

Implementation of MedDRA upgrade in SDTM: a SAS-based approach

Mahmoud ABU AZOUM, Business and Decision Life Sciences, Brussels, Belgium

ABSTRACT

It is often necessary to upgrade the values of coding variables such as AELLT, AEDECOD, etc. from the Medical Dictionary for Regulatory Activities (MedDRA) in SDTM datasets, e.g., for pooling purposes. We will present a SAS-based approach to upgrading the AE, MH, and PR domains by using two integrated programs. First, a program compares the values of the SDTM coding variable (e.g. AELLT, AEDECOD, etc.) with the new dictionary version. It flags the terms that are still current. For the outdated ones, the programs suggest automatic recoding where possible. Then, after a manual review and completion of the appropriate values for the coding variables in an excel file by a coding specialist, the updated recoded terms are merged into main datasets. A comparison report is generated for user review. We will explain our process with SAS. We will also show how we resolved some issues and potential enhancements.

Keywords: SAS-Approach, SDTM dataset, MedDRA.

INTRODUCTION

MedDRA, which is short for the Medical Dictionary for Regulatory Activities, is a standardized medical dictionary used in the field of pharmacovigilance and the regulation of medical products. It is a widely recognized and globally used terminology system that enables the exchange of information regarding the safety of pharmaceuticals, biologics, and medical devices.

MedDRA has multiple functions:

- **Standardization:** MedDRA provides a standardized code language for medical conditions, adverse events, and healthcare data. This aids communication among regulators, healthcare professionals, and researchers.
- **Pharmacovigilance:** It is important in pharmacovigilance, which oversees and evaluates medication and medical product safety. Adverse events and side effects are categorized using MedDRA terms for consistent reporting and analysis.
- **Regulatory Reporting:** Regulatory agencies, such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA), require the use of MedDRA for reporting adverse events and for evaluating the safety profile of drugs and medical devices.
- **Research and Clinical Trials:** MedDRA is used in clinical research and drug development to standardize reporting of adverse events in trials, aiding researchers and regulators in evaluating treatment safety and effectiveness.
- **Public Health Monitoring:** It helps monitor public health by providing a standardized language for reporting and analyzing health data on medications and medical devices.

MOTIVATION & OBJECTIVES

It is of utmost importance to ensure that certain variables within SDTM datasets, like AE, MH, and PR, are not coded with obsolete versions of MedDRA or left with incomplete values. In addition, Pooling studies that utilize different versions, with the sponsor intending to synchronize them, is a practice that involves combining research projects that employ various iterations, all with the purpose of aligning them.

To rectify this, it is imperative to upgrade to the most recent version of MedDRA and accurately fill in any missing values with the appropriate ones.

The information presented in the screenshot provides examples of MHDECOD and MHPTCD values which are no longer up-to-date and must be updated to conform to the latest MedDRA version, which is version 25.0.

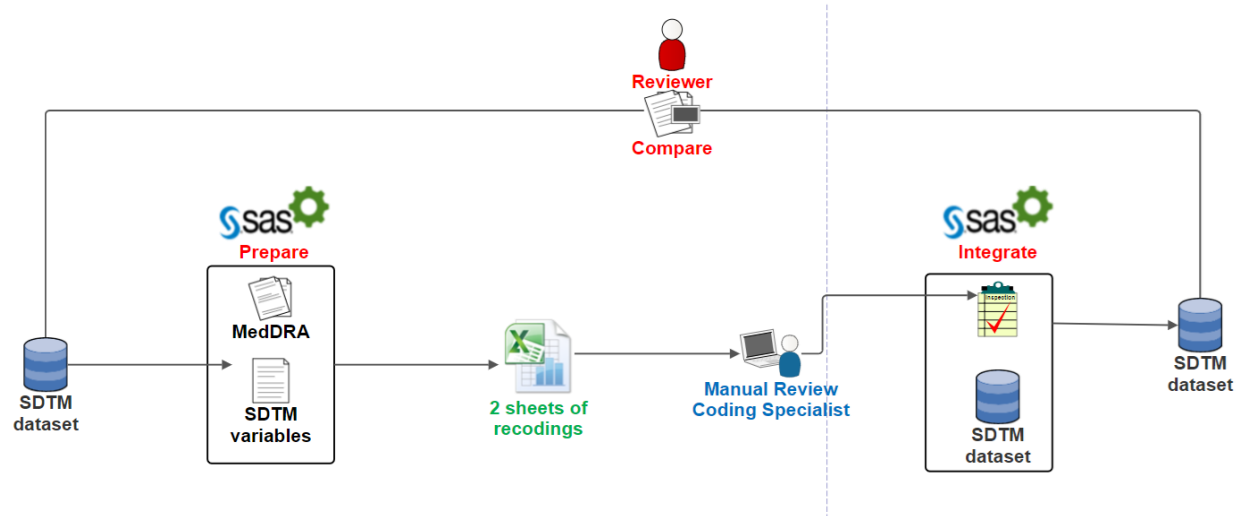
MHTERM	MHLLT	MHLLTCD	MHDECOD	MHPTCD
recurrent common cold	Common cold	10010106	Viral upper respiratory tract infection	10047482
Metrorrhagia	Metrorrhagia	10027514	Metrorrhagia	10027514
Metrorrhagia	Metrorrhagia	10027514	Metrorrhagia	10027514
flight anxiety	Phobia of flying	10034914	Phobia of flying	10034914
left shoulder pain	Shoulder pain	10040617	Musculoskeletal pain	10028391
pain left shoulder	Shoulder pain	10040617	Musculoskeletal pain	10028391
pain of left shoulder	Shoulder pain	10040617	Musculoskeletal pain	10028391
increased thyroid function	Thyroid function abnormal	10043726	Thyroid function test abnormal	10043730

OBJECTIVE

Upgrade the values of coding variables such as AEDECOD, AEPTCD from MedDRA in SDTM datasets (AE, MH,PR).

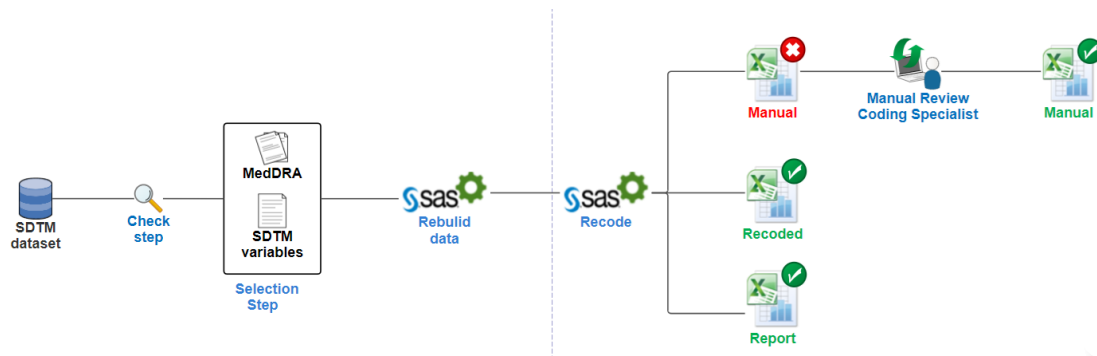
SAS-BASED APPROACH

We developed a SAS-based approach to upgrade the AE, MH, and PR domains by using two integrated programs (Prepare, Integrate).



PREPARE

This is the first program to prepare the excel sheets which will be used in the second program (integrate).



1. The validation of the parameters marks the inception of the program. In the event that the inputs or paths of the parameters contain any mistakes, the program will come to a halt and display an error message. This is done to ensure that the program runs smoothly without any hindrances.
2. The program will proceed to choose the relevant variables from the SDTM dataset and the MedDRA, unless there is a problem with the check step.
3. The macro named "Rebuild data" will create an initial dataset that contains the relevant variables from the source data that follows the SDTM standard. This dataset will then be used as the input for the next step of recoding the variables according to the specifications.
4. In the macro that is called "Recode", the program will perform a comparison between the data and the new dictionary, and it will detect the data that has no corresponding match in the new dictionary. This will generate an excel file with 3 sheets (Manual, Recoded, and Report).
 - *Manual sheet:* A coding specialist should check this sheet and update the SDTM values with the most recent dictionary.
 - *Recoded sheet:* The SDTM variables will be recoded automatically from the dictionary in this sheet.
 - *Report sheet:* The coding specialist will receive some statistics. The purpose is to do an "impact analysis" i.e. to have an idea of the amount of work/budget required to perform the upgrade.

The manual and recoded sheets have some results that are shown here. The values that marked by yellow indicate the old values, where the updated values have the blue color. One this one the LLT Name and LLT Code should also

be completed manually by the coding specialist; here the LLT Code ends up being the same and the LLT Name is different.

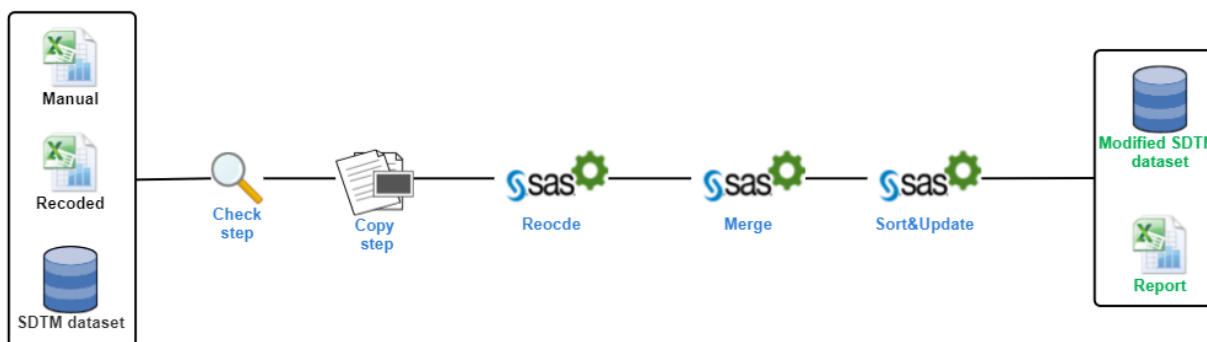
VERBATIM	LLT	LLTCD	PT	PTCD	HLT	HLTCD	LLT Name	LLT Code	PT Name	PT Code	HLT Name	HLT Code
Paraovarial cyst on Parovarian	10052456	Paraovarian cyst	10052456	Ovarian and fallopian tube cysts and neoplasms	10033121	Paraovarian cyst	10052456	Adnexa uteri cyst	10050203	Pelvis and broad ligament disorders NEC	10013344	
	MH_MANUAL	MH_RECODED	Report									

VERBATIM	LLT	LLTCD	PT	PTCD	HLT	HLTCD	LLT Name	LLT Code	PT Name	PT Code	HLT Name	HLT Code
recurrent common cold	Common cold	10010106	Viral upper respiratory tract infection	10047482	Common cold	10010106	Nasopharyngitis	10028810				
Menorrhagia	Menorrhagia	10027313	Menorrhagia	10027313	Menorrhagia	10027313	Heavy menstrual bleeding	10085423				
Metrorrhagia	Metrorrhagia	10027514	Metrorrhagia	10027514	Metrorrhagia	10027514	Intermenstrual bleeding	10022559				
flight anxiety	Phobia of flying	10034914	Phobia of flying	10034914	Phobia of flying	10034914	Aerophobia	10080300				
left shoulder pain	Shoulder pain	10040617	Musculoskeletal pain	10028391	Shoulder pain	10040617	Arthralgia	10003239				
pain left shoulder	Shoulder pain	10040617	Musculoskeletal pain	10028391	Shoulder pain	10040617	Arthralgia	10003239				
pain of left shoulder	Shoulder pain	10040617	Musculoskeletal pain	10028391	Shoulder pain	10040617	Arthralgia	10003239				
increased thyroid function	Thyroid function abnormal	10043726	Thyroid function test abnormal	10043730	Thyroid function abnormal	10043726	Thyroid disorder	10043709				
	MH_MANUAL	MH_RECODED	Report									

- Some PT might link to different SOC and that the program is designed to select the "primary" one.
- The link to dictionary is done via the old LLT, if not found in dictionary the program tries to do a match using the verbatim.

INTEGRATE

The program will handle the two sheets and update the SDTM dataset accordingly.



1. The program starts by validating the parameters and checking the existence of the excel file. If there are any errors in the inputs or paths of the parameters, the program will stop and show an error message.
2. The next step is to copy the SDTM datasets to new folder.
3. The macro named "Reocode" will take the excel sheet as an input and extract the relevant data from it. The macro will then prepare the data for the next step of the program by applying some transformations and validations.
4. The macro with the name "merge" will perform a data integration operation by taking the data from two sources as inputs. The first source is the SDTM dataset, which follows a standard format for clinical trial data. The second source is the sheet, which contains the updated variables according to the most recent version of MedDRA. The program will merge the data from both sources by matching the common variables and appending the SDTM variables.
5. After the program generates the output, it will sort the output data according to the keys that are defined in the specifications. The program will also update the output data with the latest values from the sources. In addition, the program will rework, in the SUPP-- dataset, which contains supplemental information for the output data, the IDVARVAL in case the parent dataset sorting would have changed in the process.

This is an example of SDTM before upgrading to the new version of MedDRA.

MHTERM	MHLLT	MHLLTCD	MHDECOD	MHPTCD	MHHLT	MHHLTCD
recurrent common cold	Common cold	10010106	Viral upper respiratory tract infection	10047482	Viral infections NEC	10047465
Metrorrhagia	Metrorrhagia	10027514	Metrorrhagia	10027514	Menstruation and uterine bleeding NEC	10027335
Metrorrhagia	Metrorrhagia	10027514	Metrorrhagia	10027514	Menstruation and uterine bleeding NEC	10027335
flight anxiety	Phobia of flying	10034914	Phobia of flying	10034914	Fear symptoms and phobic disorders (incl social phobia)	10068299
left shoulder pain	Shoulder pain	10040617	Musculoskeletal pain	10028391	Musculoskeletal and connective tissue pain and discomfort	10068757
pain left shoulder	Shoulder pain	10040617	Musculoskeletal pain	10028391	Musculoskeletal and connective tissue pain and discomfort	10068757
pain of left shoulder	Shoulder pain	10040617	Musculoskeletal pain	10028391	Musculoskeletal and connective tissue pain and discomfort	10068757
increased thyroid function	Thyroid function abnormal	10043726	Thyroid function test abnormal	10043730	Thyroid analyses	10043692

And here is the SDTM dataset after upgrading.

MHTERM	MHLLT	MHLLTCD	MHDECOD	MHPTCD	MHHLT	MHHLTCD
recurrent common cold	Common cold	10010106	Nasopharyngitis	10028810	Upper respiratory tract infections	10046309
Metrorrhagia	Metrorrhagia	10027514	Intermenstrual bleeding	10022559	Menstruation and uterine bleeding NEC	10027335
Methrorrhagia	Metrorrhagia	10027514	Intermenstrual bleeding	10022559	Menstruation and uterine bleeding NEC	10027335
flight anxiety	Phobia of flying	10034914	Aerophobia	10080300	Fear symptoms and phobic disorders (incl social phobia)	10068299
left shoulder pain	Shoulder pain	10040617	Arthralgia	10003239	Joint related signs and symptoms	10023226
pain left shoulder	Shoulder pain	10040617	Arthralgia	10003239	Joint related signs and symptoms	10023226
pain of left shoulder	Shoulder pain	10040617	Arthralgia	10003239	Joint related signs and symptoms	10023226
increased thyroid function	Thyroid function abnormal	10043726	Thyroid disorder	10043709	Thyroid disorders NEC	10043712

CONCLUSION

- Recoding variables in SDTM using MedDRA is a crucial step.
- The approach can handle the AE, MH, and PR domains.
- Prepare program can provide feedback for the coding specialist, especially with the excels sheets.
- Optimizing Integrate program is still worth doing to reduce the run-time.

APPENDEIX

Parts of the program:

```

1 options SASAUTOS=("\\fileserver\home\&sysuserid.\&Type._DVL_BEL-LSP_&REP_ID.\Programs\MedDRA\Compare\Macros") mrecall=YES;
2 /* This to specify P: and W: structure if it's DS or DM */
3 %macro define_DS_DM;
4     options mprint;
5     /*Sanity check*/
6     %if "&REP_ID."="" or "&STUDYID."="" or "&Type."="" or "&domainName."="" %then %do;
7         %PUT ERROR: One or more of the parameters is missing: Please check down which parameter is missing;
8         %PUT - REP_ID = "&REP_ID.";
9         %PUT - STUDYID = "&STUDYID.";
10        %PUT - Type = "&Type.";
11        %PUT - domainName = "&domainName.";
12        %PUT ERROR: Aborting program;
13        %abort cancel;
14    %end;
15
16    %global mapping_path data ;
17    %if &type = DS %then %do ;
18        %let mapping_path=Mapping;
19        %let data=Data;
20    %end;
21    %else %if &type = DM %then %do ;
22        %let mapping_path=DM Data\Mapping;
23        %let data=SDTM Data;
24    %end;
25 %mend define_DS_DM;

```

```

1 options SASAUTOS=("\\fileserver\home\&sysuserid.\&Type._DVL_BEL-LSP_&REP_ID.\Programs\MedDRA\Integrate\Macros") mrecall=YES;
2 %macro SortByKeys(data=);
3 /* IMPORT ADAM TABLE LEVEL METADATA DOMAIN SHEET */
4 filename nefmap "&mapping." ENCODING='wlatin1';
5 PROC IMPORT OUT= WORK.SDTM_TABLE_METADATA
6 DATAFILE=nefmap
7 DBMS=&DBMS. REPLACE;
8 /*SHEET="B# SDTM table metadata$";*/
9 %sysfunc(prxchange(&prxchange_sheet, 1, A. SDTM table metadata));
10 GETNAMES=YES;
11 %unquote(&import_opts);
12 RUN;
13
14 DATA SDTM_TABLE_METADATA_SHEET (rename = (A1 = DATASET A2 = DESCRIPTION A3 = CLASS A4 = STRUCTURE A5 = PURPOSE A6 = KEYS A7 = XPORT));
15 SET SDTM_TABLE_METADATA(FIRSTOBS = 2 keep = A1 A2 A3 A4 A5 A6 A7);
16 IF STRIP(A1) NE '' THEN OUTPUT;
17 RUN;
18
19 DATA SDTM_TABLE_METADATA_SHEET;
20 STUDYID = "&STUDYID" ;
21 SET SDTM_TABLE_METADATA_SHEET;
22 RUN;
23
24 /* GET SORTING KEYS, RESORT DATASET AND RECREATE --SEQ */
25 PROC SQL NOPRINT;
26 SELECT tranwrd(KEYS, " ", " ") into :keys from SDTM_TABLE_METADATA_SHEET where STUDYID = "&STUDYID" & DATASET = "&domainName";
27 QUIT;
28
29 DATA LKEYS;
30 SET SDTM_TABLE_METADATA_SHEET;
31 LKEYS=LENGTH(STRIP(KEYS));
32 KEEP DATASET KEYS LKEYS;
33 WHERE DATASET="&DOMAINNAME";
34 RUN;
35
36 PROC SQL NOPRINT;
37 SELECT LKEYS into :Lkeys from LKEYS;
38 QUIT;
39
40 %PUT NOTE: Keys are &KEYS.;

```

CONTACT INFORMATION

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

Mahmoud ABU AZOUM

Business and Decision Life Sciences-Belgium

mahmoud.abuazoum@businessdecision.be

Brand and product names are trademarks of their respective companies.