Clinical trial data conventions for the OMOP Common Data Model

Alexandra Orlova[1], Andrew Williams, PhD[2], Asiyah Yu Lin[3,4], Chris Roeder, MS[5], Cynthia Sung, PhD FCP[6], Emma Vos, MSc[7], Gregory Klebanov, MSc[1], Joshua F. Ransom, PhD[8], Katy Sadowski, BS[9], Maxim Moinat, MSc[7], Michael Kallfelz, MD[1], Mike Hamidi[10], Philip Solovyev, PhD[1], Rhonda Facile, MS[11], Shawn Dolley, MBA[12], Sonia Araujo, PhD[13], Tom Walpole[14], Vojtech Huser, MD PhD[15]

[1]Odysseus Data Services Inc., Cambridge, MA, USA; [2]Tufts Institute for Clinical Research and Health Policy Studies, Boston, MA, USA; [3]Center for Devices and Radiological Health, FDA, Silver Spring, MD, USA; [4]National Center for Ontological Research, Buffalo, New York, USA; [5]Division of Cardiology, University of Colorado, Denver/Anschutz Medical Campus, Aurora, Colorado, USA; [6]Bill & Melinda Gates Medical Research Institute, Seattle, WA, USA; [7]The Hyve, Utrecht, The Netherlands; [8]BEKHealth Inc, Wayland, MA, USA; [9]TrialSpark, New York, New York, USA; [10]CDISC, Audubon, PA, USA; [11]Elligo, Austin, Texas, USA; [12]Open Global Health, Arlington, Virginia, USA; [13]IQVIA, London, UK; [14]Trials.ai, San Diego, CA, USA; [15]National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

The abstract reflects the views of the authors and should not be construed to represent the views or policies of any of the authors' employers

Abstract

The data types that are generally collected during clinical trials have a large degree of overlap with those of observational data. For example, the same kind of lab measurements and condition occurrence reporting takes place for both data sources. However, there are some distinct features inherent to clinical trial data collection that do not have an obvious storage location or domain within the OMOP CDM, like adverse event severity and causality, and information on trial arms.

The OHDSI Clinical Trials Working Group proposes conventions for the OMOP CDM and Standardized Vocabularies to capture clinical trial specific data. Our use case is to convert clinical trial data in CDISC SDTM format to OMOP, with a view to allowing trial planning optimization.

We advocate minimum changes to the OMOP CDM and Standardized Vocabularies to minimize impact on OHDSI tools like Atlas, whilst providing a value-add SDTM-to-OMOP conversion with minimum data loss. Our proposals are built on OMOP CDM v6 and the Oncology extension, with v5.3.1 backward compatibility.

Introduction

The OMOP Common Data Model (CDM) is used for storing and analyzing observational health data from various sources, e.g. EHR records or claims data. It was not originally designed to store clinical trial data. Extending support for clinical trial data would promote collaborative research – data partners could run multiple analyses on the various data sources using standardized tools.

We will share proposals from the OHDSI Clinical Trials Working Group to allow adequate representation of clinical trial data in OMOP. Our proposals are built on OMOP CDM v6 and the Oncology extension, but CDM v5.3.1 compatibility is guaranteed through the addition of a handful of CDM fields.

Proposals

Our use case is the conversion of clinical trial data in CDISC SDTM format to OMOP, with a view to allowing trial planning optimization. SDTM was chosen as it is a clinical trials' submission standard that is "required" by the FDA and PMDA, "preferred" by the China NMPA, and "accepted" by the EMA. All our proposals assume the source data is in SDTM format and represent the final set of data from a clinical trial.

Another important assumption regards clinical trial epochs and observation periods. A clinical trial typically consists of several epochs, e.g. screening and treatment. When building cohorts or conducting characterization analysis in Atlas, only the current observation period is considered. Hence, instead of capturing clinical trial epochs as separate records in the OMOP CDM observation period table, we propose to store one observation period record only per clinical trial subject, so analysis can be completed across the clinical trial duration. The observation period start date will be the date a person gave informed consent. The observation period end date will be the last recorded date for that person.

Our work covered eight main topics – from trial information and visits, to type concept ids – for which there is currently insufficient support in the OMOP CDM and Standardized Vocabularies. Our proposals include introducing new concepts and modifiers, but no new CDM tables. Furthermore, we provide guidance for ETL developers when dealing with some data that is more complicated in nature, or certain scenarios that may be present in clinical trial submitted datasets (e.g., non-unique subject ids).

The table below summarizes our proposals across the eight topics identified. At the OHDSI US Symposium, we will expand on each of these topics and provide examples of the conventions and ETL advice.

Table 1. Proposal summary

Topic	Proposal Summary
Trial enrollment & trial outcomes	We propose to store these data as an observation for each event related to a person's trial status (e.g., informed consent or completion of trial).
Trial visits	We propose to extend OMOP CDM vocabularies to capture the different trial visit concepts across clinical trial epochs, and to have composite source values to capture time indicators within an epoch (e.g., TREATMENT:WEEK 7).
Seriousness, severity and causality	We propose to link an observation or condition to another record to capture adverse events, along with their seriousness, severity and causality to the trial subject's drug or treatment, via oncology extensions (measurement modifiers) and Observation attributes from OMOP CDM v6.
Study information and arm assignment	We propose storing information about which trial arm the individual trial subjects are in using the COHORT table. We propose storing information about the trial design and trial arms in the COHORT_DEFINITION table.
Novel concepts	Some drugs cannot be standardized as they haven't been "seen" before. For drug concepts, single new concepts can be added without substantial effort at the ingredient level. We propose an improved and simplified process to add clinical drug level drug concepts as RxNorm extensions.
Type concept ids	Type concepts in OMOP give the provenance of a record. We propose to use the newly-added standard type concept for "Case Report Form" to represent trial provenance.
Planned drug dose	To keep administered and planned drug doses in a way that makes comparing them possible, we propose to use a type concept id in the DRUG EXPOSURE table that allows that distinction.
Relative dates	In some clinical trials e.g. when a trial is anonymized, events' timepoints are given as days offset from a subject's informed consent or randomization date. If relative dates are given, we propose to calculate dates using the subject's reference date. The METADATA table can be used to record dates are derived.

Next Steps and Future Scope

The OHDSI Clinical Trials Working Group submitted its proposals to the OHDSI community in July 2020, for review and leadership approval. We hope to see these conventions ratified soon.

We have identified three future areas of interest: 1) extending to other types of data such as registries; 2) catering for scenarios where multiple trials are linked, and a person's data can be a combination of different trials; and 3) addressing OMOP-to-SDTM conversion with a view to aiding regulatory submission.

Conclusion

Whilst the OMOP CDM already supports multiple data sources, a significant gap that remains is the support of clinical trial specific data. The proposals put forward by the OHDSI Clinical Trials Working Group address that gap with minimum changes to the OMOP CDM and Standardized Vocabularies, whilst providing a value-add SDTM-to-OMOP conversion with minimum data loss. If the OMOP CDM and Standardized Vocabularies were extended to support clinical trial data in the way proposed, this would mean enhanced collaboration between clinical trials and observational research could be attained, ultimately resulting in better patient outcomes.