# Transforming the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model and Its Use for COVID-19 Research

Berta Raventós
IDIAP Jordi Gol
Barcelona
5-12-2023

# 1 Introduction: The SIDIAP database

- **Primary care EHR**
- 6M people active (June 2021)
- Representative of the Catalan population (age, sex, geographic distribution)
- Data on:
  - Demographics
  - Diseases diagnosis (ICD-10)
  - Prescriptions & dispensations
  - All -cause mortality
  - Laboratory tests
  - Measurements
  - Sick leaves
  - SES indicators

- **Linkage to hospital data and other data sources is possible.**

# 2 Objectives

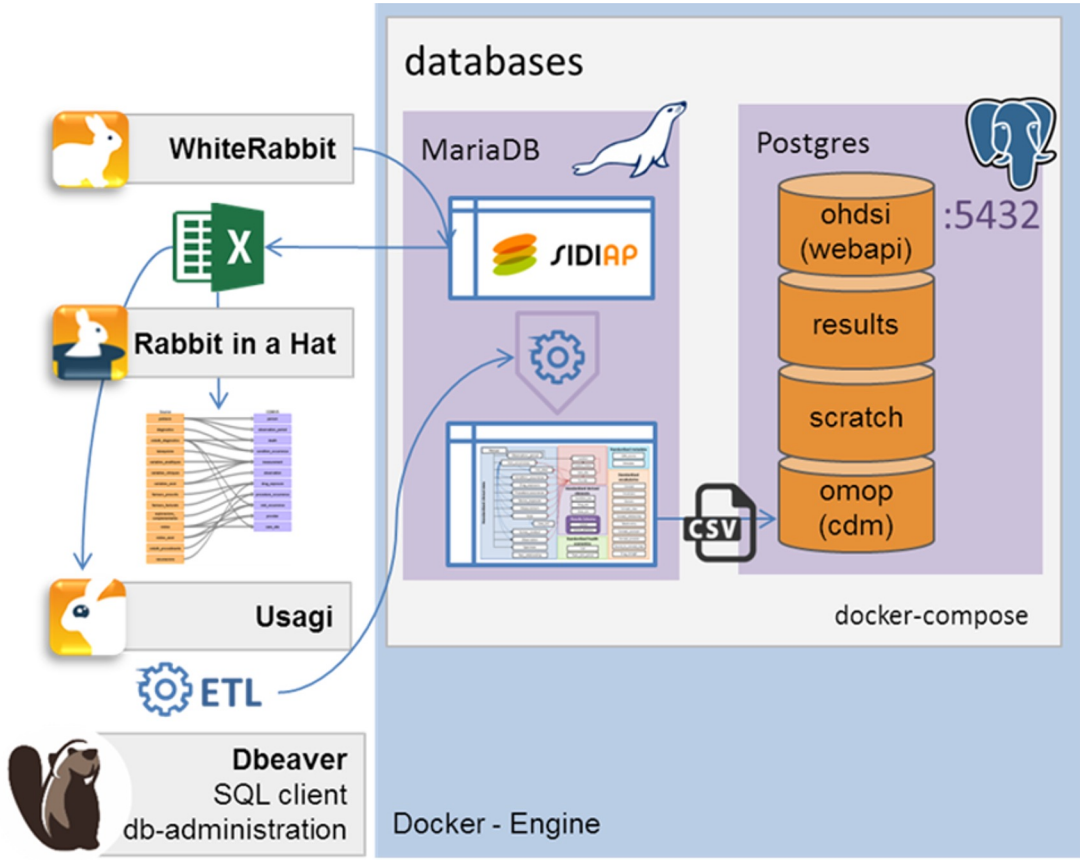
1. **Convert the SIDIAP data to the OMOP CDM** to facilitate distributed network research.

1. **Summarise the occurrence of COVID-19-related outcomes** and describe the characteristics of those affected and vaccinated against the disease.

# 3 Methods: Mapping

## Extract, Transform and Load (ETL):

1) **Design the ETL:** OHDSI WhiteRabbit tool
2) **Create the code mappings**
3) **Implemented the ETL**
4) **Quality control:** OHDSI Data Quality Dashboard




OMOP CDM v5.3.1
PostgreSQL

# 3 Methods: COVID-19

1) **Study population:** Individuals registered as of 1 March 2020

2) **Follow-up:** Until 30 June 2022

3) **Descriptive analysis:** Summary of the characteristics of the population

   - Demographics
   - Comorbidities (all prior history)
   - Symptoms (± 2 days)

# 3 Methods: COVID-19

## COVID-19-related outcomes (*):

1) **Outpatient COVID-19:**
   - Clinical diagnostic codes
   - PCR
   - PCR + Antigen tests
2) **Hospitalised with COVID-19**
3) **ICU admission with COVID-19**

Test/dx -21 to 3 days relative to the admission date

4) **Died with COVID-19:** Test/dx 28 days prior to the date of death
5) **Vaccinated against COVID-19**: First-dose (BNT162b2, ChAdOx1, mRNA-1273, and Ad26.COV2.S)

(*) Not mutually exclusive

# 4 Results: Mapping

## Source terms and registries mapped:

| Domain | Source Terms | Mapped Terms (%) | Source Registries | Mapped Registries (%) |
|---|---|---|---|---|
| Condition | 55,787 | 49,631 (89.0) | 252,201,881 | 244,070,592 (96.8) |
| Drug* | – | – | 1,623,418,192 | 1,537,021,869 (94.7) |
| Measurement | 141 | 136 (96.5) | 1,575,796,906 | 1,575,431,674 (100) |
| Observation** | 1883 | 1,879 (99.8) | 166,948,926 | 133,887,554 (80.2) |
| Procedure | 36,060 | 35,472 (98.4) | 45,632,837 | 45,402,273 (99.5) |
| Visit | 347 | 347 (100) | 802,837,844 | 802,837,844 (100) |

## Data quality:

- 3,484 data quality checks run against the database: 98.7% passed

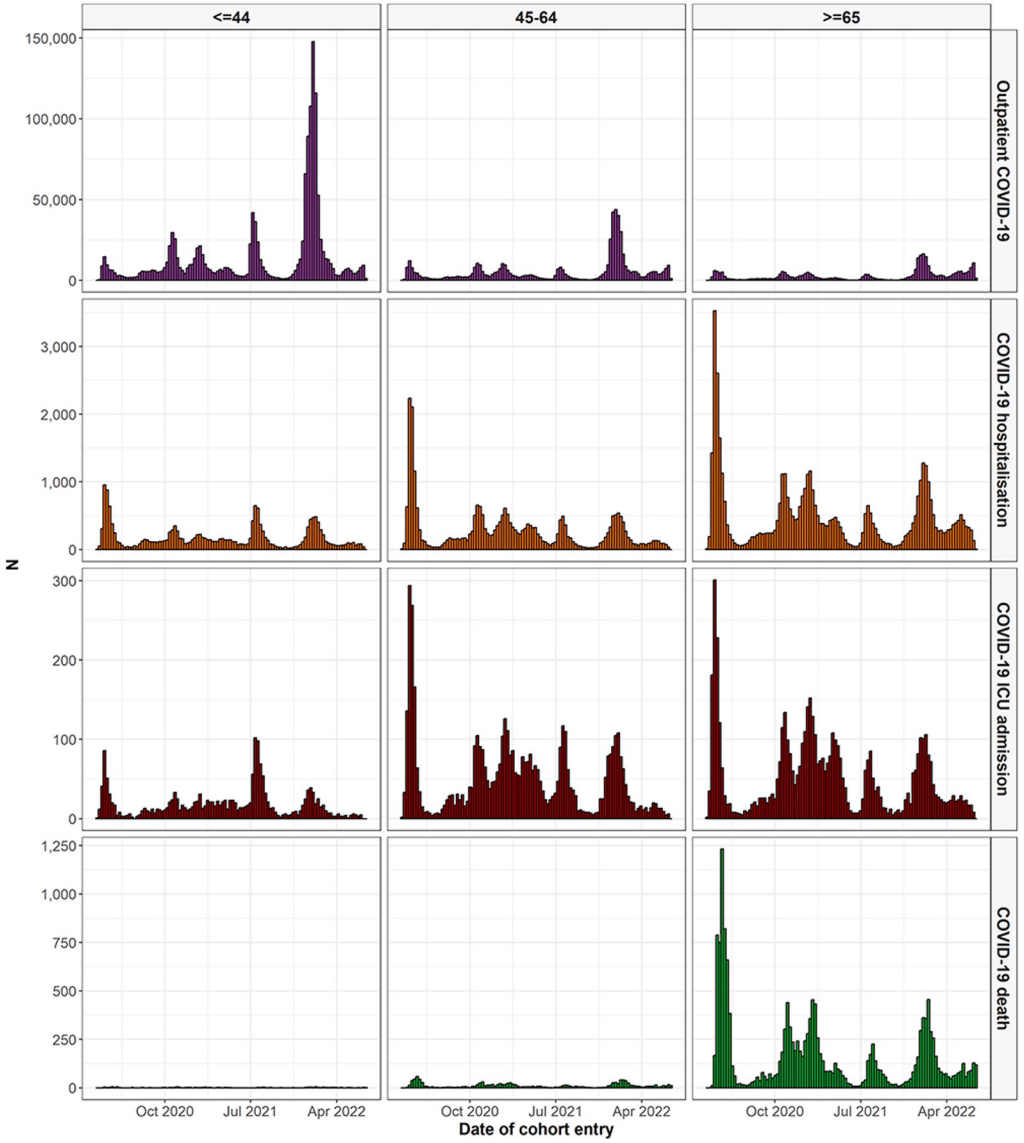
# 4 Results: COVID-19

| COVID-19-related cohorts | N | Males, % | Age, median [IQR] |
|---|---|---|---|
| General population (*) | 5,923,762 | 49.3 | 43 [25-59] |
| Vaccinated (1st dose) | 4,584,515 | 48.7 | 46 [29-61] |
| Outpatient dx or positive test | 604,472 | 46.3 | 41 [25-55] |
| Hospitalized | 58,991 | 54.5 | 65 [51-78] |
| Admitted to UCI | 5,642 | 67.2 | 63 [53-71] |
| Death | 11,233 | 50.9 | 85 [78-90] |

# 4 Results: COVID-19

| COVID-19-related cohorts | N | Males, % | Age, median [IQR] |
|---|---|---|---|
| General population (*) | 5,923,762 | 49.3 | 43 [25-59] |
| Vaccinated (1st dose) | 4,584,515 | 48.7 | 46 [29-61] |
| Outpatient dx or positive test | 604,472 | 46.3 | 41 [25-55] |
| Hospitalized | 58,991 | 54.5 | 65 [51-78] |
| Admitted to UCI | 5,642 | 67.2 | 63 [53-71] |
| Death | 11,233 | 50.9 | 85 [78-90] |

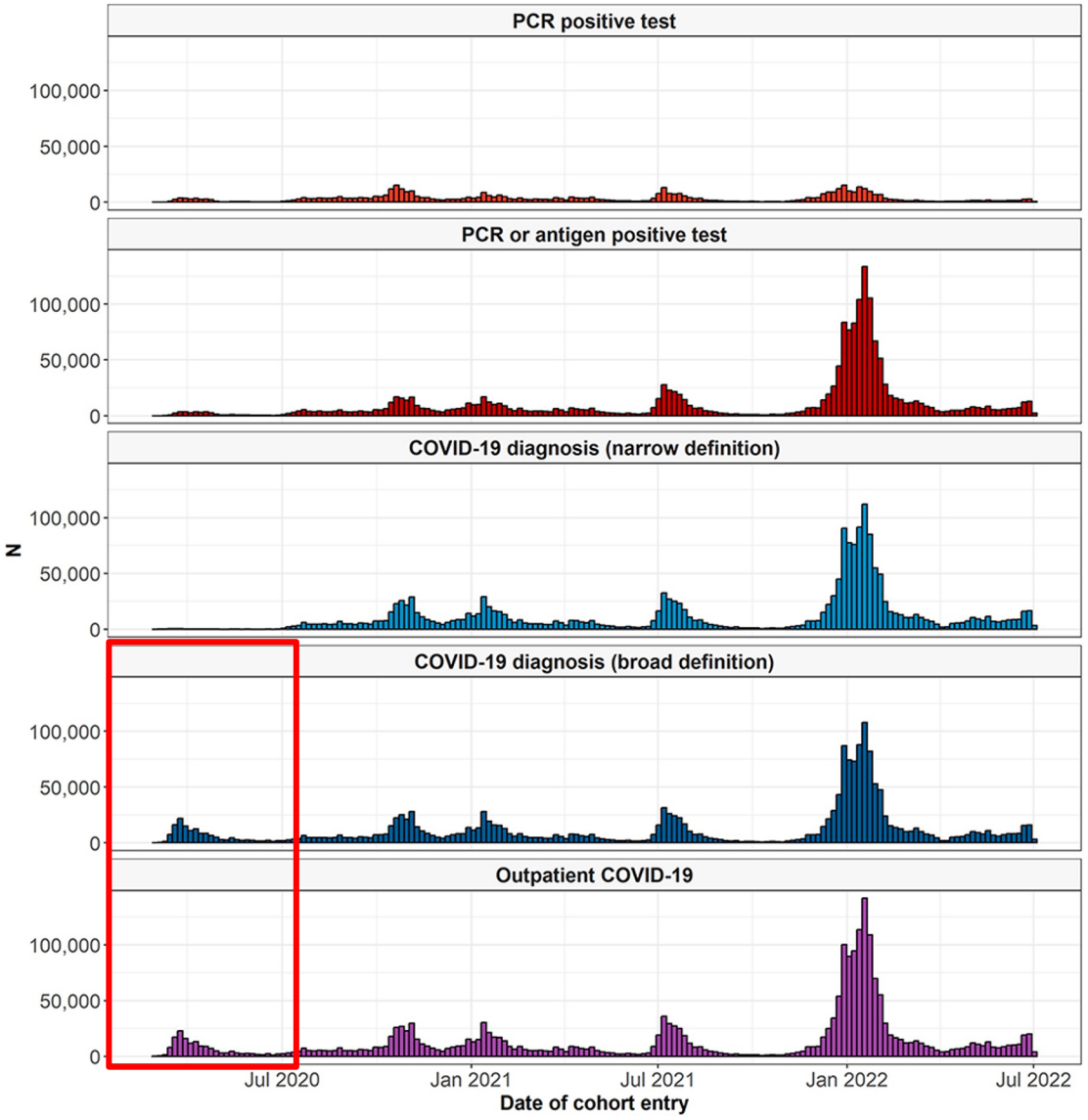
# 4 Results: COVID-19




## COVID-19 waves:

- Distinct pandemic waves.
- Increased hospitalisations/ICU admissions during the first-wave (older age groups).
- Increased cases driven by the Omicron variant (younger age groups)

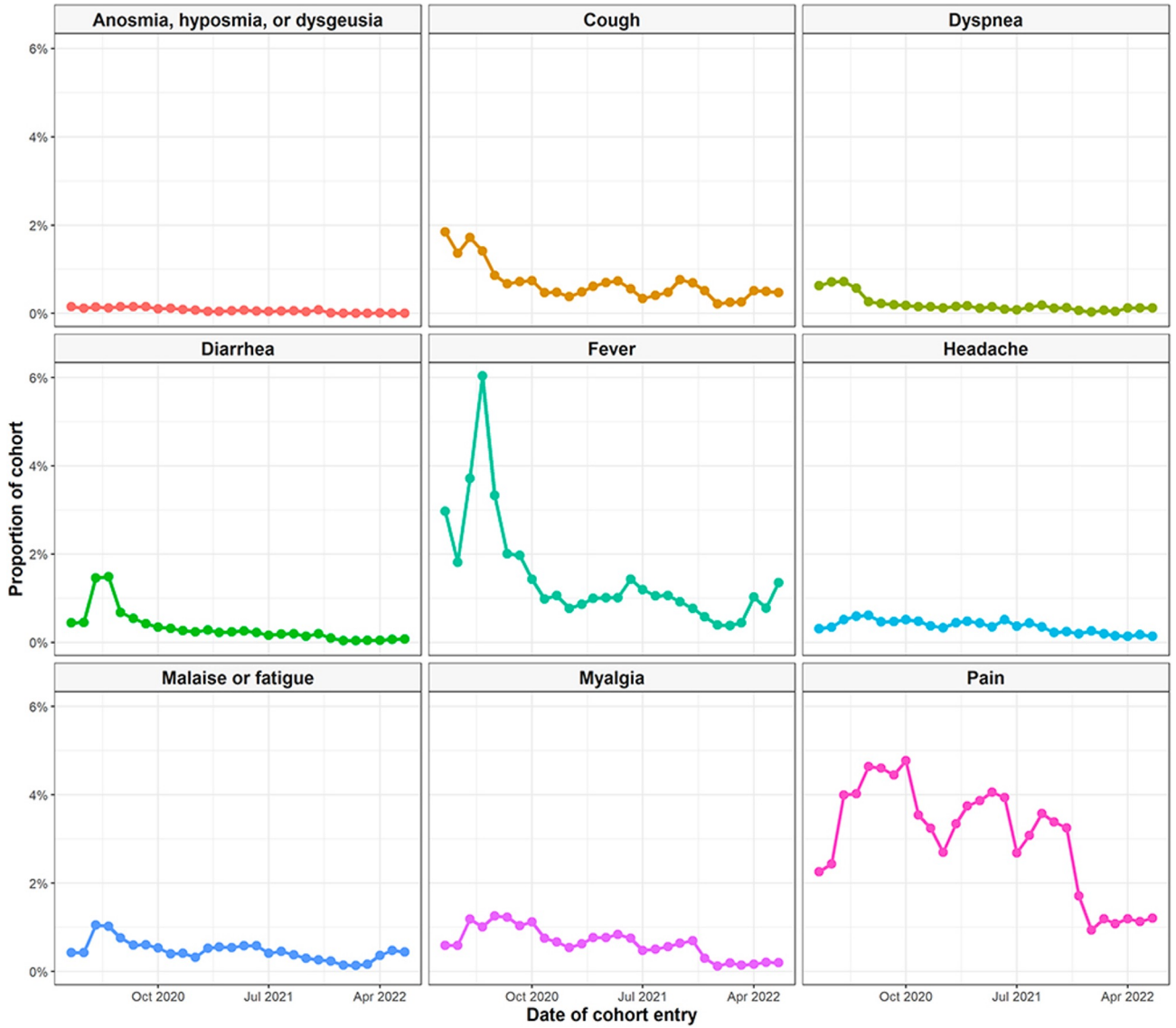
# 4 Results: COVID-19




**First wave of the pandemic:**

- Clinical diagnosis versus tests
- Tests were widely available from Sep. 2020 onwards

# 4 Results: COVID-19




## Symptoms:

- Prevalence of less than 6%
- Substantial changes over time

# 5 Strengths and limitations

- **Strengths:**
  - Breadth of data captured
  - Complete COVID-19 tests and vaccinations performed in the public health system

- **Limitations:**
  - Underreporting of cases
  - Underreporting of symptoms
  - Lack of hospital prescriptions & lab results.
  - Lack of cause of death

Since March 2022, COVID-19 testing is restricted to specific subgroups.

# 5 Conclusions

1. We **successfully harmonised SIDIAP to the OMOP CDM**, and we illustrated its potential to perform distributed network research in COVID-19 and beyond.

   → > 20 published studies

1. We have provided **insights regarding important considerations for future research in our setting**, including the impact of different outpatient COVID-19 definitions and significant testing-related information.

# Acknowledgements


- **Sergio Fernández-Bertolín**
- **María Aragón**
- Erica A Voss
- Clair Blacketer
- **Leonardo Méndez-Boo**
- Martina Recalde
- Elena Roel
- Andrea Pistillo
- Carlen Reyes
- Sebastiaan van Sandijk
- Lars Halvorsen
- Peter R Rijnbeek
- **Edward Burn**
- **Talita Duarte-Salles**


EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Generalitat de Catalunya