PharmaSUG 2021 – Paper DS-51

Coding Data Translation Automation

Chunling Xu, Quting Zhang, Yankun Zhang, Yang Song, BeiGene (Beijing) Co., Ltd.

ABSTRACT

In Jul2020, China NMPA released a guidance for submission of clinical trial data (Provisional). According to the guidance, the database in foreign language should be translated into Chinese in regarding to the dictionary coding terms including adverse events, concomitant medication and medical history. To address the new requirement, we have developed the process and tool to do the translation of SDTM and ADaM coding data.

This paper introduces the coding data translation process and the tips of translation tool development. It starts with the way of translating the English coding data based on the official Chinese dictionary including MedDRA and WHODrug, then introduces the process of communicating with UMC (Upsala Monitoring Centre) to get a formal translation, for the terms without standard translations, we also proposed a process to maintain the translations from version to version.

INTRODUCTION

In Chinese WHODrug dictionary, the information of Preferred Name (DECOD), ATC classification should be displayed in Chinese, as long as the drug has been approved in China, However, it is mentionable that there are still lots of drugs which do not. To align with the regulatory standardization related guidance issued by NMPA, the UMC recommended users not to perform ad hoc translations for these terms whose Chinese name is missing. To address this, we extract the Preferred Names that cannot be automatically translated by Chinese WHODrug dictionary and, send the preferred name list to Medical Coding team to obtain the Chinese Preferred Name translated from UMC based on the Upcoming data in WHODrug Insight.

The purpose and advantages via UMC but not ad hoc translations are:

a) Ensure the Chinese translation is complete, accurate and authorized, and also for maintain the consistency within and between companies.
b) Facilitates the establishment of the company level translation dictionary based on study level UMC translation, in the mean, it also helps to recycle the use of subsequent studies.
c) Ensure the traceability and reproducibility of historical data translation.

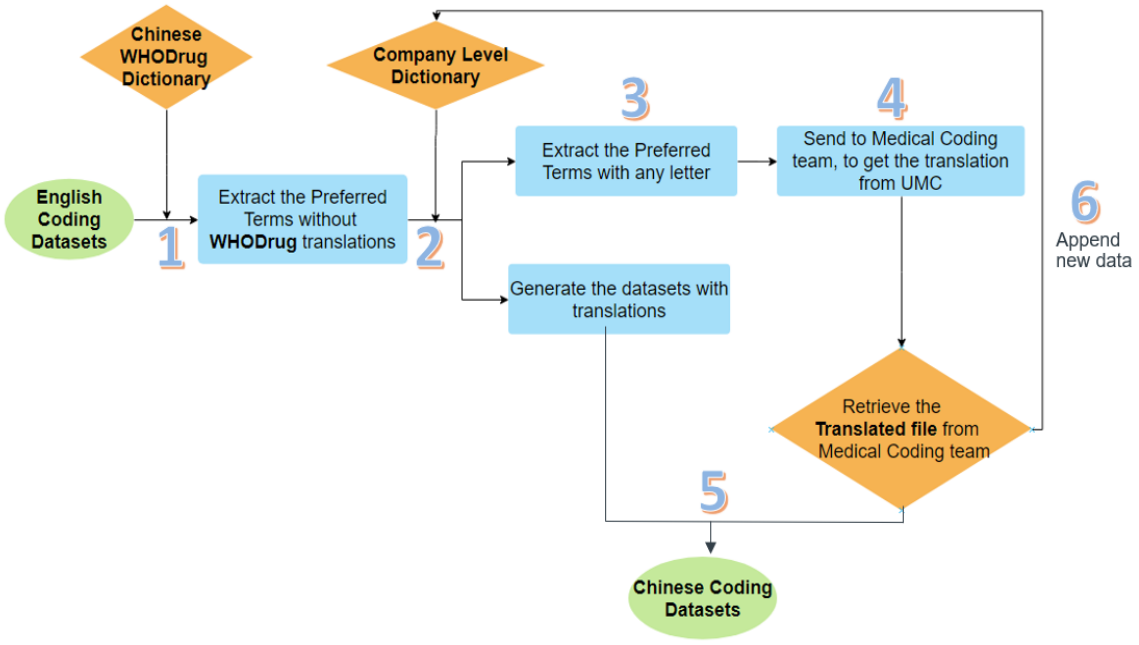
TRANSLATION PROCESS

Since the MedDRA dictionary has a complete Chinese version, the data coded based on the MedDRA dictionary can be directly translated.

For WHODrug dictionary, there are many Preferred Names missing a Chinese translation, so we did as following processing:

a) Firstly, extract the Preferred Names that cannot be automatically translated by Chinese WHODrug dictionary.
b) Send it to Medical Coding team to obtain the Chinese Preferred Name translated from UMC based on the upcoming data in WHODrug Insight.
c) Finally, retrieve the translated file from Medical Coding team, complete the study level dataset translation, and pooling them into company drug translation glossary.

Flowchart



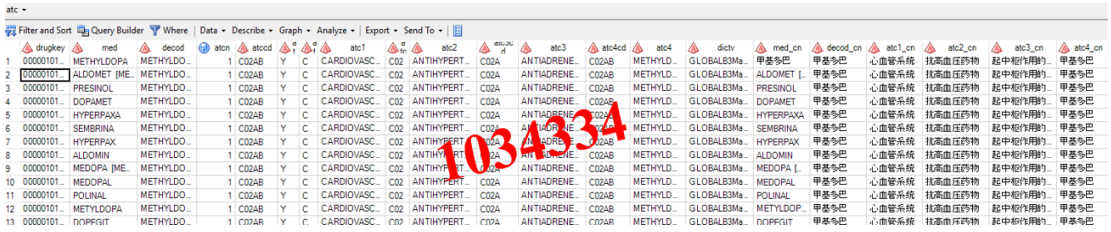
TRANSLATION TIPS

TIP 1: MAPPING ENGLISH DATASETS WITH CHINESE MEDDRA OR WHODRUG DICTIONARY

a) Covert the official dictionary into control dataset format

In Chinese WHODrug dictionary, there are more than one million records in the version which released in Mar2021 (the Dictionary structure is in screenshot 1), and more than one hundred thousand records in Chinese MedDRA dictionary version 24.0, therefore, in ordering to ensure the efficient execution of the programs, it is necessary to covert the official dictionary into the control dataset format (screenshot 2).

Screenshot 1



Screenshot 2

	fmtname	version	label	start
28	$decodcn	sep2019	(6S)-5-METHYLTETRAHYDROFOLATE...	(6S)-5-METHYLTETRAHYDROFOLATE;MECOBALAMIN;PYRIDOXA...
29	$decodcn	sep2019	(RS)-3-甲基-2-氧基戊酸钙	(RS)-3 METHYL-2-OXOVALERIANIC ACID CALCIUM
30	$decodcn	sep2019	(RS)-3-甲基-2-氧基戊酸钙;(RS)-3-甲...	(RS)-3 METHYL-2-OXOVALERIANIC ACID CALCIUM;(RS)-3-METHY...
31	$decodcn	sep2019	(RS)-3-甲基-2-氧基戊酸钙;(RS)-3-甲...	(RS)-3 METHYL-2-OXOVALERIANIC ACID CALCIUM;(RS)-3-METHY...
32	$decodcn	sep2019	(RS)-3-甲基-2-氧基戊酸钙;(RS)-3-甲...	(RS)-3 METHYL-2-OXOVALERIANIC ACID CALCIUM;(RS)-3-METHY...
33	$decodcn	sep2019	(RS)-3-甲基-2-氧基丁酸钙	(RS)-3-METHYL-2-OXOBUTYRIC ACID CALCIUM
34	$decodcn	sep2019	(RS)-3-甲基-2-氧基丁酸钙;α-酮戊二...	(RS)-3-METHYL-2-OXOBUTYRIC ACID CALCIUM;ALPHA-KETO-IS...
35	$decodcn	sep2019	(RS)-3-甲基-2-氧基丁酸钙;ALPHA-K...	(RS)-3-METHYL-2-OXOBUTYRIC ACID CALCIUM;ALPHA-KETO-IS...
36	$decodcn	sep2019	1,2-OCTANEDIOL;尿囊素;BISABOLO...	1,2-OCTANEDIOL;ALLANTOIN;BISABOLOL;BORAGO OFFICINALIS...

b) Align the Chinese terms for the same English ATC with different Chinese translations

For the WHODrug dictionary released in Mar2020, there are several records that the ATC English terms are same but the Chinese translation is different. Per NMPA's guideline, the translations should be consistent among all the documents, thus we send these terms to UMC for alignment, and then study level datasets will be translated based on the aligned version. The detailed information is in screenshot 3.

Screenshot 3

	Version	ATC Level	Code	English	Chinese	UMC Translation Future Dictionary Release
2	sep2020	$atc4cn	C02CA	ALPHA-ADRENORECEPTOR ANTAGONISTS	ALPHA-肾上腺素受体拮抗药	
3	sep2020	$atc4cn	G04CA	ALPHA-ADRENORECEPTOR ANTAGONISTS	A-肾上腺素受体拮抗药	ALPHA-肾上腺素受体拮抗药
4	sep2020	$atc4cn	R01AC	ANTIALLERGIC AGENTS, EXCL. CORTICOSTEROIDS	非皮质甾体激素类抗过敏药	非皮质类固醇抗过敏药
5	sep2020	$atc4cn	A07EB	ANTIALLERGIC AGENTS, EXCL. CORTICOSTEROIDS	非皮质类固醇抗过敏药	
6	sep2020	$atc2cn	C07	BETA BLOCKING AGENTS	BETA-受体阻断药	
7	sep2020	$atc4cn	S01ED	BETA BLOCKING AGENTS	B-受体阻断药	BETA-受体阻断药
8	sep2020	$atc3cn	S02B	CORTICOSTEROIDS	皮质甾体激素类药物	皮质类固醇类药物
9	sep2020	$atc4cn	C05AA	CORTICOSTEROIDS	皮质类固醇类药物	
10	sep2020	$atc4cn	S01BA	CORTICOSTEROIDS, PLAIN	皮质甾体激素类，单方	皮质类固醇，单方
11	sep2020	$atc3cn	D07A	CORTICOSTEROIDS, PLAIN	皮质类固醇，单方	
12	sep2020	$atc4cn	N02BG	OTHER ANALGESICS AND ANTIPYRETICS	其他解热镇痛药	
13	sep2020	$atc3cn	N02B	OTHER ANALGESICS AND ANTIPYRETICS	其它解热镇痛药	其他解热镇痛药
14	sep2020	$atc2cn	B06	OTHER HEMATOLOGICAL AGENTS	其他血液学用药	
15	sep2020	$atc4cn	B06AX	OTHER HEMATOLOGICAL AGENTS	其他血液学药剂	其他血液学用药

TIP 2: EXTRACT THE TERMS WITHOUT STANDARD TRANSLATION AND SEND TO UMC

a) Develop communication document

Since each company may have several submissions to submit each year, it is particularly critical to develop a standard model, which can not only improve work efficiency, but also reduce communication barriers, especially in cross function and external cooperation. Therefore, the programmer and the Medical Coding team established an EXCEL template with fixed header, the details information is in screenshot 4, The red dashed box is filled by UMC.

b) What needs to be set to UMC

The Preferred terms containing any letter from columns translated according to WHODrug and company-level dictionaries will be extracted, as screenshot 4 in blue dashed box, the 'NOS' in column 'Preferred term in Chinese' will be translated into '未特指' as displayed in UMC_TRANSLATION.

c) Dictionary version

For the dictionary version, the UMC recommends the format like "WHODRUG GLOBAL B3 March 1,20XX + UPCOMING B3", as UMC translation is based on Upcoming data in WHODrug Insight, and there should be an explanation in csdrg and adrg, refer to Screenshot 5 for the detail information.

Screenshot 4

CMPRDRCD	CMDECOD_EN	CMDECOD_CN	DICTIONARY_VERSION	FLAG	UMC_TRANSLATION	COMMEN
Preferred Code	Preferred term in English	Preferred term in Chinese	Dictionary Version	Is New Record?	UMC Official Translation Upcoming	Comment
148917 02 001	ACONITUM KUSNEZOFFII ROOT TUBER;ANGELICA SINENSIS ROOT;ANIMAL FECES NOS;BOSWELLIA SPP. RESIN;COMMIPHORA MYRRHA RESIN;HERBAL NOS;LIQUIDAMBAR FORMOSANA	草乌块根;当归根;动物粪便NOS;乳香;地丁树树脂;草药 NOS;枫香脂;木鳖子;麝香;地龙	WHODRUG GLOBAL B3 March 1,2020 + UPCOMING B3		草乌块根;当归根;动物粪便(未特指);乳香属树脂;没药;草药(未特指);枫香脂;木鳖子种子;麝香;地龙	
008169 01 001	ARGININE GLUTAMATE	ARGININE GLUTAMATE	WHODRUG GLOBAL B3 March 1,2020 + UPCOMING B3		精氨酸谷氨酸	

Screenshot 5

标准或词典	版本
SDTM	•SDTM v1.4 •SDTM-IG v3.2
药物词典	WHODRUG GLOBAL B3 March 1, 2020 + UPCOMING B3

UPCOMING B3 是指即将要发布的WHODD Global中文版本，纳入此版本是由于首选术语基于当前已发布的版本并无翻译，UMC依据即将发布的版本对其进行了补充翻译。

TIP 3: APPEND THE UMC TRANSLATION TO COMPANY LEVEL DICTIONARY

The company level dictionary was built as the terms that translated by UMC on study level, for confirm the manual translation can be reused well. To build the company level dictionary, several points should be considered.

a) The company level dictionary format

The dictionary including the variables that from study level EXCEL template listed in screenshot 4, and also add START and End date, that are used to control the translation version, if any translations were updated by UMC, the start and end date should be updated.
And to ensure that all translated versions can be archived, and the translation of historical study is traceable.

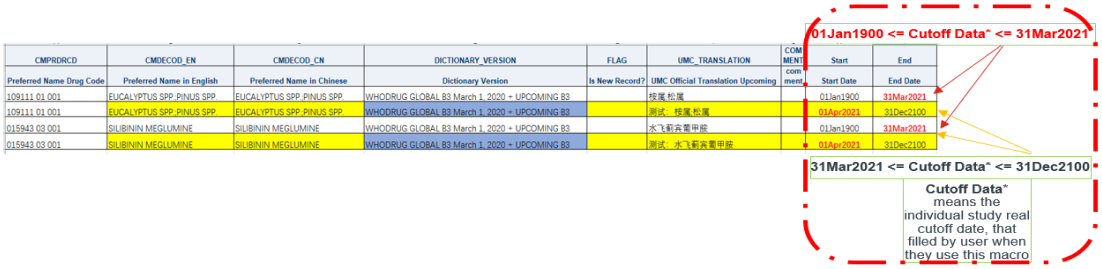
Screenshot 6

	CMDECOD_EN	CMDECOD_CN	UMC_TRANSLATION	COMMENT	CMPRDRCD	DICTIONARY_VERSION	FLAG	START	END
1	(RS)-3 METHYL-2-...	(RS)-3-甲基-2-氨基...	消旋酮异亮氨酸钙 酮缬氨酸...		115642 02 001	WHODRUG GLOBAL B3 M...		01JAN1900	31DEC2100
2	ABELMOSCHUS M...	ABELMOSCHUS MA...	黄蜀葵		070671 01 001	WHODRUG GLOBAL B3 M...		01JAN1900	31DEC2100
3	ACALABRUTINIB	ACALABRUTINIB	阿可替尼		089179 01 001	WHODRUG GLOBAL B3 M...		01JAN1900	31DEC2100

b) The dictionary maintenance

When UMC updates the translation, the previous version translation end date should be update into a date which is not have this version translation, and the new translation start date is the date that need update translation, the study translation based on which version is determined by study level cutoff date, for details please refer to the screenshot 7.

Screenshot 7



CONCLUSION

This paper is intended to introduce the way for translating the coding data, and the tips for dealing with the preferred terms that without standard translations. By applying the standard process, it can improve the overall efficiency from company level and make programmers life much easier.

REFERENCES

Guideline on the Submission of Clinical Trial Data. Center for Drug Evaluation, NMPA. July 2020
http://www.cde.org.cn/news.do?method=largeInfo&id=7a43c3abfde95950

ACKNOWLEDGMENTS

We would like to thank and acknowledge my colleagues Quting Zhang, Yankun Zhang and Yang Song, who provided us invaluable support. We are very grateful for their kind support and guidance.

CONTACT INFORMATION

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

Chunling Xu
BeiGene (Beijing) Co., Ltd.
Chunling.Xu@BeiGene.com

Quting Zhang
BeiGene (Beijing) Co., Ltd.
Quting.Zhang@BeiGene.com

Yankun Zhang
BeiGene (Beijing) Co., Ltd.
Yankun.Zhang@BeiGene.com

Yang Song
BeiGene (Beijing) Co., Ltd.
Yang.Song@BeiGene.com