Paper SI07

Pooling lab data to support FDA submission: keep it clear and simple

Joske van Schijndel, OCS Life Sciences, 's-Hertogenbosch, the Netherlands

ABSTRACT

In the preparation of an Integrated Safety of Summary of more than 30 studies, we came across hundreds of lab tests from different specimen types such as blood, urine and tissue. The lab analyses on these specimens were carried out by several laboratories each having their own standards. This resulted in 662 combinations of lab tests with different lab metadata (lab test names, categories, specimens, methods and units) for 150 unique lab test concepts.

To overcome this challenge a metalab library was created to align the metadata including the assignment of the standardised unit, for which no agreement across the industry exists yet.

This paper will describe the benefits of having the metalab library in place and will also focus on the challenges faced in the process of generating and maintaining it.

INTRODUCTION

The laboratory dataset is one of the core safety datasets. Within OCS Life Sciences, we worked on the Integrated Summary of Safety (ISS), where we prepared laboratory data from several studies dealing with the same drug to cover the safety aspects of the drug. The aim was to submit the New Drug Application (NDA) package to the Food and Drug Administration (FDA). The ISS is required by the FDA as a crucial component in any New Drug Application (NDA) submission. More than 30 studies were involved in de preparation of the study portfolio. Before we prepared the laboratory data, datasets from the clinical trials were converted to CDISC standards. SDTM domain LB, in which captures laboratory data, is one of the important domains for the ISS.

The ISS is not really a summary but rather a detailed integrated analysis of all relevant data from the clinical trials. It is a critical component of safety of the study drug. There are more ways to set up the ISS. The approach in the project was to stack the SDTM domains of each study together to create integrated SDTM datasets. Integrated ADaM datasets can then be created based on the integrated SDTM domains. From those ADaM datasets, ISS tables can be generated. This approach is chosen to align studies with each other and to make it easier to spot inconsistencies between studies.

REPORTING LABORATORY DATA

DATA COLLECTION AND PROCESSING

The 30 studies involved hundreds of lab tests from different specimen types such as blood, urine, tissue and more. Analyses on these specimens were carried out by several laboratories. These laboratories reported lab tests in their own preferred units. They also reported test name, specimen, and method using their own standards. Incoming laboratory data was transformed into SDTM for each individual study. Sponsor outsourced part of this task and the trial oversight to CROs. Therefore, a lot of different stakeholders were involved in collecting and processing the laboratory data.

Despite the fact that the studies were converted to SDTM, there were still differences in the metadata of these lab test concepts across studies. For pooling purposes, it is important that for one lab test concept, the same test name, units, specimen description and category name are used among all studies. The lab metadata should be aligned, but how to achieve this?

METALAB LIBRARY

To manage hundreds of lab tests, we extracted and combined the lab metadata from the 30 studies together and that resulted in 662 combinations of lab tests with different lab metadata across the studies. Think of corresponding lab test names, categories, specimens, methods and units. After correction, the actual number of lab tests performed across the 30 studies was around 150. We created and used a metalab library to obtain an overview of the 662 lab metadata combinations. The creation of this library was preparatory to pooling the SDTM LB domains.

For each study, a metalab was retrieved from the SDTM LB dataset. Figure 1 shows a metalab dataset of one of the studies. From left to right, the following variables were included: LBTESTCD, TYPE, LBTEST, LBCAT, LBSCAT, LBSPEC, LBMETHOD, LBORRESU, LBORNRLO, LBORNRHI, CONVFACT, LBSTRESU and LBSTNRC. Variable CONVFACT contains the conversion factor that is used to convert original results into standard results.

LBTESTCD	TYPE	LBTEST	LBCAT	LBSCAT	LBSPEC	LBMETHOD	LBORRESU	LBORNRLO	LBORNRHI	CONVFACT	LBSTRESU	LBSTNRC
ALB	N	Albumin	BIOCHEMISTRY		BLOOD		g/L	35	52	0.015	mmol/L	
ALP	N	Alkaline Phosphatase	BIOCHEMISTRY		BLOOD		IU/L	40	129	16.6667	nkat/L	
BASO	N	Basophils	HEMATOLOGY		BLOOD		10^9/L	0.01	0.2	1	10^9/L	
CA	N	Calcium	BIOCHEMISTRY		BLOOD		mmol/L	2.15	2.5	1	mmol/L	
CL	N	Chloride	BIOCHEMISTRY		BLOOD		mmol/L	98	106	1	mmol/L	
BILDIR	N	Direct Bilirubin	BIOCHEMISTRY		BLOOD		umol/L	0	5	1	umol/L	
CREAT	N	Creatinine	BIOCHEMISTRY		BLOOD		umol/L	59	104	1	umol/L	
DRUGSCR	C	Drug Screen	DRUG TEST		URINE							NEGATIVE
EOS	N	Eosinophils	HEMATOLOGY		BLOOD		10^9/L	0.03	0.7	1	10^9/L	
ETHANOL	C	Ethanol	ALCOHOL TEST		EXPIRED AIR	BREATHALYZER						NEGATIVE
GLUC	N	Glucose	BIOCHEMISTRY		BLOOD		mmol/L	4.11	5.89	1	mmol/L	
FIBRINO	N	Fibrinogen	COAGULATION		BLOOD		g/L	1.8	3.5	1	g/L	
GGT	N	Gamma Glutamyl Transferase	BIOCHEMISTRY		BLOOD		IU/L	8	61	16.6667	nkat/L	
HBSAG	C	Hepatitis B Virus Surface Antigen	SEROLOGY		BLOOD							NEGATIVE
HCAB	C	Hepatitis C Virus Antibody	SEROLOGY		BLOOD							NEGATIVE
HDL	N	HDL Cholesterol	BIOCHEMISTRY		BLOOD		mmol/L	1.46	999	1	mmol/L	
HIV12AB	C	HIV-1/2 Antibody	SEROLOGY		BLOOD							NEGATIVE
HOMOCY	N	Homocysteine	BIOCHEMISTRY		BLOOD		umol/L	0	12	1	umol/L	
HCT	N	Hematocrit	HEMATOLOGY		BLOOD		%	39	49	0.01	L/L	
HGB	N	Hemoglobin	HEMATOLOGY		BLOOD		g/L	135	180	1	g/L	
HAPTOG	N	Haptoglobin	BIOCHEMISTRY		BLOOD		g/L	0.3	3	1	g/L	
INSULIN	N	Insulin	BIOCHEMISTRY		BLOOD		pmol/L	17.8	173	1	pmol/L	
LDL	N	LDL Cholesterol	BIOCHEMISTRY		BLOOD		mmol/L	0	2.58	1	mmol/L	
LYM	N	Lymphocytes	HEMATOLOGY		BLOOD		10^9/L	1.3	4	1	10^9/L	
MCH	N	Ery. Mean Corpuscular Hemoglobin	HEMATOLOGY		BLOOD		pg	26	34	0.0621	fmol	
MCHC	N	Ery. Mean Corpuscular HGB Concentration	HEMATOLOGY		BLOOD		g/dL	31	35.8	0.6206	mmol/L	
MCV	N	Ery. Mean Corpuscular Volume	HEMATOLOGY		BLOOD		fL	81	99	1	fL	
MG	N	Magnesium	BIOCHEMISTRY		BLOOD		mmol/L	0.65	1.04	1	mmol/L	
GLUC	N	Glucose	URINALYSIS	MICROSCOPY	URINE		mmol/L	0	1.1	1	mmol/L	
PROT	N	Protein	URINALYSIS	MICROSCOPY	URINE		g/L	0	0.119	1	g/L	

Figure 1. Metalab dataset of one of the studies

The metalabs per study were combined into one Excel file using SAS programming. Data manipulations were necessary because of differences in rounding and variable type across the metalabs. Figure 2 shows an overview of the metalab library.

1	lbtestcd	type	lbtest	lbcat	lbscat	lbspec	lbmethod	lborresu	lbornrlo	lbornrhi	lbstresu	lbstnrc	CONVFA
629	UREA	N	Urea	BIOCHEMISTRY				mmol/L	1.7	8.3	mmol/L		1
630	UREA	N	Urea	BIOCHEMISTRY		BLOOD		mg/dL	7	21	mmol/L		0.357
631	UREA	N	Urea	BIOCHEMISTRY		BLOOD		mmol/L	2.1	8.9	mmol/L		1
632	UREA	N	Urea	BIOCHEMISTRY		BLOOD		mmol/L			mmol/L		1
633	URNGAL	N	Neutrophil Gelatinase-Assocd. Lipocalin	OTHER	RENAL FUNCTION MARKERS	URINE		ug/L			ug/L		1
634	URNGAL	N	Neutrophil Gelatinase-Assocd. Lipocalin	URINALYSIS		URINE		ug/L			ug/L		1
635	UROBIL	C	Urobilinogen	URINALYSIS		URINE	DIPSTICK					NEGATIVE	
636	UROBIL	N	Urobilinogen	URINALYSIS		URINE	DIPSTICK	mg/dL	0.2	1	mg/dL		1
637	UROBIL	N	Urobilinogen	URINALYSIS		URINE		mg/dL	0.2	1	mg/dL		1
638	VEGF	N	Vascular Endothelial Growth Factor	INFLAMMATORY MARKERS				ng/L			ng/L		1
639	VLDL	N	VLDL Cholesterol	LIPIDS				mg/dL		<30	mmol/L		0.0259
640	VLDL	N	VLDL Cholesterol	LIPIDS				mmol/L		<0.78	mmol/L		1
641	VLDL	N	VLDL Cholesterol	LIPIDS		BLOOD		mg/dL	8	40	mmol/L		0.0259
642	VLDL	N	VLDL Cholesterol	LIPIDS		BLOOD		mmol/L	0.21	1.04	mmol/L		1
643	VLDL	N	VLDL Cholesterol	LIPIDS		BLOOD	CALCULATED	mmol/L	0	1.04	mmol/L		1
644	VLDLPSZ	N	VLDL Particle Size	LIPIDS		BLOOD	LIPOPRINT	mg/dL	4.7	22.1	mg/dL		1
645	VOLUME	N	Volume	INSULIN SENSITIVITY	GLUCOSE CLAMP	URINE		mL			mL		1
646	VOLUME	N	Volume	URINALYSIS		URINE		mL			L		0.001
647	WBC	C	Leukocytes	URINALYSIS		URINE	DIPSTICK					NEGATIVE	
648	WBC	C	Leukocytes	URINALYSIS		URINE						NEGATIVE	
649	WBC	N	Leukocytes	HEMATOLOGY		BLOOD		10^6/L	4000	10700	10^9/L		0.001
650	WBC	N	Leukocytes	HEMATOLOGY		BLOOD		10^9/L	4	10.7	10^12/L		0.0001
651	WBC	N	Leukocytes	HEMATOLOGY		BLOOD		10^9/L	4	10	10^9/L		1
652	WBC	N	Leukocytes	HEMATOLOGY		BLOOD		10^9/L	4.00	10.00	10^9/L		1
653	WBC	N	Leukocytes	INFLAMMATORY MARKERS	MICROPARTICLES	BLOOD		up/uL			10^9/L		0.001
654	WBC	N	Leukocytes	URINALYSIS		URINE		/HPF	0	5	/HPF		1
655	WBC	N	Leukocytes	URINALYSIS		URINE		10^6/L	0	9	10^6/L		1
656	WBC	N	Leukocytes	URINALYSIS		URINE		cells/mL	0.0	10000.0	10^9/L		0.000001
657	WBC	N	Leukocytes	URINALYSIS		URINE	MICROSCOPY	HPF	0	5	HPF		1
658	WBC	N	Leukocytes	URINALYSIS		URINE	MICROSCOPY	uL	0	25	uL		1
659	WBC	N	Leukocytes	URINALYSIS	MICROSCOPY	URINE		cells/mL			10^9/L		0.000001
660	YEAST	C	Yeast Cells	URINALYSIS		URINE						NEGATIVE	
661	YEAST	C	Yeast Cells	URINALYSIS		URINE	MICROSCOPY		.	.		NEGATIVE	
662	YEAST	N	Yeast Cells	URINALYSIS		URINE	MICROSCOPY	uL	0	15	uL		1

Figure 2. Metalab library; overview

As the metalab library was stored in an Excel file, it allows you to sort the data in a relevant way or filter on specific items. The reference ranges may depend on the sex and age categories and therefore, in the actual library there were additional columns to indicated whether the references ranges were applicable to a specific gender or age range (columns were left out in figure 2). In the library there can be selected on records per study as shown in figure 3.

lborresu	lbornrlo	lbornrhi	lbstresu	lbstnrc	CONVFA	STUDY1	STUDY2	STUDY3	STUDY4
mmol/L	1.7	8.3	mmol/L		1	x			
mg/dL	7	21	mmol/L		0.357		x		
mmol/L	2.1	8.9	mmol/L		1			x	
mmol/L			mmol/L		1				
ug/L			ug/L		1				x

Figure 3. Metalab library; selection per study, not all columns have been displayed for readability purposes

For review purposes, additional checks were added: "*Standardised unit confirmed? (YES/NO)"* and *"Update needed? (YES/NO)".* Is the standardised unit confirmed by the sponsor? If not, this need to be communicated. Furthermore, we included checks whether updates are needed. The checks were based on differences and inconsistencies within lab test concepts. Within the library there is room created for comments per record and there is made use of colour coding within records, illustrated in figure 4. For example, records are marked red when discussion is needed. Because the tool generated an overview of the lab test concepts, the tool was used for communication both internally as externally (with the sponsor).

STUDY1	STUDY2	STUDY3	STUDY4	Comment
x				LBSTRESU and CONVF should be updated to pmol/Lng/mL cf AMA.
	x			Align LBSTNRC with other studies? Check.
		x		No SI Unit available.

1	lbtestcd	typ	lbtest	lbcat	lbscat
219	EPIC	C	Epithelial Cells	OTHER	RENAL FUNCTION MARKERS
220	EPIC	C	Epithelial Cells	URINALYSIS	
221	EPIC	C	Epithelial Cells	URINALYSIS	

Figure 4. Metalab library; comments and colour coding, not all columns have been displayed for readability purposes

DIFFERENCES AND INCONSISTENCIES

Figure 5 illustrates the different metadata available across the studies. Each record represents a unique combination of lab metadata. 13 different combinations of the lab test glucose were collected.

1 Row one and two show the differences between a urinary glucose test and blood glucose test. A different value for variables LBCAT ("URINALYSIS" vs. "BIOCHEMISTRY") and LBSPEC ("URINE" vs. "BLOOD") were used. The data type is different as well (character vs. numeric). From the metalab data provided here it is clear that row one and two report on actual two different glucose lab test concepts.

2 For the blood glucose tests, the results have reported in different units. See an example in row eight and nine, the difference between mmol/L vs. g/L. A different test or the same test, but differently reported?

3 Row seven to nine show different values of specimen types, stored in variable LBSPEC ("SERUM" vs. "BLOOD" vs. "CAPILLARY BLOOD"). Are these the same lab test, but differently collected and reported or really three different lab tests performed by the physician?

4 That question is also applicable to the category and subcategory, stored in variable LBCAT and LBSCAT, respectively. Is the glucose test where LBCAT = "BIOCHEMISTRY" another test than the glucose test where LBCAT = "HEMATOLOGY" or is it because of different standardisation approaches?

row	lbtestcd	type	lbtest	lbcat	lbscat	lbspec	lbmethod	lborresu	lbornrlo	lbornrhi	lbstresu	lbstnrc
1	GLUC	C	Glucose	URINALYSIS		URINE	DIPSTICK					NEGATIVE
2	GLUC	N	Glucose	BIOCHEMISTRY		BLOOD		mmol/L	3.89	5.83	mmol/L	
3	GLUC	N	Glucose	BIOCHEMISTRY		BLOOD		mmol/L	3.05	6.38	mmol/L	
4	GLUC	N	Glucose	BIOCHEMISTRY		PLASMA		mg/dL	110	126	mmol/L	
5	GLUC	N	Glucose	BIOCHEMISTRY		PLASMA		mmol/L	4.11	5.89	mmol/L	
6	GLUC	N	Glucose	GLYCEMIC AND OTHER LIPIDIC MARKERS		SERUM		mg/dL	70	110	mmol/L	
7	GLUC	N	Glucose	GLYCEMIC AND OTHER LIPIDIC MARKERS		SERUM		mmol/L	3.89	6.11	mmol/L	
8	GLUC	N	Glucose	HEMATOLOGY		BLOOD		mmol/L	3.88	5.83	mmol/L	
9	GLUC	N	Glucose	INSULIN SENSITIVITY	GLUCOSE CLAMP	CAPILLARY BLOOD		g/L			g/L	
10	GLUC	N	Glucose	PHARMACODYNAMICS		BLOOD		mmol/L			mmol/L	
11	GLUC	N	Glucose	URINALYSIS		URINE		g/L			mmol/L	
12	GLUC	N	Glucose	URINALYSIS		URINE	DIPSTICK	mg/dL			mmol/L	
13	GLUC	N	Glucose	URINALYSIS	MICROSCOPY	URINE		mmol/L			mmol/L	

Figure 5. Combinations of lab test glucose with different lab metadata across studies

In this phase of preparing the submission package for the FDA, all data was collected and mapped to SDTM on study level. It was decided to not change the SDTM datasets on study level and even in some cases it can no longer be traced back if lab test concepts that were differently reported were actually the same lab tests or not.

LAB TEST UNITS AND STANDARDISATION

There is no agreement across the industry on the units in which lab tests should be reported. Systems International (SI) units are the worldwide standard, but it's not always the units in which results are originally reported and the preferred units and naming in clinical practice. Although SI unit is the preferred unit for the standardised unit (LBSTRESU), we experienced that there are multiple SI units per lab test concept available. We even have found inconsistencies across websites and overviews reporting SI units that are being used by scientific journals.

There are best practices documents available. The PHUSE Lab Units team created the Test Unit Plan, a communicative document between the FDA and sponsor to agree on a submission level, the units to be used for each lab test concept[1]. It is a useful tool that can be used to improve the discussion between sponsor, authorities and other stakeholders when preparing a study portfolio with studies dealing with the same drug.

Standardisation of lab metadata across studies is essential to ensure that the lab test concept is being captured by the same lab metadata. This will also help in creating ADaM datasets, for instance in generating variables PARAM and PARAMCD. Standardisation of lab metadata across studies is also important for analysing the data correctly and efficiently. That is also true for pooling data from different studies within a project for analysis. As a programmer you cannot compare lab tests with different naming and units used, while the lab test is actually the same. Or the other way around, when the same naming is used for different lab tests.

OBSERVATIONS

There were a lot of differences in the use of standardised units across studies in the study portfolio. In some studies, the SI unit was used as standardised unit, while in other studies the conventional unit was used. For most lab tests SI unit is defined in the literature, but not for all lab tests. Besides, sometimes there are more than one SI units available for a lab test. An example explained. Lab test albumin in blood has 2 SI units displayed in the metalab library: g/L and mmol/L (see figure 6). g/L is the standardised unit in one study and mmol/L is the standarised unit in another study. Which SI unit is preferred as standardised unit in the pooled analysis?

lbtestcd	type	lbtest	lbcat	lbscat	lbspec	lbmethod	lborresu	lbornrlo	lbornrhi	lbstresu	lbstnrc	CONVFACT
ALB	N	Albumin	BIOCHEMISTRY		BLOOD		g/L	32	55	g/L		1
ALB	N	Albumin	BIOCHEMISTRY		BLOOD		g/L	35	52	mmol/L		0.015

Figure 6. Observation; standardised units of lab test concept albumin

Figure 7 illustrates another observation from the metalab library which is a known problem in standardisation of laboratory data and very important to correct for when pooling data from different studies. Different lab tests were converted to the same LBTESTCD/ LBTEST. The metalab library that we created provided inside in this inconsistency.

LBTESTCD	LBTEST	LBORRESU	LBSTRESU
BASO	Basophils	10^9/L	10^9/L
BASO	Basophils	%	%

LBTESTCD	LBTEST	LBORRESU	LBSTRESU
BASO	Basophils	10^9/L	10^9/L
BASOCE	Basophils/Total Cells	%	%

Figure 7. Observation; different lab tests converted to the same LBTESTCD/ LBTEST

CHALLENGES

How can you give insight in possible inconsistencies within unique lab test concepts across studies? How can you keep laboratory data clear and simple for all stakeholders? These two questions where challenges observed during the project and therefore, we created a metalab library in the preparation of the ISS.

Which standardised unit is preferred when pooling laboratory data for submission? What if there are more SI units available? This paper does not give you a clear answer to these questions. Although individual studies in the study portfolio were standardised to SDTM, it's not that easy. There were still differences in lab test concepts across studies. Laboratory data is complex and a lot of different stakeholder are involved in the process of collecting and processing laboratory data.

CONCLUSION

This paper introduced you to a metalab library, a simple approach which gives a quick and clear overview of the metalab data across studies. The approach of the metalab library can be used in a Test Unit Plan (TUP) to improve the discussion between all stakeholders about standardised lab units and other lab record characteristics to be used within a submission package. Specify standardised lab test concepts for pooling and submission purposes in an early stage of the drug development process, before data collection and processing. This makes understanding of laboratory data simpler and saves (re)work later on.

REFERENCES

1. PHUSE. “Best Practices – Standardizing Laboratory Units within a Sponsor/Submission Draft White Paper”, Published March 2015, Available at https://www.phusewiki.org/wiki/index.php?title=Lab_Units

ACKNOWLEDGMENTS

Thanks to my supervisors for giving constructive feedback and encouragement along the way. In particular, I would like to thank Lieke Gijsbers for her thoughtful review and the valuable contribution to this paper.

CONTACT INFORMATION

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

Author Name	Joske van Schijndel
Company	OCS Life Sciences
Address	Ruwekampweg 2G
City / Postcode	‘s-Hertogenbosch / 5222 AT
Work Phone	0031 (0)73 523 6000
Email	sasquestions@ocs-consulting.com
Web	www.ocs-lifesciences.com