OMOP and FHIR Data Comparison

Spencer SooHoo, Andrey Soares, Rohith Mohan, Renier Estiandan, Ryan Hoffman, Shao Chi Huang, Brian Tep, David Kreda, Dan Gottlieb, Aaron Boussina, Paul Kingsbury, Lisa Schilling

Background

The All of Us Research Program (AoURP)(1) has two methods of participation: via a participating Health Provider Organization (HPO) and via the Direct Volunteer (DV) Program. For participants consenting through an HPO, the HPO submits the participants' electronic health record (EHR) data that has modeled to the OMOP Common Data Model (CDM) (2). The DV process relies on a SMART-on-FHIR application which provides for patient-directed sharing of their EHR data via the FHIR format.(3) For both processes data are submitted to the AoURP Data and Research Center.

This project aimed to identify the high-level differences between the OMOP and FHIR data for key EHR domains (e.g., Patient, Conditions, Procedures, Medications, Observations, and Encounters).

Methods

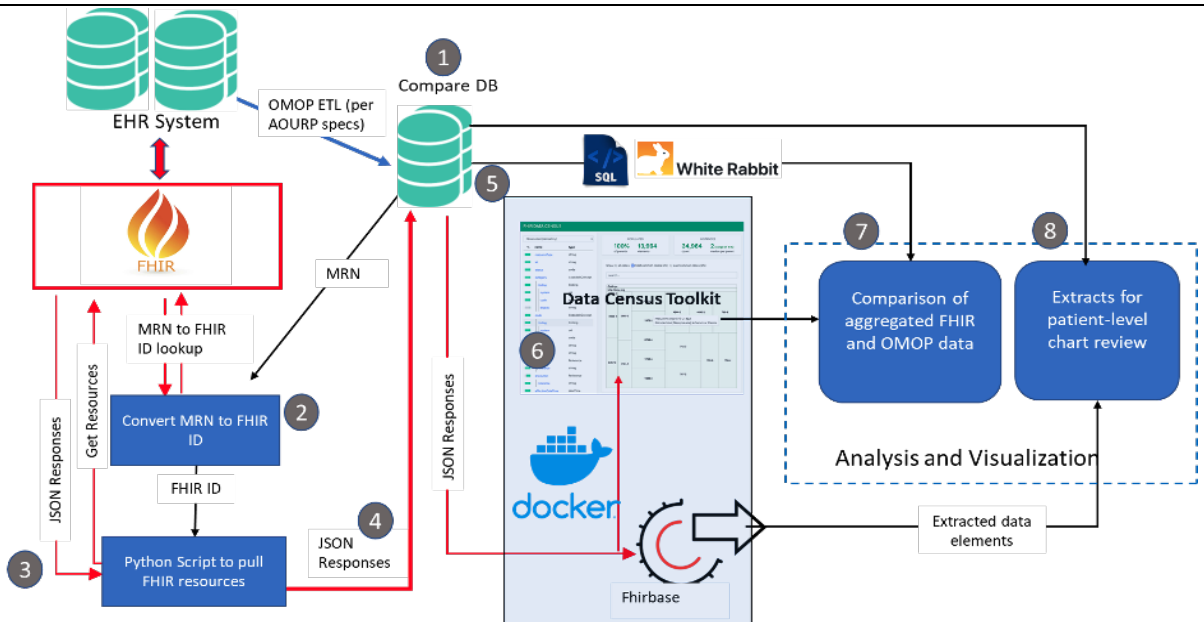
Cedars-Sinai Medical Center (CSMC) and Harvard Medical School (HMS) efforts led to a collaboration with the University of Southern California (USC), the University of California San Diego (UCSD), and the University of Colorado-Anschutz Medical Center (CU-AMC). CSMC distributed software for the creation of two final analytic datasets, OMOP and FHIR. Sites extracted data from their respective OMOP CDM using AoURP ETL specifications and EHR FHIR systems for 5000 randomly selected patients who met the eligibility criteria. Table 1 shows data sources and tools used for each site. The process developed at CSMC (Figure 1) was designed so that all patient-level data remains at each site. Only aggregate, de-identified results were shared among the collaborating sites.

Table 1. Site-specific details for OMOP and FHIR data sources.

Site	OMOP	FHIR	Tools
CSMC	ETL from Epic Clarity	Extract from FHIR endpoint	OMOP: SQL Scripts FHIR: Data Census Toolkit(4), FHIRBase(5)
CU-AMC*	ETL from Epic Caboodle	Extract from FHIR endpoint	OMOP: White Rabbit/Rabbit-in-a-Hat(6), SQL FHIR: Data Census Toolkit, FHIRBase
UCSD	ETL from Epic Clarity	Extract from FHIR endpoint	OMOP: SQL Scripts FHIR: scripts for parsing JSON resources
USC	ETL from Cerner	Unable to get organizational approval to access FHIR endpoint	

* All except CU-AMC are AoURP HPOs.

Figure 1. High-level view of process to compare OMOP and FHIR Data. The initial step was to perform a high-level characterization of the two data streams using aggregate counts of records for selected domains.



1. Generate an OMOP data mart (Compare DB) of 5000 randomly selected patients using the AoURP OMOP CDM ETL specifications.
2. Use the MRNs from the patients as input for a script that makes FHIR requests to retrieve the FHIR ID for the patients.
3. Use a Python script to get FHIR Resources and receive the JSON responses from the FHIR endpoint.
4. Store the JSON responses into NDJSON files
5. Load the NDJSON to the Compare DB.
6. Use the Data Census Toolkit to produce high-level characterization of the FHIR data.
7. Use SQL scripts and tools like OHDSI White Rabbit to generate high-level characterization for the OMOP data.
8. Use FHIRBase to parse out discrete elements from the JSON responses that can be used for patient-level data comparisons.

Results

This section provides the findings from the three sites that were able to perform a set of comparisons of their OMOP and FHIR data for their respective 5000 patient cohort samples.

Cedars-Sinai Medical Center

The analysis showed close agreement between the OMOP and FHIR demographic data at the cohort-level but varied considerably in other domains. To investigate further, CSMC performed manual chart review on ten patient charts using the EHR as the gold standard.

Identified differences between the OMOP Procedure_Occurrence and corresponding FHIR Procedure Resource revealed that on average, there were 10 times more procedure records in OMOP than in FHIR. Much of this difference is the inclusion of evaluation and management (E&M) CPT codes as procedures in OMOP and not in FHIR. Another difference was that imaging procedures were present in OMOP Procedure_Occurrence table, while in FHIR they were represented in either the ImagingStudy or DiagnosticReport Resource. At CSMC, the FHIR Procedure reflected what was found in Epic's Hyperspace

front-end Procedure Chart Review Tab.

Overall, there was good alignment in Medications, except for entries that did not have an associated RxNorm code and are excluded by the AoURP ETL rules and were therefore not in the OMOP data, Medications administered as part of a procedure, were found in the OMOP data but not in FHIR. FHIR conditions were from the Epic Problem List only due to the set category parameter. There was good correspondence between OMOP and FHIR cross mapped by LOINC code for lab results. At CSMC, FHIR contains Encounters for orders only, patient messages, and outside records encounter types but did not contain an encounter type for office visits. In contrast, OMOP data had complete visit and hospital encounter information.

University of Colorado-Anschutz Medical Campus

The OMOP Drug_Exposure concept types map to different FHIR Medication-related Resources (Table 2). The drug_type_concept_id indicates the type of drug exposure such as immunization or order. A summary of the OMOP and FHIR record counts is presented in Table 2.

Table 2. Aggregate OMOP and FHIR Records for CU-AMC with highlights of the counts based on type_concept_ids in OMOP and FHIR parameters.

OMOP Domain	OMOP Counts		FHIR Resource	FHIR Counts
Person	5,000		Patient	5,000
Condition_Occurrence	749,101		Condition (FHIR Problem List only)*	48,383
– EHR Problem List only	53,454		Observation (category=social history)	8,560
			AllergyIntolerance	9,119
Procedure_Occurrence	725,043		Procedure	357,387
			Immunization**	32,034
Drug_Exposure	823,269		Medication	403,589
– EHR Orders only	342,282		MedicationRequest	416,589
– Patient Reported only	45,205			
– Immunization (CVX codes)	9,723		Immunization**	32,034
Measurement	4,949,581		Observation (category=laboratory)	1,737,419
			Observation (category=vital-Signs)	1,300,767
Visit_Occurrence (limited to completed visits)	197,293		Encounter (and Type: finished)	401,269
Death	172		Patient (deceased)	-

* The Python script to extract FHIR data retrieved category = “problem-list-item” and clinical-status = “active,resolved”.
** Administered immunizations are recorded as drugs and procedures in the OMOP CDM

When we limited the OMOP Drug_Exposure records (n=823,269) to EHR medication orders (n=342,282), this number was fairly close to the FHIR MedicationRequest count (n=416,589). The MedicationRequest Resources includes prescribed, hospital and clinic administered, and patient reported medications. It does not include immunization which are represented in the FHIR Immunization Resource.

During the analysis of single events per patient, we identified important differences between CU-AMC

OMOP and FHIR data. While both represent basic immunization information (e.g., CVX Code, patient, and occurrenceDateTime), the FHIR data provides more information about the immunization including manufacturer, lot number, expiration date, site (e.g., left deltoid) and performer (administering and ordering providers).

University of California San Diego

Patient demographics were generally consistent between OMOP and FHIR. One notable difference was that 8 more patients were recorded as deceased in FHIR. Manual chart review of these records revealed that FHIR was properly capturing the deceased status. UCSD was excluding patient death records from OMOP, as per OMOP convention when death date was missing.

The overall counts of procedures and drug exposures between OMOP and FHIR differed substantially. A deep dive into the discrepancies revealed that the OMOP data pulled elements from a wider array of sources than FHIR. For example, the OMOP Condition_Occurrence table included diagnoses from the hospital account billing tables and conditions identified from diagnostic procedures. The FHIR APIs, by comparison, only contained items from the problem list and the encounter diagnoses.

Values from the OMOP Measurement table were overall consistent with those observed from the FHIR Observation Resource. The overall count of encounters from FHIR vs the OMOP Visit_Occurrence table were also relatively close, with the discrepancy likely the result of Office Visits being absent in the FHIR server.

Conclusion

Except for demographic data, comparison of aggregate record counts for OMOP and FHIR initially appeared dissimilar. The differences can partially be explained by inherent OMOP and FHIR differences, including 1) OMOP ETL conventions, 2) the parameters used for the FHIR requests, and 3) current FHIR conventions and site implementations. Targeted chart review was invaluable in illuminating some of the differences between the OMOP and FHIR data.

Limitations of the comparison include the inability to perform identical comparisons across all sites, to extend the detailed analyses to all other data domains, and the lack of EHR vendor diversity which would help identify differences between vendor implementations of FHIR.

Our next steps are to conduct additional detailed comparisons informed by these processes and results.

References/Citations

1. All of Us Research Hub [Internet]. [cited 2022 Jun 26]. Available from: https://www.researchallofus.org/

2. OMOP – OHDSI [Internet]. [cited 2022 Jun 26]. Available from: https://ohdsi.org/omop/

3. FHIR v4.3.0 [Internet]. [cited 2022 Jun 26]. Available from: https://www.hl7.org/fhir/

4. Data Census Toolkit [Internet]. Sync for Science; 2022 [cited 2022 Jun 26]. Available from: https://github.com/sync-for-science/data-census

5. Fhirbase - Open Source Database for HL7 FHIR® [Internet]. [cited 2022 Jun 26]. Available from: https://www.health-samurai.io/fhirbase

6. White Rabbit [Internet]. Observational Health Data Sciences and Informatics; 2022 [cited 2022 Jun 26]. Available from: https://github.com/OHDSI/WhiteRabbit

7. United States Core Data for Interoperability (USCDI) [Internet]. [cited 2022 Jun 26]. Available from: https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi

Acknowledgments

This project was funded by the All of Us project. HL7®, and FHIR® are the registered trademarks of Health Level Seven International and their use of these trademarks does not constitute an endorsement by HL7. The material in this Technical Report reflects the work of the authors and does not necessarily reflect the position of All of Us or HL7®.