Establishing a large COVID-19 cohort through mapping the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model

Edward Burn[1,2]*, Sergio Fernández-Bertolín[1]*, Erica A Voss[3,4,6], Clair Blacketer[3,4,6], Maria Aragón[1], Martina Recalde[1,5], Elena Roel[1], Carlen Reyes[1], Sebastiaan van Sandijk[6,7], Lars Halvorsen[6,8], Peter R Rijnbeek[4,5,6], Talita Duarte-Salles[1,6]

1) Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain, 2) Centre for Statistics in Medicine, University of Oxford, 3) Janssen Pharmaceutical Research and Development LLC, Titusville, NJ, USA, 4) Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands, 5) Universitat Autònoma de Barcelona, Bellaterra, Spain, 6) OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY, 7) Odysseus Data Services s.r.o., Prague, Czech Republic, 8) edenceHealth NV, Kontich, Belgium

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

Background

The Information System for Research in Primary Care (SIDIAP) was created in 2010 by the Institut Català de la Salut (ICS, Catalan Health Institute) and the IDIAPJGol Research Institute. It was created with the aim to promote research in primary care through the use of electronic health records (EHR) and other complementary databases. SIDIAP includes information recorded since 2006 by healthcare professionals during routine visits at 287 primary health care centres in Catalonia, Spain. SIDIAP has pseudo-anonymized records for more than seven million people and is representative of the Catalan population in terms of age, sex, and geographic distribution.

After the start of the 2020 outbreak of COVID-19, new data sources were made available in SIDIAP that included relevant information regarding the management of the pandemic in Catalonia. SIDIAP received a grant from the COVID-19 Rapid Collaboration Call hosted by EHDEN inviting institutions with COVID-19 data to apply for financial and technical support to convert their data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). This standardization of our data would allow our organization to research the virus more easily as well as take advantage of federated research on COVID-19. The following describes the establishment of SIDIAP's COVID-19 cohort over the course of the pandemic.

Methods

Primary care data collected in SIDIAP between 1st January 2006 and 31st December 2020 was linked, at a patient-level, to hospitalisation, COVID-19 testing, and mortality data. Hospitalisations, however, were only available up to July 2020 and will be updated when data for the full year becomes available. This data was mapped to the OMOP CDM. Using this mapped data, a cohort of the general population was followed up from 1st March 2020, with COVID-19 outcomes (outpatient clinical diagnoses of COVID-19, outpatient positive tests for SARS-CoV-2, hospitalisations with COVID-19, and COVID-19 deaths) observed over follow-up until the 31st December 2020.

Results

A total of 5,846,221 individuals were included in the general population cohort. Over observed follow up, 348,764 had an outpatient diagnosis with COVID-19, 208,697 had an outpatient positive test for COVID-19, 18,629 had a hospitalisation with COVID-19, and 8,209 had a COVID-19 death. Entry into these cohorts over time is shown in Figure 1, with cohort intersection shown in Figure 2.

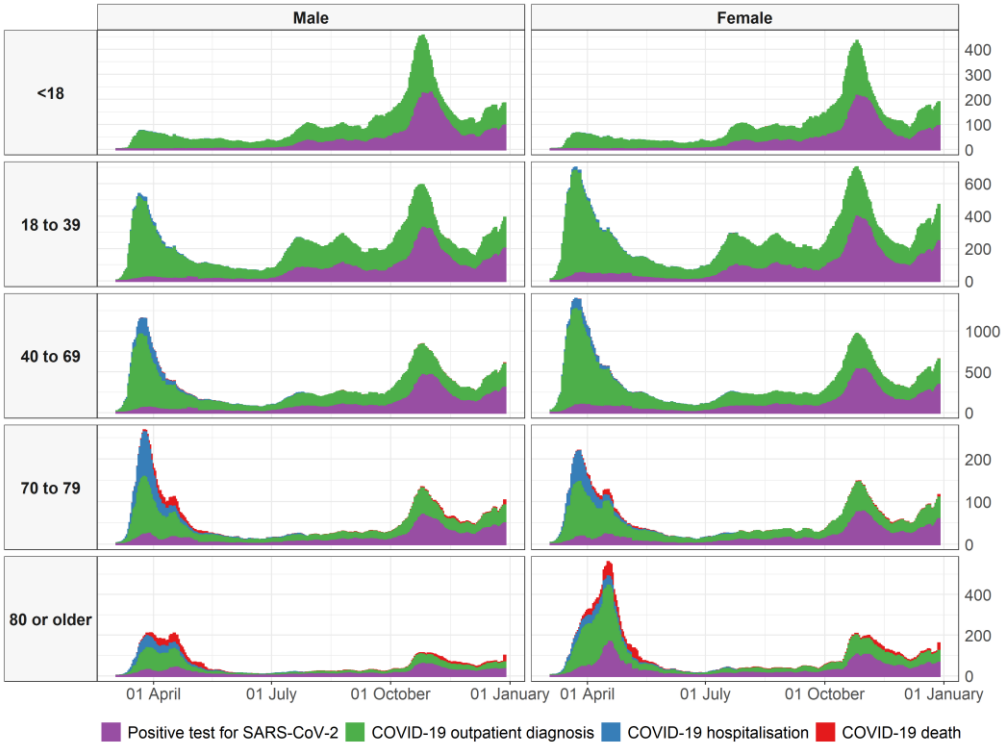

Figure 1. Histogram summarising outcome cohort entry over time. Cohort entry is shown based on a seven-day rolling average. Note, hospitalisations are only available up to July in the current release.

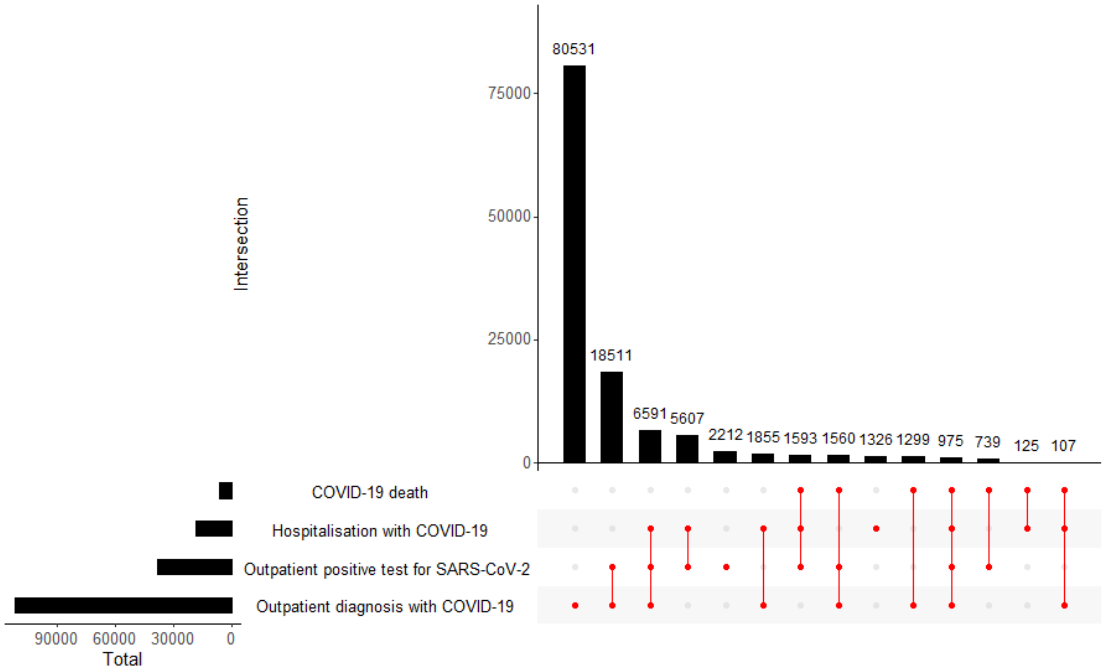

Figure 2. Intersection plot showing cohort overlap. Note, hospitalisations are only available up to July in the current release.

Conclusion

We have established a COVID-19 dataset in the OMOP CDM standard that captures COVID-19 diagnoses, test results, hospitalisations, and deaths among a general population cohort of 5.8 million individuals from Catalonia. This provides the basis for wide-ranging research relating to the COVID-19 pandemic.