Using healthcare big data in pandemic response by characterizing disease natural history and predicting patient outcomes

Patrick Ryan, PhD
Janssen Research and Development
Columbia University
Together, OHDSI has studied:

- >4.5m patients tested for SAR-COV-2
- >1.2m patients diagnosed or tested positive for COVID-19
- >249k patients hospitalized with COVID-19
Common data model to enable standardized analytics

Source 1 raw data
- Electronic health records

Source 2 raw data
- Administrative claims

Source 3 raw data
- Clinical data

Transformation to OMOP common data model

Source 1 CDM

Source 2 CDM

Source 3 CDM

Open-source analysis code

Open evidence
Driving agenda of full transparency

• All artifacts of our analytics pipeline are made available to the public

• In doing so, we are encouraging others to do the same

• Transparency is key to
  • Reproducibility
  • Interpretability
  • Trustworthiness
Complementary evidence to inform the patient journey

**Clinical characterization:**
What happened to them?

**Patient-level prediction:**
What will happen to me?

**Population-level effect estimation:**
What are the causal effects?

**Observation**

**Inference**

**Causal inference**
Findings:

– Patients hospitalized with COVID are systematically different from those hospitalized with flu

– COVID hospitalized patients, when compared to those hospitalized for influenza:
  • Greater proportion are male and slightly younger
  • Fewer comorbidities and lower medication use

• Utilized claims and electronic medical records from 10 databases across 3 different countries

Accepted in Nature-Communications
# Disease Natural History of COVID-19

## Characterization

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Medication use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CUIMC</strong> (n: 1,759)</td>
<td></td>
</tr>
<tr>
<td><strong>HIRA</strong> (n: 7,341)</td>
<td></td>
</tr>
<tr>
<td><strong>HM</strong> (n: 2,078)</td>
<td></td>
</tr>
<tr>
<td><strong>PHD</strong> (n: 5,257)</td>
<td></td>
</tr>
<tr>
<td><strong>SIDIAP</strong> (n: 16,347)</td>
<td></td>
</tr>
<tr>
<td><strong>UC HDC</strong> (n: 769)</td>
<td></td>
</tr>
<tr>
<td><strong>VA</strong> (n: 577)</td>
<td></td>
</tr>
</tbody>
</table>

- **Blood disease**
- **Cardiovascular disease**
- **Congenital disease**
- **Digestive disease**
- **Endocrine or metabolic disease**
- **ENT disease**
- **Eye disease**
- **Genitourinary disease**
- **Iatrogenic condition**
- **Infection**
- **Injury and poisoning**
- **Mental disease**
- **Neoplasm**
- **Nerve disease and pain**
- **Respiratory disease**
- **Skin disease**
- **Soft tissue or bone disease**
- **Uncategorised**

- **Alimentary tract and metabolism**
- **Antifungal agents**
- **Antineoplastic agents**
- **Antiparasitic products, insecticides and repellents**
- **Blood and blood forming organs**
- **Cardiovascular system**
- **Dermatologicals**
- **Genito urinary system and sex hormones**
- **Musculo-skeletal system**
- **Nervous system**
- **Respiratory system**
- **Sensory organs**
- **Systemic hormonal preparations excl. sex hormones and insulins**
- **Uncategorised**
- **Various**
Disease Natural History of COVID-19

• CHARYBDIS – Characterizing Health Associated Risks, and your Baseline Disease in SARS-COV-2.

• Objectives
  – Describe the baseline demographic, clinical characteristics, treatments and outcomes among those tested for SARS-CoV-2 and/or diagnosed with COVID-19 overall and stratified by sex, age and specific comorbidities
  – Describe characteristics and outcomes of patients diagnosed/tested positive for influenza as well as patients hospitalized with influenza between September 2017 and April 2018 compared to the COVID-19 population.

• Followed OHDSI’s scientific best practices
  – Made protocol and analytic code publicly available
  – Sites inspected diagnostics & results for their study before sending them to study coordinators
  – Study results made available through online interactive application
Persons tested for SARS-COV-2

Persons tested positive for SARS-COV-2

Persons hospitalized with positive test for SARS-COV-2

Persons hospitalized and requiring intensive services with positive test for SARS-COV-2

Persons with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Persons hospitalized with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Persons hospitalized and requiring intensive services with COVID-19 diagnosis record OR a SARS-COV-2 positive test

CHARYBDIS target cohorts
CHARYBDIS subgroup cohorts

- Persons tested for SARS-COV-2
- Persons tested positive for SARS-COV-2
- Persons hospitalized with positive test for SARS-COV-2
- Persons hospitalized and requiring intensive services with positive test for SARS-COV-2
- Persons hospitalized with COVID-19 diagnosis record OR a SARS-COV-2 positive test
- Persons hospitalized and requiring intensive services with COVID-19 diagnosis record OR a SARS-COV-2 positive test

Stratification cohorts:
- Age: <18, >65
- Gender: Female/Male
- Race: Black/White
- Index month
- Hypertension
- Type 2 Diabetes
- Heart disease
- Obesity
- Asthma
- COPD
- Chronic kidney disease
- End stage renal disease
- Cancer
- Autoimmune conditions
- Dementia
- HIV
- Pregnant women
CHARYBDIS time windows

Post-index characteristics for treatments and outcomes:
- Concept-based:
  - Condition groups (SNOMED + descendants), >=1 occurrence during the interval
  - Drug era groups (ATC/RxNorm + descendants), >=1 day during the interval which overlaps with at least 1 drug era
Cohort features:
- Symptoms (fever, cough, malaise, myalgia, dyspnea)
- Acute clinical events (AKI, ARDS, AMI, PE/DVT, …)
- Service utilization (hospitalization, ventilation, tracheostomy, ECMO, dialysis)

Pre-index characteristics for medical history:
Demographics:
- Age group (5-year strata)
- Sex
Concept-based:
- Condition groups (SNOMED + descendants), >=1 occurrence during the interval
- Drug era groups (ATC/RxNorm + descendants), >=1 day during the interval which overlaps with at least 1 drug era
Cohort features:
- Symptoms (fever, cough, malaise, myalgia, dyspnea)
- Acute clinical events (AKI, ARDS, AMI, PE/DVT, …)
- Service utilization (hospitalization, ventilation, tracheostomy, ECMO, dialysis)
CHARYBDIS Results Viewer

Interactive application for exploring disease natural history:

- [https://data.ohdsi.org/Covid19CharacterizationCharybdis/](https://data.ohdsi.org/Covid19CharacterizationCharybdis/)
Objective: develop and externally validate COVID-19 Estimated Risk scores that quantify a patient’s risk of hospital admission, hospitalization requiring intensive services or fatality.
• COVER interactive website to provide live risk scores

• Impact: Health minister of Catalonia Spain explicitly mentions the COVER index as one of the indicators they will use to measure the impact of a given outbreak.
COVER: COVID risk prediction

Interactive application for exploring prediction:

• https://data.ohdsi.org/Covid19CoverPrediction/
Impact of healthcare big data on the pandemic

• Governments, regulators, product manufacturers, and clinicians need to understand COVID-19 to inform its vaccine development and therapeutic evaluation

• OHDSI’s network provides the largest international collection of databases with real-world experience of patients with COVID-19

• OHDSI's data network allows evidence generation across a range of use cases:
  – Characterize the baseline characteristics of COVID patients and current treatment patterns in COVID care (CHARYBDIS)
  – Identify patients at highest risk of adverse outcomes (COVER)
  – Enable estimates of the effectiveness and safety of therapeutic interventions in COVID (SCYLLA)

• The COVID pandemic is providing the opportunity to highlight how real-world evidence can be used responsibly for regulatory decision-making