Transformation of clinical trial data to the OMOP CDM will open new opportunities to compare observational real world data (RWD) and data from randomized controlled trials (RCTs)

Transforming Clinical Trial Data to the OMOP CDM

Cynthia Sung[1], Mike Hamidi[2]*, Zhen Lin[3]*, Tom Walpole[4]*, Rebecca Baker[5], Melissa Cook[6], Shital Desai[2], Priya Gopal[7], Dan Hartley[8], Vojtech Huser[9], Priya Meghrajani[10], Tra Nguyen[11], Paul Orona[12], Katy Sadowski[13], Sebastiaan van Sandijk[9], Philip Solovyev[9], Ramona Walls[8], Kenneth J. Wilkins[14], Qi Yang[15], and the Clinical Trial Working Group; *CTWG Co-leads

[1]Duke-NUS Medical School, [2]Independent Consultant, [3]Robot Bacon Corp, [4]ZS, [5]CDISC, [6]Essex Management, [7]Flatiron Health, [8]C-PATH, [9]Odysseus [10]Talosix, [11]Boston Medical Center, [12]Kaiser-Permanente, [13]TrialSpark, [14]National Institute of Diabetes, Digestive and Kidney Diseases, [15]IQVIA

Background

The OMOP Common Data Model (CDM) was originally designed to conduct analyses on observational data, rather than clinical trial data. Yet, a large volume of patient-level data exists from clinical trials that would be informative to compare to observational data. The Clinical Trial Working Group (CTWG) is developing a general guideline to facilitate Extract-Transform-Load (ETL) of CDISC SDTM data to the OMOP CDM and proposing conventions for handling unique aspects of clinical trial data.

Methods

The Critical Path Institute (C-PATH) provided deidentified data for TB-1015 (NCT02193776) "*A Phase 2 Open-Label, Partially Randomized Trial to Evaluate the Efficacy, Safety and Tolerability of Combinations of Bedaquiline, Moxifloxacin, PA-824 and Pyrazinamide in Adult Subjects with Newly Diagnosed, Drug-Sensitive or Multi Drug-Resistant Pulmonary Tuberculosis*". The team benefitted from the diverse healthcare and informatics backgrounds of its members to understand the clinical context, CDISC standards and OMOP CDM conventions to guide the ETL process. Team members worked asynchronously to propose strategies for selecting the best standard concept. These proposals were deliberated upon during the biweekly meetings to reach consensus.

Results

Source SDTM Files to Target OMOP CDM Tables

SDTM file & description		OMOP CDM Table
dm	demographics	PERSON
sv	subject visits	VISIT_OCCURRENCE
ae	adverse events	CONDITION_OCCURENCE
ce	clinical events	
mh	medical history	
ex	drug exposure to experimental drugs	DRUG_EXPOSURE
cm	concomitant meds	
lb	laboratory test results	MEASUREMENT
ms	microbiology susceptibility findings	
mb	microbiology specimen	
vs	vital signs	
cm	concomitant meds	PROCEDURE_OCCURRENCE
pr	procedures	
ae	adverse events	OBSERVATION
dd	death details	

Selected Issues and Proposed Solutions

Issue		Proposed Solution
CR.001	**USUBJID** is the identifier linking the same patient across all **SDTM** tables. Values often contain **text** characters.	Reference **USUBJID** to autogenerate an **integer** value for **PERSON_ID.** Record **USUBJID** in **PERSON_SOURCE_VALUE**
Handling Relative Dates		
CR.002	SDTM start of a trial **STDY** = 1; OMOP CDM requires calendar dates	Use the **study start date** as the reference start date (**RFSTDTC**) for all patients . This information can be found in the trial publication or the trial registry such as ClinicalTrials.gov
	STDY is a positive number	Add (**STDY**-1) to **RFSTDTC**
	STDY is a negative number (screening visit)	Add (**STDY**) to **RFSTDTC**
CR.004	Construct a unique number for cdm.**<table name>_OCCURRENCE_ID**	Concatenate **USUBJID** and **SEQ** to autogenerate an integer value for **<table name>_OCCURRENCE_ID**
CR.008	Construct a unique number for cdm. **VISIT_OCCURRENCE_ID**	Concatenate **USUBJID** and **SVSTDY** to autogenerate integer value for **VISIT_OCCURRENCE_ID**
Find closest matching Standard Concept id		Concatenate entries in these columns using **paste(. . .);** can use Usagi to propose best matches to standard concept_ids
	lb.csv	'LBTEST','LBCAT','LBSPEC','LBSTRESU'
	cm.csv	'CMDECOD', 'CMDOSE', 'CMDOSU', 'CMROUTE'

Conclusion

We have reached consensus on several conventions and developed mapping guidelines for doing an ETL of CDISC SDTM data to the OMOP CDM. We welcome new members as we undertake the next stage of work to translate the rules and guidelines into SQL code, perform quality checks, and test the robustness of the methodology when applied to other clinical trials.

OHDSI

Join the Clinical Trial Working Group, which meets biweekly Friday at 11:30 ET.