers/percentages) of Pre-treatment and post-treatment hematological and biochemical by worst CTCAE grade.

## Data Structure for your study

Now you have some concepts in mind about laboratory data, and you could derive LB/ADLB using collected raw data. Firstly, you need to understand how laboratory data are collected in CRF. Typically, data for different categories(hematology, biochemistry, coagulation) are collected in different forms and we need to gather them together in the process of programming.

### Raw Data Collection

Here is an example of raw data(biochemistry). We could see for subject 001, biochemistry tests are performed once on Oct 2020. Raw data also specify test results and original units.

| SUBJID | LBPERF | LBDAT | SODIUM | SODIUM_UNIT |
|---|---|---|---|---|
| 001 | Yes | 28 OCT 2020 | 141 | Mmol/L |

### SDTM DERIVATION

There will be one record per lab assessment per subject. SDTM will be derived as following (use SDTM 3.3 as a reference). For this record, the information are not changed. Please note that we might need a conversion file to derive standard unit and value(if raw data are collected from central laboratories, typically raw units and values will be used for standard units and values). For below example, it is collected from central laboratory, so LBORRES and LBORRESU will be used for LBSTRESU.

| SUBJID | LBDTC | LBCAT | LBTESTCD | LBTEST | LBORRESU | LBORRES | LBORRESU | LBSTRESC |
|---|---|---|---|---|---|---|---|---|
| 001 | 2020-10-28 | BIOCHEMISTRY | SODIUM | Sodium | Mmol/L | 141 | Mmol/L | 141 |

### Tips for ADaM DERIVATION

If shift of toxicity grades are needed per Statistical Analysis Plan but not collected in raw data, for analysis purpose, we need to derive toxicity grade variables. Below is an example for derivation following CTCAE 5.0. ATOXGR is the analysis grade that we derive, BTOXGR is the baseline toxicity grade as analysis will be needed for "Change from Baseline Toxicity Grade". ANL01FL is derived for the worst post-baseline toxicity grade for each parameter. "Y" means the worst toxicity grade after baseline and it will be used for analysis. Please note that variable param is derived by concatenating LBTEST and LBSTRESU.

| SUBJID | ADT | PARAM | AVAL | ATOXGR | BTOXGR | ANL01FL |
|---|---|---|---|---|---|---|
| 001 | 28OCT2020 | Sodium(Mmol/L) | 141 | 1 | 0 | Y |

### Output Presentation

Typically laboratory analysis data will be presented for value/toxicity grade and result categories. Below is an example of change from baseline toxicity grades for Biochemistry. As we marked those with maximum Post-baseline CTCAE Grade with ANL01FL="Y" and also derived baseline CTCAE grade in Analysis Data, we could derive this table in a quick way.

| Maximum Post-baseline CTCAE Grade | Baseline[1]CTCAE Grade | | | | | |
|---|---|---|---|---|---|---|
| | Grade 0 n (%) | Grade 1 n (%) | Grade 2 n (%) | Grade 3 n (%) | Grade 4 n (%) | Total n (%) |
| Grade 0 | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) |
| Grade 1 | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) |
| Grade 2 | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) |
| Grade 3 | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) |
| Grade 4 | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) |
| Total | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx ( xx.x) | xx (100.0) |

## Summary

Laboratory analysis is very common in clinical trial. In this article we discussed the data flow that could fit the analysis purpose for Laboratory data. Good understanding of data structure could help you make the way of presenting your outputs more easily.

## REFERENCES

Referred CDISC documents are available from: http://www.cdisc.org

## CONTACT INFORMATION

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

Ruiying Liu Ruiying.Liu@fortrea.com

Room 102, B20 Universal Park, No.10 Jiuxianqiao Road, Chaoyang District, Beijing 100015, China