

A SAS Based MedDRA Coding System

Charley Z. Wu, Alexion Pharmaceuticals, CT, USA
Dona-Lyn Wales, Alexion Pharmaceuticals, CT, USA

ABSTRACT

A simple and efficient MedDRA coding system is built with SAS. This system has three major components: 1) MedDRA Dictionary and MedDRA Synonym Dictionary. 2) MedDRA coding Transaction File. 3) Auto and Manual Coding Processes. The system has the following functions: SAS MedDRA dictionary data loading, MedDRA Synonym Dictionary self-learning and self-cleaning, auto and manual Coding, audit trail, versioning, and query generation. Users can code against any version of MedDRA dictionary. The system is built on a modular basis. So it is easy to maintain and easy to extend its functionalities.

INTRODUCTION

MedDRA® (Medical Dictionary for Regulatory Activities) is dictionary of standard medical terms. It is widely used to standardize adverse events caused by drugs or medical devices. The standardized medical terms make it easier for the health authorities and the pharmaceutical companies to evaluate the safety of medical products. It also facilitates the exchange and analysis of safety data across companies. FDA requires that all adverse events in clinical studies be coded by MedDRA and the coded terms need to be reported. Therefore, many MedDRA coding systems had been developed. Most of them are very complex and integrated with EDC or CTMS systems.

We are describing a simple and efficient MedDRA coding system developed in SAS. The system consists of three major components:

- 1) MedDRA Dictionary and MedDRA Synonym Dictionary
- 2) MedDRA coding Transaction File
- 3) Auto and Manual Coding

DATA STRUCTURE

The coding system consists of a number of SAS datasets. The structure of three major ones is described here:

1. MEDDRA DICTIONARY:

MedDRA dictionary ASCII files are downloaded from MSSO (Maintenance and Support Services Organization). These files are normally used for the accompanying MedDRA browser. However, these files cannot be used directly for the coding system. So we need to merge and convert the files into a single SAS dataset. MedDRA dictionary is used globally for all studies. The structure of the SAS MedDRA dictionary is as follows. Please note a DICTVER variable is created for version control during the coding.

Table 1 **Error! Reference source not found.**: MedDRA Dictionary

SAS Variable Name	SAS Variable Format	SAS Variable Label
LLT	CHAR100	MedDRA Lowest Level Term
PT	CHAR100	MedDRA Preferred Term
HLT	CHAR100	MedDRA High Level Term
HLGT	CHAR100	MedDRA High Level Group Term
SOC	CHAR100	MedDRA System Organ Class
PRIMARY	CHAR1	Is This a Primary Path? (Y/N)
DICTVER	CHAR50	MedDRA Dictionary Version

As one LLT can be assigned to multiple SOC, the primary key for the MedDRA dictionary is a composite one which includes LLT, PT, HLT, HLGT and SOC. However, for each LLT, there is only one primary path to SOC.

A SAS Based MedDRA Coding System

2. MEDDRA SYNONYM DICTIONARY:

The Synonym Dictionary is built in house. It is used to capture manually coded terms. Initially, we also include uncoded verbatim terms from source datasets (e.g. Adverse Event). However, later on we found that there are many self-evident corrections for these uncoded terms, for example, typos, extra spaces before or after a word. All these need to be corrected before coding. Therefore, we decided to keep the self-evident corrections at the study level. So the final structure of the Synonym Dictionary is as follows. Please note DICTVER is added for version control. The Synonym Dictionary is also used globally for all studies.

Table 2: MedDRA Synonym Dictionary

SAS Variable Name	SAS Variable Format	SAS Variable Label
VTMODIFY	CHAR200	Modified Verbatim Term
LLT	CHAR100	MedDRA Lowest Level Term
DICTVER	CHAR50	MedDRA Dictionary Version
VALID	CHAR1	Is This Record Valid (Y/N)
RECUPDAT	DATETIME20	Date/Time Record Last Modified

The primary key for MedDRA Synonym Dictionary is also a composite one which includes VTMODIFY and LLT. LLT is the foreign key linked back to the LLT in the MedDRA Dictionary.

3. MEDDRA CODING TRANSACTION FILE:

The Coding Transaction File is also built in house. This is the place where a coder interacts with the coding system. Manual coding is done here. Verbatim terms which cannot be coded automatically by the coding system will be written to this file. A coder will then either modify the verbatim term to make it coded, or assign a code (LLT) directly to it. Please note this file is not global. It is implemented at the study level. There is only one transaction file for each study. Its structure is as follows:

Table 3: MedDRA Transaction File

SAS Variable Name	SAS Variable Format	SAS Variable Label	Comment
STUDYID	CHAR20	Study Identifier	
VTERM	CHAR200	Verbatim Term	
VTMODIFY	CHAR200	Modified Verbatim Term	Links to VTMODIFY in the Synonym Dictionary
LLT	CHAR100	MedDRA Lowest Level Term	Links to LLT in the MedDRA Dictionary
LLTFLAG	CHAR1	Manual (M) or Auto (A) Coding	Flag to show whether the coded LLT term is a direct hit (auto-coding) or assigned by a coder manually (manual coding)
VALID	CHAR1	Is This Term Valid (Y/N)	If the source verbatim term is updated or deleted from the original EDC system, set the flag to 'N'. Otherwise, set it to 'Y'
DSDATE	DATETIME20	Source Dataset Creation Date	
RUNDATE	DATETIME20	Coding Program Run Date	

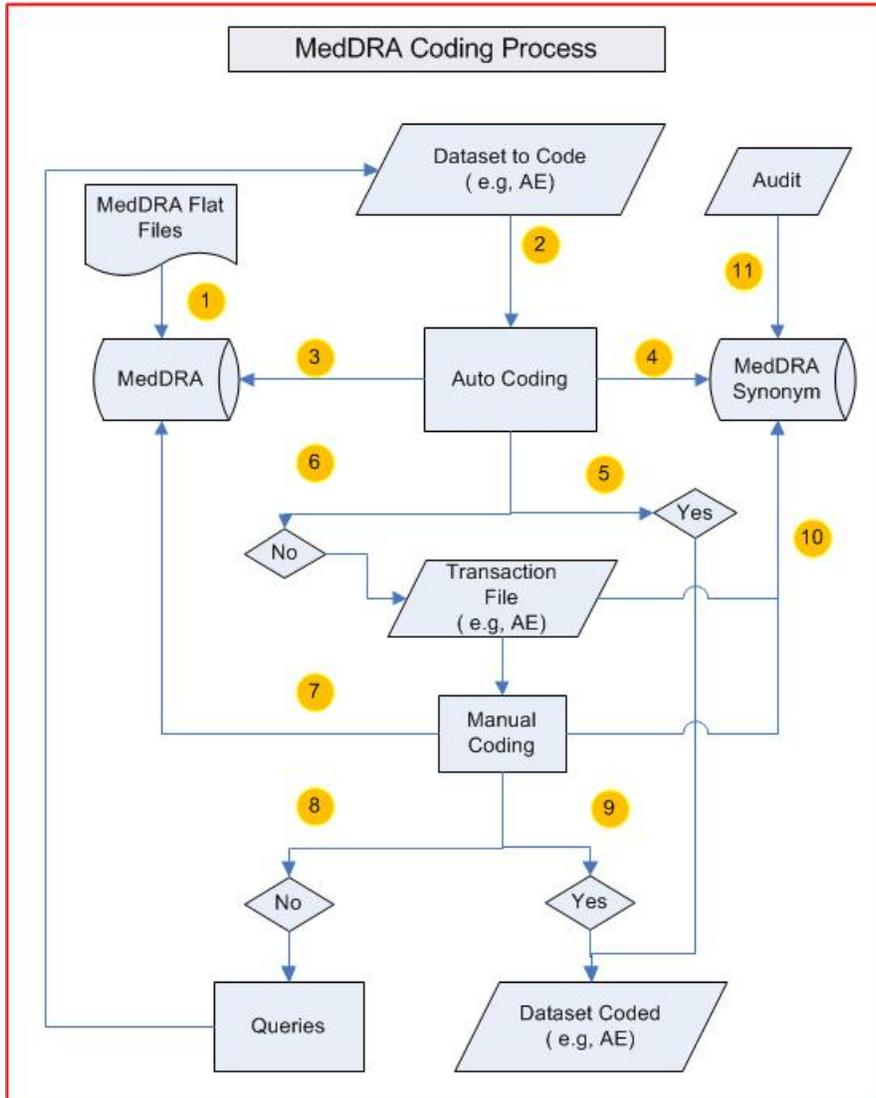
The primary key for the Coding Transaction File is a composite one which includes VTERM, VTMODIFY and LLT. VTERM is the foreign key to the verbatim term in the source dataset (e.g. Adverse Event). VTMODIFY is the foreign key to the VTMODIFY in MedDRA Synonym Dictionary. LLT is the foreign key to the LLT in the MedDRA Dictionary.

THE CODING PROCESS

Please see the diagram below for the whole coding process. The steps below matches what is in the diagram.

A SAS Based MedDRA Coding System

Figure 1: MedDRA Coding Process



1. **MedDRA Dictionary Creation:** The first step is just to create MedDRA dictionary SAS dataset from the downloaded flat files from MSSO. This step is only done once for each version of the MedDRA dictionary.

2. **Dataset to Code:** The coding system can be configured to code against any dataset. But there is one requirement. The dataset must have a verbatim variable to be coded. In the diagram, we use Adverse Event dataset (AE) as an example. Its verbatim term variable is AETERM.

The FDA requires the Preferred Term (PT) and System Organ Class (SOC) to be reported along with the verbatim terms. However, in order to find the correct Preferred Term, we need to find the Lowest Level Term (LLT) first. Once LLT is found, PT and SOC will be found automatically from the MedDRA Dictionary.

The coding system is designed to find the correct LLT for a verbatim term in two ways: Auto coding or manual coding.

3. **Auto Coding against MedDRA Dictionary:** Auto coding means the coding system will find the matching LLT for a verbatim term automatically. No human intervention is required. The auto coding process consists of two parts. The first part is the coding system will try to match a verbatim term to the LLT in the MedDRA dictionary. If a match is found, this verbatim term is coded.

A SAS Based MedDRA Coding System

4. Auto Coding against MedDRA Synonym Dictionary: If the verbatim term does not match to any of the LLT in the MedDRA dictionary, it will try to match the modified verbatim term (VTMODIFY) in the MedDRA Synonym dictionary. If a match is found, this verbatim term is coded. This is part two of the auto coding process.

5. Auto Coded: If auto coding is successful, the verbatim term is coded. So no further action is needed. Please note one LLT may have more than one path to a SOC. But there is only one primary path which the auto coding will select. In some rare occasions, if non-primary path is preferred, manual coding is required.

6. Transaction File: If a verbatim term cannot be coded during the auto coding process, it will be written to the MedDRA Transaction file where manual coding is required.

Many clinical tries will last a few months to a few years. As clinical data are collected continuously, coding is also done on an ongoing basis. The transaction file will accumulate all non-coded terms from source datasets (e.g. Adverse Event, Medical History, etc). Some verbatim terms originally existing in the EDC databases may be deleted or updated to something else. But these original verbatim terms will still stay in the Transaction File. If this happens, their 'VALID' flag will be set to 'N'.

7. Manual Coding: For a verbatim term that does not match to any of the LLT in the MedDRA coding dictionary, a coder can assign a proper LLT term to it directly, or the coder can modify the verbatim term and then assign a proper LLT to it. This process is called manual coding.

Table 3: Examples of Manual Coding

Row	Study ID	VTERM	VTMODIFY	LLT	LLTFLG	VALID
1	999-001	High Blood Pressure	HIGH BLOOD PRESSURE	HYPERTENSION	M	Y
2	999-001	Transient Diarrhea	DIARRHOEA	DIARRHOEA	A	Y
3	999-001	Edema Both Feet	EDEMA FEET	FOOT EDEMA	A	Y
4	999-001	Fever - 38.9c'	FEVER	FEVER	A	N
5	999-001	Diarrhea And Fever				

Row 1 of Table 1: "HIGH BLOOD PRESSURE" does not match to any terms in the MedDRA dictionary and Synonym Dictionary. So the coder assigned a LLT "HYPERTENSION" directly to it. As this is manually coded, the LLTFLAG is "M". The VALID column is "Y", which means the term is still in the EDC database at the time of coding.

Row 2 of Table 1: "TRANSIENT DIARRHOEA" does not match to any terms in the MedDRA dictionary and Synonym Dictionary. A coder modified "TRANSIENT DIARRHOEA" to "DIARRHOEA". Then the coding system will try to match "DIARRHOEA" against the LLT in MedDRA dictionary, and it found a match. So LLT is filled with "DIARRHOEA" and LLTFLAG is set to "A" (auto-coding).

Row 3 of Table 1: VTERM 'EDEMA BOTH FEET' does not match to any term in the MedDRA dictionary and Synonym Dictionary. So a coder modified "EDEMA BOTH FEET" to "EDEMA FEET". When the coding system runs, it found that "EDEMA FEET" matches to "FOOT EDEMA" in the Synonym Dictionary. So LLT is filled with "FOOT EDEMA" and LLTFLAG is set to "A" (auto-coding).

Row 4 of Table 1: VTERM 'FEVER - 38.9C' does not match to any term in the MedDRA dictionary and Synonym Dictionary. So the coder modified it to 'FEVER' (VTMODIFY). It is coded to 'FEVER' (LLT) automatically. However, as VALID is 'N' now, it means the term 'FEVER- 38.9C' was now either deleted or modified to something else in the original EDC database.

8. Queries: If a VTERM can neither be coded nor be modified, a query will be created to send to the investigator's site to update the verbatim term. For example, in Row 5 of Table 1, VTERM "DIARRHEA AND FEVER" cannot be coded directly. A query will be issued to ask the Investigator's site to break this term into two terms: "DIARRHEA" and "FEVER", which will then be coded correctly

After the investigator site answered the query and breaks the term "DIARRHEA AND FEVER" into two terms "DIARRHEA" and "FEVER", the original term "DIARRHEA AND FEVER" will no longer exist in the EDC database. However, it will still be in the Transaction File. So its VALID status will be set to "N".

Though the coding system has the ability to create coding queries automatically, we did not present the query structure here, as it is different from company to company. But the underlying principle is the same. That is the query should include enough information to identify the individual record with issues.

A SAS Based MedDRA Coding System

9. **Dataset Coded:** By doing both auto-coding and manual coding repeatedly, all the verbatim terms in the original dataset will be eventually all coded. All terms in the Transaction File will also be coded.

10. Update of MedDRA Synonym Dictionary:

- **This coding system is a self-learning system:**

This means once a verbatim term is manually coded in the Transaction File at the study level, it will be copied to the MedDRA Synonym Dictionary automatically. As the Synonym Dictionary is used globally for all studies, next time when the same term comes up again in the same study or other studies, the term will be coded automatically. The coder does not need to code it again.

For example, "HIGH BLOOD PRESSURE" is manually coded to "HYPERTENSION" in this study. It is then copied to the MedDRA Synonym Dictionary automatically. When the same term "HIGH BLOOD PRESSURE" appears again in the future, the term will be coded automatically to "HYPERTENSION".

The Synonym Dictionary is initially empty. As all manually coded terms are automatically copied to it, it grows every time the coding system runs. This is good as the more terms are in the Synonym Dictionary; the higher the possibility a verbatim term will be coded automatically. The more the system is used, the smarter the system will become. The more studies are coded by this system, the more time will be saved for the coder.

- **This coding system maintains its integrity by itself:**

As MedDRA Synonym Dictionary is populated with Modified Verbatim Terms (VTMODIFY) and manually assigned Low Level Terms (LLT). One requirement is that one VTMODIFY term must be assigned one and only one LLT term. Otherwise the system won't work properly. However, in reality, it is not rare that a coder may assign more than one LLT terms to the same VTMODIFY term during the coding. If that happens, the system will keep the original LLT and create a report to let the coder know that another LLT is assigned to the VTMODIFY. The coder will then review the report and decide which LLT to keep.

11. Audit Trail:

The MedDRA Dictionary does not need auditing, as it is only created once when a new version is released. Once the dictionary is created, it will never be updated or deleted.

However, it is better to have an audit trail for the MedDRA Synonym Dictionary. As it is constantly updated, we need to know when a modified verbatim term was created, who created, under which study it was created. Also sometimes, we may need to delete (soft deletion) some obsolete terms from the MedDRA Synonym Dictionary. All these information will need to be tracked in the Audit trail dataset.

The structure of the audit dataset is not presented in this paper, as it is an optional component of the system. The audit table has another function, when a record is accidentally deleted or updated in the MedDRA Synonym Dictionary, audit trail will show what has been changed and when it was changed. So it is possible to restore accidentally deleted records.

MEDDRA CODING SYSTEM VERSION CONTROL

1. **MedDRA Dictionary:** Versioning of MedDRA Dictionary is easy, as we just need to download the latest ASCII files and convert them to SAS dataset following the structures described above. MSSO normally releases two versions of MedDRA dictionary each year.

2. **MedDRA Synonym Dictionary:** Versioning of MedDRA Synonym Dictionary is not easy but has to be done. In each new version for MedDRA, new LLTs or PTs will be added. Some LLTs may be promoted to PTs. Some PTs may be demoted to LLTs. Some LLTs will be assigned to different PTs. The text of some LLTs and PPTs may also be updated.

Therefore, it is highly possible (almost certain based on our experience) that some VTMODIFY terms in the MedDRA Synonym Dictionary will be coded to some new LLTs of a newer MedDRA Dictionary. Some VTMODIFY terms may not be coded at all.

To address these issues, we developed a versioning process which includes standard reports, auto-versioning recoding, and manual-versioning recoding. Normally, auto-versioning can re-code 80-90% of existing VTMODIFY terms. But 10-20% of VTMODIFY need to be recoded manually.

MAINTENANCE OF THE CODING SYSTEM

Maintenance of the coding system includes maintenance of the MedDRA Dictionary, MedDRA Synonym Dictionary, and the coding programs.

A SAS Based MedDRA Coding System

MedDRA Dictionary is basically maintenance free. Once it is created, it stays forever as long as it is not obsolete. Once a new version is released, we just need to create the new version.

MedDRA Synonym Dictionary is a self-learning dictionary. It does not require maintenance most of time, as it maintains its integrity by itself. For example, duplicates will not be able to insert into the dictionary. Also one VTMODIFY term will never be able to code to more than one LLT terms. However, in the following scenarios, human intervention is required:

- An existing VTMODIFY term is re-coded to another LLT. Though this does not happen very often, it did happen to us. Sometimes, after a coder coded a VTMODIFY term to a LLT, he/she later found a more appropriate LLT for the VTMODIFY. But due to integrity check, the new LLT cannot be inserted in the Synonym Dictionary automatically.

A program was developed to handle this. This program will code a VTMODIFY term to another LLT as specified by the coder. It also creates an audit trail to document what has been changed.

- To retire a VTMODIFY term, we just need to set the 'VALID' flag to 'N'

Transaction File requires a lot of maintenance as many verbatim terms might be changed or deleted from the source datasets. Fortunately, the maintenance is done by SAS programs. If a verbatim is changed or deleted from the source dataset, the 'VALID' flag in the transaction file will be set to 'N' automatically.

CONCLUSION

This paper described a simple MedDRA coding system built with SAS. It is small, but it is easy to setup and can be used to code any studies. Users can also choose to code against any version of MedDRA. The system is built on a modular basis. So it is easy to maintain and also easy to extend its functionalities. The system supports both auto-coding and manual coding. It includes a synonym dictionary which has self-learning capability. The synonym dictionary greatly improves the auto-coding rate. Manual coding is easy as users only need to interact with one transaction file. The system also handles deleted or updated verbatim terms automatically. It has a built-in audit trail too.

REFERENCES

MedDRA® - the Medical Dictionary for Regulatory Activities (<http://www.meddramsso.com/>). MedDRA® is a registered trademark of the International Federation of Pharmaceutical Manufacturers and Associations.

CONTACT INFORMATION

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

Author Name: Charley Z. Wu
Company: Alexion Pharmaceutical
Address: 352 Knotter Drive
Cheshire, CT 06410
E-mail: charleyzwu@yahoo.com

Author Name: Dona-Lyn Wales
Company: Alexion Pharmaceutical
Address: 352 Knotter Drive
Cheshire, CT 06410
E-mail: WalesD@alxn.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.