

# Mapping cancer registry data to OMOP CDM: Using alternative source fields reduced unmapped cases by 95%

*Improving completeness in mapping population-based cancer registry data to OMOP CDM by using alternative source fields*

**Background:** OMOP CDM mappings do not exist for all source codes used in cancer registries. Ignoring unmapped cases reduce completeness and may bias surveillance indicators based on these mappings.

The work was done within the European Health Data and Evidence Network (EHDEN) project, and was based on the Vaud Cancer Registry's database, 1974–2018 (Switzerland).

## Results

- The OHDSI data quality tools indicated 100% correctness. Comparison against a manual extraction showed over 99% correspondence.
- Additional use of ICD-10 reduced unmapped cases by 95% compared to ICD-O alone (Figure 1).
- General OMOP codes did not increase unmapped TNM stage codes (breast cancer cases, Figure 2).
- The ICD-8 death cause codes were challenging and only 73% of them were mapped (vs 96% ICD-9 and ICD-10).

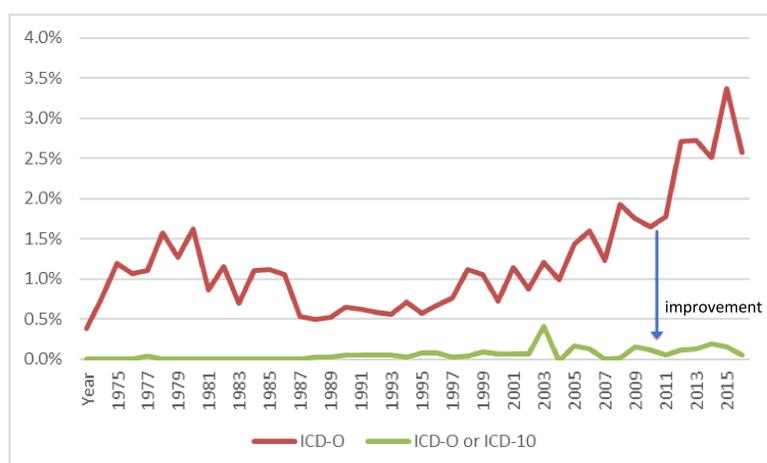


Figure 1. Annual differences (in %) between ICD-O and ICD-O or ICD-10 methods compared to the manual extraction of the source data

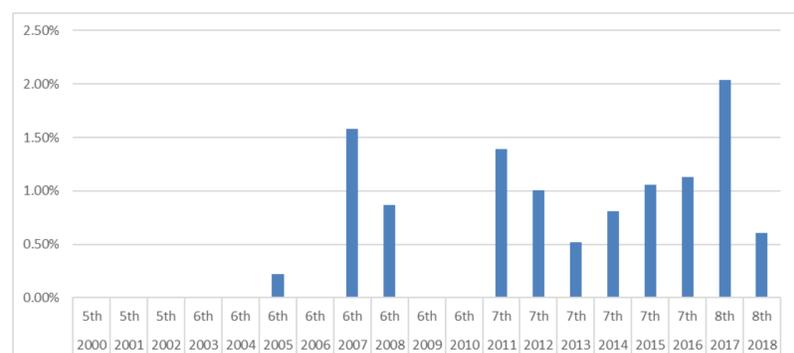


Figure 2. Percentage of unmapped TNM stage values for breast cancer cases diagnosed between 2000 and 2018 with used TNM versions

## Methods

We improved cancer registry mapping to CDM focusing on completeness by:

- 1) Using ICD-10 codes when the ICD-O code was not mapped.
- 2) Using general OMOP TNM staging codes for unmapped TNM versions.
- 3) Mapping old ICD-8 death cause codes to ICD-9 codes first.

Our validation process included two steps:

- 1) Using OHDSI data quality tools, and
- 2) Comparing CDM data to a manually extracted source data.

Example: Squamous cell carcinoma, NOS, of skin of upper limb and shoulder

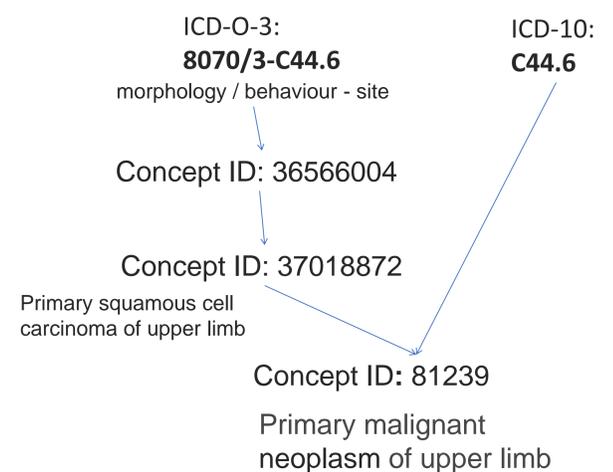


Figure 3: ICD-O vs. ICD-10 mapping

## Discussion and Conclusions

- Common cancer types and recent versions of coding systems are straightforward to map to OMOP CDM, but rare types or older versions can need more work.
- A challenge is that ICD-O or ICD-10 codes can map to different OMOP concepts. This may require applying concept hierarchies in OMOP queries (Figure 3).



www.ehden.eu



Acknowledgements: EHDEN cancer registry teams Geneva, Luxemburg, Netherlands, and Norway