

Harmonization and Interoperability between OMOP CDM and CDISC SDTM

Dhruba Sikdar

PHUSE EU Connect 2024

Agenda

- Introduction
- Main differences between OMOP-CDM and CDISC-SDTM
- Interoperability between two standards
- Consideration and Challenges
- Conclusion

** This presentation assumes the audience have basic familiarities with CDISC-SDTM and OMOP CDM*

Introduction

- Disparate data sources, disparate formats, languages, vocabularies
- Many data networks, some of them have their own CDM
 - US FDA Sentinel, EU-ADR, PCORnet, OMOP (OHDSI)
- CDM – Defined group of data elements with defined relationship
 - Common structure, format, vocabulary
- Different CDMs designed for different purpose, thus different structure
 - Difficult to compare, exchange, harmonize between them
- CDISC SDTM – de facto standard for regulatory submission
 - US FDA, other major HAs

CDISC-SDTM vs OMOP-CDM

- CDISC-SDTM: Clinical Data Interchange Standards Consortium (CDISC) – Study Data Tabulation Model (SDTM)
- OMOP-CDM: Observational Medical Outcomes Partnership (OMOP) – Common Data Model (CDM)

	SDTM	OMOP CDM
Purpose and Scope	Primarily designed for organizing and standardizing clinical trial data.	Focused on observational healthcare data. It aims to standardize and harmonize real-world data from various sources.
Data Structure	Organizes data into specific domains, such as demographics, adverse events, medical history, and laboratory findings.	Structures data into specific entities, such as patients, observations, conditions, procedures, drug exposures and establishes relationships and hierarchies.
Data Elements	Specifies a predefined set of standardized variables and controlled terminology to represent common data elements.	Does not mandate specific data elements, but rather provides a framework or schema for organizing diverse healthcare data.
Use Case	Is widely used in clinical research, pharmaceutical industry, and regulatory submissions.	Is primarily used for real-world data analysis, population health research, and comparative effectiveness studies

Interoperability and Harmonization

- FDA guidance on data standards for submission of RWD/RWE
 - Similar to submission of traditional clinical trials – data curation, mapping, transformation, traceability, documentation
- FDA supported standard – essentially CDISC-SDTM (and ADaM for analysis data)
 - Necessitates data transformation from other standards
- Purpose of data collection in RCTs and observational studies (RWD) are completely different. RCT databases are purpose built for research. RWD databases primarily serve healthcare providers. Hence the need for interoperability.
- Challenges include – management of semantic concepts, inconsistent format, inconsistent coding, changes in data collection/coding practices, handling missing information
- Benefits include – seamless transition from one standard to another, efficient and productive research, aid regulatory decision making.

Consideration and Challenges

- Data Mapping:
 - Source domain/concept to destination entities – mapping of data elements and attributes (SDTM/ OMOP to OMOP/SDTM)
 - Identify equivalent or related concept – establish mapping between SDTM and OMOP
 - Example:
 - ❖ W. Anderson et al (2023) : COVID-19 database in OMOP to SDTM mapping
 - ❖ Extract-Transform-Load (ETL) data from native to OMOP CDM
 - ❖ Map OMOP-CDM into SDTM using the OHDSI tools (Athena, Achilles, OHDSI ICD 10 vocabularies)

Variable	OMOP		SDTM CT		SDTM
	Concept ID	OMOP Domain/Column	Submission Value	SDTM C-Code	
Age*	N/A	person.birth_datetime	N/A	N/A	dm.age
Sex	8532	person.gender_concept_id	M	C20197	dm.sex
Sex	8507	person.gender_concept_id	F	C16576	dm.sex

* indicates a derived element

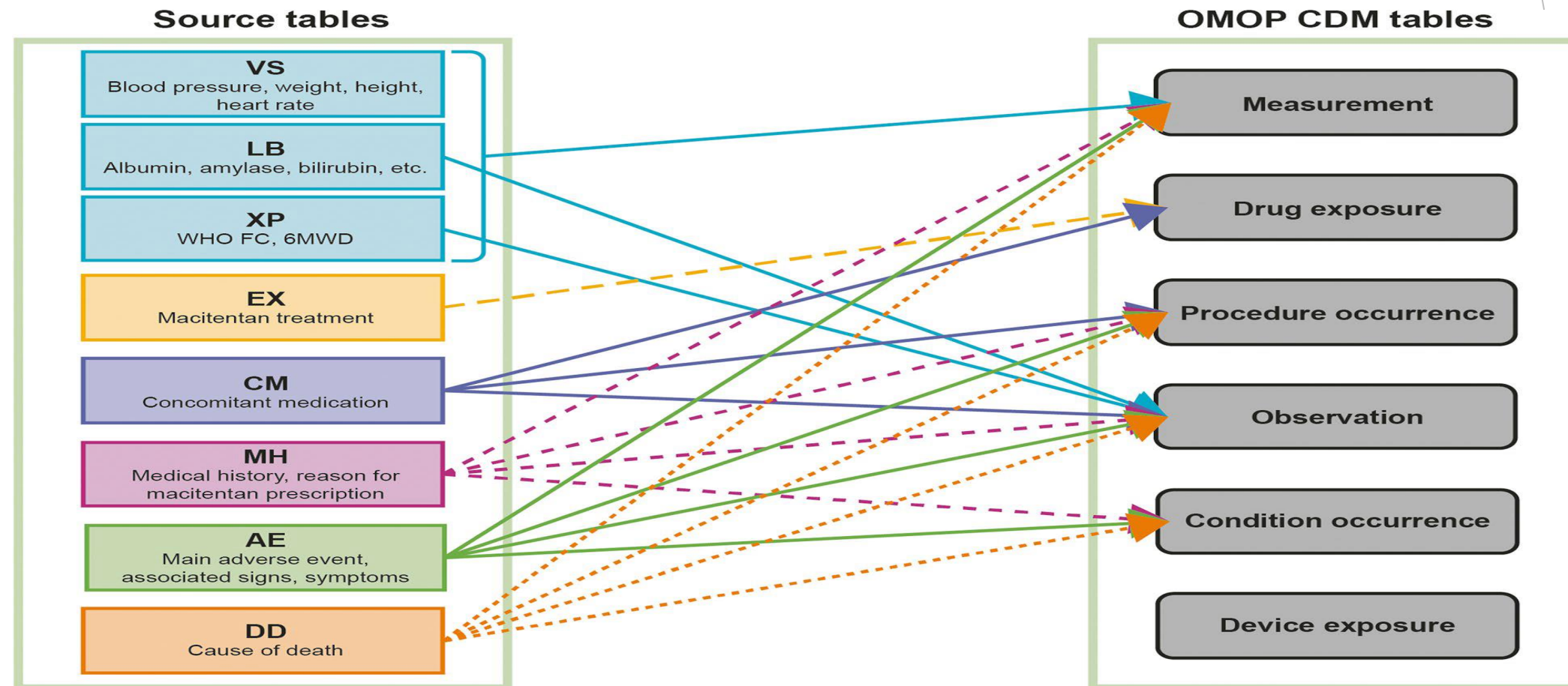
Source: Anderson W, Kurtycz R, Farid T, Hassan S, Kennon K, Dasher P, Boyce D, Roddy W, Heavner S. (2023). "From OMOP to CDISC SDTM: Successes, Challenges, and Future Opportunities of using EHR Data for Drug Repurposing in COVID-19". <https://www.ohdsi.org/wp-content/uploads/2023/10/3-AndersonBriefreport-Wes-Anderson.pdf>

Consideration and Challenges (cont.)

- Transformation and Conversion:

- Source datasets to destination format – conform to the destination data structure by restructuring source data, ensure relationships are maintained
- Example:
 - ❖ P. Biedermann et al (2021) : Registry data for PH formatted in SDTM transforming to OMOP CDM
 - ❖ Challenges – MedDRA (SDTM) codes have no direct counterpart in (OMOP) dictionaries SNOMED, LOINC
 - ❖ AE severity and causality (SDTM) have no equivalent functionality in OMOP. Custom concept needed
 - ❖ Mapping partial dates, free text entry (medication) are some other challenges.
- Example:
 - ❖ A.Y. Shanbhogue (2022) : OMOP to SDTM Mapping
 - ❖ Extract-Transform-Load (ETL) data from native to OMOP CDM
 - ❖ Used intermediary data model [Biomedical Research Integrated Domain Group (BRIDG)]
 - ❖ Fair amount of manual curation required.

Consideration and Challenges (cont.)



Source: Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, Wetherill G, Brand M, Didden E.M. (2021). "Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases". BMC Med Res Methodol Nov 2021. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8565035/>

Consideration and Challenges (cont.)

- Terminology Alignment:
 - Include but not limited to CDISC CT for SDTM; SNOMED-CT, RxNorm, and LOINC for OMOP CDM
 - Semantic interoperability, incorporating standardized terminologies, promotes consistency and accuracy in data interpretation.
 - Assumption behind semantic interoperability needs to be clearly defined and validated
 - Risk of losing information during mapping from one dictionary to other (e.g, SNOMED-CT to MedDRA)

Consideration and Challenges (cont.)

- ETL Processes:

- Extract, Transform, Load (ETL) to automate conversion processes
- Various open-source tools available
- Edge Tool suite [W. Anderson et al (2023)]
 - ❖ able to reduce the number of hours significantly to extract, transform, and load (ETL) structured components of the medical record (over 2000 hours to less than 90 hours)
- White Rabbit [A.Y. Shanbhogue (2022)] - OHDSI, Java tool
 - ❖ Scans input data present in csv files or databases and creates a scan report in the form of an excel file
 - ❖ Easy to use

Conclusion

- Interoperability between different standards in RWD space is not yet a well understood practice (yet).
- Some attempts have been made for OMOP to SDTM conversion and vice versa with varying degree of success, still a lot of attention needed on a case-by-case basis.
- Challenges lie is different structure, formats, terminologies, semantic concepts, coding vocabularies, above all differing in purposes.
- One single standard data model may not be adequate.
- In future studies with increasing complexity, a hybrid approach with well-established interoperability and harmonisation can unlock the true potential of clinical research.

References

- [1] European Medicines Agency. COVID-19: EMA sets up infrastructure for real-world monitoring of treatments and vaccines. 2020. <https://www.ema.europa.eu/en/news/covid-19-ema-sets-infrastructure-real-world-monitoring-treatments-vaccines>
- [2] Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, Wetherill G, Brand M, Didden E.M. (2021). "Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases". BMC Med Res Methodol Nov 2021. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8565035/>
- [3] FDA Guidance. (2023). "Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry". December 2023. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data>
- [4] Abolafia J, Ferko S, Holt I. (2024). "Future Clinical Data Submission Standards: CDISC, FHIR, OMOP, or Hybrid Model". Paper presented at the PHUSE US Connect 2024.
- [5] Anderson W, Kurtycz R, Farid T, Hassan S, Kennon K, Dasher P, Boyce D, Roddy W, Heavner S. (2023). "From OMOP to CDISC SDTM: Successes, Challenges, and Future Opportunities of using EHR Data for Drug Repurposing in COVID-19". <https://www.ohdsi.org/wp-content/uploads/2023/10/3-AndersonBriefreport-Wes-Anderson.pdf>
- [6] Shanbhogue A.Y. (2022). "RWD (OMOP) to SDTM (CDISC): A primer for your ETL journey". Paper presented at the PharmaSUG Annual Conference 2022, Austin, TX.

Contact Information

Dhruba Sikdar

Principal I Statistical Programming Lead

Integrated Data Analysis and Reporting

Johnson and Johnson Innovative Medicine

50-100 Holmers Farm Way

High Wycombe

HP12 4EG, UK

dsikdar@its.jnj.com

THANK YOU