Establishing a comprehensive dataset for COVID-19 in Catalonia, Spain, mapped to the OMOP CDM

BACKGROUND
- The Information System for Research in Primary Care (SIDiAP) includes information recorded since 2006 by healthcare professionals during routine visits at 287 primary health care centres in Catalonia, Spain.
- Linkage between SIDiAP and hospital and mortality data is possible. It is now also possible to link SIDiAP data to COVID-19 testing data.
- SIDiAP received a grant from the EHDEN COVID-19 Rapid Collaboration Call to convert data to the OMOP CDM.

METHODS
- Primary care data collected in SIDiAP between 1st January 2006 and 31st December 2020 was linked, at a patient-level, to hospitalisation (up to July 2020), COVID-19 testing, and mortality data.
- This data was mapped to the OMOP CDM and we present here COVID-19 related cohorts captured: 1) outpatient clinical diagnoses of COVID-19, 2) outpatient positive tests for SARS-CoV-2, 3) hospitalisation with COVID-19, and 4) COVID-19 deaths

RESULTS
- 5,846,221 individuals were included in the general population cohort.
- By the 31st December 2020, 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.
- Cohort entry over time, by age and sex, is shown in the figure.

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.
- Studies already informed by this dataset include:
  - The natural history of symptomatic COVID-19 during the first wave in Catalonia. Nature communications. 2020
  - Body mass index and risk of COVID-19 diagnosis, hospitalisation, and death: a cohort study of 2,524,926 Catalans. The Journal of Clinical Endocrinology & Metabolism. 2021
  - Unraveling COVID-19: a large-scale characterization of 4.5 million COVID-19 cases using CHARYBDIS. Research Square. 2021