OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Clinical Trials Working Group

Mike Hamidi (co-lead)
Zhen Lin (co-lead)

Working Group Update
October 5, 2021

What is the CTWG?

Clinical Trials Working Group

Objective: To allow adequate representation of clinical trial data in OMOP.

First use case: To convert clinical trial data in CDISC SDTM format to OMOP, with a view to allowing trial planning optimization.

Approach: We advocate minimum changes to the OMOP CDM and Standardized Vocabularies because we want to ensure minimum impact on OHDSI tools like Atlas, whilst providing a value-add SDTM-to-OMOP conversion with minimum data loss. We have proposed conventions introducing new concepts and modifiers, but no new CDM tables; and providing guidance for ETL developers where appropriate. Our proposals are built on OMOP CDM v6 and the Oncology extension, with v5.3.1 backward compatibility.

Status:
- Proposals submitted to the OHDSI community in July 2020.
- Currently applying our proposed conventions to the PHUSE database (a synthetic SDTM database).

Membership

WG co-leads: Mike Hamidi, Pfizer, US (Kamiar.Hamidi@pfizer.com) Zhen Lin, Robot Bacon, US (Zhen.Lin@ohdsi.org)

Current members

Name	Organization & Location
Alexander Davydov	Odysseus Data Services Inc., Cambridge, MA, USA
Alexandra Orlova	Odysseus Data Services Inc., Cambridge, MA, USA
Andrew Williams	Tufts Institute for Clinical Research and Health Policy Studies, Boston, MA, USA
Asiyah Yu Lin	National Institute of Health, Bethesda, Maryland, USA National Center for Ontological Research, Buffalo, New York, USA
Bess LeRoy	CDISC, Tucson, AZ, USA
Chris Roeder	School of Medicine, University of Colorado, CO, USA
Gregory Klebanov	Odysseus Data Services Inc., Cambridge, MA, USA
James Liddil	M2Gen, Tampa, FL, USA
Joshua F. Ransom	BEKHealth Inc, Wayland, MA, USA
Katy Sadowski	TrialSpark, New York, New York, USA
Lina Clover	SAS, Raleigh-Durham, NC, USA
Maxim Moinat	The Hyve, Utrecht, The Netherlands
Michael Kallfelz	Odysseus Data Services Inc., Cambridge, MA, USA
Mike Hamidi	Pfizer, PA, USA
Philip Solovyev	Odysseus Data Services Inc., Cambridge, MA, USA
Rebecca Baker	CDISC, Austin, Texas, USA
Sonia Araujo	IQVIA, London, UK
Ted Bebi	Medidata, Boston, MA, USA
Tom Walpole	Trials.ai, San Diego, CA, USA
Vojtech Huser	National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
Zhen Lin	Robot Bacon, TX, USA

Source: https://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:clinicalstudy
As of As of 2021-08-07

CTWG Objectives?

The Clinical Trials Working Group (CTWG) proposes conventions for the OMOP CDM and Standardized Vocabularies to capture clinical trial specific data.

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.

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 cover **eight main topics** for which there is currently insufficient support in the OMOP CDM and Standardized Vocabularies. They 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).

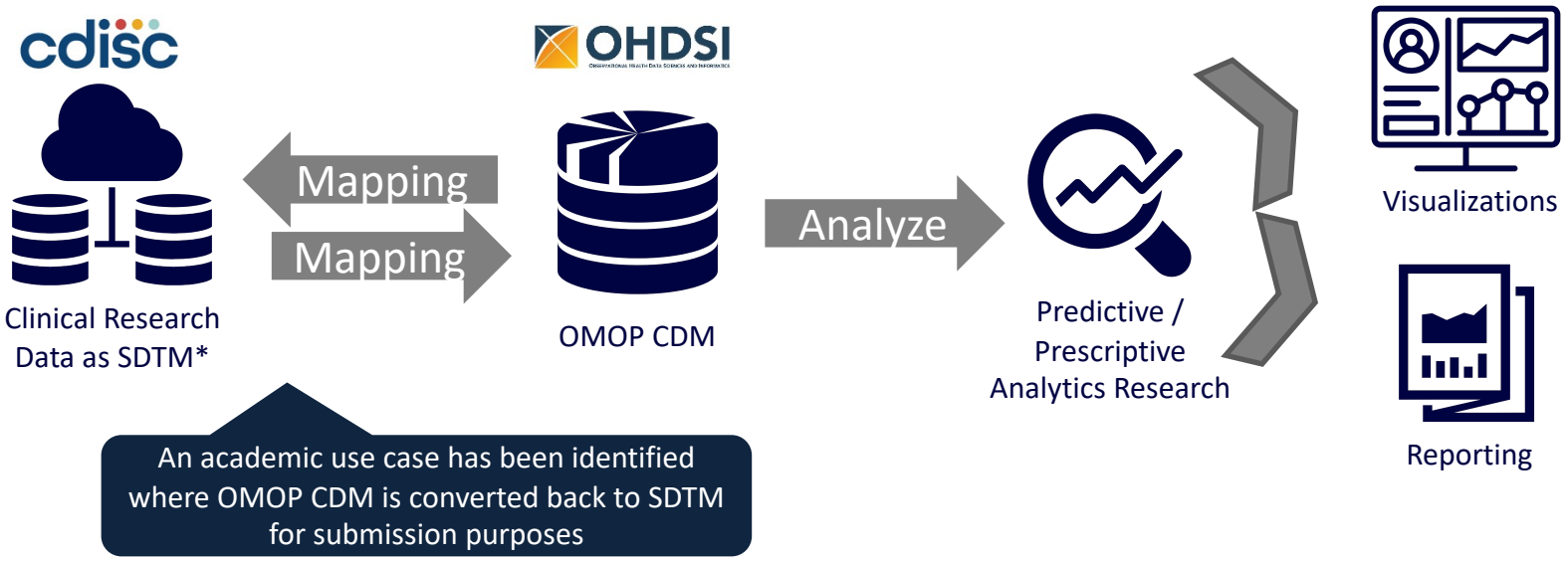
This document details our proposals for each of those eight topics, built on OMOP CDM v6 and the Oncology extension, with v5.3.1 backward compatibility.

#1 – Trial Enrollment & Outcome	#3 – Seriousness, Severity and Causality	#5 – Novel Concepts	#7 – Planned Drug Dose
#2 – Trial Visits	#4 – Study Information and Arm Assignment	#6 – Type Concept IDs	#8 – Relative Dates

Source https://github.com/OHDSI/ClinicalTrialsWGETL/wiki

SDTM→OMOP

OMOP CDM uses, for example, health records and claims databases. The CDM and vocabulary set is comprehensive to manage multiple data sources and supporting models (e.g., LOINC, SNOMED). SDTM was originally developed to support clinical research use case that are predominantly based on non-RWD terminology (e.g., CDISC, MedDRA, etc.). Although, SDTM can accommodate some aspects of observational-based research, it was not designed for this purpose. However, being able to utilize SDTM for OMOP is not beneficial for researchers but also sponsors.



Source: Mike Hamidi

*Data standards used for regulatory submissions (e.g., for the US FDA)

CTWG Activities

Clinical Research Topics

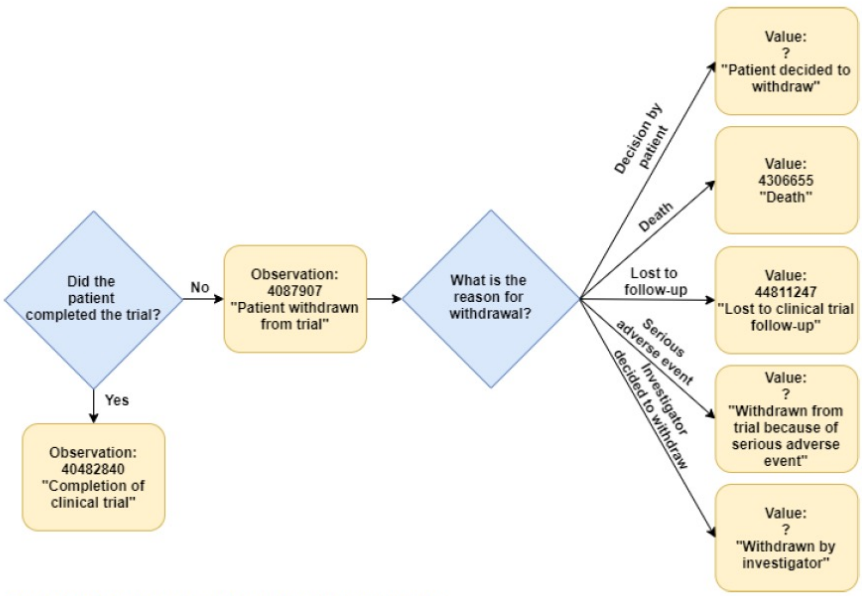
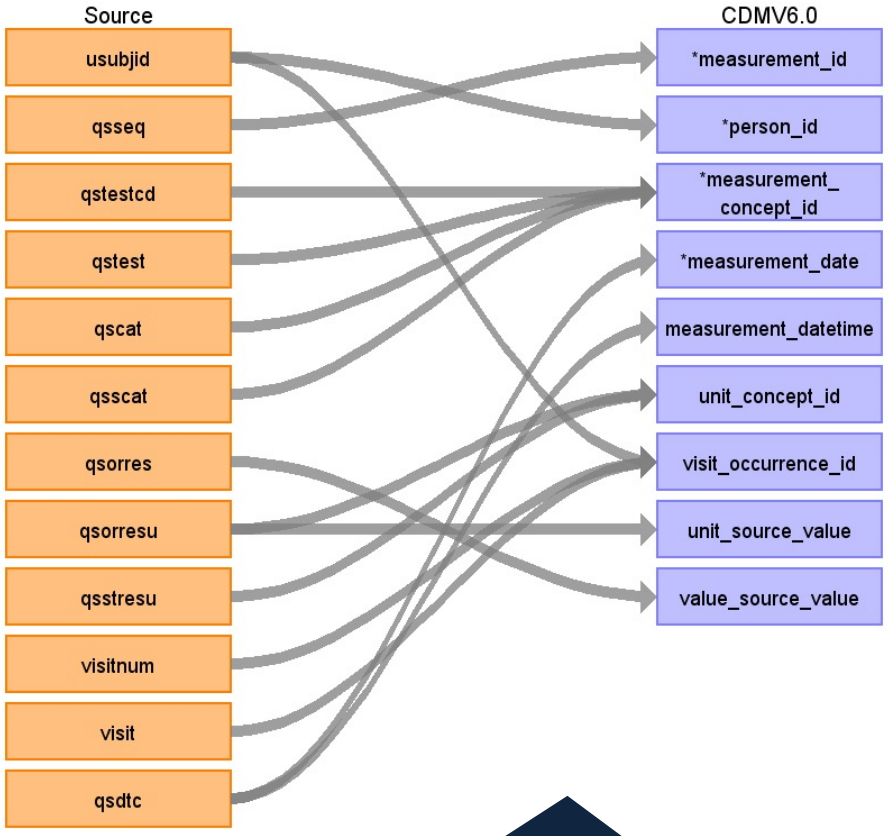


Figure 2. Trial outcomes and corresponding mappings.

Taking a conceptual thought…

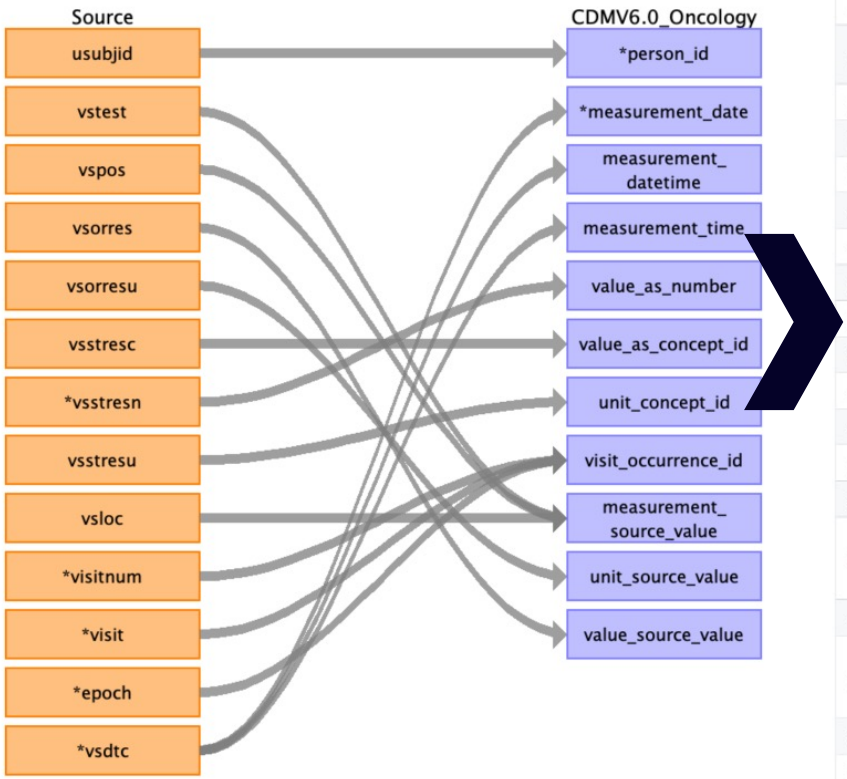
SDTM-to-OMOP CDM Mapping



…into mappings

Source: https://github.com/OHDSI/ClinicalTrialsWGETL/wiki

CTWG Activities



Destination Field	Source field	Logic	Comment field
measurement_id			Auto-increment.
person_id	usubjid		Indirectly mapped from usubjid, which will represent the subject's person_source_value.
measurement_concept_id			Derived from measurement_source_concept_id.
measurement_date	vsdtc		
measurement_datetime	vsdtc		
measurement_time	vsdtc		
measurement_type_concept_id			Hardcode as 32809 (Case Report Form).
operator_concept_id			
value_as_number	vsstresn		VSSTRESN is the standardized result value, if numeric.
value_as_concept_id	vsstresc		Derive from VSSTRESC (the standardized result value, if categorical).
unit_concept_id	vsstresu		Derive from VSSTRESU (the standardized unit).
range_low			
range_high			
provider_id			
visit_occurrence_id	visitnum visit epoch		
visit_detail_id			
measurement_source_value	vstest vspos vsloc		Concatenate VSTEST (name of test) + VSPOS (subject's position during test) + VSLOC (anatomic location of test) NOTE: may need to truncate this concatenated value in the final ETL d/t 50-character limit
measurement_source_concept_id			Derived from measurement_source_value.
unit_source_value	vsorresu		VSORRESU is the original unit as recorded.
value_source_value	vsorres		VSORRES is the original result as recorded.
modifier_of_event_id			
modifier_of_field_concept_id			

Some mappings require some additional upfront strategies

Source: https://github.com/OHDSI/ClinicalTrialsWGETL/wiki

CTWG Next Steps

- Access to additional real-world SDTM data
  - Actively working with **Vivli** and **C-Path** with a likelihood of happening Nov. 2021

- Use additional SDTM data to continue mapping activities
  - Identify gaps and limitations
  - Assess strategies for other topics not listed

- Producing an industry guidance and example package
  - A synthetic SDTM-to-OMOP data set
  - Generalized mappings with associated disclaimers/tips
  - Best practice guidance document

We need data!

Thank you!

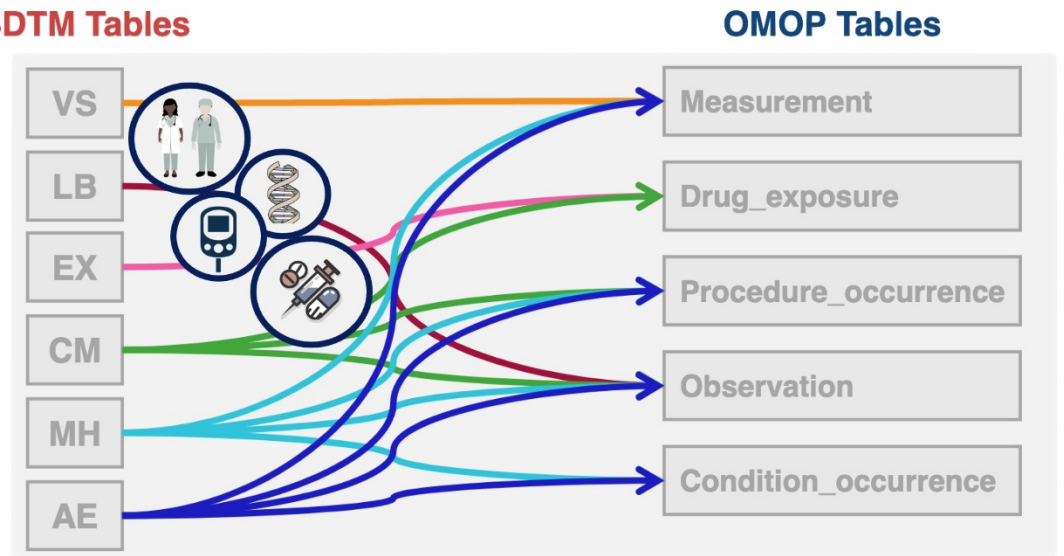
Clinical Trials WG

Fridays this Fall
7:30AM PT/9:30AM CT/10:30AM ET/
3:30PM UK/10:30PM CST

Bridging the gap between clinical trial and observational research

Join us!

http://bit.ly/OHDSI_CT



Pathways for advanced transformation of CDISC SDTM data sets into OMOP CDM, *Didden et al.*, 2020