ePaper RE05

Mapping Real-World Data (RWD) to SDTM in External Control Arm (ECA) Studies and Registry Studies

Yanhui (Kevin) Yang, Bristol Myers Squibb, New Jersey, USA
Zhouzhou (Joe) Xi, Bristol Myers Squibb, New Jersey, USA

ABSTRACT

This paper shares our experience in mapping RWD to SDTM from different perspectives including:

1. How to organize safety data into proper domains;
2. How to define the reference date in an ECA study;
3. How to handle complicated baseline information;
4. A quick review of Pn21 compliance report;

INTRODUCTION

The use of Real-World Data (RWD) to support clinical trials has become popular recently in the pharmaceutical industry. Besides conventional observation study design, the usage of RWD has become increasingly common in other types of designs such as external control arm (ECA) or registry studies, to fulfill health authority requirements.

In this situation, the standardization of RWD for the purpose of pooling or submission becomes critical. To meet the needs, all sides in the pharmaceutical industry have put their effort. For example, in Oct 2021, FDA published the draft guidance of 'Data Standards for Drug and Biological Product Submissions Containing Real-World Data, focusing on US FDA-supported data standards in drug submissions with data derived from RWD to promote compliance with relevant legal requirements. Also, CDISC recently drafted a guidance "Considerations and Examples of Using SDTM for Observational Studies" on this topic, and PHUSE also established its own committee working on future guidance.

But so far, the industry still lacks the experience of mapping and submitting RWD in SDTM/ SDTM-like or other standard formats. Recently, we have used one ECA study and four registry studies as pilot studies to map RWD to SDTM data sets following current available guidance. In this proceeding, we are going to share our experience and the lessons learned from this effort, and it is covered in the following topics.

AE (ADVERSE EVENTS) VS OTHER SUPPORTING DOMAINS

In clinical data, safety related information is usually collected by specific Case Report Form (CRF) page according to its targeted domains. E.g. adverse events along with its related information are collected by AE page, and details of clinical data collection have been guided by CDASH standards. But this is not the case with RWD, where the information of a particular safety concern is usually lumped together, and in this case, properly parsing them into different specific domains becomes necessary.

Based on experience of some registry studies, where safety data is a major concern, we propose a dichotomic approach for the parsing, which is AE + Other domain. One example we have is for infection data from one of our registry studies

In AE:

STUDYID	USUBJID	DOMAIN	AESEQ	AETERM	AEDECOD	AEBODSYS	AESER	AEACN	AEOUT	AETOXGR	AESTDTC	AEENDTC	AESTDY	AEENDY
Study-1	Study-1-001	AE	1	CLINICAL SIGNIFICANT INFECTION	CLINICAL SIGNIFICANT INFECTION	INFECTION	Y	xx	xx	xx	xxxx-xx-xx	xxxx-xx-xx	xx	xx

In IS:

STUDYID	USUBJID	DOMAIN	ISSEQ	ISGRPID	ISTESTCD	ISTEST	ISORRES	ISSTRESC	ISRESCAT	ISLOC	ISDTC	EPOCH	ISDY
Study-1	Study-1-001	IS	1	REPORTED ORGANISM 01	VIRIDENT	Viral Organism Identification	Norovirus	Norovirus	INFECTING	GI_LOWER	xxxx-xx-xx	Post Infusion	xx
Study-1	Study-1-001	IS	2	REPORTED ORGANISM 02	VIRIDENT	Viral Organism Identification	Respiratory syncytial virus (RSV)	Respiratory syncytial virus (RSV)	INFECTING	SINUS_RESP	xxxx-xx-xx	Post Infusion	xx
Study-1	Study-1-001	IS	3	REPORTED ORGANISM 03	BCTIDENT	Bacterial Organism Identification	Clostridium difficile	Clostridium difficile	INFECTING	GI_LOWER	xxxx-xx-xx	Post Infusion	xx

Figure 1: Sample of infection data in AE and IS domain

The general safety summary requires this event (clinically significant infection) to be counted as an AE, but details of each infection category are not fit for standard SDTM.AE domain (e.g., viral infection organism). After internal discussion, it is decided to split the information into two SDTM domains: AE + IS. To be more specific, the general

information about the patient's infection status including earliest start date and latest end date is kept in AE domain with AETERM='Clinical significant Infection', while the detailed infection category/site/timing go to IS, which is a BDS (Basic Data Structure) type and designed specifically for specimen assessment. Similarly, we can save CRS (Cytokine Release Syndrome)/NT (Neurotoxicity) events in AE domain, but their signs and detailed symptoms go to CE (Clinical Events) domain.

DEFINE REFERENCE/TREATMENT STARTING DATE AND MAPPING OF THERAPIES (EX VS CM)

In a typical ECA study, the real-world data is taken as control arm of the study to compare with the study medication arm. In this case, how to identify the designated therapy from real-world patient for comparison and organize this therapy and other prior/post therapies into SDTM domain becomes a key problem.

Using our current data from multiple myeloma patients as an example, we rely on either our internal clinical team, or AI/data-visualization platform to define the lines of therapy of each patient. Below is an example of one patient's line of therapies defined, such as:

Line of Therapy

			Treatment		
Index	Line	Comment	Treatment	Start Date	End Date
	1		rituximab	2016-11-08	2016-11-18
	2		bendamustine,rituximab	2016-11-29	2017-01-17
1	3		cisplatin,cytarabine,etoposide,methylprednisolone,rituximab	2017-02-10	2017-02-10
	1		rituximab	2017-02-13	2017-06-20
	4		doxorubicin,prednisone,rituximab	2017-06-13	2017-07-14
	5		cyclophosphamide,doxorubicin,prednisone,rituximab,vincristine	2017-07-28	2017-07-28

Figure 2: Sample of line of therapy definition

Then the clinical team needs to indicate one line of therapy as index line (the line taken as study medication to be compared in the study). In the example shown in fig 1, the patient's third regimen (cisplatin, cytarabine, etc. from 2017-02-10) is decided to be the first line/index line of therapy, thus its information is mapped to SDTM.EX such as:

STUDYID	USUBJID	DOMAIN	EXSEQ	EXTRT	EXCAT	EXSTDTC	EXENDTC	EXSTDY	EXENDY
Study-1	Study-1-001	EX	1	cisplatin	SYSTEMIC THERAPY	2017-02-10	2017-02-10	1	1
Study-1	Study-1-001	EX	2	cytarabine	SYSTEMIC THERAPY	2017-02-10	2017-02-10	1	1
Study-1	Study-1-001	EX	3	etopside	SYSTEMIC THERAPY	2017-02-10	2017-02-10	1	1
Study-1	Study-1-001	EX	4	methylprednisolone	SYSTEMIC THERAPY	2017-02-10	2017-02-10	1	1
Study-1	Study-1-001	EX	5	rituximab	SYSTEMIC THERAPY	2017-02-10	2017-02-10	1	1

Figure 3: Sample of SDTM.EX domain

Accordingly, the start date of this therapy is taken in DM.RFSTDTC as the overall study reference start date of this patient. At the same time, the clinical team also needs to decide the reference end date of index line of therapy, considering the overall clinical impact of this therapy. The patient's other therapies either before or after index therapy can be mapped to SDTM.CM as prior/subsequent therapies.

USE CUSTOMIZED DOMAIN TO HOLD BASELINE CHARACTERISTICS

In clinical trial data, baseline records need to be flagged out in most finding domains. A flag variable like --BLFL is expected, with the derivation logic as last non-missing result prior to first dose. This logic is clear and has been accepted in clinical SDTM mapping process, but is this still the right choice or good choice for Real-World Data? Given the data structure of RWD in our cell therapy studies, clinical characteristics were captured as pre-dosing, which is the last result before the treatment. Another consideration is that the missing exact testing dates for pre-infusion records. It is challenging to trace back the testing date and apply the derivation logic to identify the baseline record. Third consideration is that all data are lumped in one dataset, no separated raw data as clinical trial datasets in our study. Therefore, the categorization of records into each separated domain and then identifying the baseline records would be very tedious. As a result, all those pre-infusion clinical characteristics were mapped to a customized domain - baseline (BL) domain. A lot of advantages were found with this methodology: less challenges from categorizing each clinical characteristics information, simplified the programming process, and beneficial to the subsequent baseline analysis.

STUDYID	DOMAIN	USUBJID	TESTCD	TEST	BLORRESC
Study01	BL	Study01-001-00000001	ECOGPR	ECOG performance status prior to CT	1
Study01	BL	Study01-001-00000002	ECOGPR	ECOG performance status prior to CT	1
Study01	BL	Study01-001-00000003	ECOGPR	ECOG performance status prior to CT	3
Study01	BL	Study01-001-00000004	ECOGPR	ECOG performance status prior to CT	1
Study01	BL	Study01-001-00000001	EN1PLPR	>1 extranodal site prior to CT	N
Study01	BL	Study01-001-00000002	EN1PLPR	>1 extranodal site prior to CT	Y
Study01	BL	Study01-001-00000003	EN1PLPR	>1 extranodal site prior to CT	N
Study01	BL	Study01-001-00000001	RXSYSPR	Prior systemic therapy	Y
Study01	BL	Study01-001-00000002	RXSYSPR	Prior systemic therapy	Y
Study01	BL	Study01-001-00000003	RXSYSPR	Prior systemic therapy	Y

Figure 4: Sample of SDTM.BL

SAMPLE OF PINNACLE21 COMPLIANCE CHECK USING RWD

Completed SDTM domains were uploaded to Pinnacle 21 to check for CDISC compliance. There are a lot of issues identified by P21, and here we emphasize several key issues related to real-world data.
In AE domain, one of major error is SD0008, which says that values for AEDECOD were not found in MedDRA dictionary. As we know this is a challenge for RWD, since AE terms usually are input as a mixture of control terms, multiple terms combined in one, and free text. To solve this issue, the MedDRA coding team is actively involved, and each AE term was manually mapped to correct Controlled Term.

SD0021 is another major issue in AE domain, which means there are records with Missing End Time-Point values. This is a known challenge for RWD too, since a lot of adverse event records with events ending time not provided. The same issue was found in CE and CM domains too.

In DD domain, issue SD1117 (Duplicate records) was identified by P21. According to IG, the structure of Finding class domains should be one record per finding result per subject. While in RWD data, multiple secondary causes of death are captured, and they must be mapped to the DD domain. But they are not duplicate records.

STUDYID	DOMAIN	USUBJID	DDSEQ	DDTESTCD	DDTEST	DDORRES	DDSTRESC	DDDTC	DDDY
Study01	DD	Study01-001-00000001	1	COCDTH	Contributing Cause of Death	Cardiac failure	Cardiac failure	2022-12-23	10
Study01	DD	Study01-001-00000001	2	COCDTH	Contributing Cause of Death	Pulmonary failure	Pulmonary failure	2022-12-23	10
Study01	DD	Study01-001-00000001	3	COCDTH	Contributing Cause of Death	Cytokine release syndrome (CRS)	Cytokine release syndrome (CRS)	2022-12-23	10
Study01	DD	Study01-001-00000001	4	PRCDTH	Primary Cause of Death	Recurrence/persistence/progression of p	Recurrence/persistence/progression of primary disease	2022-12-23	10

Figure 5: Sample of Pn21 compliance check on DD

In LB domain, the major issue is TS0057, which means that LBSTRESN is populated, but LBSTNRHI is not populated. For laboratory tests where normal ranges are applicable, if a numeric result is provided, a normal range high should be provided. But in our RWD, no normal ranges were captured.

With consideration of all these issues, we can clearly conclude that Real-World Data is quite different from Clinical trial data, and it is challenging to map RWD into SDTM as standards for clinical trial data. New standards fit for RWD are needed.

CONCLUSION

To support data source standardization and future submission to health authority, we are conducting SDTM mapping for real-world data. But due to complexity of RWD and lack of clear industry guidance/experience, this mapping requires a lot of exploration through dark tunnel. This paper shares some of our experience in mapping RWD to SDTM from different perspectives.

This experience reflects the issues and our initial effort to solve from our current RWD studies, which can be preliminary and not complete. We hope our sharing in this paper can help better shaping of the industry standards for mapping RWD into CDISC SDTM format and getting ready for submitting them to health authority.

REFERENCES

CDISC 2023-06 *Considerations and Examples of Using SDTM for Observational Studies*
FDA 2021-10 *Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry*

CONTACT INFORMATION

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

Author Name: Kevin Yang
Company: Bristol Myers Squibb
Email: yanhui.yang2@bms.com

Author Name: Joe Xi
Company: Bristol Myers Squibb
Email: joe.xi@bms.com