**ETL instructions for mapping ICD-O to SNOMED**

**COMPLETE ICD-O SOURCE CODES**

Cancer diagnoses are usually represented by a combination of ICD-O-3 histology and topography codes. To map this combination to SNOMED follow these steps:

1. Transform diagnosis SOURCE VALUE
	1. Histology code. In the source, it is normally formatted like this: 8070/3, where 8070 is histology type and 3 is tumor behavior. If histology type and behavior are stored separately, concatenate them to get one histology concept, e.g. 8070/3.
	2. Topography code. the source, it is normally formatted like this: C50.2. Be aware of the dot. if the source doesn't have the dot, insert it after the 3d character: C502 -> C50.2. If the source code contains only 3 characters, the dot is not required: C50 -> C50.
	3. Source value. Concatenate histology code and topography code using hyphen: 8070/3-C50.2. This value will be stored in the CONDITION\_OCCURRENCE.CONDITION\_SOURCE\_VALUE field.
2. Extract value of diagnosis SOURCE CONCEPT ID

Concept ID for the combined histology/topography code is stored in the CONCEPT table. The following SQL shows how to extract its value for the above example:

SELECT CONCEPT\_ID

FROM CONCEPT

WHERE CONCEPT\_CODE = ‘8070/3-C50.2’

 AND VOCABULARY\_ID = ‘ICDO3’

The resulting value 36517865 will be stored in the CONDITION\_OCCURRENCE.CONDITION\_SOURCE\_CONCEPT\_ID field and will be used in mapping to a standard SNOMED code (next section).

1. Extract value of STANDARD CONCEPT ID

Source concept ID of the combined histology/topography code is mapped to a standard concept ID in the CONCEPT\_RELATIONSHIP table. The following SQL shows how to extract its value for the above example:

SELECT CONCEPT\_ID\_2

FROM CONCEPT\_RELATIONSHIP

WHERE CONCEPT\_ID\_1 = 36517865

 AND RELATIONSHIP\_ID = 'Maps to'

The resulting value [36517865] will be stored in the CONDITION\_OCCURRENCE.CONDITION\_ CONCEPT\_ID field.

**INCOMPLETE ICD-O SOURCE CODES**

In some cases when the source data are incomplete, apply the following approach.

1. Tumor behavior is not known

Use 1 (uncertain behavior) to making your code complete: 8070 -> 8070/1

1. Topography is unknown.

Use mappings from this file <https://seer.cancer.gov/tools/conversion/ICD03toICD9CM-ICD10-ICD10CM.xls> (last 3 tabs of this file) to obtain topography if you have ICD-10 code for this diagnosis. Note, if you have long ICD-10CM code, you need to cut it off to have only 5 symbols (including dot): C50.211 -> C50.2

In case when a patient has several cancer diagnoses, use ICD-10 from the date closest to the ICD-O histology date.

**REFERENCES**

Information about ICDO3 vocabulary is here: <http://codes.iarc.fr/usingicdo.php>

Information about our approach to mapping is her:

<http://www.ohdsi.org/web/wiki/lib/exe/fetch.php?media=documentation:oncology:poster2018-improvement_of_cancer_diagnosis_representation_in_omop_cdm3_1_.pdf>