The OMOPCAN Study: Preliminary Insights on Cancer Patient Characterization Across 34 Cancer Types from SIDIAP database with planned OHDSI Network Collaboration

Irene López-Sánchez[1, 2], Anna Palomar-Cros[1], Agustina Giuliodori[1], Laura Granés[1], Elena Roel[1], Vlad Korsik[3], Anton Barchuk[4], Asieh Golozar[5] Talita Duarte-Salles[1,4]

[1]IDIAP Jordi Gol, Barcelona, Spain
[2]Programa de Doctorat en Metodologia de la Recerca Biomèdica i Salut Pública, Universitat Autònoma de Barcelona, Barcelona, Spain
[3]EPAM Odysseus, US
[4]Erasmus MC University Medical Center, Rotterdam, The Netherlands
[5]Nemesis Health, US

Background

Cancer poses a significant global health threat. Real-world evidence from routine clinical care provides timely insights into cancer outcomes, addressing gaps in clinical trial data. Electronic Health Records further enable the characterization of pre-diagnostic clinical history, which is often lacking in cancer registries and has the potential for identifying prodromes[1]. These data can complement traditional sources for cancer epidemiology studies and contribute to the understanding of the entire patient journey.

The OMOPCAN study is a multinational cohort study that aims to characterize patients and evaluate time trends of incidence, prevalence and survival across 34 cancer types, using data from different real world data sources mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) from the OHDSI community. While this study is still in the recruitment phase, preliminary characterisation results from the SIDIAP database are already available.

Methods

For this poster, we present preliminary data from SIDIAP database, a primary care electronic health records database covering close to 80% (> 8 million people since 2006) of the population of Catalonia, Spain, representative of the Catalan population.

In addition, twenty-two databases from the OHDSI Network, including primary care, hospital care, health claims and cancer registries mapped to the OMOP-CDM, from seventeen countries (UK, Spain, Estonia, Netherlands, Norway, Switzerland, Belgium, Romania, Italy, Finland, France, Croatia, Hungary, Canada, EEUU, Corea and Taiwan) have agreed to participate in this study.

Individual-level data was used from individuals with at least one year of prior history and a cancer diagnosis of one of 34 cancer types. Cancer definitions were based on the classification used in the GLOBOCAN 2020 database[2] (ICD-10 codes C00- C97) and adapted for the databases using SNOMED and ICD-O-3 diagnostic codes. OMOP standardized tools, including ATLAS, CodelistGenerator and CohortDiagnostics R packages, were used to develop and evaluate the phenotyping. The cohort study period spanned from January 1, 2006, to June 2023. Patient characteristics, including demographic and prevalence of clinical factors, were described.

Results

In SIDIAP, breast, colorectal, prostate, lung, and bladder cancers were the most common cancers, each affecting over 40,000 individuals. Vaginal and nasopharyngeal cancers were the least common, with less than 1,000 individuals. The prevalence of non-sex-specific cancers was higher in males, especially bladder and respiratory cancers (oesophagus, hypopharynx, larynx, lung, nasopharynx, and oropharynx) for which over 70% of cases occurred in males. Additionally, individuals diagnosed with bladder and respiratory cancers exhibited a higher percentage of former smokers (over 40%) compared to other cancers. Individuals under 60 years were more commonly diagnosed with Hodgkin lymphoma, thyroid, cervix, nasopharynx, Kaposi sarcoma, and breast cancers, while individuals over 70 years were more frequently diagnosed with colorectal, stomach, vulva, pancreatic, non-melanoma skin cancer, mesothelioma, multiple myeloma, prostate, and gallbladder cancers. Obesity (clinical diagnosis or BMI over 30) was most prevalent in individuals with corpus uteri, gallbladder, and vulva cancers (>40%). Hypertension was the most common pre-diagnosis condition overall, while COPD was most prevalent in lung, larynx, hypopharynx, bladder, esophagus and oropharynx cancers (>40%), and type 2 diabetes among pancreatic, gallbladder and liver cancer patients (>20%).

Conclusions

This study offers a comprehensive demographic and clinical description of patients with 34 different cancer types in Catalonia, serving as a valuable resource for future research and healthcare planning. The use of the OMOP-CDM framework ensures standardized data integration, facilitating robust cross-national comparisons and enhancing our understanding of cancer epidemiology on a global scale.

References

1. Lancet T. Cancer registries: the bedrock of global cancer care. The Lancet. 2025 Feb 1;405(10476):353.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.