

Challenges in Processing Clinical Lab Data

Alan Meier, MedImmune, LLC., Gaithersburg, MD

ABSTRACT

Processing Clinical Laboratory data can be among the most complex tasks during analysis data set creation, particularly if the data are coming from local labs. This paper will explore some of these complexities within the CDISC SDTM LB data set structure, including:

- Determination of how to identify and group the lab tests: Starting with assigning test codes and high-level groupings (hematology, chemistry, urinalysis, etc.), deciding which codes to use, especially if the tests are not in the standard terminologies, and the need to sub-group the tests (proteins, WBC differentials, etc.).
- Dealing with different lab units: Where can conversion factors be found and how an organization can maintain and reuse them?
- Safety labs “pesky” urine dipstick textual results (trace, 1+, etc.): What can be done with them?
- Abnormality Criteria: What criteria can be used to identify abnormalities and what are the complexities involved with them?

INTRODUCTION

It seems that Lab data is a mystery to many clinical/statistical SAS programmers, who blindly run other programmers' code without understanding the data and the analysis that's taking place. This can lead to inaccurate results and make it nearly impossible to assure the validity of the programming. With the goal of providing a better understanding of lab data, it's hoped the reader will be able to produce better code.

The Clinical Data Interchange Standards Consortium (CDISC) Standard Data Tabulation Model (SDTM) data standards are a recognized industry standard for storing clinical data and the FDA will soon be requiring data to be submitted in this format. References to this standard are used throughout this paper and are based on the SDTM version 1.3 and Implementation Guide version 3.1.3. For more information about CDISC and the SDTM, visit the CDISC web site at www.cdisc.org.

Any reference to web sites or books is for example purposes and should not be seen as an endorsement by the author for the value or content of the material presented therein.

LAB DATA BASICS

Clinical lab data begins with the collection of a biological specimen from a study subject. A specimen can be a biological fluid (e.g. blood, urine, spinal fluid, semen, nasal wash), tissue sample (e.g. skin, hair, nail) or other biological sample (e.g. feces). That specimen is sent to a laboratory where it is analyzed. The lab can be a “central lab” used by multiple investigational sites within a study that consolidates the data for a sponsor or it can be a “local lab” within the investigator's office or associated with the investigator's site.

The results from that analysis are sent to the investigator and eventually to the study sponsor. The investigator will use the results to monitor and treat the subjects. The sponsors will use the results to assess the safety (and potentially efficacy) and they may need to submit the results to regulatory agencies.

Each collected specimen can produce one or more lab results. When the specimens are collected, they are usually given some type of identification by the processing lab. This id is sometimes referred to an accession number. Some type of requisition is included with the specimen when it is sent to the lab that indicates which tests are to be run. A “panel” is often used to refer to a pre-specified set of tests. The requisition can indicate one or more panels to be run.

A minimum of three fields beyond the subject and specimen identification are needed when lab results are sent to the sponsor: the lab test, the result and the units used to report the result. Other useful fields would include the high and low reference ranges and specimen condition. The reference ranges (also referred to as “lab normals”) are usually the range in which most lab results would fall for a large percentage of people.

IDENTIFYING LAB TESTS

At a minimum, there are two fields that are required to identify a lab result: a name and a code. In the SDTM LB domain, these variable names are LBTEST and LBTESTCD. CDISC has developed a control terminology for these variables to help ensure standardization across the industry.

The control terminology is managed by CDISC and is maintained and distributed by the National Cancer Institute (NCI) Enterprise Vocabulary Services (EVS). It can be downloaded in a multitude of file formats including Excel, text, odm.xml, pdf and html. A link to the NCI Thesaurus download web site is available on the CDISC web site under Foundation Standards, Terminology. It can also be accessed directly at <http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc>. CDISC manages a number of sets of control terminologies for the various Foundational Standards. The terminology is updated on a quarterly basis. Requests for new terms can be submitted for consideration using the New Term Submission Request at the NCI download web page. Below is a screen-shot example of the Lab Test Code terminology.

| Code | Codelist Code | Codelist Extensible (Yes/No) | Codelist Name | CDISC Submission Value | CDISC Synonym(s) | CDISC Definition | NCI Preferred Term |
|---------|---------------|------------------------------|----------------------|------------------------|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| C65047 | | Yes | Laboratory Test Code | LBTESTCD | Laboratory Test Code | Terminology used for Laboratory Tests of the CDISC Standard Data Tabulation | CDISC SDTM Laboratory Test |
| C74797 | C65047 | | Laboratory Test Code | ANISO | Anisocytes | A measurement of the inequality in the size of the red blood cells in a whole | Anisocyte Measurement |
| C79602 | C65047 | | Laboratory Test Code | POIKILO | Poikilocytes | A measurement of the odd-shaped erythrocytes in a whole blood specimen. | Poikilocyte Measurement |
| C38037 | C65047 | | Laboratory Test Code | BILI | Bilirubin; Total Bilirubin | A measurement of the total bilirubin in a biological specimen. | Total Bilirubin Measurement |
| C38462 | C65047 | | Laboratory Test Code | APTT | Activated Partial Thromboplastin Time | A measurement of the length of time that it takes for clotting to occur when reagents are added to a plasma specimen. The test is partial due to the | Partial Thromboplastin Time |
| C105585 | C65047 | | Laboratory Test Code | GLUC | Glucose | A measurement of the glucose in a | Glucose |
| C45997 | C65047 | | Laboratory Test Code | PH | pH | A measure of the acidity or alkalinity of a fluid on a scale of 0 to 14.(NCI) | pH |
| C51946 | C65047 | | Laboratory Test Code | RBC | Erythrocytes; Red Blood Cells | A measurement of the total erythrocytes in a biological specimen. | Erythrocyte Count |
| C51947 | C65047 | | Laboratory Test Code | RETI | Reticulocytes | A measurement of the reticulocytes in a biological specimen. | Reticulocyte Count |
| C51948 | C65047 | | Laboratory Test Code | WBC | Leukocytes; White Blood Cells | A measurement of the leukocytes in a biological specimen. | Leukocyte Count |
| C51949 | C65047 | | Laboratory Test Code | LYM | Lymphocytes | A measurement of the lymphocytes in a biological specimen. | Lymphocyte Count |
| C51951 | C65047 | | Laboratory Test Code | PLAT | Platelets | A measurement of the platelets in a | Platelet Count |

Display 1. CDISC SDTM Lab Test Code Terminology

For ease of use, the Lab Test Code terminology is identified as Code List C65047 and the Lab Test Name is Code List C67154. For each test code, there should be a test name in the terminology. It should be noted that the test codes and test names are not specimen specific (e.g. Glucose can be measured in both blood and urine).

When the data comes in to the sponsor, the lab tests need to be mapped to the correct test code and test name. This can sometimes be difficult as there are synonyms and acronyms that are commonly used. For example, Red Blood Cells, RBC and Erythrocytes are all the same analyte. The CDISC terminology has most of the synonyms/acronyms included in the table.

If a lab test is done that is not in the control terminology, there are two courses of action. The first is to submit a request for a new lab test code and name. When requesting a new lab test code and name, the requestor is required to provide contact information, information on whether the request is for a new code or a new code list, and provide a proposed lab test code and corresponding test name. The CDISC Terminology team will review the request internally first, perhaps soliciting input from various SDTM sub-teams and if accepted, will send the proposal out for review by the entire user community, before adding it to the terminology. The submitter may be asked to participate in a meeting where the new term or code list is discussed. Unfortunately final resolution could take 3 months. The second course is to create a test and name combination. The two terminologies are extensible, so there is no issue for a sponsor to add to them for a submission. It is advisable to still submit a request for a new code and name even if the second option is selected.

One place to look for new codes is the lab that is supplying the data or another lab site. One site that has been useful is the ARUP Lab site (www.aruplab.com). ARUP Laboratories is a National Reference Laboratory and it's an enterprise of the University of Utah and its Department of Pathology. They perform more than 3,000 tests and test combinations that cover routine screening to highly esoteric assays. The web site's Laboratory Test Directory

includes an mnemonic that can be used in many cases as a starting point for a new test code. The web site will sometimes supply the analyte's reference ranges and a standard unit of measurement. Another potential site to use for new test codes is the Cleveland Clinic Laboratory (portals.clevelandclinic.org/reflab/Home/tabid/4036/Default.aspx).

ARUP LABORATORIES | Laboratory Test Directory

Laboratory Test Directory | Search Site

Browse A-Z | Search the LTD.

Pancreatic Polypeptide

0099436

Ordering Recommendation

Aids in the diagnosis and monitoring of pancreatic neuroendocrine tumors.

| | |
|------------------|-------------------------------|
| Mnemonic | Methodology |
| PAN POLY | Quantitative Radioimmunoassay |
| Performed | Reported |
| Wed | 3-11 days |

New York DOH Approval Status

This test is New York DOH approved.

Specimen Required

Patient Preparation: Patient should be fasting for 10 hours prior to collection of specimen.

Collect: Serum separator tube or plain red.

Specimen Preparation: Allow specimen to sit in collection tube for 15-20 minutes at room temperature for proper clot formation. Centrifuge and separate serum from cells ASAP or within 2 hours of collection. Transfer 1 mL serum to an ARUP Standard Transport Tube. (Min: 0.5 mL)

Storage/Transport Temperature: Frozen.

Unacceptable Conditions: Plasma. Severely hemolyzed or lipemic specimens.

Remarks:

Stability: After separation from cells: Ambient: 24 hours; Refrigerated: 24 hours; Frozen: 2 months

Reference Interval

0-435 pg/mL

Interpretive Data

Display 2. ARUP Laboratory Test Directory Sample

Cleveland Clinic

Cleveland Clinic Laboratories

TRUST in us for everything you need in a reference lab

Search Results

Ordering Mnemonic Search Results

Located 4 tests containing 'panc' in (and sorted by) Ordering Mnemonic ...

| No. | Test Name | Ordering Mnemonic | Billing Code | CPT Code |
|-----|--------------------------------------------|-------------------|--------------|--------------------------|
| 1. | Pancreatic Polypeptide | PANC | 75675 | 83519 |
| 2. | Pancreatic Elastase, Fecal | PANCEF | 83044 | 83520 |
| 3. | Pancreatitis Panel | PANCPL | 83724 | 81223, 81224, 81404 (x2) |
| 4. | Pancreastatin | PANCST | 88584 | 83519 |

Search Test Information

ABCDEFGHIJKLMNOPQRSTUVWXYZ

1 To 9

Search

Test Name

Submit

Search Tips

Please contact Client Services (216.444.5755) if you are unable to locate a test in the directory.

Display 3. Cleveland Clinic Laboratory Test Code Sample

When creating new lab test codes, there are a number of rules that need to be followed.

1. Test names must be limited to 8 characters, and cannot start with a number, nor can they contain characters other than letters, numbers, or underscores. This is a SAS V5 Transport file limitation, and since use of SAS V5 is specified in the current guidance this limitation will be in effect until the use of other formats (such as XML) becomes acceptable to regulatory authorities.
2. Variable descriptive names (labels), up to 40 characters, should be provided as data variable labels.
3. Use the entire test name for the variable if the text name is 8 characters or less.
4. Test names do not contain the matrix of the specimen. This information is populated in the specimen type field (LBSPEC).
5. Lab tests where results can be expressed as qualitative, semi-quantitative or quantitative should all have the same variable name.
6. If there are multiple tests with the same base, (e.g, Bilirubin, Direct Bilirubin, Indirect Bilirubin) the base text goes first and the qualifiers are suffixes; BILI, BILDIR, BILIND.
7. For all differential test names the absolute count is a short defined term (e.g, EOS, BASO) and the ratio/percentage has the same short mnemonic with a second short mnemonic for the denominator e.g, LE for leukocytes (EOSLE, Eosinophils/Leukocytes), LY for lymphocytes (LYMMCELY, Lymphoma Cells/Lymphocytes), RBC for erythrocytes (RETIRBC, Reticulocytes/Erythrocytes). LY will be used when there is a denominator. When lymphocyte(s) is part of the core test name use LYM.
8. Rule 7 above states that LY will be used when there is a denominator. When lymphocyte(s) is part of the core test name use LYM.
9. Use HPO for hypo and HPR for hyper as prefix.
10. Use 'biological specimen' instead of 'biological sample' in the definition
11. Use BLST as a prefix for blasts so that all blasts sort together
12. Use CY as a prefix for crystals
13. Use CS as a prefix for casts
14. Use AB as a suffix for antibody
15. Use AG as a suffix for antigen
16. Use AC as an abbreviation for acid
17. Use CE as an abbreviation for cells
18. Use EPI as an abbreviation for epithelial
19. Will save CR for 'corrected'. This would apply to test results that are corrected
20. Use FR an abbreviation for Free
21. Use FG an abbreviation for Fragment
22. Use LG an abbreviation for Large
23. Use CREAT or CRT for creatinine
24. For multiple related tests place the core concept first so all will sort together
25. For alpha use A or AL
26. For any item that has components and is also expressed as a total, the components will have LBTESTCD defined as the item with a suffix and the total value will be the item code with no "total" in LBTESTCD - total will be in the definition
27. Use VLD or VL as a suffix for viral load as needed to fit the 8 character rule.
28. Use AT for active and use IT for inactive
29. HEX will be the shortened text for the notion of six
30. Calculations that require two or more separate collections or time points will not be assigned controlled lab test codes

31. Use the first letter of the genus and 2 letters for the species
32. Use TXN for toxin
33. Use A for activity
34. Use IDX for index

Once test codes and test names are assigned, it is useful to look to higher level groupings that can aid in the review and reporting of the lab results.

GROUPS AND SUB-GROUPS

The most common grouping of lab specimens are Chemistry, Hematology and Urine, but this list is far from complete. Additional groups include Drug Screening, Fertility, Genetics, Toxicology, Immunology, Virology, Histology, Cytology and Pathology, to name a few. These higher level groupings are usually stored in the LBCAT variable and should use a sponsor defined terminology that is consistent across a submission.

A number of the groups have sub-groups. For example, Hematology can have Coagulation, Complete Blood Counts (CBC) and WBC Differentials and Chemistry can have Lipids, Endocrine, Liver (Hepatic) Function Tests, Renal (Kidney) Function Tests, Electrolytes, Proteins, Cardiovascular and Thyroid Function Tests. These values are usually stored in the LBSCAT variable and should use a sponsor defined terminology that is consistent across a submission.

REFERENCE RANGES

A reference range, also known as a normal range, is a range of values where a specified percentage of samples fall for a given analyte and partition. A common interval is 95%, but may vary. Partitions subset the population and usually include factors such as the subject's age, gender and sometimes race (e.g. PSA and eGFR in African Americans).

Reference ranges are usually specific to a laboratory site, so even though a sponsor is using a central lab, if samples are shipped to more than one laboratory location, the reference ranges can vary.

Reference ranges can also change over time. Labs are required to review their reference ranges on a regular basis. Reference ranges can also change when the lab changes or recalibrates machinery or if a method for analysis is changed.

Most central labs have the ability to attach reference ranges to test results when they are sent to the sponsor. Smaller and specialized labs may not have the computing resources to provide that service and it falls on the sponsor to attach the reference ranges for reporting. When an electronic version of reference ranges comes to the programmer, it can be a challenge as reference range partitions for the analytes may vary. One approach in dealing with this is to duplicate all reference ranges for all partitions, but that may not work as some may have different age partitions and varying start dates. With the exception of the race partition, using simple PROC SQL code works well in this application. (Reference ranges for PDA and eGFR vary for African Americans and a temporary flag variable can be created to indicate African American or not. That variable is then added to the JOIN ON statement.)

To illustrate, we can use a small subset of data. For simplicity it's assumed that there is only one lab site. If more than one site is used, the lab site variable will need to be added to both data sets and included on the PROC SQL code below.

First the sample reference ranges. The coloring and first column, which is a row number, are used to help in tracking the steps and are not part of the actual data set.

| | LBTESTCD | STARTDTC | ENDDTC | SEX | AGE_LO | AGE_HI | LO_RANGE | HI_RANGE |
|----|----------|------------|------------|------|--------|--------|----------|----------|
| A1 | ALP | | 2014-03-31 | both | 25 | 60 | 13 | 131 |
| A2 | ALP | 2014-04-01 | | both | 25 | 60 | 15 | 135 |
| A3 | ALP | | 2014-03-31 | both | 61 | | 51 | 153 |
| A4 | ALP | 2014-04-01 | | both | 61 | | 55 | 150 |
| A5 | HGB | | | F | | | 12 | 15 |
| A6 | HGB | | | M | | | 13.5 | 16.5 |

Table 1. Lab Reference Range Sample Data

Next are the data. Again the first column is a row number for help tracking the progress.

| LB | | | | | | | |
|----|----------|----------------------|-----|-----|------------|----------|---------|
| | LBTESTCD | LBTEST | SEX | AGE | LBDBC | USUBJID | LBORRES |
| B1 | ALP | Alkaline Phosphatase | F | 30 | 2014-03-21 | S001-001 | 12 |
| B2 | ALP | Alkaline Phosphatase | F | 30 | 2014-04-05 | S001-001 | 51 |
| B3 | ALP | Alkaline Phosphatase | M | 65 | 2014-03-21 | S001-002 | 152 |
| B4 | ALP | Alkaline Phosphatase | M | 65 | 2014-04-05 | S001-002 | 151 |
| B5 | HGB | Hemoglobin | F | 30 | 2014-03-21 | S001-001 | 13.5 |
| B6 | HGB | Hemoglobin | F | 30 | 2014-04-05 | S001-001 | 14.2 |
| B7 | HGB | Hemoglobin | M | 65 | 2014-03-21 | S001-002 | 15.1 |
| B8 | HGB | Hemoglobin | M | 65 | 2014-04-05 | S001-002 | 14.9 |

Table 2. Lab Results Sample Data

Once the data are read in, the next step is to duplicate the reference ranges that are the same for both genders. In this example, they are identified with a code of "both". If the end date and/or high age values are missing, values greatly beyond possibility are supplied. A simple DATA step handles this.

```
**** Duplicate Genderless Reference ****;
**** Ranges for Females and Males ****;
data LABNORM;
  set LABNORM;
  **** if no end date - create one ****;
  if enddtc = '' then enddtc = '2525-01-01';
  **** if no hi age range - create one ****;
  if age_hi = . then age_hi = 9999;
  if sex = 'both' then do;
    sex = 'F';
    output;
    sex = 'M';
    output;
  end;
  else output;
run;
```

Simple PROC SQL logic from Brenda Barber's SUGI22 paper is used to combine the two data sets.

```
**** Combine and Set Reference Ranges ****;
proc SQL;
  create table LB1 as
  select l.*, n.lo_range as lbornlo, n.hi_range as lbornhi
  from LB as l left join LABNORM as n
  on l.lbtestcd = n.lbtestcd and l.sex = n.sex and
  (n.age_lo <= l.age <= n.age_hi) and (n.startdtc <= l.lbdtc <= n.enddtc)
  order by lbtestcd, sex, age, lbdtc
  ;
quit;
```

The resulting data set is illustrated in the following table.

| | LBTEST CD | LBDBC | SEX | AGE | USUBJID | LBORRES | LBORNLO | LBORNHI |
|----|-----------|------------|-----|-----|----------|---------|---------|---------|
| B1 | ALP | 2014-03-21 | F | 30 | S001-001 | 12 | 13 | 131 |
| B2 | ALP | 2014-04-05 | F | 30 | S001-001 | 51 | 15 | 135 |
| B3 | ALP | 2014-03-21 | M | 65 | S001-002 | 152 | 51 | 153 |
| B4 | ALP | 2014-04-05 | M | 65 | S001-002 | 151 | 55 | 150 |
| B5 | HGB | 2014-03-21 | F | 30 | S001-001 | 13.5 | 12 | 15 |
| B6 | HGB | 2014-04-05 | F | 30 | S001-001 | 14.2 | 12 | 15 |
| B7 | HGB | 2014-03-21 | M | 65 | S001-002 | 15.1 | 13.5 | 16.5 |
| B8 | HGB | 2014-04-05 | M | 65 | S001-002 | 14.9 | 13.5 | 16.5 |

Table 3. Lab Results Sample Data with Reference Ranges

If a lab is unable to provide reference ranges, there are some acceptable published sources. A search of the internet will provide a host of results for common and not so common lab tests. www.globalrph.com/labs_home.htm is referenced by several other sites. The National Institute of Health (NIH) gives some ranges in its A.D.A.M. Medical Encyclopedia at <http://www.nlm.nih.gov/medlineplus/encyclopedia.html>. The Merck Manual also publishes reference ranges on-line at www.merckmanuals.com/professional/appendixes/normal_laboratory_values/normal_laboratory_values.html. There is also a book by Young and Huth titled "SI Units for Clinical Measurements" which supplies reference ranges and unit conversion factors for hundreds of analytes from various specimens.

LAB UNITS

Another major headache associated with lab data and the one that leads to many issues are the units that the analytes are reported in. Many large central labs can report lab test results in just about any unit (assuming there is a conversion factor). Sponsors just need to determine ahead of time which units they want to use when reporting the data. The issue again is when dealing with smaller and specialty labs or if a decision is made later to report in some unit other than that originally reported by the lab.

The SDTM LB domain has two sets of results and reference range variables. One for the original units reported by the lab and one for the sponsor's standardized units used for reporting. Programming the conversions is not difficult if the basics are understood and a conversion factor is available.

There are two major groupings for units, Standard International (SI), also known as metric, and Conventional Units. Conventional Units refer to what most U. S. trained physicians are accustomed to seeing. Within each, there are further variations in amounts. These types of variations account for the majority of conversions.

DO THE UNITS MAKE SENSE?

To really be able to process lab data, a rudimentary understanding of the lab tests is required. While it is usually the responsibility of Data Management and/or Data Monitoring, it may be difficult for personnel in these functions to see where issues exist. In some Database Management Systems (DBMS), the reference ranges are entered once for a lab site and the DBMS automatically merges the units onto the lab results. If a data entry error occurred with the entry of these units or if the person entering the units has no knowledge of lab data and doesn't know to query questionable units, improper units may be entered. Issues are usually first seen when processing the lab data.

Most numeric results are a count of something in some volume or a ratio. Issues occur when the volume is left off or when the units indicate a count is being given where a ratio is expected or vice-versa. If the test code indicates an absolute monocyte count and the units indicate "%", there's an issue. If the units for a Prothrombin Time (PT) indicate a "RATIO", an International Normalized Ratio (INR) result is probably being transmitted in error. Be aware that PT and INR are not the same result. The web sites cited above for lab test codes and reference ranges are good places to get an idea of what units should look like for an analyte. Another good place is the "SI Units for Clinical Measurement" book.

Things become tricky when sites use synonyms for the same units. For Albumin, common Conventional Units are grams per deciliter. Synonyms include G/dL and gm/dL. If a unit comes across as mg/dL, is it a really milligrams per deciliter or a typo ('m' and 'g' inverted)? If the results appear to be a thousand times greater than the gm/dL results then it is probably correct.

The conversion factor for electrolytes in milliequivalent (mEq) to millimoles (mmol) is often 1, but the two units are not the same. Equivalence is based on the Moles of electrons gained or lost in the ionized form from a Mole of the substance. Potassium has a +1 valence, so Moles and equivalents are the same. For calcium, which has a valence of +2, one Mole would be two equivalents.

Another common issue occurs with urine samples when results are reported in amounts per time rather than amounts per volume. Unless the volume per the time interval is also reported, there is no way to convert the units. If the total volume is supplied, calculate what percentage of the total volume the reported volume is and take that percentage of the original result for the standardized result.

CONVERSION FACTORS – CREATING STANDARD UNITS

Per the SDTM LB domain, sponsors are to supply lab results in original units and in standard units. The standard units are the units that the sponsor intends to use in analyzing and reporting the results. If the original and standard units are the same, then the conversion factor is 1 and the original results and reference ranges are copied into the standard variables.

The majority of unit conversions involve a count per volume unit. Conversion can be made to either the count or the volume or both and it often gets confusing.

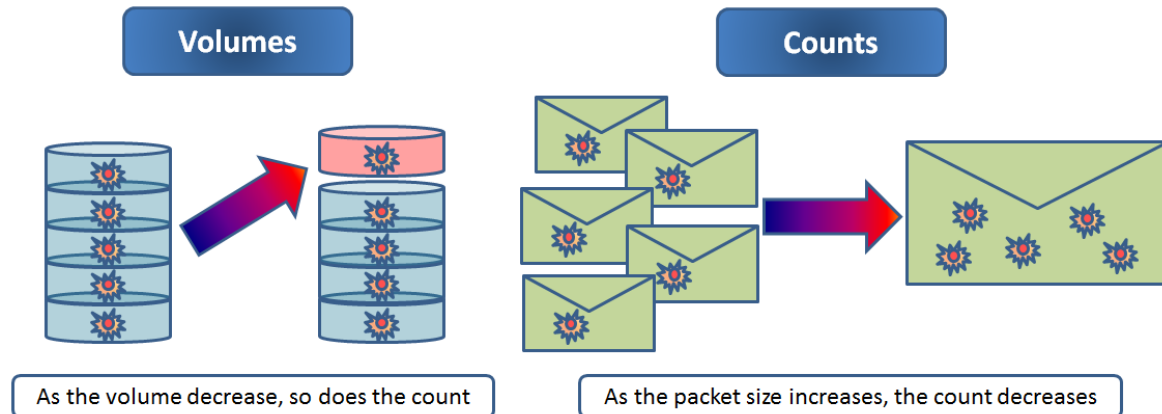


Figure 1. Volume and Count Conversions

When going from a larger volume to a smaller volume, the result should get smaller (e.g. when going from a liter to a milliliter, divide the result by 1000). Counting is done on packages of items and can be as small as a single item (e.g. a cell) or can be millions of an item. As the size of the packet increases (more of the individual items in a packet), the result will get smaller (e.g. when going from milligrams to grams, divide by 1000). Results are multiplied by the conversion factor when converting in the opposite direction. When modifying both count and volume, the modification is cumulative (e.g. converting from milligrams per liter to grams per milliliter, the result is divided by 1000 for the volume conversion and divided by 1000 again for the count conversion). Sometimes the conversions cancel each other out and the units are synonymous (e.g. converting from grams per liter to milligrams per milliliter).

Another major area of conversion happens when converting from Conventional Units to SI units. Besides the mEq to mmol conversion from above, the major changes are reporting mass in Moles (when the molecular weight is known) and enzyme activity in katal (kat) instead of Enzyme Units (U).

A Mole is the number of atoms contained in 12 grams of carbon-12, the result, referred to as Avogadro's number, is approximately 6.022×10^{23} . Following the generalization that the mass in grams of one Mole of atoms of any element is equivalent to its atomic mass, 1 gram of an element is equal to $(1 / \text{atomic mass unit (amu) of the element})$ equals the amount in Moles. To convert a bilirubin sample, 1 gram equals $(1 / 584.69)$ or 0.00171 Moles and 1 milligram is $1000 * 0.00171$ or 1.71 Moles. Finding the amu of an analyte is often a challenge. In the case of bilirubin, one molecule is made up of 33 carbon (12.01 amu), 26 hydrogen (1.01 amu), 4 nitrogen (14.00 amu) and 6 oxygen (16.00 amu) atoms. So the amu of bilirubin is $(33 * 12.01) + (26 * 1.01) + (4 * 14.00) + (6 * 16.00)$ or 584.69.

A katal (KAT) is the SI unit that measures catalytic activity and replaces Enzyme Units (U), though results reported in U are still more prevalent. An Enzyme Unit is the amount of an enzyme that causes 1 μmol of substrate to catalyze in a minute under specific conditions. A katal is the amount of an enzyme that causes 1 mol of substrate to catalyze in a second under specific conditions. The conversion factor is $1\text{U} = 1/60 \text{ micro-kat}$. Note that an International Unit (IU) is not the same as an Enzyme Unit. An IU is the mass or volume of a substance and the amount varies by the substance with the variation based on the biologic activity of the substance. IUs are mainly used to measure substances with different forms or preparations (e.g. vitamins, hormones, some medications, vaccines, etc.).

Another major headache in conversions is trying to keep track of how many zeroes to use in the conversion factors. A simple chart can help. The following conversion chart may be helpful.

| | pico- | nano- | micro- | milli- | deci- | zero | thousand | million | billion | trillion |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| pico- | 1.00 | 1.00E+03 | 1.00E+06 | 1.00E+09 | 1.00E+11 | 1.00E+12 | 1.00E+15 | 1.00E+18 | 1.00E+21 | 1.00E+24 |
| nano- | 1.00E-03 | 1.00 | 1.00E+03 | 1.00E+06 | 1.00E+08 | 1.00E+09 | 1.00E+12 | 1.00E+15 | 1.00E+18 | 1.00E+21 |
| micro- | 1.00E-06 | 1.00E-03 | 1.00 | 1.00E+03 | 1.00E+05 | 1.00E+06 | 1.00E+09 | 1.00E+12 | 1.00E+15 | 1.00E+18 |
| milli- | 1.00E-09 | 1.00E-06 | 1.00E-03 | 1.00 | 1.00E+02 | 1.00E+03 | 1.00E+06 | 1.00E+09 | 1.00E+12 | 1.00E+15 |
| deci- | 1.00E-11 | 1.00E-08 | 1.00E-05 | 1.00E-02 | 1.00 | 1.00E+01 | 1.00E+04 | 1.00E+07 | 1.00E+10 | 1.00E+13 |
| zero | 1.00E-12 | 1.00E-09 | 1.00E-06 | 1.00E-03 | 1.00E-01 | 1.00 | 1.00E+03 | 1.00E+06 | 1.00E+09 | 1.00E+12 |
| thousand | 1.00E-15 | 1.00E-12 | 1.00E-09 | 1.00E-06 | 1.00E-04 | 1.00E-03 | 1.00 | 1.00E+03 | 1.00E+06 | 1.00E+09 |
| million | 1.00E-18 | 1.00E-15 | 1.00E-12 | 1.00E-09 | 1.00E-07 | 1.00E-06 | 1.00E-03 | 1.00 | 1.00E+03 | 1.00E+06 |
| billion | 1.00E-21 | 1.00E-18 | 1.00E-15 | 1.00E-12 | 1.00E-10 | 1.00E-09 | 1.00E-06 | 1.00E-03 | 1.00 | 1.00E+03 |
| trillion | 1.00E-24 | 1.00E-21 | 1.00E-18 | 1.00E-15 | 1.00E-13 | 1.00E-12 | 1.00E-09 | 1.00E-06 | 1.00E-03 | 1.00 |

Table 4. Conversion Chart

A few examples on how to use the chart. Example 1, changing volumes – liter to milliliter. Find zero on the vertical (6th row) and go horizontally until the column with the milli- heading (4th column). The result is 1.00E-03. This indicates that the result should be reduced by dividing by 1,000 (3 zeroes) (or multiplying by 0.001). Going from milliliters to liters, go to the milli- (4th row) over to the zero (6th) column. The result is 1.00+03, indicating an increase in the volume and result by multiplying by 1,000. Example 2, changing counts – cells to millions of cells. Starting on the zero (6th) column and going down to the million (8th) row, the result is 1.00E-6, indicating a decrease in the result as the size of packets increase and the number of packets decreases. The conversion is to divide the result by 1,000,000.

There are many on-line tools and books that will supply conversion factors. If a web-site won't provide the actual conversion factor, requesting the conversion of value 1 will provide the conversion factor. Molecular weights for elements and numerous molecules are easy to find on-line. If a particular molecular weight can't be found, hopefully the elemental structure will be available and the weight calculated based on the process described above. If all else fails, contact the lab that generated the results. Hopefully they can provide conversion factors or provide the results in alternate units.

One final conversion topic – dealing with White Blood Cell (WBC) Differentials. There are five types of WBCs: Basophils, Eosinophils, Lymphocytes, Monocytes and Neutrophils. Neutrophils are made up of Segmented Neutrophils and Banded Neutrophils. Lymphocytes are made up of B-cells, T-cells and natural killer cells. A WBC result will be the sum of the absolute count of the five different types. Results for the differentials will be either reported as absolute counts or the percentage of the cell type in relation to the total WBC result.

Conversion of the results of the differentials from absolute to percent and vice-versa is possible. The first step is to merge the WBC result with the subject / time point matched differential. To convert from absolute to percent, divide the absolute differential result by the WBC count and multiply by 100. To convert from percent to absolute, divide the percent differential by 100 and multiply by the WBC count. Be aware that the converted results are only estimates and may not add up to the WBC count or 100%.

It is not possible to convert the reference ranges. The matched WBC count is the result for the subject at that specific time point, not the averaged WBC count from the multitude of reference specimens used to generate the reference ranges. If the converted reference ranges are needed, they must be obtained from the original lab.

URINALYSIS RESULTS

Urinalysis results are many times evaluated using a dipstick testing method that produces qualitative (positive / negative) or semi-quantitative (1+, 2+, 3+ 4+) results. If a microscopic examination is done, cast, cell and crystal results can be supplied in a semi-quantitative manner (e.g. none, trace, few, some, many, too numerous to count). Some sponsors code these results to make them easier to work with. A table is built for each analyte and the possible character results are mapped to numeric equivalents. See the table below for an example.

| Analyte | 0 | 0.5 | 1 | 2 | 3 | 4 |
|---------|---------------------------------------|-------|-------------------------|---------------------------|-----------------------------------------|-------------------------------------------------------|
| xxx | 0, NONE, NONE SEEN, NEGATIVE | TRACE | 1+, +1, +, FEW | 2+, +2, ++, SOME | 3+, +3, +++, MANY, POSITIVE | 4+, +4, ++++, TNTC, TOO NUMEROUS TO COUNT |

Table 5. Sample Urine Numeric Conversion Chart

This type of chart should be constructed with medical personnel who are familiar with urine results and the methods of analysis used by the sponsor on the results.

Once there are numeric representations, it is easier to do change from baseline and shift table analyses with the results.

LABORATORY ABNORMLITY CRITERIA

The analysis of lab data usually involves the application of some type of abnormality criteria to the lab results in order to evaluate the safety and sometimes the efficacy of a drug. There are numerous sets of criteria, including the National Cancer Institute's *Common Terminology Criteria for Adverse Events* (NCI CTCAE), the FDA's *Toxicity Grading Scale for Healthy Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials*, the Drug-Induced Liver Injury Network (DILIN) *Severity Grading In Drug Induced Liver Injury (DILI)*, and the NIH's *Division Of Aids Table For Grading The Severity Of Adult And Pediatric Adverse Events*. There are many more published criteria and modified versions of the published criteria for disease specific expected lab values. Sponsors may also have

their own set of abnormality criteria. Finally there is simply the evaluation of results being outside the reference ranges (Out-of-Normal Range – ONR).

If using one of the published Adverse Event criteria, some detective work is required as the criteria covers more than lab abnormalities and the lab criteria may be buried. Certain conditions may also involve multiple lab analytes that would lead to a toxicity grading greater than zero. One example seen below in Display 5 is the Chronic Kidney Disease that looks at GFR (estimated Glomerular Filtration Rate), creatinine clearance and urinary protein. Other criteria require a knowledge of the equipment used to analyze the specimen as seen in the NCI CTCAE criteria for Cardiac Troponin I (see Display 4 below.).

| | | | | | |
|------------------------------|--------------------------------------------------------------------------------------------------------------------|---|-----------------------------------------------------------------------------|---|---|
| Cardiac troponin I increased | Levels above the upper limit of normal and below the level of myocardial infarction as defined by the manufacturer | - | Levels consistent with myocardial infarction as defined by the manufacturer | - | - |
|------------------------------|--------------------------------------------------------------------------------------------------------------------|---|-----------------------------------------------------------------------------|---|---|

Definition: A laboratory test result which indicates increased levels of cardiac troponin I in a biological specimen.

Display 4. NCI CTCAE Criteria for Cardiac Troponin I

| Renal and urinary disorders | | | | | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-------|
| Adverse Event | Grade | | | | |
| | 1 | 2 | 3 | 4 | 5 |
| Acute kidney injury | Creatinine level increase of >0.3 mg/dL; creatinine 1.5 - 2.0 x above baseline | Creatinine 2 - 3 x above baseline | Creatinine >3 x baseline or >4.0 mg/dL; hospitalization indicated | Life-threatening consequences; dialysis indicated | Death |
| Definition: A disorder characterized by the acute loss of renal function and is traditionally classified as pre-renal (low blood flow into kidney), renal (kidney damage) and post-renal causes (ureteral or bladder outflow obstruction). | | | | | |
| Bladder perforation | - | Extraperitoneal perforation, indwelling catheter indicated | Intraperitoneal perforation; elective radiologic, endoscopic or operative intervention indicated | Life-threatening consequences; organ failure; urgent operative intervention indicated | Death |
| Definition: A disorder characterized by a rupture in the bladder wall. | | | | | |
| Bladder spasm | Intervention not indicated | Antispasmodics indicated | Hospitalization indicated | - | - |
| Definition: A disorder characterized by a sudden and involuntary contraction of the bladder wall. | | | | | |
| Chronic kidney disease | eGFR (estimated Glomerular Filtration Rate) or CrCl (creatinine clearance) <LLN - 80 ml/min/1.73 m ² or proteinuria 2+ present; urine protein/creatinine >0.5 | eGFR or CrCl 59 - 30 ml/min/1.73 m ² | eGFR or CrCl 29 - 15 ml/min/1.73 m ² | eGFR or CrCl <15 ml/min/1.73 m ² ; dialysis or renal transplant indicated | Death |
| Definition: A disorder characterized by gradual and usually permanent loss of kidney function resulting in renal failure. | | | | | |

Display 5. CTCAE Criteria

Another issue is that some analytes have bi-directional criteria (hyper-/hypo-). One example in the NCI CTCAE is blood glucose criteria.

| NCI CTCAE Criteria for Glucose (version 4.03) | | | | | | |
|-----------------------------------------------|--------------|------------------------|------------------------|----------------------|-----------------|---------|
| Directionality | Grade 0 | Grade 1 | Grade 2 | Grade 3 | Grade 4 | Grade 5 |
| Hyperglycemia | value <= ULN | value >ULN - 160 mg/dL | value >160 - 250 mg/dL | >250 - 500 mg/dL | >500 mg/dL | Death |
| Hypoglycemia | value >= LLN | value <LLN - 55 mg/dL | value <55 - 40 mg/dL | value <40 - 30 mg/dL | value <30 mg/dL | Death |

Table 6. Table of NCI CTCAE Glucose Criteria

There are two basic ways of dealing with bi-directional criteria. The first is to indicate the direction of the abnormality (Low/Hypo or High/Hyper) in the grading. A second way is to create new test codes that indicates the directionality (e.g. GLUC-L and GLUC-H). Unfortunately, neither of these methods conform with SDTM requirements. Per the SDTM Implementation Guide version 3.1.3, LBTOXGR, which records the toxicity grade if one exists, is only supposed to contain the numeric value if the scale is numeric, as is the case in NCI CTCAE. New test codes for glucose are not acceptable as there is already a code defined (GLUC).

The reason this is important is for analysis when evaluating shifts in the toxicities. When going from a hypoglycemia of grade 1 to a hyperglycemia of grade 2, it is a 3 step change when using directionality in the toxicity grading and a two-step change if using directionality in the test code (hyperglycemia goes from a baseline of 0 to a grade 2). A final resolution to this issue should be defined by the sponsor's statistical analysis plan (SAP) and implemented in the analysis data set.

BASIC CDISC SDTM LAB STRUCTURE

The Clinical Data Interchange Standards Consortium (CDISC) Standard Data Tabulation Model (SDTM) data standards are a recognized industry standard for storing clinical data and the FDA will soon be requiring data to be submitted in this format. Some basic mapping guidelines to supplying the data in this format are included. This information is based on the SDTM version 1.3 and Implementation Guide version 3.1.3. For more information about CDISC and the SDTM, visit the CDSIC web site at: www.cdisc.org.

IDENTIFIERS

The first set of variables in the LB domain are Identifiers. These are identified in the table below. The **Variable Name** is the required SAS variable name in the SDTM LB domain and the **Variable Label** is the required SAS variable label. Note that variable names in SDTM are limited to 8 characters and labels to 40 characters to comply with SAS V5 Transport format. Character variables are limited to 200 characters, but should only be the maximum length needed for the data in a submission. The **Core** column indicates which variables are required ("Req"), meaning they must be in the data set and cannot be missing, expected ("Exp"), meaning the variable must be in the data set, but can contain missing values and permissible ("Perm"), meaning they are optional in the data set and may contain missing values. If a control terminology, codelist or format is associated with any variable it will be indicated in the **Controlled Terms, Codelist or Format** column. There is a specified order of variables in an SDTM data set and the LB domain order is indicated in the **Variable Order** column.

| Variable Order | Variable Name | Variable Label | Type | Controlled Terms, Codelist or Format | Core |
|----------------|---------------|----------------------------|------|--------------------------------------|------|
| 1 | STUDYID | Study Identifier | Char | | Req |
| 2 | DOMAIN | Domain Abbreviation | Char | LB | Req |
| 3 | USUBJID | Unique Subject Identifier | Char | | Req |
| 4 | LBSEQ | Sequence Number | Num | | Req |
| 5 | LBGRPID | Group ID | Char | | Perm |
| 6 | LBREFID | Specimen ID | Char | | Perm |
| 7 | LBSPID | Sponsor-Defined Identifier | Char | | Perm |

Table 7. LB Domain Identifiers

STUDYID and USUBJID are sponsor defined variables to identify the sponsor's study and uniquely identify a subject across a submission (if a subject is in more than one study in a submission, the USUBJID should be the same for that subject across all studies). DOMAIN is set "LB" and has a length of 2 characters. LBSEQ is a unique record identifier within a subject's set of observations within the LB domain. LBGRPID is used to group records within a subject and should not have any significance across subjects (unlike LBCAT and LBSCAT). For example, a value of "MICRO" could be used to identify additional records for urinalysis results when a dipstick result was abnormal, requiring a manual, microscopic examination of the specimen. In this case, LBGRPID would be included as a "key" variable. LBREFID would contain an identifier used by the lab to uniquely identify and track a specimen sample (e.g. accession number on a test tube) . LBSPID is another tracking number, usually defined by the sponsor's data collection system.

TOPIC AND QUALIFYING VARIABLES

The topic and qualifying variable within the LB domain are identified in the table below.

| Variable Order | Variable Name | Variable Label | Type | Controlled Terms, Codelist or Format | Core |
|----------------|---------------|------------------------------------|------|--------------------------------------|------|
| 8 | LBTESTCD | Lab Test or Examination Short Name | Char | (LBTESTCD) | Req |
| 9 | LBTEST | Lab Test or Examination Name | Char | (LBTEST) | Req |
| 10 | LBCAT | Category for Lab Test | Char | * | Exp |

| Variable Order | Variable Name | Variable Label | Type | Controlled Terms, Codelist or Format | Core |
|----------------|---------------|------------------------------------------|------|--------------------------------------|------|
| 11 | LBSCAT | Subcategory for Lab Test | Char | * | Perm |
| 12 | LBORRES | Result or Finding in Original Units | Char | | Exp |
| 13 | LBORRESU | Original Units | Char | (UNIT) | Exp |
| 14 | LBORNRL0 | Reference Range Lower Limit in Orig Unit | Char | | Exp |
| 15 | LBORNRLHI | Reference Range Upper Limit in Orig Unit | Char | | Exp |
| 16 | LBSTRESC | Character Result/Finding in Std Format | Char | | Exp |
| 17 | LBSTRESN | Numeric Result/Finding in Standard Units | Num | | Exp |
| 18 | LBSTRESU | Standard Units | Char | (UNIT) | Exp |
| 19 | LBSTNRLO | Reference Range Lower Limit-Std Units | Num | | Exp |
| 20 | LBSTNRHI | Reference Range Upper Limit-Std Units | Num | | Exp |
| 21 | LBSTNRC | Reference Range for Char Rslt-Std Units | Char | | Perm |
| 22 | LBNRIND | Reference Range Indicator | Char | * | Exp |
| 23 | LBSTAT | Completion Status | Char | (ND) | Perm |
| 24 | LBREASND | Reason Test Not Done | Char | | Perm |
| 25 | LBNAM | Vendor Name | Char | | Perm |
| 26 | LBLOINC | LOINC Code | Char | * | Perm |
| 27 | LBSPEC | Specimen Type | Char | * | Perm |
| 28 | LBSPECND | Specimen Condition | Char | * | Perm |
| 29 | LBMETHOD | Method of Test or Examination | Char | * | Perm |
| 30 | LBBLFL | Baseline Flag | Char | (NY) | Exp |
| 31 | LBFAST | Fasting Status | Char | (NY) | Perm |
| 32 | LBDRVFL | Derived Flag | Char | (NY) | Perm |
| 33 | LBTOX | Toxicity | Char | * | Perm |
| 34 | LBTOXGR | Standard Toxicity Grade | Char | * | Perm |

Table 2. LB Domain Topic and Qualifying Variables

LBTESTCD is the topic variable and identifies the analyte being reported. The length of this variable is limited to 8 characters and has an extendible CDISC control terminology. Note that the lab test code may not be unique within a subject's visit. Protein ("PROT") is tested in both blood chemistry and urine. LBSPEC is needed to differentiate between the two. In the example above about microscopic results being included in addition to dipstick, LBMETHOD would be needed to differentiate these two sets of results. Both LBSPEC and LBMETHOD should have a sponsor defined control terminology used consistently across a submission). LBTEST is a synonym qualifier for LBTESTCD and is a decode or label LBTESTCD. The maximum length of LBTEST is 40 characters and also has an extendible CDISC control terminology. If a sponsor chooses to use LOINC codes to identify lab tests, the codes are stored in the LBLOINC variable. Lab test results can be grouped using the LBCAT and LBSCAT variables. LBCAT is an expected variable and usually contains information about the overall panel used for testing (e.g. Hematology, Chemistry, Urinalysis, etc.). LBSCAT is a permissible variable and contains information on the sub-panels (e.g. CBC, Proteins, etc.).

The test results and reference ranges are reported in two sets (original units and standard units) and in two formats (character and in standard units – numeric). The original results reported from the lab are stored as character in LBORRES and the reference ranges are stored in LBORNRL0 and LBORNRLHI. The original units are stored in LBORRESU. Once the original value is converted to the sponsor's standard reporting units, the character version of the results are stored in LBSTRESC and a numeric version, if applicable, is stored in LBSTRESN. The numeric version of the reference ranges in standard units are stored in LBSTNRLO and LBSTNRHI. Non-numeric reference ranges are stored in LBSTNRC. This is commonly used for urinalysis results (e.g. NEGATIVE TO TRACE). There is CDISC control terminology for the two units fields that is extendible. LBNRIND is used to indicate where the result is in relation to the reference ranges. It is suggested that if the lab doesn't provide this analysis that the programming is

done using the original result value and reference ranges as rounding and hardware implementations may adversely affect the analysis.

LBSTAT and LBREASND are used to indicate when a test was not performed and why. These values can apply to an entire panel (e.g. 'TUBE BROKEN IN SHIPMENT') or to a subset of tests (e.g. 'QUANTITY OF SAMPLE INSUFFICIENT'). LBSPCCND also describes an entire panel and is used to report the condition of the specimen (e.g. 'HEMOLYZED'). LBFASST is used to flag whether the subject was fasting at the time the specimen was collected. LBNAM is used when the sponsor wants to indicate the name of the lab that performed the analysis. This becomes important in cases where two sets of specimens are collected with one set being performed at a local lab and the second at a central lab. In this case, LBNAM would also be a "key" variable.

The LB domain defines a number of analysis variables that may be included with the SDTM version of submission data sets as opposed to the analysis (ADaM) data sets. LBLFL is used to flag which record is the sponsor-defined baseline record for an analyte. LBTOXGR is used to indicate the toxicity grade for a result. If the scale is numeric (e.g. Grades 1 to 5) then the value of LBTOXGR should be the numeric value. LBTOX is used to indicate the scale and version used to determine the value of LBTOXGR (e.g. "NCI CTCAE v. 4.03"). If an entire observation is derived by the sponsor (for example, a relative value is calculated using an absolute differential and WBC count), the new record should be flagged as being derived using the LBDRVFL variable.

TIMING VARIABLES

The date/time and visit variables are identified in the table below.

| Variable Order | Variable Name | Variable Label | Type | Controlled Terms, Codelist or Format | Core |
|----------------|---------------|------------------------------------------|------|--------------------------------------|------|
| 35 | VISITNUM | Visit Number | Num | | Exp |
| 36 | VISIT | Visit Name | Char | | Perm |
| 37 | VISITDY | Planned Study Day of Visit | Num | | Perm |
| 38 | LBDMTC | Date/Time of Specimen Collection | Char | ISO 8601 | Exp |
| 39 | LBENDTC | End Date/Time of Specimen Collection | Char | ISO 8601 | Perm |
| 40 | LBDY | Study Day of Specimen Collection | Num | | Perm |
| 41 | LBTPPT | Planned Time Point Name | Char | | Perm |
| 42 | LBTPPTNUM | Planned Time Point Number | Num | | Perm |
| 43 | LBELTM | Planned Elapsed Time from Time Point Ref | Char | ISO 8601 | Perm |
| 44 | LBTPPTREF | Time Point Reference | Char | | Perm |
| 45 | LBRFTDTC | Date/Time of Reference Time Point | Char | ISO 8601 | Perm |

Table 3. LB Domain Timing Variables

At a minimum the date (unless a specimen was not analyzed) and/or a visit number should be supplied using LBDMTC and VISITNUM. A character description of the visit can be supplied using VISIT. If a specimen is collected across multiple days (e.g. 24 hour urine sample) then the start date/time is recorded in LBDMTC and the ending date/time is recorded in LBENDTC. The planned study day (being the number of days since the reference start date indicated in the DM domain plus 1) can be stored in the VISITDY variable. The actual study day can be recorded in LBDY. If multiple specimens are collected in relation to a specific event (e.g. time since dose last dose) LBTPPT can be used to describe the interval (e.g. '4 HOURS POST DOSE'), a numeric value used for sorting can be stored in LBTPPTNUM, the description of the relative time point (e.g. 'LAST DOSE') can be stored in LBTPPTREF and the actual date/time of the relative time point can be stored in LBRFTDTC.

CREATING AND MAINTAINING STANDARDS

Maintaining test codes, determining standard units and maintaining conversion factors are three areas that can lead to increases in efficiency and enhancing the validity of data programming. If these three items are standardized significant amounts of programming can be written and reused across multiple studies. Validation can be done once at a higher, generic level using more rigorous techniques that result in more accurate and robust code. The higher the level of standardization (project level, therapeutic area level, company-wide level), the greater the savings.

At a minimum, standards need to be set at the project level, as test codes, units and conversion factors need to be maintained across all studies in a submission. If standards are set at a therapeutic area or company level, representatives from multiple projects should be included when creating and maintaining the standards. The team

should consist of someone with a medical background (to help the team understand the significance of the lab tests and what makes medical sense), a statistician (to help the team understand the analysis and presentation needs as well as to help validate conversion factors), someone from the data collection area (to help identify new tests and new units) and someone from the programming area (to help the team understand the data processing needs and to facilitate the implementation of the standards). If an “officially” sanctioned team can’t be created, it is still possible to recruit a group of like-minded individuals to meet on an ad hoc basis to discuss new tests and/or units and maintain standards in a non-formal manner.

If possible, these standards can be programmed into a clinical database management system (DBMS) and automatically update the lab results with the new test codes and/or standard results. Where not possible, spreadsheets can be maintained and read by SAS programs when creating the SDTM and analysis data sets. In keeping with the SDTM normalized structure, two spreadsheets are required, one for keeping track of test codes and standardized units and another for conversion factors of actual units found in the data. The first spreadsheet can keep track of additional items like groups, sub-groups, print order, number of significant decimal places, etc. and the second could contain validation information by keeping the date and name of the person who entered the new units and conversion factors and the date and name of the person who reviewed and/or validated the conversion factors.

Another task for the standards team would be to interface with the CDISC Terminology Team. The team would be the official point of contact for submitting new terms and justifying them when needed. They would also review the terminology quarterly updates and identify changes or additions that affect lab processing – disseminating the information to the appropriate functional areas.

CONCLUSION

Lab data can be a mystery, but by coming to understand the intricacies and areas of concern when working with lab data, anyone programming or analyzing lab data can produce better code and more accurate results. With further study and usage, a deeper understanding, resulting in better and more accurate code is possible.

REFERENCES

- “CDSIC Standards and Implementation.” 2014. Available at <http://www.cdisc.org/standards-and-implementations>.
- “CDISC Terminology.” NCI Enterprise Vocabulary Services (EVS), Dictionaries, and Partnerships. Available at <http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc>.
- “ARUP Laboratories: A National Reference Laboratory.” 2013. Available at <http://www.aruplab.com>.
- “Cleveland Clinic Laboratories.” 2010. Available at <http://portals.clevelandclinic.org/reflab/Home/tabid/4036/Default.aspx>.
- “Common Laboratory (LAB) Values.” 1993-2014. Available at http://www.globalrph.com/labs_home.htm.
- “A.D.A.M. Medical Encyclopedia.” 9-Jan-2014. Available at <http://www.nlm.nih.gov/medlineplus/encyclopedia.html>.
- “Normal Laboratory Values.” 2010-2013. Available at www.merckmanuals.com/professional/appendixes/normal_laboratory_values/normal_laboratory_values.html
- Young, Donald S. and Edward J. Huth. 1998. “*SI Units for Clinical Measurements*.” Philadelphia, PA : American College of Physicians.
- Jones, Graham, and Antony Barker. “Reference Intervals.” Aug, 2008. Available at www.ncbi.nlm.nih.gov/pmc/articles/PMC2556592/.
- Barber, Brenda M. 1997. “Overcoming Kainophobia : Replacing Complex Merges with PROC SQL”. *Proceedings of the 22nd SAS® Users Group International*. Available at <http://www2.sas.com/proceedings/sugi22/POSTERS/PAPER198.PDF>
- “Potassium conversion.” OnlineConversion.com Forums. 2007. Available at <http://forum.onlineconversion.com/showthread.php?t=9032>.
- “Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03”. National Cancer Institute. June 14, 2010. Available at http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf
- “Guidance for Industry: Toxicity Grading Scale for Healthy Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials”. FDA. September, 2007. Available at <http://www.fda.gov/biologicsbloodvaccines/guidancecomplianceregulatoryinformation/guidances/vaccines/ucm074775.htm#TOXICITYGRADINGSCALETABLES>
- “Severity Grading In Drug Induced Liver Injury”. Drug-Induced Liver Injury Network (DILIN). 3-Mar-2014. Available at <http://livertox.nih.gov/Severity.html>
- “Division Of Aids Table For Grading The Severity Of Adult And Pediatric Adverse Events”. NIH Division of AIDS. December, 2004”, Available at

<http://www.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/Documents/daidsaegradingtable.pdf>

ACKNOWLEDGMENTS

I'd like to acknowledge and thank those individuals who reviewed this paper and supplied technical consultation.

RECOMMENDED READING

- CDISC SDTM version 1.3 .
- CDISC SDTM Implementation Guide version 3.1.3.

CONTACT INFORMATION

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

| | |
|------------------|------------------------|
| Name: | Alan Meier |
| Enterprise: | MedImmune LLC |
| Address: | One MedImmune Way |
| City, State ZIP: | Gaithersburg, MD 20878 |
| Work Phone: | 301-398-4613 |
| Fax: | 301-398-9613 |
| E-mail: | meiera@medimmune.com |

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.