

Detailed Operation to Process MedDRA/WHODrug Dictionary Up-Version with SAS Macros

Yejin Huang, ClinChoice

ABSTRACT

MedDRA and WHODrug dictionary are always on the processing of upgrading. Usually, for integrated studies or legacy studies, team may request to re-map the dictionary variables based on one new dictionary version. Therefore, the up-version process is definitely useful and important. To do up-version efficiently, we aim to standardized the process and use SAS macros to implement the process. The difference between the MedDRA and WHODrug up-version processes will also be discussed in this paper.

INTRODUCTION

The MedDRA (Medical Dictionary for Regulatory Activities) Terminology is the international medical terminology developed under the auspices of the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use. It is a multilingual dictionary of standard terminology that is used to code medical events in humans, including patients' medical histories and adverse events. Figure 1 shows five hierarchies in MedDRA dictionary:

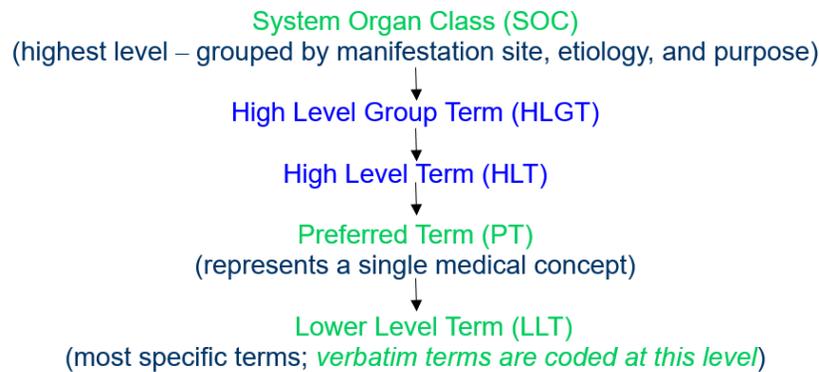


Figure 1 MedDRA Hierarchies

The WHODrug dictionary is used worldwide, which provides a reporting structure used to classify drug data and is a source of international drug names. With its unique drug code hierarchy and extensive coverage, it provides a consistent drug dictionary with exact terminology when coding concomitant medications. The dictionary is used to identify drug names and evaluate medicinal product information, including active ingredients and products' anatomical and therapeutic classifications, from nearly 150 countries. Figure 2 shows five hierarchies in WHODrug dictionary:

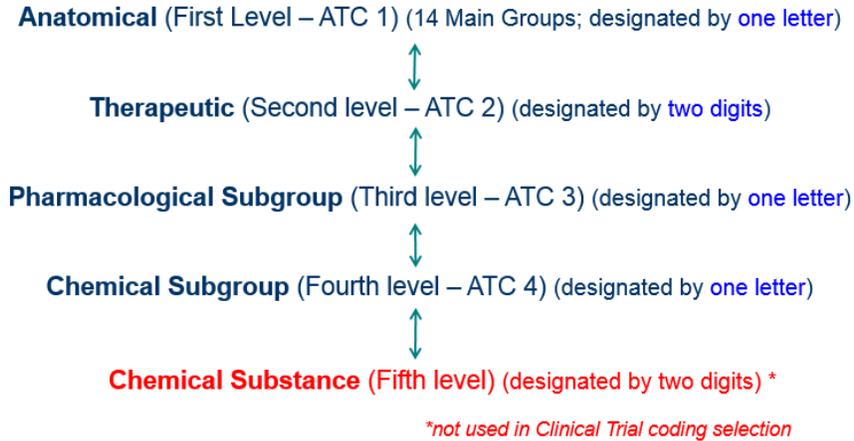


Figure 2 WHODrug Hierarchies

MedDRA and WHODrug dictionaries have undergone multiple revisions since it first began, and both are updated every six months (March and September).

The detailed operations below will take MedDRA as main example given the up-version process for MedDRA and WHODrug have many similarities, differences for WHODrug will also be mentioned.

PREPARE THE BASIC DATASETS

When up-version need be performed, that means the standard dataset(s) (SDTM/ADaM) is ready with old version coding information, therefore, the process of up-version is to leave the coding which are not changed compared with new version dictionary alone and extract the uncoded ones for manually coding, so selection of implementing the up-version process in SDTM or ADaM level will be the very first step. It is clear to know the lowest level term in MedDRA is the key and unique information. That is to say whether which level used to do up-version, --LLT must be included in datasets. For WHODrug, ATCCODE and Key_Drugname (which is CMDECOD in CM data), are key variables to make coding unique.

CODING FILE PREPARATION

The second step is pulling all data together with all coding variables after deciding SDTM or ADaM level. Then merge dataset with dictionary selecting LLT_Currency='Y' (LLT_Currency='N' means this LLT Name should not be used currently) by LLT Name and LLT Code, records will be classified into two parts:

- Merged records will be output as Auto coding terms into one tab named as "Auto_Coding". Usually, there are two scenarios: 1) all coding variables in dataset are perfectly matched with dictionary. 2) some coding variables in dataset are not matched with dictionary which means the coding for such LLT Name are changed, then keep the new coding from dictionary as up-versioned coding information and flag out which coding variables are changed in variable ITEM_CHANGED. See the Example 1 below:

LLT_CODE	PT_CODE	LLT_CUR REN CY	PT_NAME	HLT_CODE	HLT_NAME	HLGT_CODE	HLGT_NAME	SOC_CODE	SOC_NAME	ITEM_CHANGED
10056519	10056519	Y	Abdominal infection	10017967	Abdominal and gastrointestinal infections	10021879	Infections - pathogen unspecified	10021881	Infections and Infestations	HLGT Name
10068719	10033295	Y	Overdose	10076292	Overdoses NEC	10076290	Product use issues	10022117	Injury, poisoning and procedural complications	HLT Code, HLT Name, HLGT Name, HLGT Code
10000381	10000381	Y	Accidental overdose	10064291	Maladministrations	10064289	Medication errors	10022117	Injury, poisoning and procedural complications	HLT Code, HLT Name

Example 1 Variable ITEM_CHANGED in Auto_Coding Tab

- Unmerged records will be output as Uncoded terms in a separate tab named as “Uncoded”, only keep the verbatim information and set all coding variables as blank. Two possible situations are: 1) the LLT Name coded before can not be found in new dictionary any more, obviously they need coder’s manually coding with new dictionary. 2) in ongoing studies, coding file is probably not in same progress with raw data collected so some verbatims are not coded which means LLT Name are missing and need coder’s manually coding for them. See the Example 2 below:

VERBATIM	SPLIT	LLT_CODE	LLT_NAME	PT_CODE	LLT_CURRENCY	PT_NAME
ABDOMINAL DISTENSION						
ABDOMINAL PAIN (HYPOGASTRIUM)						
ALLERGIC RHINITIS, INTERMITTENT						
ALP PHOS						
ASYSTOLE						

Example 2 Uncoded Tab

It would be better to sort records by Verbatim and Study Identifier when outputting the coding preparation file. Sometimes same verbatim may have different coding depending on coders’ knowledge or habits, sponsor requests or other factors. Put same verbatim together will make it clear for coder to unify the coding information for same verbatim across many studies.

Different with MedDRA structure, WHODrug dictionary includes three parts (dd, dda, ina) so a preprocess is needed to make WHODrug dictionary be merge-ready (see Figure 3 below). And it is necessary to keep indication and route as merge by variables since them two will also affect the coding.

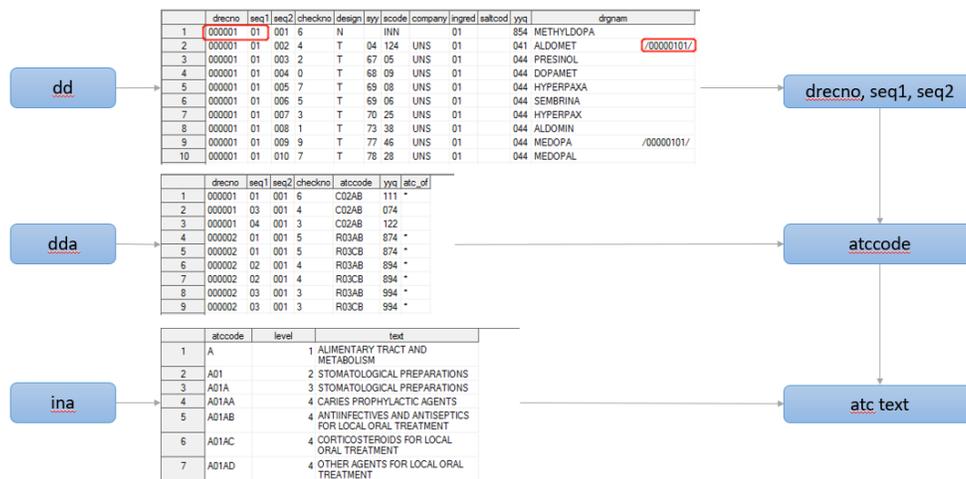


Figure 3 Three datasets for merge-ready WHODrug Dictionary

The macros are created to process this step automatically, user-need input parameters are as following. (See Figure 4 and Figure 5)

```

/* Up-version for AE or MH domain */
%let domain_name=AE;
/* New MedDRA version needed */
%let dictver=v23_0;
/* Insert all studies needed, number of studies is dynamic */
%let study1=%str(XXX);%let study2=%str(XXX);%let study3=%str(XXX);%let study4=%str(XXX);/**/
%let studyn=%str(4);
/* Output data path for QC */
%let outputdata=%str(XXX);
/* Output excel name */
%let outputfile=%str(XXX);

```

Figure 4 Macro variables for MedDRA preparation

```

/* Up-version for CM domain */
%let domain_name=CM;
/* New MedDRA version needed */
%let dictver=v2022_03_01;
/* Insert all studies needed, number of studies is dynamic */
%let study1=%str(xxx);%let study2=%str(xxx);/**/
%let studyn=%str(xxx);
/* Output data path for QC */
%let outputdata=%str(xxx);
/* Output excel name */
%let outputfile=%str(xxx);

```

Figure 5 Macro variables for WHODrug preparation

VERIFICATION OF MANUALLY CODED FILE

The third step 'Verification' will have two sides: verify the consistency between coded file and new dictionary; verify the uniqueness of coding information when SPLIT is 'No'.

SPLIT: multiple adverse events/medical histories were collected in one Verbatim in raw data, then the one record will be split into multiple records (one symptom corresponds to one coding) by coder.

CONSISTENT WITH DICTIONARY OR NOT?

This is to ensure the coded information are consistent with the new dictionary, with no typos, truncations or other manual mistakes. Merge coded file with new dictionary by LLT Name and do exact comparison for each coding variable. All mismatched records will be output into a check file and flag out mismatched variables for coders to check. For WHODrug, Key_Drugname (CMDECOD), ATCCODE will be the merge variables.

BE UNIQUE FOR EACH VERBATIM OR NOT?

This step is team decided. If same verbatim having different coding is acceptable, the step will be skipped. Otherwise, non-unique coding will be flagged out and need coder to select unique one.

The macros are created to process this step automatically, user-need input parameters are as following. (See Figure 6 and Figure 7)

```

%macro chk_meddra_coding(
  domain=      , /* Verification for AE or MH domain */
  version=     , /* The right version need be verified with */
  path_file=   , /* The coded coding file location */
  sheet1=     , /* The Auto_Coding sheet name */
  sheet2=     , /* The Uncoded sheet name */
  llt_n=      , /* Set the right name with variable name from coding file */
  llt_c=      ,
  pt_n=       ,
  pt_c=       ,
  hlt_n=      ,
  hlt_c=      ,
  hlgt_n=     ,
  hlgt_c=     ,
  soc_n=      ,
  soc_c=      ,
  checkdup=   ); /* Set the check duplicate coding file name for same Verbatim (same indication same route) */

```

Figure 6 Macro parameters for MedDRA verification

```

%macro chk_whodurg_coding(
  version= , /* The right version need be verified with */
  path_file= , /* The coded coding file location */
  sheet1= , /* The Auto_Coding sheet name */
  sheet2= , /* The Uncoded sheet name */
  verb_term= , /* Set the right name with variable name from coding file */
  ref_term= ,
  ref_code= ,
  atc1_de= ,
  atc2_de= ,
  atc3_de= ,
  atc4_de= ,
  atc1_co= ,
  atc2_co= ,
  atc3_co= ,
  atc4_co= ,
  atc_co= ,
  checkdup= ); /* Set the check duplicate coding file name for same Verbatim (same indication same route) */

```

Figure 7 Macro parameters for WHODrug verification

IMPLEMENT CODING FILE INTO ANALYSIS DADASET

Read in the coding file which is passed verification and set the Auto_Coding and Uncoded records together as a whole coding dataset. Implement them back into basic dataset, through proc sql statements, left join by Verbatim.

For WHODrug, indication and route also need be included as by variables.

DOUBLE CHECK AFTER ALL PROCESS

After the whole up-version process, it is essential to make sure: 1) no uncoded terms exist. 2) the number of records does not change if there are no SPLIT coding, or the number of records increases accordingly because of SPLIT coding terms.

EXPECTATION

Actually, the whole process above is enough for coding up-version for DBL data (legacy studies), however, if studies are ongoing which means verbatims would increase, more actions should be taken during the process. We aim to create SAS macros to do the duplicate check steps in the future.

- a) In first run: operate whole process above.
- b) In second/third... run:
 - Flag out verbatims which have been already up-versioned to give coder a reminder.
 - Compare the new coded file with previous one, if any verbatims are removed or changed by mistakes.
- c) In final run, all the verbatims should have coding information (except ones are not in coding scope) and pass the verification.

VISUAL TOOL TO USE THE MACROS

For the purpose of using the macros in up-version process easier, a simple and visual VBS tool is generated to call the macros, and user only need to input the necessary parameters.

- Step 1: choose the section you want to process (see Figure 8)
- Step 2: fill in the blanks with parameters (see Figure 9, Figure 10)
- Step 3: click 'batch' button and check results

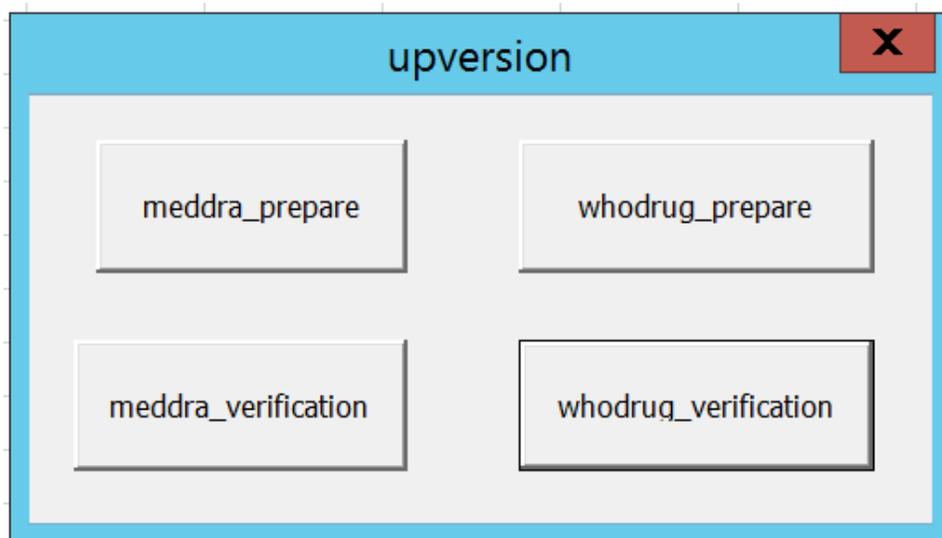


Figure 8 Step 1 Process Selection

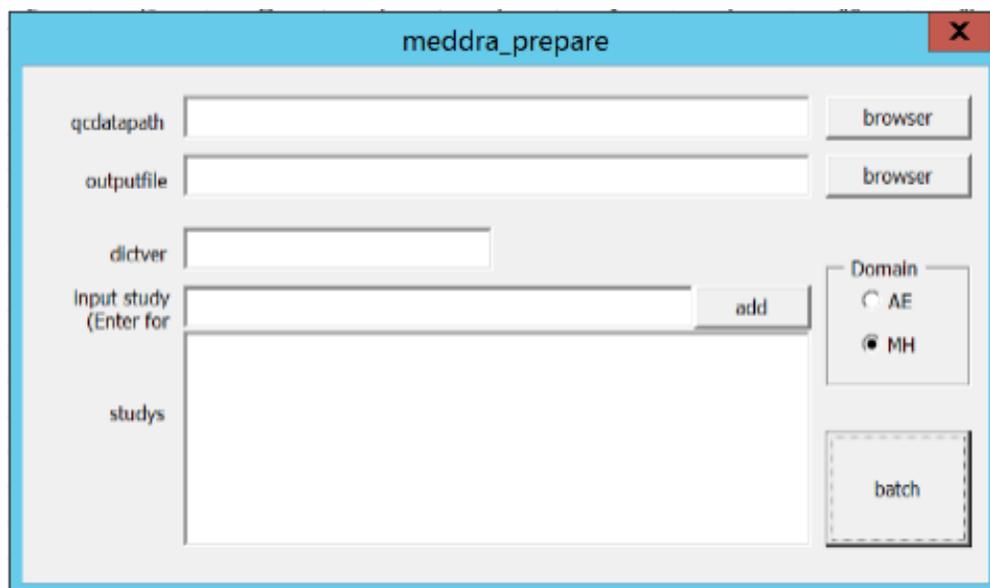


Figure 9 Step 2 Fill in Parameters in Preparation Process

The screenshot shows a software window titled "meddra_verification" with a close button (X) in the top right corner. The interface contains several input fields and buttons:

- chkdup**: An empty text input field with a "browser" button to its right.
- file:**: An empty text input field with a "browser" button to its right.
- dictver**: A text input field containing the value "v23_0".
- sheet_name1:** A text input field containing "Auto_Coding".
- sheet_name2:** A text input field containing "Uncoded".
- llt_name:** A text input field containing "LLT NAME".
- llt_code:** A text input field containing "LLT CODE".
- pt_name:** A text input field containing "PT NAME".
- pt_code:** A text input field containing "PT CODE".
- hlt_name:** A text input field containing "HLT NAME".
- hlt_code:** A text input field containing "HLT CODE".
- hlgt_name:** A text input field containing "HLGT NAME".
- hlgt_code:** A text input field containing "HLGT CODE".
- soc_name:** A text input field containing "SOC NAME".
- soc_code:** A text input field containing "SOC CODE".
- Domain:** A group box containing two radio buttons: "AE" (unselected) and "MH" (selected).
- batch:** A button located below the Domain group box.

Figure 10 Step 2 Fill in Parameters in Verification Process

REFERENCES

Challenges of different MedDRA versions for integrated data and databases, Peter Bonata, Statistical Analyst, Bayer Pharma, Wuppertal, Germany. Available at

https://www.lexjansen.com/phuse/2014/dh/DH09_ppt.pdf

ACKNOWLEDGMENTS

I would specifically like to acknowledge Zhilong Jiang for his assistance and Ping Ni, Xin Wang for being my reviewer and offering great feedback.

CONTACT INFORMATION

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

Yejin Huang, ClinChoice
yejin.huang@clinchoice.com