Paper PP19

How to map lab data into the LB domain and ADLB

Bert Parmet, IQVIA, Amsterdam, Netherlands

ABSTRACT

Capturing laboratory data in compliance with CDISC standards can be challenging due to various factors. Understanding the data requires a basic knowledge of medical concepts to assess validity. It is necessary to convert the diverse units used by different sites into single standard units to ease comparing the data and facilitate data analysis. Some lab values must be derived from existing measurements, like the Creatinine Clearance. Flags must be applied to out-of-range values, sometimes involving data from multiple time points. In oncology studies, lab values need to be graded according to the NCI CTC standard.

This presentation offers a detailed guide on mapping lab test results to CDISC standards, starting with a medical context overview, followed by mapping lab results to SDTM. Finally, creating ADLB is discussed, covering key tasks like deriving calculated lab values, setting flags for clinical concerns, and applying NCI CTC grading.

INTRODUCTION

The statistical analysis of laboratory data is an important part of the safety analysis. It provides valuable information on the subject's condition during the trial, as well as focused information on the condition of important organs. Capturing laboratory data in compliance with Clinical Data Interchange Standards Consortium (CDISC) standards[1] presents several challenges though. Understanding the data often requires a basic understanding of medical concepts to assess their validity. Another complexity arises from the diverse units used by sites to record lab results, necessitating conversions of lab results into standard units to ease comparing the lab values and prepare the data for analysis. Additionally, certain lab values must be derived from existing measurements (for example, calculating the Creatinine Clearance from serum Creatinine levels). Furthermore, to identify potential clinical concerns, flags need to be applied to out-of-range values, sometimes requiring consideration of multiple lab results across different time points. And finally, in oncology studies, lab values must be graded according to the Common Terminology Criteria for Adverse Events (CTCAE)[2] grading standard.

This paper provides a comprehensive overview of the process for mapping lab test results according to CDISC standards, including a brief introduction to the medical context of lab tests, the mapping of laboratory data into the CDISC Study Data Tabulation Model (SDTM)[1] standard and converting lab data into their standard units, and creating the Analysis Data Model (ADaM)[1] dataset ADLB including key tasks such as deriving lab values that need to be calculated from other lab values, setting flags for lab values of potential clinical concern, and applying CTCAE grading where applicable.

LABORATORY TESTING

There are a lot of different lab tests that can be performed, but there are basically two types of lab results:

- Quantitative: Numeric concentrations, numeric cell counts, fractions etc.
  Quantitative results are obtained when it is needed to know the exact concentration of a substance or the exact number of a certain type of cells, e.g. to determine whether the amount you see is a normal amount, too high or too low.
- Qualitative: Descriptions of the presence or absence of cells, substances, etc.
  Qualitative results are obtained when it is not needed to know the exact concentration of a substance or the exact number of cells, but only needed to know whether the substance occurs or the cells occur. For instance, the mere presence of glucose in the urine is already indicative for diabetes, and the mere presence of red blood cells in the urine is already indicative for structural damage in the urinary tract.
  There are several ways to qualitatively describe samples, but the most common ones are:
  - Is the substance or cell of interest present or not (positive versus negative).
  - Amount of the substance or cell, but not as exact numbers but indicated with for example plus and minus signs (-, +/-, +, ++, +++, >+++).
  - Colour of the specimen (e.g. yellow, orange, red).
  - Opaqueness of the specimen (e.g. clear or opaque).

MAPPING THE LAB DATA INTO SDTM: THE LB DOMAIN

CATEGORISATION OF THE LAB TESTS: LBCAT

The lab results are generally categorised in the clinical database as “Haematology” (blood cell counts), “Chemistry” (blood concentrations of substances), “Urinalysis”, “Immunology” (virus antigens and antibodies) etc. This is essentially a grouping based on the sample to be analysed. This categorisation based on the analysed substance is commonly mapped into LBCAT, but setting LBCAT according to a categorisation based on the method of analysis is also in use. A categorisation based on the method of analysis can look like the following:

- Chemistry: Determining the concentration of a substance in the specimen.
- Haematology: Counting the blood cells per volume unit.
- Urinalysis: Urine dipstick tests, as well as determining the urine’s colour and opaqueness.
- Coagulation: Measuring how long it takes before blood coagulates.
- Immunology: Checking for the presence or absence of antigens and antibodies against viruses, bacteria, parasites etc. This can be done as a positive/negative assessment, but also a titer (for example 1:2, 1:4, 1:8). From SDTM IG version 3.4 onwards, these data should be mapped into the IS domain.
- Pregnancy test.

DESCRIBING THE LAB TESTS: LBTESTCD, LBTEST, LBSPEC, LBMETHOD AND LBLOINC

The interpretation of a lab test result requires to know not only the substance tested or the cell counted, but also the specimen the result was obtained from. The variables LBTESTCD, LBTEST and LBSPEC are used for this purpose.

The variables LBTESTCD and LBTEST indicate the lab test performed, i.e. the substance whose concentration or presence was determined, the counted blood cells, aspect of the sample etc. These variables must be filled according to the codelists LBTEST and LBTESTCD in the SDTM controlled terminology.

It should be noted that some substances and blood cells can be measured in both blood samples and urine. LBTEST and LBTESTCD do not differentiate that out though. Whether Glucose is measured in blood serum or in urine, LBTESTCD is always equal to “GLUC” and LBTEST is always equal to “Glucose”. But the result requires a different interpretation depending on the source, i.e. blood serum vs urine, therefore it is necessary to capture the specimen in LB. Therefore, it is needed to capture the specimen where the result comes from as well. That happens in the variable LBSPEC. This variable contains the specimen where the results were obtained from. This can be whole blood, serum, plasma, urine, sputum and many more. This variable must be filled according to the codelist SPECTYPE in the SDTM controlled terminology.

In some cases, a fourth variable is needed to differentiate the lab tests even further. As an example, urine Glucose can be detected by means of a dipstick (yielding a qualitative result like “Positive”, “Negative” or the plus and minus signs) or by means of a concentration measurement. LBCAT indirectly references the method as dipstick results are captured in “URINALYSIS” while concentration measurements are captured in “CHEMSTRY”. Using LBMETHOD, the method can be referenced directly, e.g. “DIPSTICK MEASUREMENT METHOD”, as per the codelist METHOD in the SDTM controlled terminology.

A way to group the information in LBCAT, LBTEST, LBSPEC and LBMETHOD together is using a standard controlled terminology identifying medical laboratory observations. The controlled terminology most commonly used for this task is the Logical Observation Identifiers Names and Codes (LOINC)**Error! Reference source not found.**. Its values are provided with the lab data in the clinical database and mapped into the variable LBLOINC.

MAPPING THE ORIGINAL VALUES: LBORRES AND LBORRESU

The values as captured in the clinical database are mapped into LBORRES and LBORRESU. This should be done without modifying the results, but there is an exception. The units captured in LBORRESU should be units from the UNIT codelist in the SDTM controlled terminology. If the unit used in the clinical database does not occur in the UNIT codelist, an equivalent that does not require a conversion of the original result has to be mapped into LBORRESU (for example, if a blood cell count was measured in 10^6/ml but the codelist only foresees the units 10^6/L and 10^9/L, then the unit 10^9/L must be mapped into LBORRESU, because that unit does not require the original unit to be converted). Only if there is no equivalent in the UNIT codelist, a unit not occurring in this codelist can be mapped into LBORRESU.

MAKING THE LAB VALUES COMPARABLE AMONG EACH OTHER: LBSTRESC, LBSTRESN AND LBSTRESU

These variables are meant to show the collected lab values in only one unit (i.e. the “standard unit”, provided by the sponsor, although this still needs to follow SDTM controlled terminology) per combination of LBCAT, LBTEST, LBSPEC and LBMETHOD (i.e. the combination of lab test, specimen and testing method).

In the case of qualitative lab results, this is mostly done by setting LBSTRESC to LBORRES, LBSTRESN to missing and LBSTRESU to LBORRESU.

In the case of quantitative lab results, LBSTRESN should contain the numeric representation and LBSTRESC the character representation of the value in LBORRES, and LBSTRESU should contain the standard unit. If the standard unit differs from the original unit, the value to go into LBSTRESC and LBSTRESN must be converted into the standard unit by means of multiplying or dividing the original result by a conversion factor. This can lead to a different

value in LBSTRESC and LBSTRESN than in LBORRES, for instance when the original result was stored in mg/L, but the standard unit is mmol/L.

If LBORRES contains a “>” or “<” sign to indicate that the lab value is above or below a certain value (mostly because the indicated value is the highest or lowest value to measure), then the numeric part of LBORRES has to be converted to the standard unit, and that converted result has to be mapped to LBSTRESC with the “>” or “<” in front of it. Some study teams choose to set LBSTRESN to the numeric part of LBSTRESC in such cases.

To prevent an increasing rounding error in the tables, listings and figures, it is better to not round the lab results in SDTM or ADaM. However, because SAS stores numeric values in floating-point representation, which can lead to very small imprecisions[7], it is advisable to round at 8 decimals.

The units used as standard units can be divided into two groups: Units from the *Système International d’Unités* (International System of Units; abbreviation SI)[4], and the Conventional Units or United States Customary Units[5] [6]. More unit groups are in use, but the SI units and the Conventional Units are the two groups most commonly used in clinical research.

It is common practice to use SI units as standard units in the LB domain. However, as per the FDA’s Study Data Technical Conformance Guide of March 2025,[8], the FDA requires to receive the lab data both in SI units and in Conventional Units. The FDA suggests delivering the data in SI units in the LB domain and create an LC domain according to the structure of the LB domain to capture the lab results with Conventional Units as standard units.

NORMAL OR ABNORMAL LAB RESULTS AND THEIR CLINICAL SIGNIFICANCE

The purpose behind capturing lab values is to determine whether they change over time. If they do, it is important to see how much they change and whether the lab values become a health issue. Assessing whether lab values are a health issue or not requires lab reference ranges, which are mapped into SDTM like the following:

Variable	Contents
LBORNRLO	Lower bound of the reference range in the original unit stored in LBORRESU.
LBORNRHI	Upper bound of the reference range in the original unit stored in LBORRESU.
LBSTNRLO	Lower bound of the reference range in the standard unit stored in LBSTRESU. This must be equal to the value in LBORNRLO converted into the standard unit.
LBSTNRHI	Upper bound of the reference range in the standard unit stored in LBSTRESU. This must be equal to the value in LBORNRHI converted into the standard unit.
LBNRIND	Determination whether the lab value is inside, above or below the lab reference range. The values used are “NORMAL”, “ABNORMAL”, “LOW” and “HIGH”. It is up to the sponsor to decide whether the lab result and the reference range in the original unit or those in the standard unit should be used, and whether “NORMAL” includes or excludes the bounds of the reference range.
LBCLSIG	Supplemental variable to determine whether the lab value abnormality is clinically significant or not, as per the investigator’s assessment. If a lab value is clinically significant, it should be recorded on the Medical History form if measured at screening, and on the Adverse Events form as an Adverse Event if measured after screening.

It is up to the sponsor to decide whether local or central lab reference ranges should be mapped into LBORNRLO and LBORNRHI, but the values in LBSTNRLO and LBSTNRHI should always be a conversion of LBORNRLO and LBORNRHI respectively from the local unit to the standard unit.

PREPARING THE LAB VALUES FOR STATISTICAL ANALYSIS: THE ADAM DATASET ADLB

IDENTIFYING LAB TESTS: PARAM, PARAMCD AND PARCAT1

The variables to use for identifying the lab tests are PARAM, PARAMCD and PARCAT1. Each value of PARAM can be related to only one value in PARCAT1. This can be achieved by creating PARAM as a combination of LBTEST, LBSPEC and LBMETHOD. For Glucose, the result could be the following.

LBCAT	LBTESTCD	LBTEST	LBSTRESU	LBSPEC	LBMETHOD	PARAMCD	PARAM	PARCAT1
CHEMISTRY	GLUC	Glucose	mmol/L	SERUM		GLUCSER	Serum Glucose (mmol/L)	CHEMISTRY
CHEMISTRY	GLUC	Glucose	mmol/L	URINE		GLUCUR	Urine Glucose (mmol/L)	CHEMISTRY
URINALYSIS	GLUC	Glucose	Arbitrary U	URINE	DIPSTICK MEASUREMENT METHOD	GLUCDIP	Urine Glucose (arbitrary U; dipstick measurement method)	URINALYSIS

Because PARAM is mostly the value used in the tables to indicate the lab test, it is recommended to add the unit to the text in PARAM. This is also commonly done.

Several sponsors have controlled terminology systems in place to assign a unique code to each combination of lab test, lab specimen and lab method. These unique codes are mapped into PARAMCD and associated with unique texts to go into PARAM.

THE LAB TEST RESULTS: AVAL, AVALC, BASE, BASEC, CHG AND PCHG

In most cases, the unit to be used at ADaM level is the same one as the standard unit (LBSTRESU) in SDTM. In that case, the result in AVAL comes from LBSTRESN, and the result in AVALC comes from LBSTRESC. It can happen, however, that the unit to be used for AVAL and AVALC is a different one than the standard unit in LBSTRESU. In that case, the values in LBSTRESN and LBSTRESC must be converted to the unit to be used for AVAL and AVALC.

Caution has to be applied when filling AVAL and AVALC, because the ADaM standard requires a 1:1 relation between the values in AVAL and those in AVALC.

BASE and BASEC are used for capturing the baseline values as defined at ADaM level, i.e. the AVAL and AVALC values from the record where ABLFL is equal to “Y”.

CHG is the variable containing the change from baseline, calculated as AVAL – BASE.

PCHG is the variable containing the percentage change from baseline, calculated as (CHG / BASE) x 100.

DERIVING LAB VALUES

The most common kinds of derivations performed on lab values are the calculations of average values and derivations of lab parameters based on already available lab test results. For example, when the total number of white blood cells is known but not the absolute numbers of the individual types of white blood cells, then those absolute numbers can be derived from the total number of white blood cells and the differential blood cell counts in percentages.

Another lab test result that can be derived is the Creatinine Clearance (a measure for the kidney function). This is calculated by means of the Cockcroft-Gault Formula[9], using the serum Creatinine concentration and the subject’s age, weight and biological gender at birth as factors. There are two versions of the Cockcroft-Gault Formula: One to be used when the serum Creatinine concentration is given in µmol/L, and another one to be used when the serum Creatinine concentration is given in mg/dL.

When a lab value has been derived, the variable DTYPE (derivation type) must be filled, for example with “AVERAGE” for average records.

TRACING LAB VALUES

To facilitate tracing the values in ADLB back to SDTM and the clinical database, it is better to keep all SDTM variables from LB in ADLB. In derived records, however, the variables coming from SDTM.LB should be emptied, although keeping VISITNUM, VISIT and LBDTC filled makes it easier to see on which lab values the derived results are based. When a derivation is based on lab test results not obtained on the same date and time, however, it is better to empty LBDTC as well.

REFERENCE RANGES

While only one reference range can be stored in SDTM.LB, ADLB has the possibility to store multiple reference ranges. ANRLO, ANRHI and ANRIND are usually used for capturing the values in LBSTNRLO, LBSTNRHI and LBNRIND respectively, while BNRIND contains the ANRIND value of the baseline record (ABLFL=”Y”). When the reference ranges in SDTM are local reference ranges, which can differ between sites, the sponsor can decide to use uniform reference ranges for each lab test and store those uniform reference ranges in A1LO, A1HI, A1IND and

B1IND. More than one of such additional sets of reference ranges can be added as well (A*y*LO, A*y*HI, A*y*IND and B*y*IND). The following table explains these variables.

Variable	Purpose
AyLO	The lower bound of the *y*th reference range.
A*y*HI	The upper bound of the *y*th reference range.
A*y*IND	Determination whether the lab value is inside, above or below the *y*th lab reference range. The values used are “NORMAL”, “ABNORMAL”, “LOW” and “HIGH”. It is up to the sponsor to decide whether the lab result and the reference range in the original unit or those in the standard unit should be used, and whether “NORMAL” includes or excludes the bounds of the reference range.
B*y*IND	The A*y*IND value from the baseline record (ABLFL=”Y”).

CTCAE GRADING

As mentioned before, CTCAE toxicity grades should be assigned to the lab tests. This is a requirement from the National Cancer Institute (NCI) in the case of oncology trials. Not all lab tests need to be graded, but it is highly recommended to grade important ones like Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Bilirubin, Glucose, blood cell counts etc.

The NCI CTCAE version 5 documentation provides a list of how adverse events and abnormal lab values should be graded[2].

Grading lab values is mostly done by determining how far the lab value is outside the reference range, but in some cases by checking based on absolute values, or by checking how much bigger or smaller a lab value is compared with a baseline value that itself is outside the reference range.

There are a total of six variables containing the toxicity grades, and also six variables containing the toxicity grades at baseline. The following table explains the variables.

Character variable	Numeric variable	Contents of the variables
ATOXGR	ATOXGRN	Toxicity grade assigned.
ATOXGRL	ATOXGRLN	Toxicity grade assigned if AVAL is below the reference range.
ATOXGRH	ATOXGRHN	Toxicity grade assigned if AVAL is above the reference range.
BTOXGR	BTOXGRN	ATOXGR and ATOXGRN from the baseline record (ABLFL=”Y”).
BTOXGRL	BTOXGRLN	ATOXGRL and ATOXGRLN from the baseline record (ABLFL=”Y”).
BTOXGRH	BTOXGRHN	ATOXGRH and ATOXGRHN from the baseline record (ABLFL=”Y”).

CHECKING THE LIVER FUNCTION: HY’S LAW

Hy’s Law is a rule of thumb used for identifying subjects at risk of serious liver damage while they are on study treatment[10]. It consists of three components:

1. ALT and AST levels both elevated to ≥3 times above the upper bound of the reference range.
2. When the ALT and AST levels is ≥3 times above the upper bound of the reference range, also the Bilirubin values is ≥2 times above the upper bound of the reference range.
3. No other reason (like alcohol abuse or hepatitis) can explain these elevated ALT, AST and Bilirubin values.

The evaluation whether Hy’s Law applies to a subject or not is not performed in ADLB. What needs to happen in ADLB, however, is to flag the subjects who fulfil the Hy’s Law criteria regarding the ALT, AST and Bilirubin values. This should be done by checking for the first occurrence of ALT and AST values ≥3 times above the upper bound of the reference range, and then checking if there are any records with Bilirubin levels ≥2 times above the upper bound of the reference range chronologically on or after the elevated ALT and AST values. If that happens, then all the subject’s records on ALT, AST and Bilirubin need to be flagged, because they are all needed for the evaluation whether Hy’s Law applies or not.

CONCLUSION

Mapping the lab data into CDISC datasets is a challenging task requiring not only the data entered into the clinical database, but also the input of several external data sources for classifying and grading data. Furthermore, lab values and their reference ranges need to be converted from local units into standard ones to facilitate statistical analysis and comparability. Finally, a recent change in the FDA requirements has created the need to produce two lab datasets (one with SI units as standard units, the other with conventional units as standard units). This paper outlines key points to consider while creating SDTM.LB and ADLB.

REFERENCES

1. CDISC standards:
   https://www.cdisc.org/standards/foundational/sdtm
   https://www.cdisc.org/standards/foundational/sdtmig
   https://www.cdisc.org/standards/foundational/adam

2. https://dctd.cancer.gov/research/ctep-trials/for-sites/adverse-events/ctcae-v5-8x11.pdf

3. https://loinc.org/get-started/

4. https://www.bipm.org/en/measurement-units/

5. https://onbiostatistics.blogspot.com/2013/08/laboratory-tests-us-conventional-units.html

6. https://en.wikipedia.org/wiki/United_States_customary_units

7. https://documentation.sas.com/doc/en/workbenchcdc/v_001/vwblepg/n0o6e1t72yad42n10y9wf8pgmat8.htm

8. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document

9. https://en.wikipedia.org/wiki/Glomerular_filtration_rate#Cockcroft%E2%80%93Gault_formula

10. https://en.wikipedia.org/wiki/Hy%27s_law

CONTACT INFORMATION

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

Author Name: Bert Parmet
Company: IQVIA
Address: Herikerbergweg 181 – 342, 1101 CT Amsterdam, Netherlands
Email: bert.parmet@iqvia.com
Website: https://www.iqvia.com/