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

Hamed Abedtash, PharmD, PhD[1,2], Jon D. Duke, MD, MSc[3]

[1] Eli Lilly & Company, Indianapolis, IN; [2] Indiana University School of Informatics and Computing, Indianapolis, IN; [3] Georgia Tech Research Institute, Atlanta, GA

Background

HL7 CCD messages are widely being used for exchanging patient medical records across healthcare IT systems. However, no solution is available to transform CCD data to OMOP CDM.

Objectives

To assess the **feasibility** and **performance** of transforming HL7 CCD data to OMOP CDM.

Hypothesis: OMOP CDM is capable of accommodating HL7 data.

Significance

- **Enhances OMOP CDM extensibility** by enabling deployment of OHDSI analytics tools on patient medical data in a real-time manner.
- **Reduces time and cost** of data transformation
- **Allows automated, real-time processing** of target cohorts data.

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

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

Methods

The CCD-TO-OMOP ETL module:
- Extracts data from HL7 C-CDA-based CCDs,
  - The ETL pipeline located data elements by **Template IDs** in the CCDs as defined in HL7 implementation guide document (1).
- Maps concepts to OMOP vocabulary,
- Transforms the data into OMOP data model,
- Loads the transformed data into a PostgreSQL repository.

CCD-TO-OMOP Package (**Figure 1**)
- **CCD Parser** module extracts data values from CCDs
- **OMOP Mapper** transforms the data into intermediate OMOP tables
- **Loader** loads the transformed data from the intermediate tables into OMOP CDM database
- **Database Connector** connects other modules to OMOP CDM

Assessments
- The performance and accuracy of extraction, concept-mapping, and loading processes, and validity of derived elements (e.g., drug day supply, drug and condition era constructions).
- Assessed on 250 CCDs obtained from Regenstrief Institute.
- Gold reference: Expert manual review

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

Table 2. Challenges and recommendations to enhance ETL performance

Challenge	Recommendation
In case of active allergy, CCD provides the type of allergy and the causing agent, but CDM does not accommodate both data values (**Figure 3**).	Option 1: Add a new field in Condition Occurrence table for the causing agent. Option 2: Map active allergy to the concept that contains both values.
Unmapped procedure codes were due to deprecated relationships to standard concepts.	Add updated relationships to vocabulary.
No placeholder exists in CCD schema to provide drug's days of supply.	Use daily dose interval and quantity of drug per dose to calculate days of supply
Sparsely reported daily dose interval and quantity of drug per dose to calculate days of supply.	The CCD schema supports the data elements, so make sure they are included in patient CCDs.
Few records of medications coded in local codes or only drug name was provided.	Ask CCD provider to record drug information using RxNorm concepts.
CVX codes did not exist in OMOP vocabulary at the time of study.	As of August 2017, CVX codes are included in OMOP vocabulary.
Observation entries were optional in CCD schema, such as Family History Section, Functional Status Section, Problem Section, Allergies Section, etc.	If important, ask CCD provider to include the entries in the documents.

Results

- OMOP CDM could **successfully** accommodate HL7 CCD data; however, some tweaks were needed in Drug Exposure and Condition Occurrence tables.
- 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**).

98.5% Recall   **100%** Precision
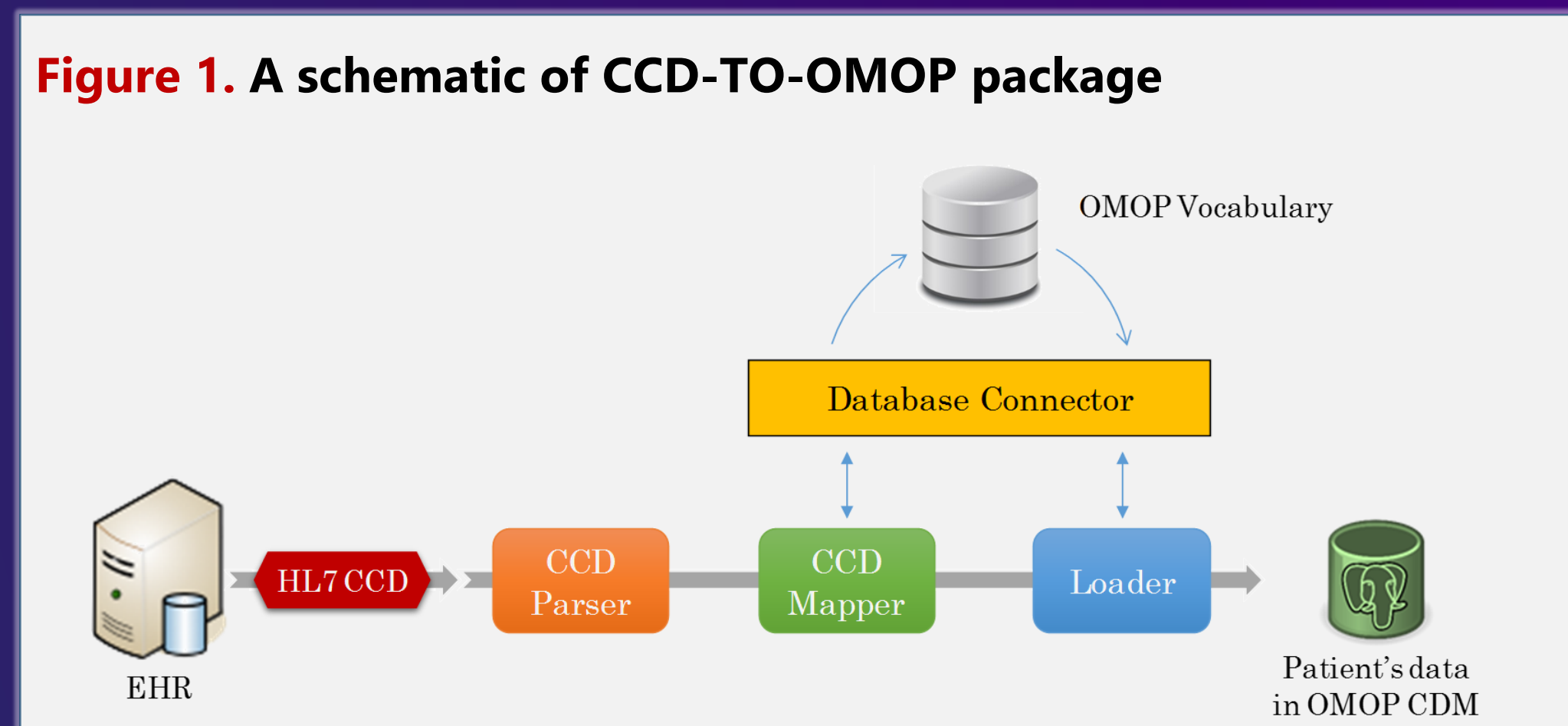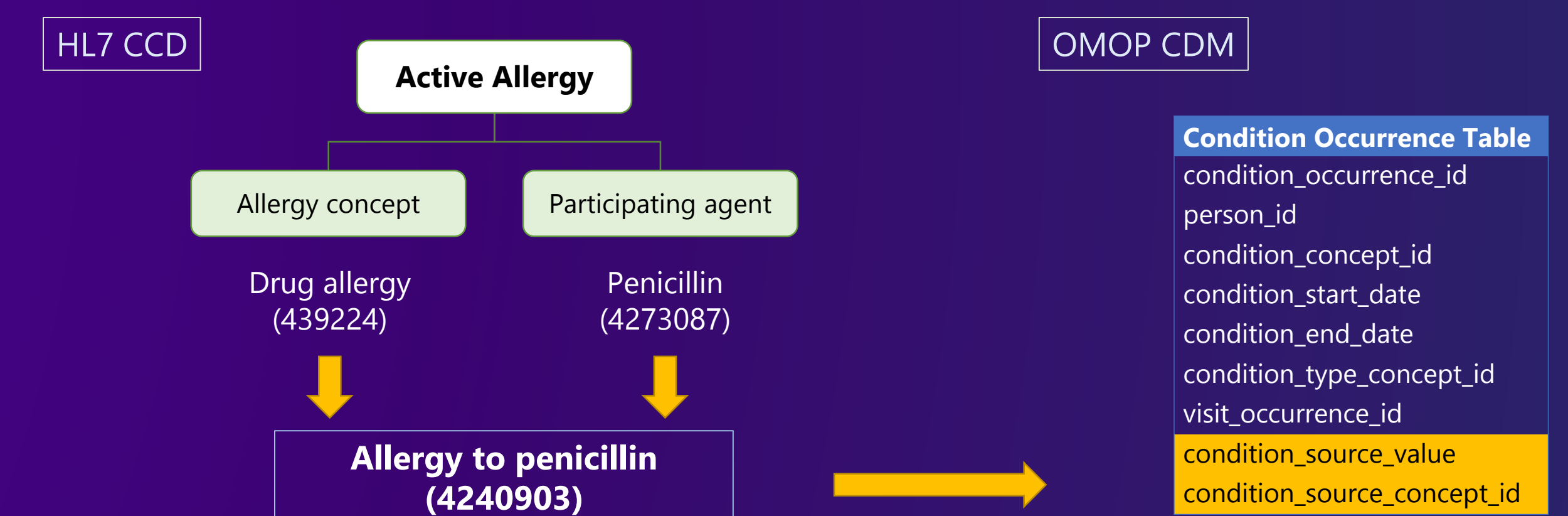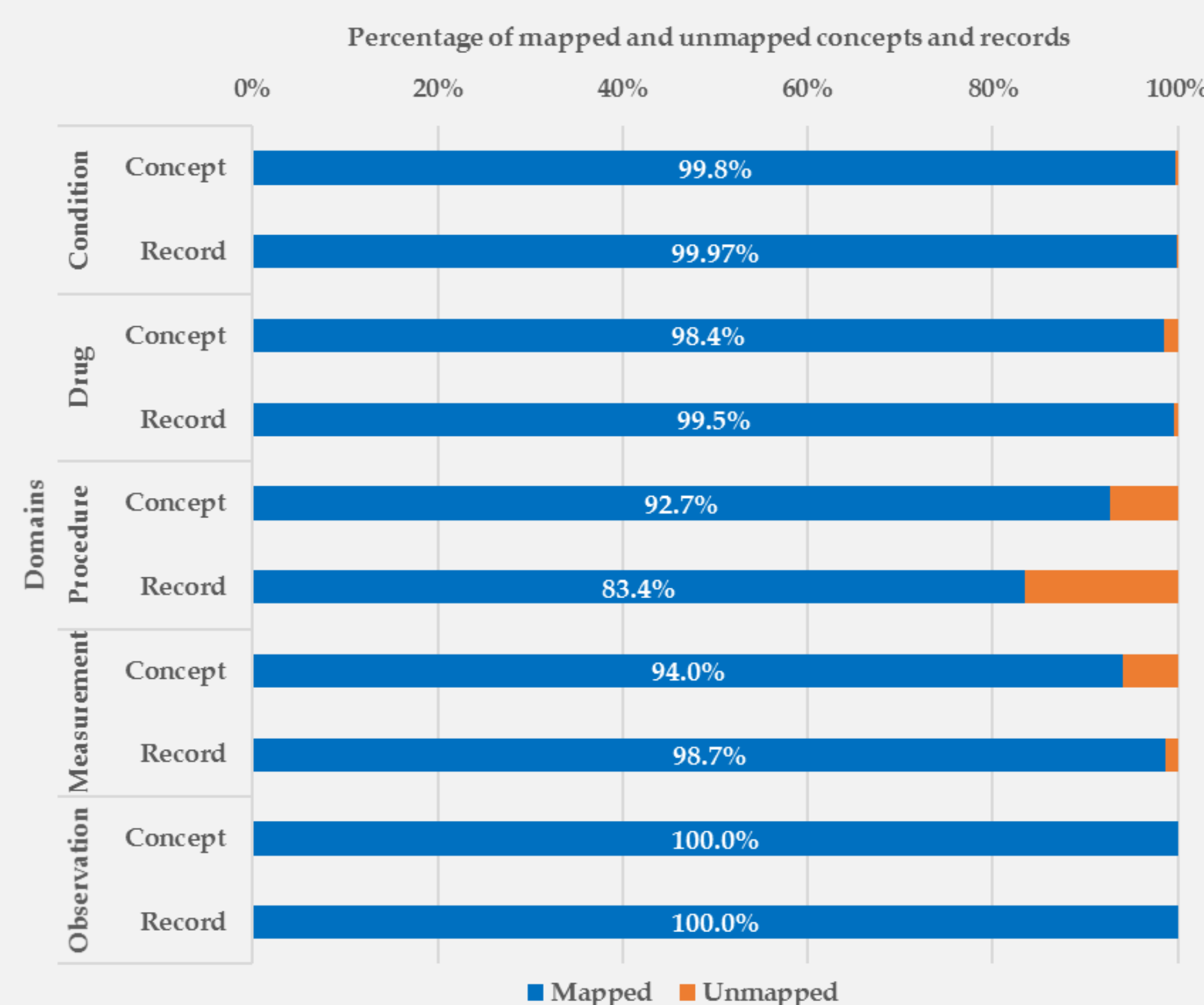
Concept-mapping performance

Figure 3. An example of mapping active allergy from CCDs to standard concept. The CDM does not accommodate both allergy type and causing agent.

Conclusion

- The OMOP CDM accommodates patient EHRs transferred through HL7 CCDs.
- To maximize the ETL performance, we addressed some challenges and recommended solutions in **Table 2**.
- The yielded repository can be used for drug safety surveillance, comparative effectiveness, and other observational studies.
- The module may facilitate involvement of clinical institutions in large-scale observational data analyses using OHDSI analytics tools even with non-OMOP-based data warehouses.
- The module allows automated, real-time transformation of EHR data to OMOP CDM.

Reference
1. 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.