

Lateral specific concept mapping from ICD-10 to SNOMED CT: addressing information loss

Vojtech Huser¹, Ilya Pyatin¹, Tatsiana Skuhareuskaya¹, Maria Khitrin¹, Sarah Seager¹, Sebastiaan van Sandijk¹, Polina Talapova², Anna Ostropolets³, Oleg Zhuk¹
¹EPAM, ²SciForce, ³Columbia University

BACKGROUND

Medical terminologies, such as SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), play an important part in data harmonization. Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) includes an important terminology layer in the model and uses the principle of non-standard terms being harmonized to standard terms. Any gaps and challenges in this principle and underlying terminologies may impact research use of the OMOP-shaped data.

Pre-coordination is a frequently adopted approach in medical terminologies. SNOMED International allows creation of pre-coordinated, lateral side specific disease concepts in SNOMED CT.[1] In current SNOMED CT version, this approach may not be fully implemented (i.e., not all expected concepts may currently exist). Another important terminology, ICD-10CM, which is a source data terminology, also uses pre-coordination to capture left or right location detail about a diagnosis.

Current mapping of ICD10-CM concepts for diagnoses within OMOP model terminology layer (standardized vocabularies) often maps side specific diagnoses to standard concepts (from SNOMED CT terminology) that are not side specific (narrow to broad type of mapping). This leads to information loss and it is not optimal for some research questions. Our goal is submit new concept requests to SNOMED International (standard developing organization for SNOMED CT) for side specific condition concepts and to better quantify to what extent a mapping that “drops laterality” occurs. In this study, we focus on diagnoses in the ophthalmology domain.

MATERIALS

SNOMED International provides an Excel batch submission template. Each new proposed concept must be linked to a parent concept. A pilot submission of two concepts made by us was accepted by SNOMED and provides guidance for future (and larger) batch submission targeted by this study.

METHODS

We used published concept counts [2] to arrive at ICD-10CM ophthalmology conditions that include left or right laterality. We define DiseaseConceptWithLaterality (e.g., <https://athena.ohdsi.org/search-terms/terms/37200408>; Exudative age-related macular degeneration, left eye; H35.322) as concept that describes a disease and includes a specific left vs. right side location. Similarly, we define DiseaseConceptWithoutLaterality (e.g., <https://athena.ohdsi.org/search-terms/terms/376966>; Exudative age-related macular degeneration; SNOMED CT concept code 414173003) as concept that describes a disease only and does not include left or right side as location. We used regular expressions to identify DiseaseConceptsWithLaterality in non-standard and standard concept subgroups.

We decided that the submission of new concepts will be split into batch A that will contain requests for new concepts where usage (count of events on one or multiple datasets) of the ICD10 term is over some percentile. In other words, batch A will target the most important diagnoses and avoid overloading SNOMED team with a large number of new concept requests. Remaining new concepts will be submitted in a possible later batch B. Results of the ophthalmology domain implementation will inform later phase 2 global scope expansion of the approach (for all diagnoses not limited to ophthalmology).

For batch submission, we generated new concepts with fully specified names (FSNs) using a script. We added ‘of left eye’ and ‘of right eye’ to the existing DiseaseConcept WithoutLaterality. This strategy was used in our pilot submission.

RESULTS

Pilot submission was made in October 2025 and appeared in production in December 2025 (SNOMED CT US Edition; facilitated by the monthly release cycle). Study repository at github.com/informaticsrepo/omop-lat contains result files referenced below. We found 3116 ICD10-CM source ophthalmology diagnosis concepts of type DiseaseConceptWithLaterality (see supplemental file s1.csv) mapped to 730 standard SNOMED CT concepts (file s2.csv). Study repository contains analogous results and supplemental files for phase 2 scope for all diagnoses, other expanded results and additional discussion points. For batch A, number of SNOMED CT parent concepts where lateral-specific terms will be requested is: 20 (for top 3 percentile), 34 (for top 5 percentile) and 132 (for top 20 percentile) (file s3.csv).

CONCLUSION

New SNOMED CT terms allow for preservation of laterality information contained in the source data and source concept (e.g., H35.322). In phase 2 context, we quantified that laterality is preserved in 17.2% and dropped in 92.8%.

DISCUSSION

There are several related considerations. First, OMOP model currently relies fully on pre-coordinated terms. Laterality addition is a poster-child example of possibly using a post-coordination approach. SNOMED CT expression language is a fully developed post-coordination framework. The HL7 FHIR standard also mostly works with pre-coordination approaches. Changes to OMOP model must be carefully considered (balanced against the virtue of relative model stability over time) and we concluded that large paradigm shift for *_concept_id columns (introducing post-coordination) is not something the OMOP research community may prefer as of 2026.

Second, solving laterality for diagnoses is related to solving related problem in procedural data. E.g., intraocular injection of anti-VEGF agents to the affected eye (left vs. right) would possibly have to be similarly addressed in procedural terminology. We provide some expanded discussion on this topic in study repository. We define analogous scopes of phase 3 (ophthalmology procedures) and phase 4 (general procedures) for procedures.

Third, we fully acknowledge that for the majority of research analyses, the laterality information is not needed. However, in the current SNOMED-CT it is clearly implemented in some diseases/symptoms and not implemented in others. It is a somewhat subjective decision to declare it important for some concepts and not for other concepts. Different researchers will place the implementation boundary differently.

Fourth, OMOP vocabularies as of January 2026 consist of 447430 non-standard condition concepts and 167045 standard concepts (with SNOMED CT accounting for 105324 concepts and ICD-O-3 for 56858). Addition of several thousands of pre-coordinated terms within SNOMED CT is unlikely to disrupt existing approaches, tools or software packages. OMOP tools and model infrastructure for concept relationships can easily accommodate such an expansion. For users not interested in lateral side location detail, they can continue using existing concepts (with subsumed concepts included) as before.

REFERENCES

1. van Berkum MM, Rallins MC, Spackman KA. Choosing Sides. Assigning Laterality as an Attribute in SNOMED CT. Proc AMIA Symp. 2002:1184.
2. Datasets (Center for Clinical Observational Investigations)
<https://lhncbc.nlm.nih.gov/CCOI/datasets/datasets.html> [accessed Dec 3, 2025]