Identifying and harmonising data on presence and location of metastasis across three diverse European hospitals to improve care quality

Defining international approaches for the detection of emergent metastasis and the classification of location of metastasis from hospital EHR

**Background:** A multi-centre, European DigiONE study, investigating treatment and outcomes in metastatic non-small cell lung cancer (mNSCLC), required the identification of metastatic locations at diagnosis and at relapse to define clinical groups of importance for statistical analysis. We leveraged OMOP databases across three centres to identify locations of metastases and devise a common classification system within the OMOP framework.

### Methods & Results

#### 1. Detection of emergent metastases

- **i. Oslo University Hospital (OUH):**
  - Only non-curated source data included in the local OMOP database.
  - Metastasis events captured in three ways: (1) a hospital cancer episode in the EHR system with a metastasis diagnosis in the form of an ICD-10 code, (2) a recording of TNM staging with M1, or (3) a pathology report specifying malignant histology originating from a metastatic location.

- **ii. Leeds Teaching Hospitals NHS Trust (LTHT):**
  - Staging data and location of metastases for each site-specific cancer diagnosis and recurrence event available in in-house built EHR system.
  - NSCLC staging data was manually curated from source data by clinical review of relevant imaging reports, pathology reports and clinical letters.

- **iii. Maastricht University Medical Centre+ (MUMC+):**
  - Implemented Natural Language Processing (NLP) for data structuring with manual validation by an oncology nurse.
  - NLP performed on clinical notes using CTCue.
  - NLP output included presence, date and location of metastasis.

#### 2. Classification of metastasis location data

- Metastatic locations categorised by medical oncologists into six groups: brain, liver, adrenal glands, bone, lung, and “other” anatomical structures.
- Variations identified in coding systems across centres were:
  - (1) brain and leptomeningeal metastases grouped at OUH but reported separately at LTHT.
  - (2) bone and bone marrow metastases grouped at OUH but reported separately at LTHT, and
  - (3) adrenal metastases specified at OUH but classed as “other” at LTHT.
  - Common coding was agreed to harmonise LTHT and OUH data.
  - MUMC+ altered NLP rules to capture metastasis location with the highest granularity.