PharmaSUG China 2022 - Paper AD-114

Lab Unit Conversions - Dictionary and Macro Made Easy

Michael Yang, David Fielding, Kent Letourneau, Tobias Bottor, Jessica Kueker and Amy Welsh, ICON plc

ABSTRACT

Unit conversions can be tedious and time consuming - original test names and units need to be mapped to CDISC submission values (LBTESTCD/LBTEST/LBORRESU), then conversion factors found for each test to get results in standard units (LBSTRESU/LBSTRESN/LBSTNRLO/LBSTNRHI).

In an effort to improve and simplify our process for conversion of lab data into a consistent set of standardized units, ICON has created a lab data dictionary. The dictionary is stored and maintained in a central database for version control purposes. A unit conversion macro takes extracted data from the lab dictionary and merges it with the source lab data to create standardized results for both System International (SI) and Conventional units. Converted results can be stored as one set in the LB dataset and the other set in the SUPPLB dataset.

This paper will introduce the structure of the lab data dictionary and describe how to develop the unit conversion macro step-by-step.

INTRODUCTION

In general, clinical trials using the same central lab to do all the lab testing do not require conversion of test results. However, a central lab is often too expensive or not very practical, so more projects may choose to use multiple local labs to measure samples. However, the problem is that the same kind of test can vary in unit across different countries, or even across different investigators within the same country. In such case, the importance of data conversion is self-evident, which provides a crucial basis for subsequent statistical analysis of relevant data. And how to quickly and accurately achieve data conversion, is often a difficult problem for programmers.

Lab data conversion factors are usually available from the following sources.

1. The EDC system or vendor data comes with it (the conversion factors may be inconsistent and precision maybe lost)
2. Provided by Sponsor (relatively efficient, but many sponsors do not provide this documentation)
3. Provided by DM team (sponsor approval required, some conversion factors may be inaccurate)
4. Prepared by programming team (sponsor approval required)

This paper will focus on the fourth scenario above and discuss the following issues:

- Dictionary - essential lookup tables
    - Table of standard mapping of test names
    - Table of normalization of original test units
    - Table of standard unit conversion factors
- SAS macros - batch conversion tools
    - Standard mapping of test names
    - Normalization of original test units
    - Standard unit conversion
- Limitations of this method and solutions to them

Incompleteness of the dictionary at the beginning of application

Inability to provide a single set of mixed conversion factors

Inability to directly convert character test results

DICTIONARY - ESSENTIAL LOOKUP TABLES

LB data conversion is not just the conversion of units. Before performing data conversion on the project, we also need to consider two other practical issues - the standardization of test names, and the normalization of original units (CDISC controlled terminology synonyms, central lab data, sponsor libraries, and historical lab data can be used here).

TABLE OF STANDARD MAPPING OF TEST NAMES

Different projects and different labs may have different expressions on name for the same test. In order to unify different test names that with the same meaning, we need to create the following look-up table:

Synonym	LBTESTCD	LBTEST
Protein (PROT) [random urine]	PROT	Protein
Protein (PROT) [Serum Protein Electrophoresis]	PROT	Protein
Protein (PROT) [Total Protein]	PROT	Protein
Protein (PROT) [Urine Protein Electrophoresis]	PROT	Protein
Protein	PROT	Protein
TOTAL PROTEIN	PROT	Protein
PROTEIN	PROT	Protein
PROTEIN, TOTAL	PROT	Protein
PROTEIN, TOTAL, RANDOM UR	PROT	Protein

Table 1. Standard Mapping of Test Names

With this table, we are able to unify all the different representations of the same test shown in the first column into the same CDISC submission value, that is, the standard LBTESTCD/LBTEST are obtained by the original test name.

TABLE OF NORMALIZATION OF ORIGINAL TEST UNITS

Different projects and different labs may have different expressions on name for the same test unit. In order to normalize (but not convert, since there is no calculation or conversion of any results only a change to the unit) original test units that with the same meaning, we need to create the following look-up table:

Synonym	LBORRESU
g%	g/dL
g/dL	g/dL
Gram per Deciliter	g/dL
g/L	g/L
Gram per Liter	g/L
kg/m3	g/L
Kilogram per Cubic Meter	g/L

mg/mL	g/L
Microgram per Microliter	g/L
Milligram per Milliliter	g/L
ug/uL	g/L
mg%	mg/dL
mg/dL	mg/dL
Milligram per Deciliter	mg/dL
mg/100mL	mg/dL
g/m3	mg/L
Gram per Cubic Meter	mg/L
mcg/mL	mg/L
mg/L	mg/L
Microgram per Milliliter	mg/L
Milligram per Liter	mg/L
ng/uL	mg/L
ug/mL	mg/L

Table 2. Normalization of Original Test Units

With this table, we are able to unify all the different representations of the same unit shown in the first column into the same CDISC submission value, that is, normalize the original unit to obtain LBORRESU.

TABLE OF STANDARD UNIT CONVERSION FACTORS

When the test name and the original test unit are unified, we can start to convert the measurement results to standard values.

Lab Specialist Committee use the following references to identify a standard Conventional and standard SI unit:

- Mosby: Mosby’s Diagnostic and Laboratory Test Reference – e-book: Version 14 Website
- Fischbach: A Manual of Laboratory and Diagnostic Tests, Frances Fischbach & Marshall B. Dunning III, 2015 Website
- Harrison: Harrisons Manual of Medicine 18th Edition
- AMA: American Medical Association (AMA)

The most common unit among the 4 sources is used to identify the preferred SI and Conventional unit, which will be mapped to CDISC controlled terminology. Conversion factors are then populated using various sources.

Now, the following table can be generated:

LBTESTCD	LBORRESU	LBSTRESU (SI)	Conversion Factor (SI)	LBSTRESU (Conventional)	Conversion Factor (Conventional)
PROT	g/dL	g/L	10	g/dL	1
PROT	g/L	g/L	1	g/dL	0.1

PROT	mg/dL	g/L	0.01	g/dL	0.001
PROT	mg/L	g/L	0.001	g/dL	0.0001

Table 3. Standard Unit Conversion Factors

Using the LBTESTCD(or LBTEST) and LBORRESU obtained earlier as key variables, the corresponding conversion factors can be obtained directly.

Note: The conversion factors here correspond to two sets of units - International System of Units (abbreviated as SI, modern form of the metric system) and U.S. Conventional Standard Units (abbreviated as Conventional, used in the United States). “FDA may require laboratory data using conventional units for reviewing submissions and labeling. Sponsors should discuss with the review divisions what laboratory data should utilize conventional units prior to submission (FDA Study Data Technical Conformance Guide v4.9).”

Based on this factor, we will consider generating two sets of conversion results for both units, one set stored in the main domain and the other set stored in the SUPP domain.

SAS MACROS - BATCH CONVERSION TOOLS

Once the first three basic tables are ready, it is time to consider how to use SAS to automate the data conversion.

STANDARD MAPPING OF TEST NAMES

The first is the standardization of test names, which can be achieved with the following code:

```
*** REMOVE DUPLICATE RECORDS FROM LAB_TEST_SYNONYM ***;
proc sort data=Lab_Test_Synonym out=Lab_Test_Synonym_s nodupkey;
    by SYNONYM_ TEST_SUBMISSION_VALUE;
run;

data __x_lab0;
    set &dsin.;
    _ORD=_n_;
run;

proc sql noprint;
    create table __x_lab1 as select
        a.*,
        b.TESTCD_SUBMISSION_VALUE as LBTESTCD,
        b.TEST_SUBMISSION_VALUE as LBTEST
    from __x_lab0 as a left join Lab_Test_Synonym_s as b
        on a.&raw_test_var.=b.SYNONYM_
    order by _ORD;
quit;
```

NORMALIZATION OF ORIGINAL TEST UNITS

The next step is the normalization of units, which can be achieved by the following code:

```
*** REMOVE DUPLICATE RECORDS FROM UNIT_SYNONYM ***;
proc sort data=unit_Synonym out=unit_Synonym_s nodupkey;
    by SYNONYM_ CDISC_SUBMISSION_VALUE;
run;

proc sql noprint;
    create table __x_lab2 as select
        a.*,
        b.CDISC_SUBMISSION_VALUE as LBORRESU
    from __x_lab1 as a left join Unit_Synonym_s as b
        on a.&raw_unit_var.=b.SYNONYM_
    order by _ORD;
quit;
```

STANDARD UNIT CONVERSION

Finally, the conversion of the test results is achieved by the following code:

```
proc sql noprint;
    create table __x_lab3 as select
        a.*,
        b.CONVERSION_FACTOR_TO_SI  as __v_CVFACTOR_SI,
        b.STRESU_SI  as __v_LBSTRESU_SI,
        b.CONVERSION_FACTOR_TO_CNV as __v_CVFACTOR_CNV,
        b.STRESU_CNV as __v_LBSTRESU_CNV,
        case when a.LBTEST^='' then b.__v_CONVFLAG
            else '' end as __v_CONVFLAG
    from __x_lab2 as a left join __x_conversion_sub as b
        on a.LBTEST=b.TEST and a.LBORRESU=b.ORRESU
    order by _ORD;
quit;

data __x_lab4;
    set __x_lab3;
    if LBORRES^="" and LBORRES=compress(LBORRES,".","kd") and
__v_CVFACTOR_SI^=. then do;
        __v_LBSTRESN_SI=input(LBORRES,best.)*__v_CVFACTOR_SI;
        if LBORNRLO^="" then
__v_LBSTNRLO_SI=input(LBORNRLO,best.)*__v_CVFACTOR_SI;
        if LBORNRHI^="" then
__v_LBSTNRHI_SI=input(LBORNRHI,best.)*__v_CVFACTOR_SI;
    end;
    if LBORRES^="" and LBORRES=compress(LBORRES,".","kd") and
__v_CVFACTOR_CNV^=. then do;
        __v_LBSTRESN_CNV=input(LBORRES,best.)*__v_CVFACTOR_CNV;
        if LBORNRLO^="" then
__v_LBSTNRLO_CNV=input(LBORNRLO,best.)*__v_CVFACTOR_CNV;
        if LBORNRHI^="" then
__v_LBSTNRHI_CNV=input(LBORNRHI,best.)*__v_CVFACTOR_CNV;
    end;
    %if %upcase(&SI_or_Conventional.)=SI %then %do;
        LBSTRESU = __v_LBSTRESU_SI;
        LBSTRESN = __v_LBSTRESN_SI;
        LBSTNRLO = __v_LBSTNRLO_SI;
        LBSTNRHI = __v_LBSTNRHI_SI;
        LBSTRSU2 = __v_LBSTRESU_CNV;
        LBSTRSN2 = __v_LBSTRESN_CNV;
        LBSTNRL2 = __v_LBSTNRLO_CNV;
        LBSTNRH2 = __v_LBSTNRHI_CNV;
    %end;
    %else %if %upcase(&SI_or_Conventional.)=CONVENTIONAL %then %do;
        LBSTRESU = __v_LBSTRESU_CNV;
        LBSTRESN = __v_LBSTRESN_CNV;
        LBSTNRLO = __v_LBSTNRLO_CNV;
        LBSTNRHI = __v_LBSTNRHI_CNV;
        LBSTRSU2 = __v_LBSTRESU_SI;
        LBSTRSN2 = __v_LBSTRESN_SI;
        LBSTNRL2 = __v_LBSTNRLO_SI;
        LBSTNRH2 = __v_LBSTNRHI_SI;
    %end;
    if __v_CONVFLAG='Y' then
__v_SI_or_Conventional="%upcase(&SI_or_Conventional.)";
run;
```

LIMITATIONS OF THIS METHOD AND SOLUTIONS TO THEM

The "dictionary + macro" approach described above has the following advantages:

- Dictionaries can be created, validated, maintained, and expanded by professionals with medical backgrounds, making the base content more accurate and reliable.
- The macro can be applied to different projects, which is easy to reuse and improves programming efficiency.

Nevertheless, this tool has the following limitations.

INCOMPLETENESS OF THE DICTIONARY AT THE BEGINNING OF APPLICATION

At the early stage of the dictionary's launch, it is inevitable that the contents it contains are not comprehensive, and it cannot take into account all the mapping requirements of all projects. For example, new test names and original units may appear, which makes it impossible to complete automatic information mapping and data conversion successfully.

The solution can be considered from two aspects:

- Macro generates an excel report to summarize these new test name(s) and unit(s), and then output a warning message in the SAS log to remind the user to communicate with the dictionary maintenance team and require them to add the new contents to the dictionary.
- Macro outputs new test name(s) and unit(s) encountered during each execution to a central file, and the dictionary maintenance team verifies them periodically and takes the initiative to add new contents to the dictionary.

INABILITY TO PROVIDE A SINGLE SET OF MIXED CONVERSION FACTORS

When the sponsor needs only one set of conversion results, with some tests using SI units, and the others using conventional units, this method cannot directly meet this requirement.

An indirect solution is to set the macro to generate a summary report table after initially run, which contains all test names, original units, and two sets of standard units and conversion factors for the project. Then we can mark the required unit and conversion factor for each test on the report to get a draft version of conversion table. Next step we send the draft version to sponsor for confirmation and approval to obtain the final version of the conversion table, which can then be used as a new dictionary, and be used by manual code to complete the lab data conversion.

INABILITY TO DIRECTLY CONVERT CHARACTER TEST RESULTS

Text results, or the mix of text and numeric results (e.g. LBORRES = "> 5") cannot be converted directly. This situation requires the programmer to add an extra code to implement the conversion (e.g. LBSTRESC = "> "||"5*conversion factor").

CONCLUSION

Lab data dictionary and conversion macro can make a SAS programmer’s life easier. The dictionary can be managed easily by lab specialists, and it's handy to be used for lab data conversion for different studies. The macro can help SAS programmers to simplify their codes, and improve both programming efficiency and the quality of output LB datasets.

REFERENCES

U.S. Food and Drug Administration. 2022. “Study Data Technical Conformance Guide (Version 4.9)”. March, 2022. https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources.

Kathleen Deska Pagana, Timothy James Pagana, Theresa Noel Pagana. 2019. *Mosby's Diagnostic and Laboratory Test Reference*. 14th ed. St. Louis, MO: Elsevier

Frances Talaska Fischbach, Marshall Barnett Dunning, III. 2015. *A manual of laboratory and diagnostic tests*. 9th ed. Philadelphia, PA: Wolters Kluwer Health.

Dan Louis Longo, Tinsley Randolph Harrison. 2013. *Harrison's manual of medicine*. 18th ed. New York, NY: McGraw-Hill Medical.

Sandra Minjoe. 2005. “Lab Unit Conversions: Taking Advantage of "Expert" External Resources”. *PharmaSUG 2005 - Paper DM03*.

ACKNOWLEDGMENTS

The authors would like to thank ICON plc for facilitating and supporting this paper.

CONTACT INFORMATION

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

Michael Yang
ICON plc
Michael.Yang@iconplc.com

David Fielding
ICON plc
David.Fielding@iconplc.com

Kent Letourneau
ICON plc
Kent.Letourneau@iconplc.com

Tobias Bottor
ICON plc
Tobias.Bottor@iconplc.com

Jessica Kueker
ICON plc
Jessica.Kueker@iconplc.com

Amy Welsh
ICON plc
Amy.Welsh@iconplc.com