PharmaSUG 2017 - Paper AD10

Laboratory Data Standardization with SAS

Charley Z. Wu, Dona-Lyn Wales

1 ABSTRACT

In clinical trials, laboratory data usually come from central laboratories and/or many different local laboratories. Different laboratories usually have different original reported laboratory test names, units and reference ranges. In order to pool and analyze these reported data efficiently and correctly, central/local laboratory test names, units and results must be converted to a standard format. A SAS-based laboratory standardization application was built in order to standardize original reported laboratory data to SDTM standards and associated controlled terminology. This application will harmonize original reported central/local laboratory data by mapping original reported lab test names and units to SDTM standard test names and units, assigning standard units to lab tests, maintaining conversion factor tables for converting original reported units and numeric results to standard units and results, and performing the conversion of original reported numeric results and reference ranges to standard numeric results and reference ranges. This application will also standardize and maintain original reported character lab results and reference ranges. Finally, a set of lab reports are created based on the standardized data to facilitate lab data review and cleaning.

2 INTRODUCTION

Lab data collection, cleaning, and analysis are one of the most important aspects in clinical studies. As in many studies, efficacy data and safety data can come directly from lab data.

In most clinical studies, there are hundreds of lab tests from different specimen types such as blood, urine, biopsy tissue, central nervous system fluids, etc. Analyses on these specimens may be performed in one or more central or local laboratory. In the end, a study may have thousands of original reported lab data records from different laboratories.

Different central/local laboratories will usually report lab tests in their own original preferred units and ranges. Furthermore, they may also report the same test name, specimen, and method using their original preferred nomenclature. For example, the following three original reported test names from different laboratories all refer to the same lab test: HBsAb, Hepatitis B Virus Surface Antibody, and Anti-HBs Antibody.

In order to analyze data efficiently and correctly, original reported lab test names, categories, specimens, methods, units, results, and ranges must be standardized. That is also true for pooling data from different studies within a project for analysis.

Additionally, original reported laboratory data must be standardized for regulatory submissions.

The lab standardization application will take original reported laboratory data and standardize it to SDTM requirements for:

1) Lab test name, 2) Lab test unit, 3) Lab test range, 4) Numeric lab test result, 5) Character lab test result 6) Lab category, subcategory, specimen, and method

3 ENTITY-RELATION MODEL (ER MODEL) OF THE STANDARDIZATION SYSTEM

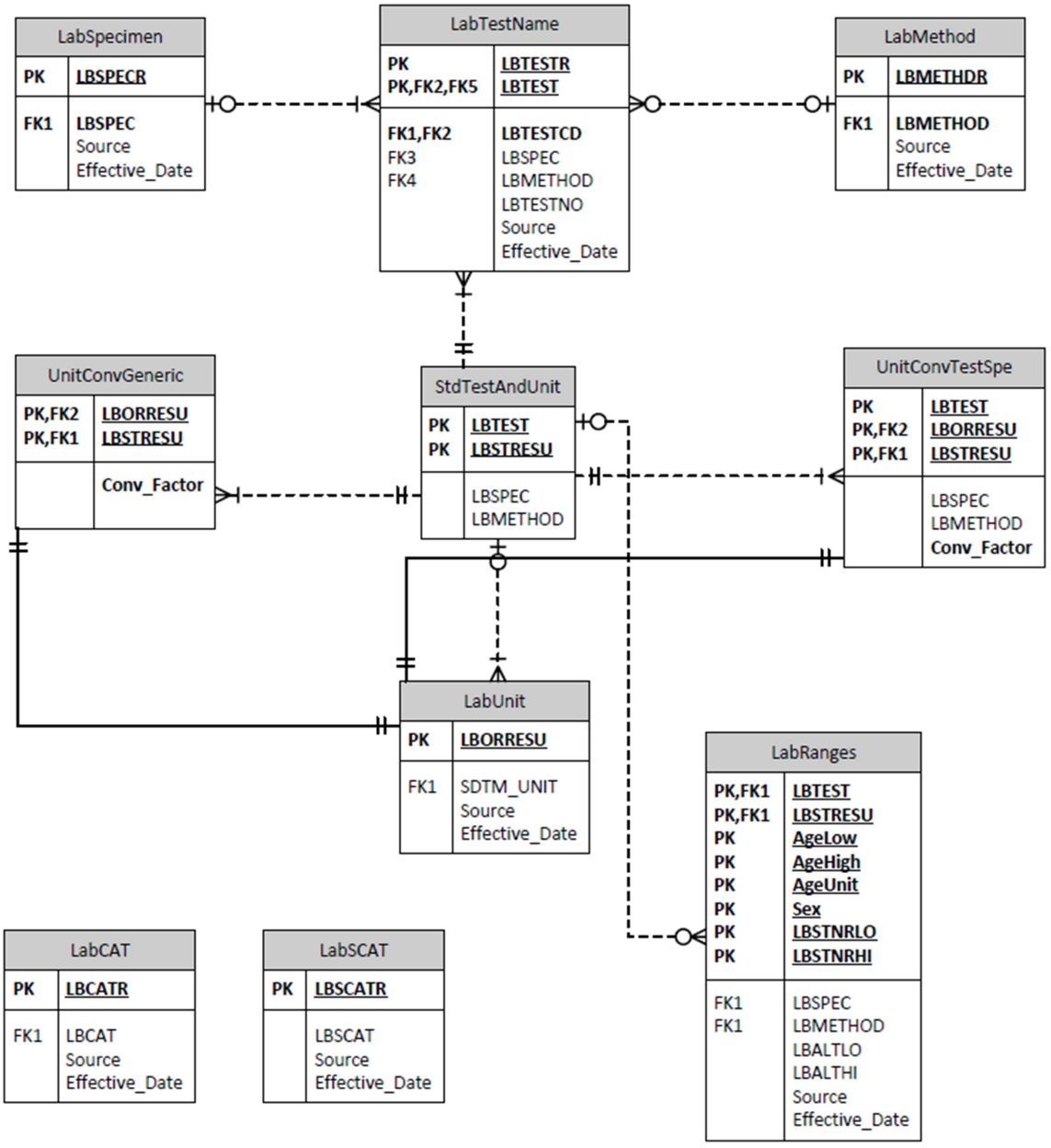


Figure 1. ER Model of the lab standardization system

3.1 LabTestName – Lab Test Name

This table will map the original reported lab test names to SDTM standard lab test names and test code. LBTESTR is the original reported lab test name from the local/central laboratory. LBTEST is the SDTM standard laboratory test name from the codelist, LBTEST. For example, if LBTESTR is **'WBC'**, then LBTEST will be mapped to **'Leukocytes'** according to the SDTM standard.

Variable	Label	Core	Example
LBTESTR	Lab Test Name, Reported	Required	WBC
LBTEST	SDTM Lab Test Name	Required	Leukocytes
LBTESTCD	SDTM Lab Test Short Name	Required	WBC
LBSPEC	SDTM Specimen Type	Expected	Serum
LBMETHOD	SDTM Lab Test Method	Optional	Auto count

For reasons unknown, SDTM does not distinguish some lab tests. For example, SDTM has a lab test name 'pH'. SDTM does not tell whether it is 'Serum pH' or 'Urine pH', which has different normal ranges and medical significances. Therefore, labs cannot use this test name directly. In this case, labs have to use LBSPEC (Specimen Type) and/or LBMEHTOD (Method of Test or Examination) to distinguish them during mapping. However, in most cases, LBSPEC and/or LBMETOD are optional.

3.2 LabSpecimen, LabMethod – Lab Specimen, Lab Method

These tables will map original reported lab specimens and methods to SDTM standard lab specimens and methods. LBSPECR and LBMETHDR are the original reported specimen and method from the local/central laboratory. LBSPEC and LBMETHOD are the SDTM standard specimen and method from code lists, SPECTYPE and METHOD respectively.

For example, if LBSPECR is **'Venous Blood Sample'**, then LBSPECR will be mapped to **'Venous Blood'** according to the SDTM standard. Likewise, if LBMETHDR is **'Enzyme Immunoassay (EIA)'**, then LBMETHDR will be mapped to **'EIA'** according to the SDTM standard.

Variable	Label	Core	Example
LBSPECR	Specimen Type , Reported	Expected	Venous Blood Sample
LBSPEC	SDTM Specimen Type	Expected	Venous Blood
LBMETHDR	Lab Test Method, Reported	Optional	Enzyme Immunoassay
LBMETHOD	SDTM Lab Test Method	Optional	EIA

3.3 LabUnit – Lab Test Units

This table will map original reported lab units to SDTM standard units in codelist, UNIT.

Variable	Label	Core	Example
LBORRESU	Original Lab Test Unit	Required	x10E9/L
SDTM_UNIT	SDTM Unit	Required	10^9/L

Many laboratories report their units following their own naming conventions. These units are not necessarily SDTM standard terminology. For example, some laboratories report **'x10E9/L'** as the unit for blood Leukocytes. But based on SDTM standards, the unit should be **'10^9/L'**. Therefore, the original reported unit must be mapped to the SDTM standard.

Lab tests with character results usually do not have units. For example, qualitative urine protein usually has results such as '**negative, trace, +, ++, +++, ++++**'. In this case, **'Not Applicable'** is assigned as the test unit to be consistent that all tests have a reported unit. Some companies prefer to leave the units blank. But that will have some cleaning complications as it is hard to tell whether a null unit means that a unit is not appropriate for the test or that the lab forgot to enter the unit.

Some lab tests with numeric results may also not have units. For example, pH and urine specific gravity do not have units. In this case, **'Not Applicable'** is assigned as the test unit.

3.4 StdTestAndUnit – Standard Lab Test Name and Standard Unit

This table defines the adopted company-wide standard test names and associated SI standard units.

Variable	Label	Core	Example
LBTEST	SDTM Lab Test Name	Required	Hemoglobin
LBSTRESU	SDTM SI Standard Unit	Required	g/dL
LBSPEC	SDTM Specimen Type	Expected	Serum
LBMETHOD	SDTM Lab Test Method	Optional	

Once the original reported lab test name and unit are mapped to the SDTM standard, the standard SI unit for a lab test can be defined. STDM does not have a guideline for choosing a standard SI unit, so the company will adopt a standard SI unit using the following criteria:

1) The company identifies references to use for establishing standard units
2) Prevailing use of the unit by laboratories
3) The standard SI unit is in SDTM controlled terminology
4) The lab results won't be two large or too small. For example, a normal range for hemoglobin is 13.5-17.5 g/dL. 'g/dL' will be adopted as the standard unit. 'g/L' or 'mg/L' will not be selected as its standard unit
5) Specimen type is usually required. The same lab test with different specimen types may have different standard unit
6) Lab test method is an optional field

3.5 UnitConvGeneric: Generic Unit Conversion

This table records all the unit conversion factors that apply to all lab tests regardless of test name, specimen, and/or method. So, the conversion factors in this table are generic and independent of test name/specimen/method.

Variable	Label	Example
SDTM_UNIT	SDTM Unit	10^9/L
LBSTRESU	SDTM SI Standard Unit	10^6/L
CONV_FACTOR	Conversion factor	1000

For example, converting an original reported unit from 10^9/L to the standard SI unit 10^6/L, the conversion factor is 1000 regardless whether the lab test is WBC or RBC or platelets.

Many companies tend to create the conversion table by lab test. Although that approach is straight forward and will work, it is just not efficient. For the example above, it would be necessary to maintain a conversion record for every lab test (i.e. RBC, WBC, platelet, etc.) that had the same conversion factor for converting 10^9/L to 10^6/L. That will make the conversion table unnecessarily big and very inefficient to maintain. Therefore, it is more efficient to break the conversion table into two separate tables where one table is the generic table just discussed, and the other table is the lab test-specific conversion table below.

3.6 UnitConvTestSpe: Test Specific Unit Conversion

This table records all the unit conversion factors that are test-specific. Lab test names matter here. This typically happens when the unit is based on the molecular composition of the lab test, such as mole to gram conversion. For example, converting blood Glucose from the original reported unit of g/L to the standard SI unit mmol/L, the conversion factor is 5.551 while converting BUN from g/L to mmol/L, the conversion factor is 35.7.

LBSPEC and LBMETHOD are optional. They are required only when there is a need to distinguish lab tests.

Variable	Label	Example
LBTEST	Lab Test Name	Glucose
SDTM_UNIT	SDTM Unit	g/L
LBSTRESU	SDTM SI Standard Unit	mmol/L
LBSPEC	Specimen Type	
LBMETHOD	Lab Test Method	
CONV_FACTOR	Conversion factor	5.551

3.7 LabTestRange: Company-wide Generic Lab Test Ranges

Central laboratories and most local laboratories can usually provide normal lab ranges for their lab tests. However, it can happen that many local labs will omit lab ranges for some lab tests despite attempts to query the laboratories for this information.

For example, some local laboratories cannot provide lab ranges for "Eosinophils (Differential, percent)"but this is a common test that does have normal ranges.

Medical reviewers need normal ranges to determine whether a lab result is normal/abnormal, or medically significant. Statisticians also need normal ranges to finish their safety analysis and CSR reports. To address this issue, a company-wide lab range table for standard units is defined. So, if a laboratory omits some ranges, the generic lab ranges can be used and applied to the lab data.

The table below describes the structure of the company-wide lab range table for all lab tests, and by age and gender when appropriate. These ranges are identified from standard references adopted by the company, and are reviewed and approved by medical doctors.

For some new or proprietary biomarker or PD tests, the ranges from the laboratory that performs the tests are used.

Variable	Label	Example
LBTEST	Lab Test Name	ALT
LBSTRESU	SDTM SI Standard Unit	U/L
AgeLow	Starting Age	0
AgeHigh	Ending Age	2
AgeUnit	Age Unit	Years
Sex	Sex	M
LBORNRLO	Reference Range Lower Limit in Orig Unit	6
LBORNRHI	Reference Range Upper Limit in Orig Unit	45

4 PROCESS

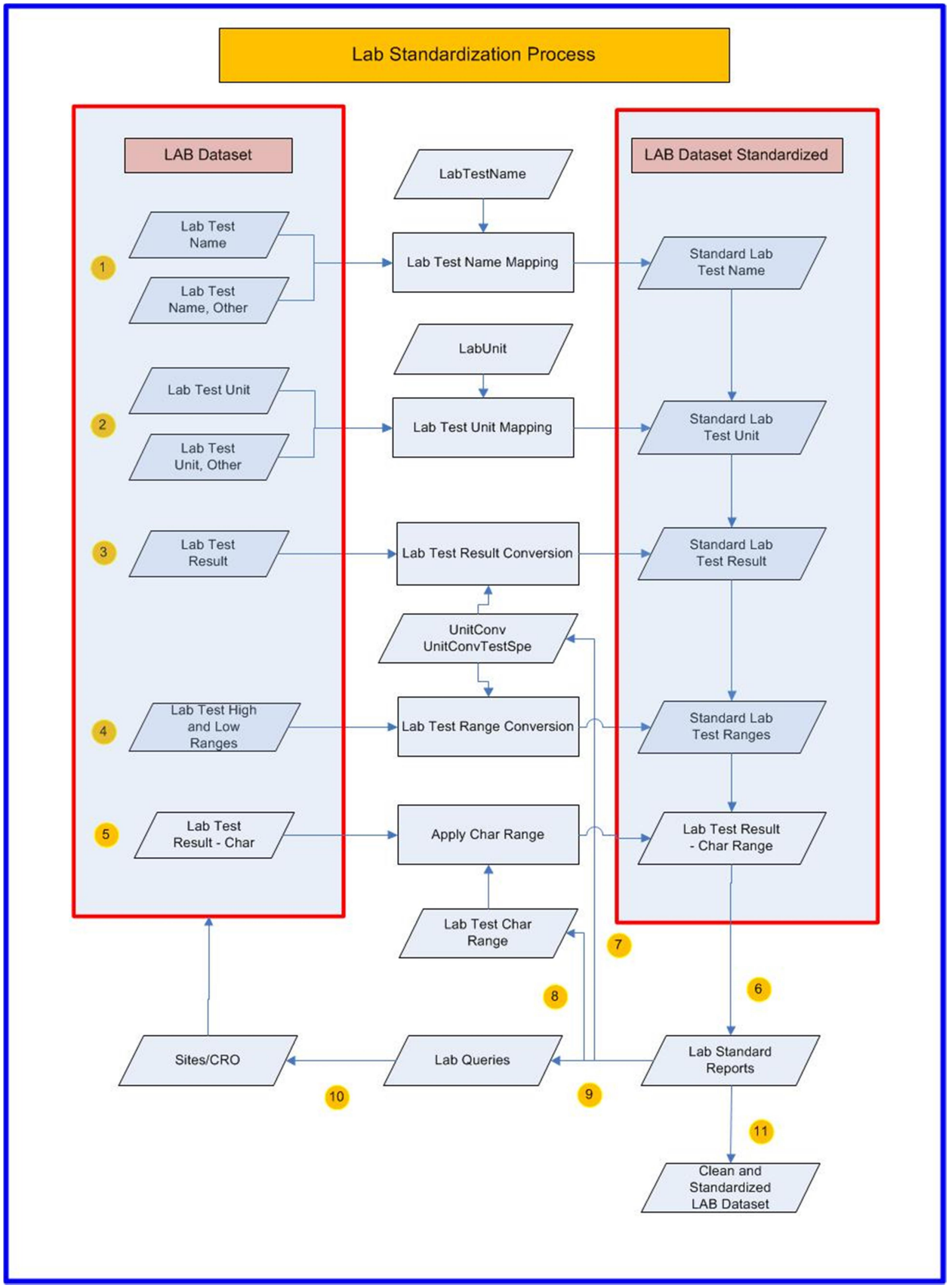
Figure 2 below outlines the laboratory standardization process. It consists of 5 sub-processes: (1) lab test name standardization, (2) unit standardization, (3) numeric result, (4) numeric range standardization, and (5) character result/range standardization.

SAS programs were developed to create and populate the tables in **Figure 1**. SAS macros were also developed to perform the 5 sub-processes in **Figure 2**. The programming follows a modular-based approach. Each module only completes one of the sub-processes. There are also SAS modules that create the standard reports for data review. All modules work together to complete the standardization of the data

4.1 Lab Test Name Mapping (Figure 2 Step 1)

As discussed before, test names from local/central laboratories do not necessarily follow SDTM standards. Figure 2 step 1 shows how to map these non-standard original reported test names to the SDTM standard using the **LabTestName** table. Examples are in the table below (Not all columns are shown)

Laboratory Data Standardization with SAS



Row	LBTESTR	LBTESTRO	LBTEST	LBSPEC
1	AST		Aspartate Aminotransferase	
2	SGOT		Aspartate Aminotransferase	
3	K		Potassium	
4	RBC		Erythrocytes	Serum
5	RBC – Urine		Erythrocytes	Urine
6	Other	SSC	S-Sulfocysteine	

Row 1 shows that **AST** is mapped to SDTM **Aspartate Aminotransferase**

Row 2 shows that **SGOT** is mapped to SDTM **Aspartate Aminotransferase** as well

Row 3 shows that **K** is mapped to SDTM **Potassium**.

Row 4 and 5 show **RBC** and **RBC – Urine** are mapped to the same SDTM **Erythrocytes**. But they differentiate themselves by specimen type, LBSPEC. Please note, SDTM LBTEST is limited to 40 characters.

Row 6: Some CRFs allow sites to enter additional lab tests such as SSC. It will also be mapped to a standard lab test (S-Sulfocysteine).

If a lab test is neither in SDTM controlled terminology nor in the **LabTestName** table, that means this is a new lab test. As LBTEST is an extensible codelist, the new lab test can be added to this table manually. Then a lab specialist will review the test name and also assign LBTEST/LBTESTCD following SDTM guidelines.

4.2 Mapping of Lab Specimen and Method (Not shown in Figure 2)

Mapping of lab specimen and method to the SDTM standard is the same as mapping of lab test name but using the tables **LabSpecimen** and **LabMethod.**

4.3 Lab Unit Mapping (Figure 2 Step 2)

Figure 2 step 2 shows how to map original reported lab units (or Lab Reported Unit, Other) to the SDTM standard unit using the **LabUnit** table. Examples are in the table below.

Again, as it is impossible to pre-print all units into the CRFs, sites sometimes will need to enter units in the "Unit, Other field", these units need to be reviewed and mapped to SDTM standard units as well (Row 5 below).

Row	LBORRESU (Original Reported Lab Unit)	LBORREUO (Reported Lab Unit, Other)	SDTM_UNIT (SDTM Unit)
1	10*12/L		10^12/L
2	10E12/L		10^12/L
3	µMOL/L		umol/L
4	Giga per Liter		10^9/L
5	Other	milligram/100 mL	mg/dL

4.4 Lab Test and Standard SI Unit Mapping (Not shown in Figure 2)

Once an original reported lab test and its unit are mapped to the SDTM controlled terminology, a standard SI unit can be assigned to the lab test and stored in the **StdTestAndUnit** table. Examples are in the table below. Please note, pH does not have a unit, so 'Not Applicable' is assigned to it.

LBTEST (SDTM Lab Test Name)	LBSPEC (SDTM Specimen)	LBSTRESU (SDTM SI Standard Unit)
Albumin	Serum	g/L
Bilirubin	Serum	umol/L
Erythrocytes	Blood	10^12/L
pH	Urine	Not Applicable

4.5 Numeric Data Conversion (Figure 2 Step 3)

Lab result conversion is the heart of the standardization process. Once the lab test name, original unit, standard unit, and conversion factor are available, conversion of the original result to the standard result is straight forward. Tables used for this process are **StdTestAndUnit, UnitConvGeneric,** and **UnitConvTestSpe.**

Row	LBTESTR (Original Reported Test Name)	LBTEST (SDTM Test Name)	LBORRES (Original Reported Result)	LBORRESU (Original Reported Unit)	LBSTRESN (Standard Numeric Result)	LBSTRESC (Standard Char Result	LBSTRESU (SDTM SI Standard Unit)	Conversion Factor
1	RBC	Erythrocytes	3.3	10^6/uL	3.3	3.3	10^12/L	1
2	Basophils Absolute	Basophils	>0.01	10^6/uL		>10	10^6/L	1000

3	BUN	Blood Urea Nitrogen	41	mg/dL	14.64	14.64	mmol/L	0.3571
4	Amylase	Amylase	50	mg/L			U/L	

Row 1: Conversion factor from 10^6/uL to 10^12/L is 1 from UnitConvGeneric table. Please note the conversion factor here is independent of the lab test "Erythrocytes"

Row 2: Please note the result is a character and cannot be converted to a numeric. So, after conversion, the standard numeric result is empty but the standard char result is '>10'

Row 3: Conversion factor from mg/dL to mmol/L is 0.3571 from UnitConvTestSpe table. Please note, the conversion factor here is specific to the lab test "Blood Urea Nitrogen" due to the conversion to mmol.

Row 4: Amylase cannot have a unit mg/L. So, it cannot be converted to its standard unit U/L. Therefore, this record needs query.

4.6 Reference Range Lower Limit and High Limit (Figure 2 Step 4)

Conversion of low and high ranges to standard is the same as converting numeric results. When ranges are omitted by the local laboratory, the table **LabRanges** is used to assign standard ranges.

4.7 Character Data Conversion(Figure 2 Step 5)

The system can standardize character data, for example, to convert 'Neg' to 'Negative', 'Pos' to 'Positive', 'Trc' to 'Trace'. However, as character results are rarely analyzed by statistics and are usually not as important as numeric results, standardization of character data is optional. Please note the ER model in Figure 1 does not include this mapping table.

4.8 Standard Reports (Figure 2 Step 6):

Once lab data is mapped to the SDTM standard and results are converted, many reports are developed to check the lab data quality.

1) **Lab test without standard unit report:** This report will output all lab tests without standard units. So lab the administrator can add standard units to the StdTestAndUnit table.
2) **Lab test with multiple standard units report:** This report will output any lab test that has more than one standard unit. Based on this report, the StdTestAndUnit table will be corrected to keep one and only one standard unit.
3) **Lab numeric result with no range report:** This report will output all lab tests with numeric results but without lab ranges. Numeric lab tests usually have normal ranges, so a query will be sent to the laboratory to report the normal ranges. However, if these labs are truly

not able to provide the ranges after querying, then the company-wide lab ranges in the **LabRanges** table will be applied to the data.

4) **Lab test name check report:** Check whether all lab tests follow the SDTM standard as recorded in the table **LabTestName.**
5) **Lab unit check report:** Check whether all lab units follow the SDTM standard as recorded in the table **LabUnit.**
6) **Lab specimen check report:** Check whether all specimens follow the SDTM standard as recorded in the table **LabSpecimen.**
7) **Lab method check report:** Check whether all methods follow the SDTM standard as recorded in the table **LabMethod.**
8) **Lab test result missing report:** If a lab test has no results, a query is sent to the laboratory requesting this data.
9) **Lab original unit missing report:** If a lab test does not have a reported unit, a query is sent to the lab to provide the data when required for the appropriate lab test.
10) **Lab char results review report:** Check whether all character results are logical for the given lab test.
11) **Lab test low range greater than high range report**: Report any lower ranges that are greater than higher ranges.
12) **Lab results top 5 and low 5 by SI units:** This is the most important report that identifies outlying results from thousands of lab records. Based on test name and its corresponding original units, standard units, and standard results, the 5 highest and 5 lowest values can be identified. By comparing these values to the normal ranges, and also by reviewing the differences between the highest and the lowest values, outliers can be easily identified. Some outliers may be valid abnormal values due to underlying diseases that don't require query. But some outliers are too large or too small to be valid abnormal values. These outliers are usually due to data entry errors such as entering wrong units (e.g. entering 10^6/uL for 10^9/uL, or mg for g) or confusion between qualitative and quantitative results. Therefore, outliers will need to be queried.

5. CONCLUSION

Lab data are usually complex and huge. Many efficacy and safety endpoints are buried in the lab data; therefore, high quality lab data are crucial for accurate statistical analysis.

To identify data issues within thousands of lab records is usually very challenging. Many companies depend on medical review. That is important but it is usually manual and very tedious. Medical staff usually prefer more aggregated data reports such as mean, median, standard deviation, p-value etc. But these kinds of reports usually require relative clean data in the first place.

The mapping of data to SDTM standards is done up-front during the ongoing conduct of the trial. By standardizing lab test names, units, results, and ranges, many edit checks and reports across the lab records can be performed. By reviewing these checks/reports, data issues can be identified easily and early. So, this application has greatly improved data review and cleaning

efficiency. Additionally, up-front standardization of lab data to SDTM standards eliminates the need for this effort to occur during the development of SDTM and ADaM datasets; statistical analysis can proceed without further data conversion.

In summary, the design, implementation, and use of a SAS-based laboratory standardization application has been described to take original reported central/local laboratory data and standardize it to SDTM requirements. The input data for the application is lab data from local and/or central laboratories. The output data is lab test names, units, results, and ranges that are standardized to SDTM format.

6. SAS CODE FOR THIS APPLICATION

Please contact authors for details

7. CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the authors at:

Author Name: Charley Z. Wu

E-mail: charleyzwu@yahoo.com

Author Name: Dona-Lyn Wales