

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

Peter Prinsen¹, Maaïke van Swieten¹, Chiara Attanasio¹, Espen Enerly², Siri Larønningen³, Elisabetta Rapiti⁴, David Marcic^{5,6}, Evelyne Fournier⁴, Pierre Künzli⁷, Michael Schnell^{5,6}, Sophie Couffignal^{5,8}, Claudine Backes^{5,8}

¹ Netherlands Comprehensive Cancer Organisation (KNKI), Utrecht, The Netherlands; ² Department of Research, Cancer Registry of Norway, Oslo, Norway; ³ Department of Registration, Cancer Registry of Norway, Oslo, Norway; ⁴ Geneva Cancer Registry, Global Health Institute, University of Geneva, Switzerland; ⁵ National Cancer Registry (Registre National du Cancer (RNC)), Luxembourg Institute of Health (LIH), Strassen, Luxembourg; ⁶ Data Integration and Analysis Unit, Luxembourg Institute of Health (LIH), Strassen, Luxembourg; ⁷ Scientific Computing Support (SciCos), University of Geneva, Switzerland; ⁸ Department of Precision Health, Luxembourg Institute of Health (LIH), Strassen, Luxembourg

INTRODUCTION

To best map population-based cancer registry (PBCR) data to the **Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM)**, PBCRs 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)) joined forces while working under the umbrella of the **European Health Data & Evidence Network (EHDEN)**¹.

This study investigated how many ICD-O-3 codes in PBCRs data could not be mapped to the OMOP ICD-O-3 vocabulary (OMOP-ICD-O-3).

METHODS

ICD-O-3 codes for invasive cancers in the four included PBCRs were compared and put in parallel with the OMOP-ICD-O-3 codes and a list of all possible ICD-O-3 codes, i.e. codes with a valid topography and a valid morphology. We used the **IARC/IACR Cancer Registry Tool v3.12** to check for unlikely ICD-O-3 codes. These are combinations of morphology, behavior and topography where the topography and morphology are both valid (so the code is possible), but the combination is unlikely to occur.

RESULTS

Of all possible ICD-O-3 codes, **82% were absent in OMOP-ICD-O-3**. For all likely codes this proportion was 38%. **2-14% of the used ICD-O-3 codes could not be mapped** to OMOP-ICD-O-3, corresponding to approximately 1% of cancer diagnoses. For specific subgroups this percentage is significantly higher. In the NCR, for example, 15% of leiomyosarcoma diagnoses and 20% of angiosarcoma diagnoses cannot be mapped. PBCRs use unlikely codes as well: in the NCR 14% of codes are unlikely, accounting for 0.7% of diagnoses.

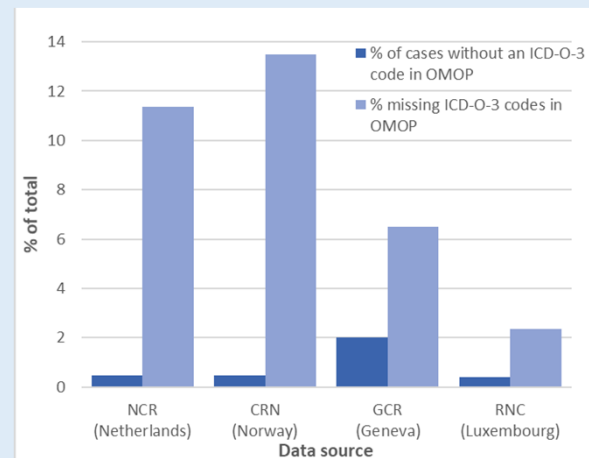
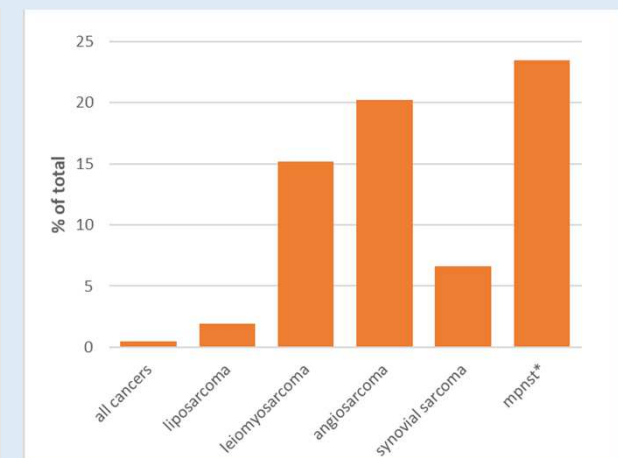


Figure 1. Percentage of cases without or with missing ICD-O-3 codes in OMOP by PBCR



* = malignant peripheral nerve sheath tumor

Figure 2. Percentage of cancer type cases in the NCR (Netherlands) without an ICD-O-3 code in OMOP

CONCLUSION

Less common cancer types are underrepresented in OMOP-ICD-O-3 codes. Although the number of diagnoses that cannot be mapped is small, this will have a disproportionately large impact on studies on rare cancers. Thus, there is a clear need to expand OMOP-ICD-O-3 with the recommended set of ICD-O-3 codes to allow for studies on specific patient subpopulations.