

IMPROVEMENT of CANCER DIAGNOSIS REPRESENTATION in OMOP CDM

Dmitry Dymshyts MD¹, Anna Ostropolets MD¹, Christian Reich MD, PhD²; Rimma Belenkaya, MA, MS³; Michael Gurley BA⁴

1-Odysseus Data Services, Cambridge MA USA; 2-IQVIA, Cambridge MA USA; 3-Memorial Sloan Kettering Cancer Center, New York, NY, USA; 4-Northwestern University, Chicago, IL, USA

1. BACKGROUND

In a source data, neoplastic disorders can be encoded in multiple coding systems including ICD9CM, ICD10, ICD10CM, and ICD-O-3. SNOMED is uniformly used to represent diagnosis in OMOP CDM. Currently, only ICD9CM, ICD10, and ICD10CM concepts are mapped to SNOMED. Moreover, SNOMED does not always provide granularity sufficient to represent cancer diagnoses and available in ICD-O-3 coding.

We propose a solution for representing ICD-O-3 in OMOP without changes of the existing Common Data Model (CDM) or conventions. Such a solution needs to add the granularity of the histology and topology information ICD-O-3 can provide, and to have these represented in a single Concept. In addition, these Concepts have to be incorporated into the hierarchical system of SNOMED while still preserving its description logic.

2. METHODS

To represent cancer diagnosis in the OMOP CDM Condition domain (Condition_Occurrence), we propose to perform a pre-coordination of the ICD-O-3 axes, topography and histology, to a single concept representing unique cancer diagnosis. We propose to map each combination to a corresponding SNOMED concept with respective anatomic site and morphology attributes. If a concept does not exist in SNOMED, we propose to create an OMOP standard concept.

Concept pre-coordination

In ICD-O, cancer diagnosis represented by a combination of ICD-O-3 topography and histology.

ICD-O-3 topography example: in C18.7, the C18 indicates that the site is the colon and the 7 indicates that the subsite is the sigmoid colon.

ICD-O-3 histology example: in 8140/3, the "8140" indicates the Adenocarcinoma, "3" indicates "malignant, primary site", this way we have "Adenocarcinoma, NOS".

We propose to represent a combination of ICD-O-3 topography and histology as one pre-coordinated concept and map it to a pre-coordinated SNOMED concept.

Example: 8140/3-C18.7 Adenocarcinoma of Sigmoid colon Maps to 301756000 Adenocarcinoma of Sigmoid colon

Demand-driven generation of concepts



Mapping between ICD-O-3 and SNOMED

Pre-coordinated histology-topography Concepts already exist in many cases in SNOMED-CT.

In order to detect those we mapped:

- 1 ICD-O-3 topography to SNOMED anatomy
- 2 ICD-O-3 histology to SNOMED morphology
- 3 Find a pre-coordinated SNOMED concept at the intersection of anatomy and morphology.

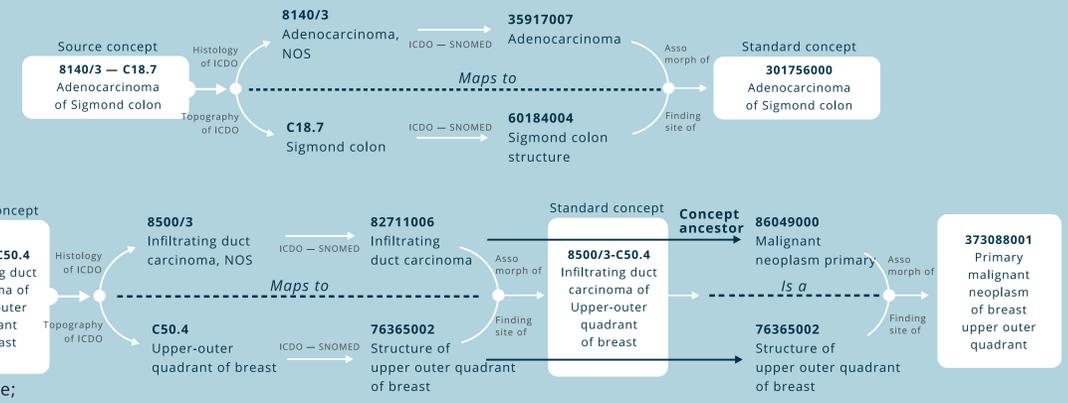
Mapping Challenges and Solutions

Missing pre-coordinated concepts

We represented a missing SNOMED pre-coordinated concept with a new standard OMOP concepts

Missing equivalents of ICD-O-3 topography in SNOMED

1. Overlapping lesions. SNOMED doesn't define the Overlapping lesion as the separate attribute;
 2. General categories like connective, subcutaneous and other soft tissues of upper limb and shoulder and similar concepts.
- We need to map ICD-O-3 combinations manually.



3. RESULTS

We added 6,554 ICD-O-3 combinations taken from the real patient data out of 50,000 possible combinations reported by SEER. 2,608 of these combinations had the SNOMED equivalent and 3,946 did not. We represented those missing pre-coordinated concepts by new OMOP standard concepts. The coded combinations were based on combinations occurring in one of our first available data set. When tested on another data set, the coverage was 75%.

Example of cancer registry data ETL using ICD-O-3 vocabulary

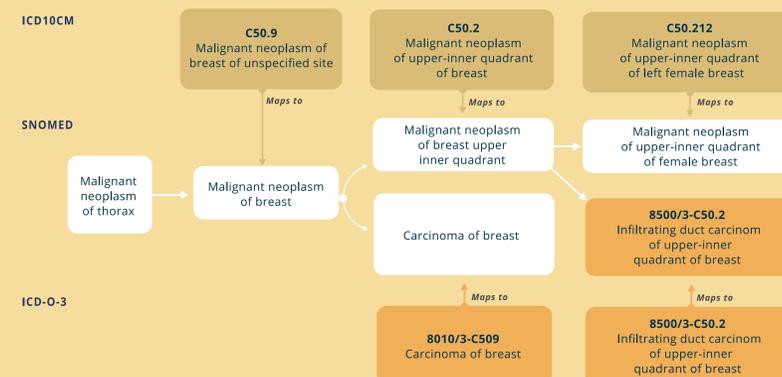
Cancer Registry data

	Visit 1	Visit 2
ICD10CM	C50.9 Malignant neoplasm of breast of unspecified site	C50.212 Malignant neoplasm of upper-inner quadrant of left female breast
ICD-O-3 Topography	C50.9, Breast, NOS	C50.2, Upper-inner quadrant of breast
ICD-O-3 Histology	8010/3, Carcinoma, NOS	8500/3, Infiltrating duct carcinoma, NOS

Condition Occurrence table

condition_occurrence_id	person_id	condition_concept_id	condition_start_date	condition_end_date	condition_type_concept_id	visit_occurrence_id	condition_source_value	condition_source_concept_id
2222	1111	<SNOMED: Malignant tumor of breast>	11/12/2016		<cancer registry>	3333	ICD10CM: C50.9	1567562
2222	1111	<SNOMED: Carcinoma of breast>	11/12/2016		<cancer registry>	3333	ICD-O-3: 8010/3,C50.9	44505310
2255	1111	<SNOMED: Malignant neoplasm of upper-inner quadrant of female breast>	12/7/2016		<cancer registry>	4444	ICD10CM: C50.212	35211522
2255	1111	<ICDO: Infiltrating duct carcinoma of Upper-inner quadrant of breast>	12/7/2016		<cancer registry>	4444	ICD-O-3: 8500/3, C50.2	45557035

Relational graph between SNOMED, ICD10CM, ICDO for the concepts from the example above:



4. CONCLUSION

We extended OMOP standardized vocabularies to support more granular representation of cancer diagnoses. To extend the coverage of mappings and improve the quality, we are presently converting several data sets containing ICD-O-3 cancer codes into OMOP CDM.

To perform mappings of the entire SEER reported set, we will improve mappings of ICD-O-3 histology and topography attributes to respective SNOMED attributes. We will formally propose that new pre-coordinated concepts are implemented in SNOMED and improved mappings between ICD-O-3 and SMOMED become part of UMLS.

The remaining challenge is to reconcile in OMOP heterogeneous cancer coding granularity inherited from the source systems. We intend to use SNOMED hierarchy to tackle this challenge.