

A Lightweight FHIR-to-OMOP Software Demonstration Using pyOMOP

Sanghati Basu, Bell Raj Eapen

University of Illinois Springfield, Springfield, IL, USA

University of Illinois Springfield, Springfield, IL, USA

Code: <https://github.com/SANGHATI23/ohdsi-fhir-omop-showcase-demo>

Background

FHIR supports interoperable exchange of clinical data, whereas the OMOP Common Data Model (CDM) supports standardized observational analysis. Moving from FHIR Bulk Data to a queryable OMOP database remains challenging for new OHDSI collaborators. We demonstrate a lightweight, local workflow using Synthea, pyOMOP—a tool developed by our team, Athena vocabularies, SQLite, and basic analytics [1–5]. We audited patient counts and concept mappings and added reproducible post-import corrections. The intended users are students, clinical informatics researchers, and developers learning vocabulary mapping, source reconciliation, and initial cohort queries. This is an educational software demonstration, not a production ETL benchmark.

Methods

Synthea-generated FHIR R4 Bulk NDJSON included Patient, Encounter, Condition, Observation, MedicationRequest, Medication, Immunization, and related resources. pyOMOP 6.4.0 was used to transform the FHIR data into a local SQLite OMOP CDM using OHDSI Athena vocabularies. Custom Python and SQL scripts were used to inspect the transformed OMOP tables, generate demographic and clinical summaries, evaluate vocabulary mappings, and run a constrained natural-language-to-SQL demonstration.

The workflow also included source-to-target reconciliation between FHIR resources and OMOP records, validation of Drug and Visit concept mappings, and automated regression testing. Mapping issues identified during this validation were corrected in a separate working copy while preserving the original database for comparison. The repository includes the transformation workflow, validation scripts, correction utilities, automated tests, and instructions for preparing the required Athena vocabulary files.

Results

The retained input contains 27 unique FHIR Patient resources, with a one-to-one match to 27 OMOP PERSON records. The input also contains 405 Immunizations and 1,386 Encounters. OMOP contained 983 conditions, 1,275 drug exposures, 14,150 measurements, and 14,168 observations.

The submitted nonzero-concept coverage was 100% for conditions, 78.12% for drugs, 99.99% for measurements and observations, and 0% for visits (Figure 1).

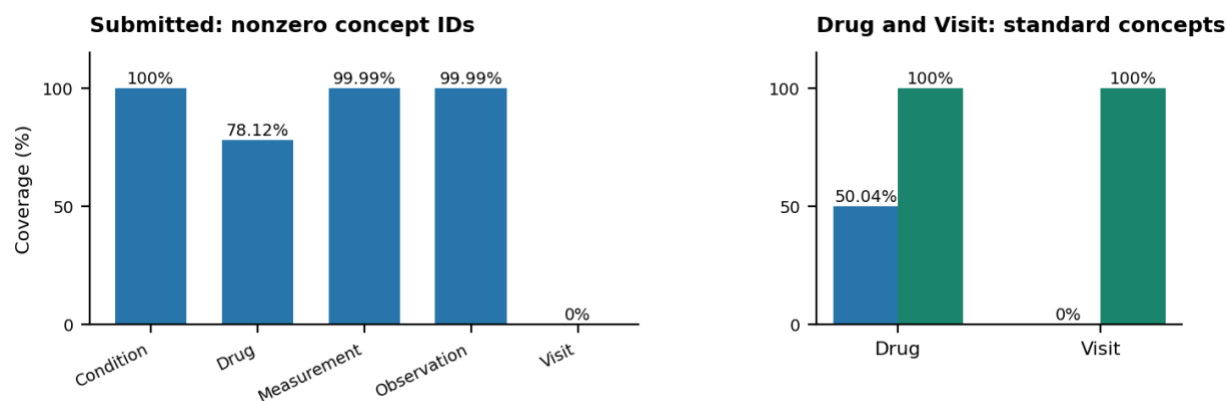


Figure 1. Submitted nonzero concept IDs (left) and retrospectively audited valid-standard coverage (right). Blue indicates baseline and green indicates corrected records; the panels use different metrics.

Of 1,275 baseline drug records, 996 had nonzero concept IDs, but 358 were immunizations incorrectly assigned to unrelated payer or billing vocabularies. The other 279 had zero IDs: 232 medication-reference records and 47 immunizations. The retrospective audit therefore identifies 638 baseline records with valid standard Drug concepts (50.04%); 78.12% was nonzero-ID coverage, not validated mapping. The 47 initially unmapped and 358 incorrectly mapped immunizations together account for all 405 FHIR Immunizations.

Targeted corrections on a separate SQLite copy resolved the medication references, applied CVX-specific vaccine mappings, corrected vocabulary collisions, and mapped Encounter classes to standard Visit concepts. All 1,275 drugs and 1,386 visits passed local valid-standard concept checks. Nine medication records have reconstructed-equivalent assignments because unique historical linkage within same-time groups was not recoverable. Cross-vocabulary mappings for CVX 207, 208, and 212 were checked in Athena's interface but were absent from the CVX-only download.

The original and corrected domain results are summarized in Table 1.

Table 1. Submitted nonzero-ID coverage versus corrected valid-standard coverage. Other domains were not separately revalidated.

OMOP domain	Submitted: nonzero ID	Corrected: valid standard
Condition occurrence	100.00%	Not separately revalidated
Drug exposure	78.12%	1,275/1,275 (100%)
Measurement	99.99%	Not separately revalidated
Observation	99.99%	Not separately revalidated
Visit occurrence	0.00%	1,386/1,386 (100%)

Encounter-class assignments comprised AMB (1,338) to Outpatient Visit (9202), EMER (31) to Emergency Room Visit (9203), IMP (13) to Inpatient Visit (9201), VR (2) to Telehealth (722455), and HH (2) to Home Visit (581476). The corrected showcase completed six analytics stages and returned 19 male, eight female, and 27 total patients. Four synthetic replay/safety tests and three SQL-query tests passed; a fresh full-cohort import has not yet been repeated.

Conclusions

This educational demonstration shows how a local FHIR-to-OMOP workflow can expose and repair mapping failures while preserving its submitted baseline. Complete valid-standard concept coverage does not establish clinical mapping accuracy or full preservation of source meaning. A fresh full-cohort pyOMOP import and stage-level execution timings remain pending; portability beyond SQLite and comparative scalability have not been validated.

References

1. OHDSI. The Book of OHDSI. <https://ohdsi.github.io/TheBookOfOhdsi/>
2. HL7 International. FHIR Bulk Data Access Implementation Guide. <https://hl7.org/fhir/uv/bulkdata/>
3. Walonoski J, Kramer M, Nichols J, et al. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J Am Med Inform Assoc.* 2018;25:230–238. doi:10.1093/jamia/ocx079.
4. OHDSI. Athena Standardized Vocabularies. <https://athena.ohdsi.org/>
5. pyOMOP. Python package for OMOP CDM creation and FHIR import. <https://github.com/dermatologist/pyomop>