The OMOP ICDO3 vocabulary is incomplete. Give us your missing codes!

Mapping of ICD-O-3 codes to standard concepts: Lessons learned from four population-based cancer registries

Background: ICD-O-3 contains lists of histology and topography codes. A complete list of conditions, i.e., valid combinations of a histology and a topography, does not exist. The OMOP ICDO3 vocabulary consists of the most commonly occurring conditions in a limited set of data sources: several conditions are missing from the vocabulary.

Result 1: A significant fraction of ICD-O-3 codes are missing from the OMOP ICDO3 vocabulary, but they represent only a small fraction of tumors in each registry.

Result 2: For certain cancers, this fraction is significant, potentially affecting studies on those cancers.

Result 3: Most missing codes occur in only one of the four registries. Therefore, to create a list of missing codes that is as complete as possible, we need missing codes from as many sources as possible.

Methods

We compared the ICD-O-3 codes for invasive cancers in the OMOP ICDO3 vocabulary with codes used in four population-based cancer registries from Geneva (Geneva Cancer Registry (GCR)), Luxembourg (Registre National du Cancer (RNC)), The Netherlands (Netherlands Cancer Registry (NCR)) and Norway (Cancer Registry of Norway (CRN)).

Limitation: We believe that with enough data sources we can capture most of the missing ICD-O-3 codes and make the OMOP ICDO3 vocabulary sufficiently complete. How many data sources are needed must be determined.