OMOP Mapping of Real-World Data From Brazil & Pakistan Towards Management of COVID-19 In the Global South

Authors
Elzo Pereira Pinto Junior¹, Priscilla Normando¹, Renzo Flores-Ortiz¹, Muhammad Usman Afzal², Muhammad Asaad Jamil², Sergio Fernandez Bertolin³, Vinícius de Araújo Oliveira¹, Valentina Martufi¹, Edward Burn⁴, Maria Yury Ichihara¹, Maurício L. Barreto¹, Talita Duarte Salles³, Daniel Prieto-Alhambra⁴, Haroon Hafeez², Sara Khalid⁴

¹ The Center for Data and Knowledge Integration for Health (CIDACS – FIOCRUZ), Brazil
² Shaukat Khanum Memorial Cancer Hospital and Research (SKMHR&C), Pakistan
³ The Information System for Research in Primary Care (SIDIAP), Spain
⁴ Centre for Statistics in Medicine, University of Oxford, UK.
Background

The COVID-19 pandemic highlighted need for rapid, reliable, representative evidence generation.
The OHDSI COVID-19 Data Network

- Rapid
- Reliable
- Representative?

<table>
<thead>
<tr>
<th>USA (8)</th>
<th>EUROPE (7)</th>
<th>ASIA-PACIFIC (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premier (National – Hospital Billing)</td>
<td>CPRD (UK – Electronic Health Records)</td>
<td>HIRA (South Korea – Administrative Claims)</td>
</tr>
<tr>
<td>HealthVerity (Claims linked to diagnostic testing)</td>
<td>SIDIAP (Spain – Electronic Health Records)</td>
<td>DCMC (South Korea – Electronic Health Records)</td>
</tr>
<tr>
<td>Optum EHR (National – Electronic Health Records)</td>
<td>SIDIAP-H (Spain – EHR hospital linkage)</td>
<td>Nanfang Hospital (China – Electronic Medical Records)</td>
</tr>
<tr>
<td>IQVIA Open Claims (National – Administrative Claims)</td>
<td>HM Hospitales (Spain – Hospital Billing)</td>
<td></td>
</tr>
<tr>
<td>Department of Veterans Affairs (National – Electronic Health Records)</td>
<td>ICPI (Netherlands – Electronic Health Records)</td>
<td></td>
</tr>
<tr>
<td>Stanford University (CA – Electronic Health Records)</td>
<td>LPD France (France – Electronic Health Records)</td>
<td></td>
</tr>
<tr>
<td>Tufts University (MA – Electronic Health Records)</td>
<td>Germany DA (Germany – Electronic Health Records)</td>
<td></td>
</tr>
<tr>
<td>Columbia University (NY – Electronic Health Records)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Together, OHDSI has studied (to date):
- >4.5m patients tested for SAR-COV-2
- >1.2m patients diagnosed or tested positive for COVID-19
- >250k hospitalized for COVID-19
The OHDSI COVID-19 Data Network
Data Partners – Pakistan

- **Data source:**
  - De-identified electronic health records
- **Period:**
  - 1994 – 2022 (ongoing)
- **Unique records:**
  - 8.3 million individuals
- **Regional COVID-19 hub**

**SKMHR&C Hospital Network in South Asia**
- 3 Hospitals
- 3 Diagnostic Centres
- 3 Walk-in Clinics
- 180 Labs in 70 cities

**Country Facts**
- Population: 220 million
- 5 Provinces

**Clinical activity snapshot (2021)**
- 12,018 New Registrations
- 210,667 OPD Visits
- 14,840 Admissions
- 63,725 Chemotherapy Sessions
- 19,128 Surgical Procedures
- 77,809 Radiotherapy Sessions
- 198,393 Imaging Studies
- 6,277,572 Pathology Tests
Data Partners – Brazil

- **Data source/s:**
  - Brazil Ministry of Health Influenza Surveillance System (SIVEP-Gripe)
- **Period**
  - 2020 – 2022 (ongoing)
- **COVID-19 records:**
  - 2.6 million individuals
  - ~67,000 hospitalisations
  - ~27,000 deaths

CIDACS-FIOCRUZ COVID-19 Data Platform

- Brazil MoH Surveillance System
  - Case notification
  - Hospitalisation
  - Vaccinations
- CIDACS
  - eSUS
  - SRAG
  - VAC

Bahia State Facts:
- Population: 15 million
- 417 municipalities
Harmonisation to OMOP

ETL Process

1. Perform scan on source database using White Rabbit.
2. Perform (case/ column) level mapping from source database to CDM (Table is a FIBA).
3. Export source data to CSV files after extraction of required variables / populations.
4. Import CSV files in a staging database via Oracle / Pentaho Transformations.
5. Concept Mapping (EGAC and Athena4).
6. Apply pipeline preprocessing and transformation on staging database.
7. Load the data into CDM using PostgreSQL, record procedures.

Software used during ETL:
- Source database – Oracle 19c
- Target database – PostgreSQL 12.2
- CDM v5.3.1
- DBeaver v21.3.3
- Pentaho v9.2.0

Mapping in Numbers
- >100K (source) to 108K (CDM) concepts
- >33M measurements
- >2M procedures
- >600K observations
- <1% missing matching concepts
**COVID-19 cohorts – cases over time**

### Distribution of cases over time (Jan/March 2020 – April 2022)

<table>
<thead>
<tr>
<th>Category</th>
<th>Tested population</th>
<th>COVID-19 diagnosis or positive test</th>
<th>Hospitalised with COVID-19</th>
<th>ICU admission with COVID-19</th>
<th>COVID-19 death</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brazil</strong></td>
<td>N=1,312,832</td>
<td>752,699</td>
<td>34,699</td>
<td>17,041</td>
<td>13,877</td>
</tr>
<tr>
<td><strong>Pakistan</strong></td>
<td>N=341,505</td>
<td>88,771</td>
<td>447</td>
<td>33</td>
<td>117</td>
</tr>
</tbody>
</table>

[https://doi.org/10.1093/jamia/ocac180](https://doi.org/10.1093/jamia/ocac180)
COVID-19 cohorts – baseline characteristics

- COVID-19 outcomes were more severe in men, elderly, and those with co-morbidities

Distribution of cases by age and sex in each cohort

JAMIA 2022 https://doi.org/10.1093/jamia/ocac180
COVID-19 cohorts – baseline characteristics

General population tested for COVID-19:
• Age: median (IQR) was 36 (25-75) and 38 (27 - 50) for Pakistan and Brazil
• Sex: 45.5% and 55% were female in Pakistan and Brazil
• Ethnicity/race: 1.2% Pakistan individuals had “Afghan” ethnicity. In Brazil, 52.3% had “Mixed” ethnicity.
Two health databases covering 8.3 million people from Pakistan and 2.6 million people from Bahia, Brazil were analysed.

109,504 (Pakistan) and 921 (Brazil) medical concepts were harmonised to OMOP CDM.

341,505 (4.1%) people in the Pakistan dataset and 1,312,832 (49.2%) people in the Brazilian dataset tested for COVID-19 between 1st Jan 2020 and 30 April 2022.

In agreement with international findings, COVID-19 outcomes were more severe in men, elderly, and those with underlying health conditions.

This proof-of-concept study demonstrates potential for OMOP-harmonised data from under-represented regions for global knowledge mobilisation and clinical translation for timely response to healthcare needs in pandemics and beyond.
What’s next

• COVID-19 – variant and vaccine surveillance study
• Communicable, NCDs
  – Cancer (OHDSI Oncology WG)
• Environment, equity, and artificial intelligence
  – Social deprivation dashboard (OHDSI GIS WG, OHDSI Equity WG)
• Data science ecosystem
  – Capacity building
  – Data re-use projects
  – Data governance
Thank You

Acknowledgements
OMOP Mapping of Real-World Data From Brazil & Pakistan Towards Management of COVID-19 In the Global South

Authors
Elzo Pereira Pinto Junior¹, Priscilla Normando¹, Renzo Flores-Ortiz¹, Muhammad Usman Afzal², Muhammad Asaad Jamil², Sergio Fernandez Bertolin³, Vinícius de Araújo Oliveira¹, Valentina Martufi¹, Edward Burn⁴, Maria Yury Ichihara¹, Maurício L. Barreto¹, Talita Duarte Salles³, Daniel Prieto-Alhambra⁴, Haroon Hafeez², Sara Khalid⁴

¹ The Center for Data and Knowledge Integration for Health (CIDACS – FIOCRUZ), Brazil
² Shaukat Khanum Memorial Cancer Hospital and Research (SKMHR&C), Pakistan
³ The Information System for Research in Primary Care (SIDIAP), Spain
⁴ Centre for Statistics in Medicine, University of Oxford, UK.