Name:	Hamed Abedtash
Affiliation:	Indiana University School of Informatics and Computing, Eli Lilly and Co.
Email:	abedtash_hamed@lilly.com
Presentation type (s):	**Poster**

Assessing Mapping Performance from HL7 Continuity of Care Documents to the OMOP CDM

Hamed Abedtash, PharmD, PhD[1,2], Jon D. Duke, MD, MSc[3]
[1]Indiana University School of Informatics and Computing, Indianapolis, IN; [2]Eli Lilly & Company, Indianapolis, IN; [3]Georgia Tech Research Institute, Atlanta, GA

Abstract

This paper presents the assessment of OMOP CDM to accommodate exchanged electronic medical records (EHRs) through HL7 continuity of care documents (CCD). The developed Extract-Transform-Load (ETL) Python Package extracted data values by HL7 Template IDs, and transformed data into OMOP CDM. The studies CDM tables could successfully accommodated CCD data, and yielded very good to excellent performance in mapping source codes to OMOP standard concepts. This work complements previous efforts to standardize data format and representation of medical records for clinical and epidemiological research purposes. This package will support OMOP CDM extensibility through integrating longitudinal data from diverse EHR systems, and help deploy OMOP-based analytics tools on select cohorts of patients where transformation of medical records occurs in a real-time manner.

Introduction

OHDSI community has developed Extract-Transform-Load (ETL) pipelines to transform administrative claims and electronic medical records (EHRs) to OMOP common data model (CDM) to detect post-marketing drug safety signals (1-3), comparative effectiveness analyses (4), and share registry data (5). However, no solution is available that transforms electronic medical records (EHRs) via HL7 messages as the widely accepted standard to exchange medical records in healthcare system. Such solution will be significant to enhance OMOP CDM extensibility by enabling deployment of OHDSI analytics tools on patient medical data in a real-time manner. It will also reduce the time and cost of data formatting as it allows setting up an automated process on the target cohorts data.

This study aimed to assess the feasibility and performance of transforming exchanged EHR data via HL7 CCD messages to OMOP CDM in a real-time manner. We tested the hypothesis that OMOP CDM is capable of accommodating HL7 data through evaluating data extraction process, concept-mapping performance, accuracy of derived elements, and loading process of formatted data.

Methods

We developed the CCD-TO-OMOP ETL module that extracts data from HL7 C-CDA-based CCDs, maps concepts to OMOP vocabulary, transforms the data into OMOP data model, and loads them into a PostgreSQL repository. The ETL pipeline locates data elements in CCDs by Template IDs and specifications defined in HL7 implementation guide document (6). The package consists of four modules (**Figure 1**). The *CCD Parser* module extracts data values from CCDs; the *OMOP Mapper* transforms the data into intermediate OMOP tables; the *Loader* loads the transformed data from the intermediate tables into an OMOP CDM database. The *Database Connector* connects other modules to OMOP CDM.

The scope of this study was to explore certain tables of OMOP CDM that are key to conduct observational studies, including Person, Observation Period, Visit Occurrence, Condition Occurrence, Condition Era, Procedure Occurrence, Drug Exposure, Drug Era, Measurement, and Observation tables. The accuracy and performance of the pipeline was assessed on 250 CCDs obtained from Regenstrief Institute. The data processing pipeline was manually reviewed to examine the accuracy of data extraction, concept mapping, calculated fields (e.g., drug day supply), derived elements validity (e.g., drug and condition era constructions), and loading process.

Results

The ETL could successfully extract proper data elements from CCDs, mapped the data to OMOP vocabulary, and loaded the transformed data into CDM tables accordingly. All CDM tables assessed in this study could successfully accommodated CCD data; however, some tweaks were needed in Drug Exposure and Condition Occurrence tables

to optimize data transformation. The ETL pipeline yielded very good to excellent performance in mapping source codes to OMOP standard concepts (**Figure 2**). A total of 12,648 records and 1,459 concepts of diagnoses and reported conditions were retrieved from the CCDs (Table 1). The concept-mapping pipeline yielded an overall recall of 98.5% and precision of 100%. In general, CCD documents provided minimum required data elements of patient information to feed OMOP CDM.

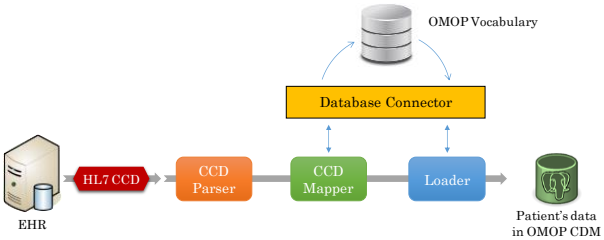

Figure 1. A schematic of CCD-TO-OMOP package

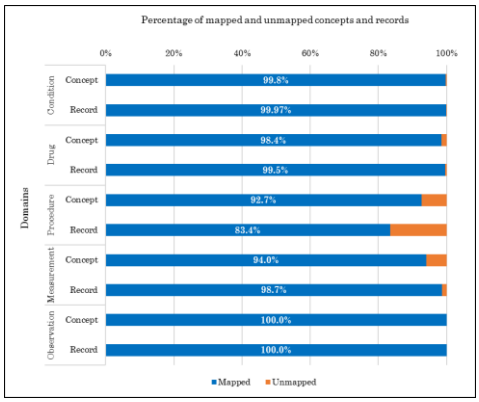

Figure 2. Overall mapping performance of concepts and records to OMOP CDM vocabulary by domain.

Table 1. Mapping performance of source codes to standard concepts of OMOP CDM vocabulary.

Domain and Code System	Concepts			Records		
	Mapped, *n* (%)	Unmapped, *n* (%)	Total	Mapped, *n* (%)	Unmapped, *n* (%)	Total
Condition	1,456 (99.8)	3 (0.2)	1,459	12,644 (99.97)	4 (0.03)	12,648
Drug	758 (98.4)	12 (1.6)	770	7,230 (99.5)	33 (0.5)	7,263
Procedure	367 (92.7)	29 (7.3)	396	941 (83.4)	187 (16.6)	1,128
Measurement	642 (94.0)	41 (6.0)	683	35,724 (98.7)	460 (1.3)	36,184
Observation	71 (100)	–	71	280 (100)	–	280
Race	4 (4)	–	4	74 (100)	–	74

Conclusion

The OMOP CDM demonstrated the ability to accommodate patient EHRs transferred by HL7 C-CDA-based CCD, and the final dataset can be used for drug safety surveillance, comparative effectiveness, and other observational studies. This work complements previous efforts to standardize data format and representation of medical records for clinical and epidemiological research purposes. It may also facilitate involvement of clinical institutions in large-scale observational data analytics using OHDSI software tools even with non-OMOP-based data warehouses

References

1. Trifiro G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? Journal of internal medicine. 2014;275(6):551-61.
2. Rho MJ, Kim SR, Park SH, Jang KS, Park BJ, Choi IY, editors. Development common data model for adverse drug signal detection based on multi-center EMR systems. 2013 International Conference on Information Science and Applications, ICISA 2013; 2013.
3. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. Journal of the American Medical Informatics Association : JAMIA. 2012;19(1):54-60.
4. FitzHenry F, Resnic FS, Robbins SL, Denton J, Nookala L, Meeker D, et al. Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. Applied clinical informatics. 2015;6(3):536-47.
5. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. Journal of biomedical informatics. 2016;64:333-41. doi: 10.1016/j.jbi.2016.10.016
6. Health Level Seven International. HL7 Implementation Guide for CDA Release 2: IHE Health Story Consolidation, DSTU Release 1.1 (US Realm) Draft Standard for Trial Use: Health Level Seven, Inc.; 2012.