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

[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
*Clinical Trial Working Group Co-Leads

Background

The OMOP Common Data Model (CDM) was originally designed to conduct analyses on observational data,[1] 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. Questions that can be answered by using clinical trial data with observational data include the following: Are outcomes from controlled clinical trials comparable to those in a diverse population after marketing authorization? Can those comparisons inform trial design for pragmatic trials? Is the adverse event profile from clinical trials similar to those in the real world? Are serious adverse events detectable in a larger population that were not found in the more limited populations in the clinical trials? Do specific factors account for differences between clinical trial results and post-market observational studies? Additionally, access to individual level patient data yields valuable information about clinical trial design that is not evident from summary descriptions in publications and trial registries.

Clinical trial data have several unique aspects compared to observational data, such as the use of relative study days rather than calendar dates, trial arms that represent "standard of care," experimental drugs that are not yet registered, and phases (screening, treatment, follow-up). The Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) was chosen as the prototypical clinical trial data standard because it is the digital standard recommended by drug regulatory authorities.[2,3] The Clinical Trial Working Group (CTWG) is working on a general guideline to facilitate the transformation of SDTM data to the OMOP CDM and proposing conventions for handling unique aspects of clinical trial data.

Methods

After an extensive search for full access to clinical trial data in SDTM data, the CTWG received permission to access the Critical Path Institute's Tuberculosis Clinical Trial Data (https://c-path.org/tools-platforms/tb-pacts/). The first trial being examined is 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*". We searched for additional documentation needed to interpret the SDTM data entries. Mapping efforts were coordinated based on the target OMOP CDM table, identifying the SDTM tables serving as sources for these targets. Team members worked asynchronously to propose strategies for selecting the best standard concept. These proposals were subsequently deliberated upon during biweekly meetings. Many common rules were discussed to reach consensus, such as choice of a start and end date when only relative dates are available.

Results

SDTM documentation of table names and column headings is essential for decoding the data tables. As SDTM evolves with time, one should obtain the CDISC standard when the data were created. Important documentation for interpreting the TB-1015 study data were the study protocol (which included a detailed description of each trial arm and the medications, drug forms and dosages), information posted at ClinicalTrials.gov, the peer-reviewed publication about the trial, associated supplemental material, and the statistical analysis plan. An annotated CRF is also a useful document because it shows the exact questions asked and the variable in SDTM table used to capture the data, but it is not always available publicly, and we did not have it for TB-1015. Sometimes, more than one SDTM table would map to a target OMOP CDM table. Here the team benefitted from the knowledge of members from CDISC. Like the OMOP CDM, SDTM table names and column headings are standardized and contain an entry for every subject and every visit and a foreign key that links the multiple tables. Unlike the OMOP CDM, controlled vocabularies are applied to only certain variables and not always strictly adhered to. It was necessary to identify and prioritize the use of columns intended for controlled vocabulary. Nonetheless, we found misspellings and different representations of the same concept even in columns designated for controlled vocabulary. Finding the best OHDSI standard concept often requires information from multiple columns in the source data. To date, we have proposed mapping conventions for the following OMOP CDM tables: PERSON, OBSERVATION_PERIOD, VISIT_OCCURRENCE, CONDITION_OCCURRENCE, DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, and MEASUREMENT. These are currently under review by a team member different from the one who prepared the proposal. Subsequently, sample computer code will be written as templates for other users.

Conclusion

The CTWG is benefitting from having a diversity of backgrounds: data scientists, clinical trialists, regulators, and those with extensive experience performing OMOP ETLs, as well as representatives from C-PATH and CDISC. While we have reached consensus on several conventions, ETLs of other trials will be needed to ensure that the common rules and mapping conventions can be reused with high fidelity. After completing the ETL of TB-1015, we will make use of the Data Quality Dashboard, ACHILLES and ARES to check the data quality and share the findings with C-PATH. The team focused on conversion of SDTM data to the OMOP CDM, though equally helpful would be to establish a general guideline for converting data in OMOP CDM to CDISC SDTM. Doing so will enable research on signals detected from observational data to be studied in controlled clinical trials. Setting up of pragmatic clinical trials will also benefit from such a guideline. Promoting the use of standardized ontologies (e.g. SNOMED, LOINC, drug dictionaries) in clinical trial data collection will facilitate the reuse of those data by making it easier to transform them to a common data model and reduce ambiguity in the meaning of terminology during an ETL. Going forward, we will apply the guideline to other TB trials at C-PATH, which may elucidate common and discordant features of trials for the same indication. By reusing clinical trial data for multiple analyses, greater value can be derived from the cost and time expended to conduct the original trial, and the contributions of patients who volunteered for such studies will generate beneficial knowledge beyond the goal of the original trial. We welcome new members interested in bringing clinical trial data into the OMOP ecosystem and/or those who have clinical trial data in SDTM format who would like to apply the guidelines that the CTWG is developing to create an OMOP CDM instance of the trial data.

References

1. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, Welebob E, Scarnecchia T, Woodcock J. Advancing the science for active surveillance: rational and design of the Observational Medical Outcomes Partnership. Ann Intern Med. 2010 Nov 2;153(9):600-6.

2. Study Data Standards Resources. US Food and Drug Administration, 2023. Accessed 14 March 2024.

3. European Medicines Regulatory Network Data Standardization Strategy, 2021 Accessed 14 March 2024