Paper RE02

Usage of OMOP-CDM along with CDISC in Clinical Trials: Challenges and Opportunities

Parag Shiralkar, Sumptuous Data Sciences, LLC, USA
Mary Anne Rutkowski, Merck, USA

Abstract

With growing importance of RWD in getting clinical evidence, many companies are finding ways to make the best use of data from real world sources. Claims and healthcare data obtained from EHR systems is a major source of such data. OHDSI organization has developed a Common Data Model (CDM), which can take large observational data from multiple EHR and claims data sources around the world, and read and process such data in a consistent manner. Usage of such data in randomized clinical trials poses challenges as regulatory bodies prefer the submission of the data in CDISC format. How are traceability and provenance issues dealt with in OMOP-CDM? Can such data be easily transformed into SDTM? Would regulatory bodies accept the OMOP CDM as a submission standard? This manuscript provides detailed assessments of efforts taken by industry in usage, and transformation of OMOP CDM along with FHIR to CDISC SDTM.

Introduction

Observational data in healthcare sectors constitutes the patient data collected while receiving the healthcare. Such data can be collected at healthcare facilities, clinics, critical care units, and other diagnostic facilities. The insurance claims data as well as data collected at pharmacies can also be considered as an observational data. Such data is truly 'real-world data' as it captures the patient information in the real-world setting. Such observational data can help healthcare practitioners make informed decisions about patient care and helps them identify and decide appropriate course of actions pertaining to further treatments which need to be given to the patients. For researchers, such observational data can provide lots of insight about patient response to treatments, patient biomarkers and phenotypes, as well as patient and clinician reported outcomes. In the case of health insurance and pharmacy data, such data can drive the assessments related to pharmaco-economics, outcomes research, and overall cost related to healthcare assessments.

In the clinical world, such observational data can play a critical role in providing insight about patient populations considered in clinical studies. The biomarkers and phenotypes assessed through such observational data can help the pharmaceutical industry in selection of appropriate patient population for the clinical studies. The patient reported outcomes and clinician reported outcomes are important aspects of effectiveness and safety of the therapy or drug in assessment. Besides this, the observational data can also provide lot of information about possible new biomarker definitions and assessments and can also help with exploratory analysis.

Need for Common Data Model

The observational data can come from various sources. Such sources include small clinics as well as large healthcare facilities. Small clinics and healthcare facilities often do not have

necessary resources to ensure that the data is collected, stored, managed, and transferred in an industry acceptable format. As a result, such facilities often collect the data in proprietary format and in many cases such data is collected in no format. When such data is aggregated or collected from various facilities, integration of such data often poses challenges. In order to address this challenge, the Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") was established. OHDSI established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University, NY, USA[1]. This program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. The Common Data Model (CDM) developed and maintained by OHDSI is referred as the Observational Medical Outcomes Partnership (or OMOP) Common Data Model.

The OMOP CDM provides a common data model for the data coming from various healthcare sources. Healthcare facilities can adopt and implement the common data model in order to ensure interoperability, and usage of such data by wider research network. It can also help healthcare facilities, academia, and researchers to leverage existing systems of open-source tools, technologies, and processes. The current common data model from OMOP is maintained and managed by OHDSI consortia.

Design Principles Common Data Model

Since its inception, the common data model is evolved by accommodating input from industry, academia, regulatory bodies, and researchers. With input from user community, the standards governing body at OHDSI consortia continued to modify and upgrade the OMOP Common Data Model and released updated versions of this model. For more details about this mode, please visit the data standards page at: https://ohdsi.org

In the pharmaceutical industry, clinical data standards used by Clinical Data Interchange Standards Consortia (CDISC) are used for submission of clinical data to regulatory authorities[2]. The CDISC standards are built based on certain principles. Similarly, OMOP CDM standards need to be implemented by following certain design principles. These principles are considered to make sure that the data model is secure, robust, compatible, and uses vocabularies in an optimum way.

Currently, the Common Data Model uses 37 tables, out of which 17 are used to standardize clinical data, 10 are used to standardize the vocabularies, and other 10 are used for derived variables, health system specific data, and health economics specific data[3]. Backward compatibility and scalability are important factors while designing the Common Data Model. As the healthcare sector evolves, accommodation of additional data types is an important factor for successful use of the Common Data Model. Keeping this in mind, the scalability of model, and its compatibility with prior models is an important design principle while designing the CDM.

Traceability and Provenance Aspects

In real-world data, traceability and provenance of the data is found to be an important challenge for usability and validity of the data. While using real world data in clinical studies, the source data may need to undergo a series of steps involving normalization and curation. This could result into traceability and provenance issues. If the source observational data uses OMOP CDM, the robust set of processes, use of standardized vocabularies, and adherence of model designing principles can reduce the challenges pertaining to traceability and provenance. Looking at the usage of OMOP CDM by the researcher and healthcare community, the observational data in CDM format can be readily used for analysis purposes. The set of modelling, extract transform and load (ETL), and vocabulary usage in CDM is thoroughly evaluated through the OHDSI

consortia.
In the context of drug development and approval process in clinical studies, the provenance and traceability is much more thoroughly reviewed and scrutinized by the regulatory bodies. Such scrutiny is important to ensure that safety evaluation of a new drug or use of an existing drug for additional indications by a wider patient population. It is unclear as of yet if the provenance and traceability provisions in OMOP CDM can suffice the requirements and expectations from regulatory bodies in clinical data submission. Regardless, comparing to other available data standards and models in healthcare data, OMOP CDM has very robust structure and set up rules for establishment, which can help address the challenges of traceability and provenance.

Regulatory Submissions Requirements and Common Data model Compatibility

Acceptance of Standards

As per the regulatory guidances issued by major regulatory bodies, Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) by CDISC are the widely acceptable and in some cases required data standards for regulatory submission[4]. In case of submission of real-world data, in many cases, current CDISC models do not have necessary provisions to accommodate the data fields or vocabularies used in real world data. Despite that, because of existence of available system for review of submitted data, the majority of regulatory bodies have recommended submission of real-world data in preferably CDISC format. Food and Drug Administration (FDA) from USA in their guidance for submitting real-world data, have recommendations pertaining to early discussion with regulatory body when it comes to submission of real-world data[4]. Since most of the observational data is used in post marketing clinical studies, submission of such data in the context of regulatory communication can be feasible. However, for mainstream clinical development involving new drug applications, OMOP CDM may pose difficulties in submission process.

Compliance with Model

Currently OMOP CDM model is based on certain design principles. The usage and effectiveness of the model is dependent on how effectively those design principles are being followed by the implementation team. Since the model implementation can happen at a wider user base, it is a bit difficult to develop a unified set of compliance which can be utilized by all user stakeholders of OMOP CDM. OHDSI system has lot of open -source tools for processes to build and validate the model. These tools can guide the users for effective implementation of the CDM.
Contrary to this, the submission model (SDTM) developed by the CDISC organization has a good set of compliance rules. The data as well as submission compliance can be tracked effectively through these set of processes and tools available.

Usage of Vocabularies

OMOP CDM accommodates and uses a large number of standardized vocabularies and can also accommodate a lot of non-standard vocabularies. This helps users in healthcare sector in dealing with wide variety of usage of vocabularies. Contrary to this, CDISC models use a handful of vocabularies which are primarily used in clinical studies and drug development processes. As we accommodate more and more healthcare data as real-world data in clinical studies, we are bound to get exposed to more vocabularies used in healthcare sector. This in turns poses more challenges in terms of conversion of dictionary terms to acceptable dictionaries in clinical studies by regulatory bodies. As an example, the SNOMED CT[5] is widely used dictionary to report clinical terms in healthcare. Since MEDDRA[6] is widely accepted dictionary in clinical studies and its submission, now we need to look into SNOMED to MEDDRA conversion. The other

option could be acceptance of SNOMED by regulatory body for submission processes pertaining to new drug applications.

The OMOP CDM can accommodate several other vocabularies as it deals with a wide variety of healthcare data including pharmacy and health insurance data as well. Many of these vocabularies may not be related to the core clinical tables in CDM, but there are several vocabularies related to clinical tables. For such vocabularies there may not be sufficient provision in CDISC SDTM to map or align those terms when it comes to the regulatory submission process. This poses further need to establish some interoperability between these two standards while dealing with regulatory submission of observational data in OMOP CDM.

Further Opportunities to Establish Interoperability

Data standards are developed by considering specific stakeholders and users in mind. In case of clinical data, CDISC has taken efforts for more than two decades to develop, maintain, mature and grow the standards which can be used by the pharmaceutical industry. Similarly, OHDSI consortium has taken extensive efforts in establishment of a common data model for observational data used in healthcare. Both these standards are developed to support and help different stakeholders and user groups. While CDISC is more focused on clinical data collection, analysis, and submission, OMOP CDM is focused on healthcare data analysis, research, and effective usage.

As the healthcare world uses more and more clinical data produced by the pharmaceutical industry, there is a growing need of transformation of CDISC data to OMOP CDM. Likewise, as healthcare data (specifically observational data) gets incorporated into regulatory communication and submission, there will be need of forward transformation of OMOP CDM to CDISC.

There have been efforts to establish interoperability between CDISC SDTM and OMOP CDM. Such efforts showed that SDTM can't accommodate the OMOP CDM V5 vocabulary specifications completely[7]. At the same time from submission of data from randomized clinical trials (RCT), we may not need all vocabularies accommodated through and used in OMOP CDM. At this moment, the recommended approach which industry can take is to have early discussions with regulatory authorities in case of submission of real-world data involving OMOP CDM and assess the viability of conversion of such data to CDISC SDTM.

Conclusion

As we accommodate more RWD into RCTs, we are likely to get at the intersection of interoperability needs between different healthcare data standards and regulatory submission requirements in the pharmaceutical industry. Large healthcare institutions are following the Fast Health Interoperability Network (FHIR)[8] platform for data exchange. This platform facilitated by HL7 has framework for data standards. However, the input data going to this framework may follow various data models including OMOP CDM. There have already been efforts taken to harmonize OMOP CDM with FHIR platform and such efforts endorsed a need to establish more robust mapping mechanisms of FHIR W5 ontology with real-world data models like OMOP CDM[9]. In order to accommodate and effectively use the observational data collected through the healthcare system, further efforts in modelling and alignment of OMOP CDM with FHIR platform and with CDISC model will be required in future.

REFERENCES

1) Observational Health Data Sciences and Informatics, www.ohdsi.org

2) Clinical Data Interchange Standards Consortium, www.cdisc.org

3) OMOP-CDM Data Standards, https://ohdsi.org/data-standardization/

4) Data Standards for Drug and Biological Product Submission Containing Real- World Data, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data

5) SNOMED International, https://www.snomed.org/

6) Medical Dictionary for Regulatory Activities, MedDRA, https://www.meddra.org/

7) Leveraging the OMOP CDM V5 for CDISC SDTM RCT Data, by Joshua R., et. all, https://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=projects:workgroups:shyft-ohdsi-poster.pdf

8) Fast Health Interoperability Network, https://www.hl7.org/fhir/

9) A Consensus-based Approach for Harmonizing the OHDSI Common Data Model with HL7 FHIR, by Guojian J. et. all, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5939955/

ACKNOWLEDGMENTS

Authors are very much thankful to PHUSE Organization and community for providing a platform for industry collaboration and for sharing the research findings.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Parag Shiralkar
Company: Sumptuous Data Sciences, LLC
Email: Parag.shiralkar@sdatasci.com

Author Name: Mary Anne Rutkowski
Company: Merck
Email: mary_anne_rutkowski@merck.com