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
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.
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.
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
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)
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**

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
Results: Mapping

Source terms and registries mapped:

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

Data quality:
- 3,484 data quality checks run against the database: 98.7% passed
## Results: COVID-19

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

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

First wave of the pandemic:
- Clinical diagnosis versus tests
- Tests were widely available from Sep. 2020 onwards
Results: COVID-19

Symptoms:
- Prevalence of less than 6%
- Substantial changes over time
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.
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.

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