OHDSI Workgroup
Objectives and Key Results (OKR)

May 2024 Update

Clinical Trials Workgroup leads: Mike Hamidi, Zhen Lin, Tom Walpole
Objective: Our objective is to facilitate the integration of clinical trial data, formatted according to the CDISC SDTM standard, into the OMOP framework.

Approach: Our approach prioritizes minimal alterations to the OMOP CDM and Standardized Vocabularies. This ensures that OHDSI tools such as Atlas remain unaffected while enhancing the conversion process from SDTM to OMOP with minimal data loss. We propose introducing new concepts and modifiers through established conventions, without necessitating the creation of new CDM tables. Additionally, we offer guidance for ETL developers as needed. Originally based on OMOP CDM v6 and its Oncology extension, our proposals maintain backward compatibility with v5.3. With the integration of Oncology extension features into the standard v5.4, our modifications remain minimal, ensuring full compatibility with this latest version.
CTWG Accomplishments To Date

2024: Mapping specifications and ETL for the initial SDTM study are completed. Subsequent studies will adhere to the established conventions, with OMOP CDM guidelines being defined for implementation.

2023: Transitioned from Vivli real-world SDTM data due to environmental constraints, opting instead for tuberculosis study data sourced from C-Path, spanning over three clinical studies. Additionally, potential data sources from NIDA Data Share, including several real-world studies in SDTM format, and synthetic data options are identified. Mapping efforts have commenced for one of the C-Path TB clinical studies.

2022: The CTWG gained access to 20 Vivli clinical study packages in SDTM format. An inventory of these packages is underway to prioritize SDTM-to-OMOP mappings. Existing CTWG guidance topics are under review, with new topics being identified as needed.

2021: CTWG conducted an evaluation of clinical trial data providers offering SDTM data access, leading to engagements with Vivli for general data usage agreements and platform feasibility assessments.

2020: Utilized synthetic representations of CDISC SDTM data via PHUSE Test Data. Initial guidance topics were established but require validation with diverse real-world SDTM data. CTWG proposals were submitted to the OHDSI community in July 2020.
CTWG Challenges

Challenges arise from the Working Group's expertise limitations in mapping and reviewing SDTM-to-OMOP CDM tables. Additionally, there is a scarcity of technical SMEs available to develop ETL scripts and validate outcomes effectively. Despite these constraints, efforts are focused on maintaining a consistent set of rules applicable to forthcoming studies, balancing the conceptual and physical data application levels.
CTWG OKR

**Objective #1:**
To define the conceptual mappings and guidance to support CDISC SDTM-to-OMOP conversion

**Key Result #1:**
Identify >=3 real-world SDTM clinical studies
– Sept. 2023: This has been achieved.

**Key Result #2:**
Develop SDTM-to-OMOP mapping specifications using a prioritized set of common SDTM domains (adverse events, vital signs, demographics, concomitant medications, laboratory test results, medical history, and procedures)
– May 2024: The first study has been mapped and currently undergoing further reviews. The next step is to develop the ETL scripts.

**Key Result #3:**
Publish draft SDTM-to-OMOP guidance by Q1 2024
- Conceptual mappings on key domains of interest
- Define OMOP CDM conventions
- Identified gaps, issues, and challenges
– May 2024: The mapping specifications and conventions are being document. The current activity is developing ETL scripts to test the mappings already defined. After which, the team will work on the guidance document.
CTWG Ask

- Additional sources of real-world clinical studies in SDTM format
- Any volunteers to support SDTM-to-OMOP ETL mappings
- Any organization actively working on SDTM-to-OMOP conversions that have lessons learned, supporting documentation, or ETL code
CTWG Mapping Overview

Tracker

CDM Table

<table>
<thead>
<tr>
<th>CDU</th>
<th>SDTM Datasets</th>
<th>Mapping Owner(s)</th>
<th>Reviewer(s)</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSON</td>
<td>CM</td>
<td>Q Yang</td>
<td>Ready for Review</td>
<td></td>
</tr>
</tbody>
</table>

OBSERVATION PERIOD

| TV, SY | V1, V2, V3, V4, V5 | Q Yang | Ready for Review |

OBSERVATION V1

| TV, SY | Q Yang | Ready for Review |

DRUG EXPOSURE

| EX, CM | Mike Hamidi, Q Yang | Ready for Review |

PROCEDURE OCCURRENCE

| PR | Katy, Tien | Ready for Review |

DEVICE EXPOSURE

| Device | Do not populate |

MEASUREMENT

| Measurement | Do not populate |

DEATH

| Death | Do not populate |

NOTE

| Note | Do not populate |

NLP

| NLP | Do not populate |

SPECIMEN

| Specimen | Do not populate |

DATA ELEMENTS

| Data Elements | Do not populate |

LOCATION

| Location | Do not populate |

CARE SITE

| Care Site | Do not populate |

PAYER PLAN PERIOD

| Payer Plan Period | Do not populate |

COST

| Cost | Do not populate |

---

Common Rules

CDM Mapping Specification

For each subject, assign an arbitrary PERIODIC_F but then define further calculation of a start_date value from each start_period.

*PERIODIC_F* is the number of days between two consecutive start_dates. If PERIODIC_F is a positive number, the calculation of start_date is accurate to the last day of the start_period. However, if PERIODIC_F is negative, the calculation is accurate to the first day of the start_period.

For each subject, assign an arbitrary next_start_period. Then define further calculation of a start_date value from each next_start_period.

*next_start_period* is a negative PERIODIC_F but then define further calculation of a start_date value from each next_start_period.

For each subject, assign an arbitrary last_start_period. Then define further calculation of a start_date value from each last_start_period.

*last_start_period* is a positive PERIODIC_F but then define further calculation of a start_date value from each last_start_period.