Power in numbers: overcoming the scarcity of rare cancer data by harmonizing European sarcoma registries

Title: Standardizing European sarcoma registry data to the OMOP Common Data Model

Background: Research in rare cancers is hampered by low patient numbers, dispersed clinical data and tumor samples, and a limited number of experts per rare cancer diagnosis. The ERN EURACAN (European Reference Network for Rare Adult Solid Cancers) was established to bring together data and knowledge of European Healthcare professionals. The Blueberry project, funded by the Dutch Cancer Society (KWF), is part of EURACAN and aims to develop a blueprint for a harmonized sarcoma registry using the OMOP Common Data Model.

Result 1: Current overview of data within the Blueberry network provided by ARES.

- 4 Data sources
- ~ 80k people
- 226 Data Quality Issues*
- 4 Data Source Releases

* Failed checks are summed over all data sources. Quality checks need to be optimized to remove checks for data that do not exist within the network (e.g. tables for claims data)

Result 2: Distribution of the number of patients of 14 histological sarcoma subgroups across four data sources.

Methods

1. 7 European Data Partners
2. Data conversion to OMOP-CDM
3. Quality checks & data analysis

Limitation: The data conversion is not yet complete due to implementation issues resulting from missing ICDO-3 codes in the OMOP vocabulary, invalid ICDO-3 codes in the source data, and differences in how data was originally coded. Clinical input will be needed to finalize the data mappings across the network.