

Challenges in implementing standards for Real-world data

Ritesh Dhimmar

Programming Leader

Clinical Programming and Business Excellence

GSK, Collegeville, PA

Raman Bassi

Director, eData and Submission Support

Statistical Programming and Analysis

Bayer, Whippany, NJ

Background

The utilization of Real-World data in clinical and healthcare research is not a new concept.

- Post-marketing safety monitoring and surveillance
- Efficiency or effectiveness claims in rare cases
- Planning for an RCT
- Managing and improving a healthcare practice

Recent US Legislations supporting RWD/RWE utilization

The 21st Century Cure Act, December 13, 2016

- Designed to accelerate the development and delivery of medical products to the patients
- Mandated FDA to develop a framework for evaluating the potential of using RWD/RWE

The 6th Prescription Drug User Fee Act (PDUFA VI), August 18, 2017

- Designed to enhance efficiencies of development and approval of medicines by timely regulatory review
- Required FDA to initiate activities to address key issues in using RWE in regulatory decision-making

The Health Information for Economic and Clinical Health (HITECH), February 17, 2009

- Promoted increase adoption of EHR resulting in vast RWD in electronic format
- Prior to HITECH Act, only 10% of hospitals had adopted EHRs

Definitions

Real World Data (RWD)

- data related to patient health status and/or the delivery of health care routinely collected from a variety of sources such as
 - electronic health records (EHRs)
 - claims and billing data
 - product and disease registries
 - Patient generated data
 - mobile devices, wearables, and biosensors

Real World Evidence (RWE)

- clinical evidence about the usage and potential benefits or risks of a medicinal product derived from analysis of RWD
 - observational studies
 - hybrid “pragmatic” studies
 - External control arm
 - Randomized clinical trials

RWE frameworks

Framework for FDA's Real-World Evidence Program, December 2018

- Evaluate the potential use of RWE to support
 - approval of new indications for drugs or biological products already approved by FDA
 - postapproval study requirements
- Framework considerations
 - whether RWD is fit for use
 - whether the study design can provide adequate evidence
 - whether the study conduct meets regulatory requirements

EU's OPERational, TechnIcal, and MethodologicAl (OPTIMAL) framework for regulatory use of RWE, 2019

- Appropriate use of valid RWE for regulatory purposes (e.g., safety, efficacy, benefit-risk monitoring)
- Framework considerations
 - desired criteria for acceptability of RWE
 - challenges with use of RWD to generate acceptable RWE
 - Possible solutions (EU context)

Mandated by these frameworks, both FDA and EMA released various guidance documents

FDA Guidance regarding RWD submission

Data Standards for Drug and Biological Product Submissions Containing Real-World Data – Guidance for Industry, October 2021

Purpose: provides recommendations to sponsors for complying with section 745A(a) of FD&C by using data standards specified in the data standards catalog when submitting study data from RWD sources in a drug submission

- FDA recognizes the challenges to standardize the data from RWD sources and plans to provide further guidance and/or update the catalog with standards for RWD
- currently, absent a waiver, in accordance with section 745A(a), study data (including RWD) in a specific submission type must be submitted electronically in a format supported by FDA
- with adequate documentation on data transformation, conversion, or mapping, study data from RWD source can be converted into SDTM and submitted to FDA

Recommended Documentation:

- data transformation approaches to FDA supported standards in study protocol, DM Plan, and/or final study report
- description of general approach and anticipated impact of data mapping to CDISC elements in Study Data Reviewer's Guide
- technical details about data transformation and mapping in define.xml
- any challenges encountered in data transformation and their solutions in Study Data Reviewer's Guide

FDA Data Standards Catalog

- Submission Data Exchange Standards

Data Exchange Standard	Supported Version	Supported IG Version	Date Support Begins	Date Requirement Begins
SDTM	1.4	3.2	08/17/2015	03/15/2018[1] 03/15/2019[2]
SDTM	1.7	3.3	03/15/2021	03/15/2023
ADaM	2.1	1.1	03/15/2018	03/15/2019[1] 03/15/2020[2]
ADaM	2.1	1.2, 1.3	07/18/2022	03/15/2024

[1] For NDAs, ANDAs, and BLAs

[2] For INDs

- Submission Terminology Standards

- CDISC-CT
- MedDRA
- WHO DD

Mapping of Optum database into CDISC SDTM- An assessment

- Assess the magnitude of mapping the sample Optum database into CDISC SDTM structure
- The Optum® Clinformatics® Data Mart is a database comprised of administrative health claims for members of a large nationally managed care company affiliated with Optum
- Anonymized patient level data is organized in the following tables

Dataset Name	Dataset Description
MEMBER_CONTINUOUS_ENROLLMENT	Member Eligibility
MEMBER_ENROLLMENT	Member Detail Eligibility
MEDICAL	Medical Services
MED_DIAGNOSIS	Medical Diagnosis
MED_PROCEDURE	Medical Procedure
CONFINEMENT	Inpatient Confinement
RX	Prescription Claims
LABRESULT	Lab Results
PROVIDER	Provider
PROVIDER_BRIDGE	Provider Bridge
DOD	Date of Death
LU_DIAGNOSIS	Diagnosis Lookup
LU_NDC	NDC Lookup
LU_PROCEDURE	Procedure Lookup

Optum to SDTM-IG mapping

Member Detail Eligibility Data Fields					
Variable	Description	Type	Example value	SDTM Variable(s)	Mapping Specifications
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	DM.USUBJID	
GDR_CD	A code identifying the sex of the member M=male, F=female, U=unknown	Char	M	DM.SEX	Transform the non-standard data to standardized data using CDISC CT
RACE	Rolled Up Ethnic Code.	Char	W	DM.RACE DM.ETHNIC	Transform the non-standard data to standardized data using CDISC CT
YRDOB	The member's year of birth, capped at 90 years.	Num	1994	DM.BRTHDTC	
STATE	The two-character US Postal Service Abbreviation for the state name.	Char	NJ	COUNTRY="USA"	
Date of Death Data Fields					
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	DM.USUBJID DS.USUBJID	
YMDOD	Month and Year of Death on the Death Master File where the Individual has been matched to the file.	Num	2022-12-25	DM.DTHDTC DM.DTHFL DS.DSSTDTC	

Any other study-specific variable from Optum dataset will be included in the respective supplemental domain.

Optum to SDTM-IG mapping

Prescription Claims Data Fields					
Variable	Description	Type	Example value	SDTM Variable(s)	Mapping Specifications
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	DM.USUBJID EX.USUBJID CM.USUBJID	
FILL_DT	Date the prescription was filled by the pharmacy.	Num	2020-11-23	DM.RFSTDTC DM.RFXSTDTC EX.EXSTDTC CM.CMSTDTC	
GNRC_NM	Contains the drug ingredient name for a specific drug as adopted by the United States Adopted Names.	Char	IBUPROFEN	EX.EXTRT CM.CMTRT	Data transformation may be needed to facilitate standard dictionary coding (e.g., WHO DD)
STRENGTH	Description of drug potency in units of grams, milligrams, percentage, and other terms.	Char	600 MG	EX.EXDOSE EX.EXDOSU CM.CMDOSE CM.CMDOSU CM.CMDOSTXT	Information can also be extracted from LU_NDC
NDC	The unique code that identifies a drug product as defined by the National Drug Data File	Char	53746046505	EX.EXDOSFRM CM.CMDOSFRM EX.EXROUTE CM.CMROUTE	Information Extracted from LU_NDC

Any other study-specific variable from Optum dataset will be included in the respective supplemental domain.

Optum to SDTM-IG mapping

Medical Procedure Data Fields					
Variable	Description	Type	Example value	SDTM Variable(s)	Mapping Specifications
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	DM.USUBJID	
PROC	Claim Level 1 ICD-X procedure code off of the header portion of the claim.	Char	0JX50ZC	PR.PRTRT	Decode from LU_PROCEDURE
FST_DT	The beginning date for the service, event, or confinement being billed by the provider.	Char	2019-10-23	PR.PRSTDTC	
Medical Diagnosis Data Fields					
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	AE.USUBJID MH.USUBJID	
DIAG	Level 1 ICD-X as entered on the claim (without decimal point).	Char	80316	AE.AETERM MH.MHTERM	Decode from LU_DIAGNOSIS AETERM if FST_DT is GE RFXSTDTC MHTERM if FST_DT is LT RFXSTDTC Data transformation may be needed to facilitate standard dictionary coding (e.g., MedDRA).
FST_DT	The beginning date for the service, event, or confinement being billed by the provider.	Char	2016-12-23	AE.AESTDTC MH.MHSTDTC	AESTDTC if FST_DT is GE RFXSTDTC MHSTDTC if FST_DT is LT RFXSTDTC

Any other study-specific variable from Optum dataset will be included in the respective supplemental domain.

Optum to SDTM-IG mapping

Medical Service Data Fields					
Variable	Description	Type	Example value	SDTM Variable(s)	Mapping Specifications
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	DM.USUBJID	
CONF_ID	Confinement Identifier.	Char	K4LZ4TZOT	PR.PRREFID	
PROC_CD	CPT/HCPCS Procedure code that describes the service provided.	Char	0JX50ZC	PR.PRTRT	Decode from LU_PROCEDURE
FST_DT	The beginning date for the service, event, or confinement being billed by the provider.	Char	2019-10-23	PR.PRSTDTC	
DRG	The Diagnosis Related Group (DRG) Code assigned by the source system.	Char	775	PR.PRINDIC	Decode from DRG codelist

Any other study-specific variable from Optum dataset will be included in the respective supplemental domain.

Optum to SDTM-IG mapping

Lab Results Data Fields					
Variable	Description	Type	Example value	SDTM Variable(s)	Mapping Specifications
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	LB.USUBJID	
FST_DT	The beginning date for the service, event, or confinement being billed by the provider.	Num	2020-11-23	LB.LBDTC	
LOINC_CD	A universal identification number for a particular test or analytic, based on Logical Observation Identifiers Names and Codes (LOIN); a registered trademark of the Regenstrief Institute.	Char	718-7	LB.LBLOINC	
TEST_DESC	Internal analytic name, specific to each laboratory vendor.	Char	HEMOGLOBIN	LB.LBTEST	Transform the non-standard data to standardized data using CDISC CT
HI_NRML	Numeric value for high-end of the normal range for a specific test or analytic.	Num	17	LB.LBORNRHI	
LOW_NRML	Numeric value for low-end of the normal range for a specific test or analytic.	Num	12.5	LB.LBORNRLO	
RSLT_NBR	Contains a numeric value if such is applicable for a particular test result. Please note that not all vendors give a separate numeric field. Numeric results may also be in the Result Text field.	Num	14.8	LB.LBORRES	
RSLT_TXT	Result value of the test. This can be a numeric or alpha numeric value.	Char	NEGATIVE	LB.LBORRES	Transform the non-standard data to standardized data using CDISC CT
RSLT_UNIT_NM	Contains the units upon which the result is based (milliliters, deciliters, percentage, et al). Please note that not all vendors give this field.	Char	g/dL	LB.LBORRESU	Transform the non-standard data to standardized data using CDISC CT

Any other study-specific variable from Optum database will be included in the respective supplemental domain.

Optum to SDTM-IG mapping

Inpatient Confinement Data Fields					
Variable	Description	Type	Example value	SDTM Variable(s)	Mapping Specifications
PATID	An encrypted, system-generated number that identifies an individual across multiple groups/policies.	Num	1111111111	HO.USUBJID	
CONF_ID	Confinement Identifier.	Char	K4LZ4TZOT	HO.HOREFID	
				HOTERM="Inpatient Confinement"	
ADMIT_DATE	Admission Date of Confinement (YYYYMMDD)	Num	2020-11-23	HO.HOSTDTC	
DISCH_DATE	Discharge Date of Confinement (YYYYMMDD)	Num	2020-11-25	HO.HOENDTC	
LOS	Length of Stay from start of first confinement record to end of last confinement record.	Num	3	HO.HODUR	

Any other study-specific variable from Optum database will be included in the respective supplemental domain.

Challenges in mapping RWD to standards

- Information unavailable to populate many required/expected variables
- Duplicate data
- Missing data
- Incomplete data
- Start dates might not be an actual start date
- End dates of events and interventions are unavailable
- Some diagnosis and procedure codes do not match with the look-up tables
- No standards used in collecting the lab test name, character results, and units
- TEST_DESC column is missing in many lab observations
- Lab result units are missing in some cases
- A lot of information is mapped to the SUPP domains
- Challenges in creating Trial design domains
- Preparation and maintenance the data transformation/mapping documentation

Thank you!