



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

OHDSI COVID-19 Data Network

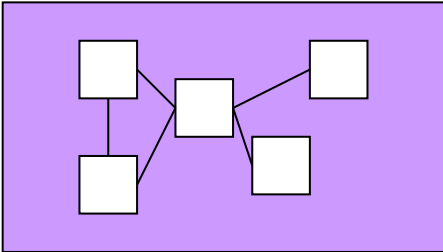


USA (8)	EUROPE (7)	ASIA-PACIFIC (3)
Premier (National – Hospital Billing)	CPRD (UK – Electronic Health Records)	HIRA (South Korea – Administrative Claims)
HealthVerity (Claims linked to diagnostic testing)	SIDIAP (Spain – Electronic Health Records)	DCMC (South Korea – Electronic Health Records)
Optum EHR (National – Electronic Health Records)	SIDIAP-H (Spain – EHR hospital linkage)	Nanfang Hospital (China – Electronic Medical Records)
IQVIA Open Claims (National – Administrative Claims)	HM Hospitales (Spain – Hospital Billing)	<div>Together, OHDSI has studied:</div> <ul style="list-style-type: none">>4.5m patients tested for SAR-COV-2>1.2m patients diagnosed or tested positive for COVID-19>249k patients hospitalized with COVID-19
Department of Veterans Affairs (National – Electronic Health Records)	ICPI (Netherlands – Electronic Health Records)	
Stanford University (CA – Electronic Health Records)	LPD France (France – Electronic Health Records)	
Tufts University (MA – Electronic Health Records)	Germany DA (Germany – Electronic Health Records)	
Columbia University (NY – Electronic Health Records)		



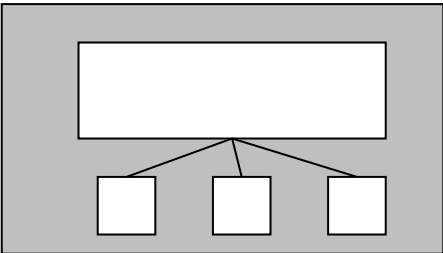
Common data model to enable standardized analytics

Source 1 raw data



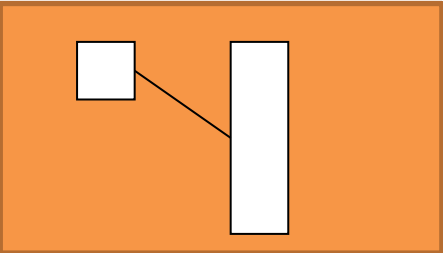
Electronic health records

Source 2 raw data

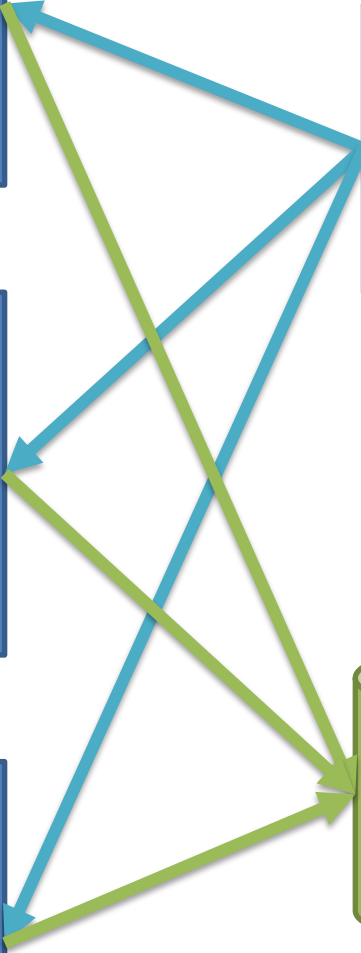
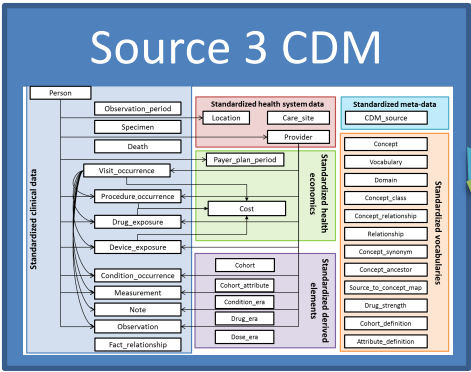
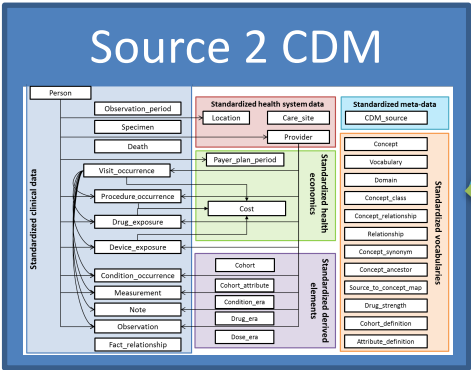
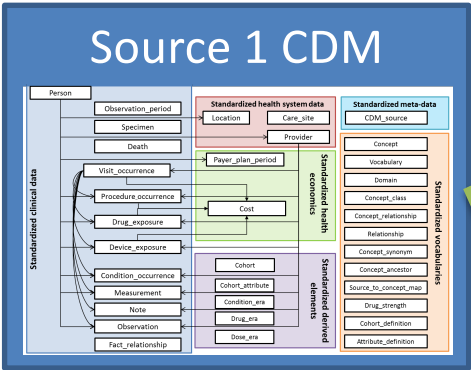
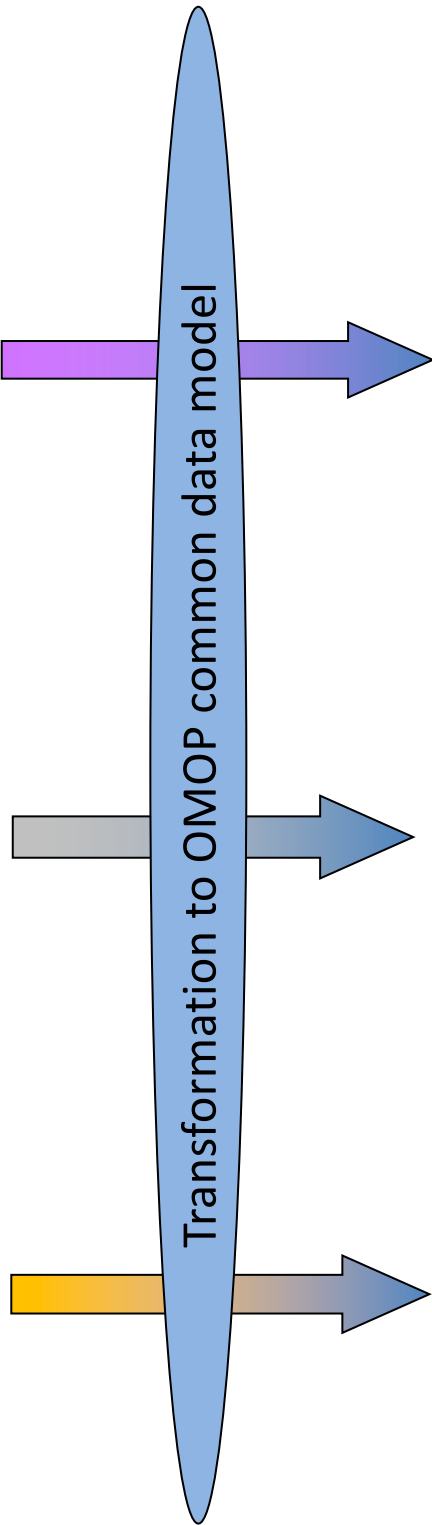


Administrative claims

Source 3 raw data



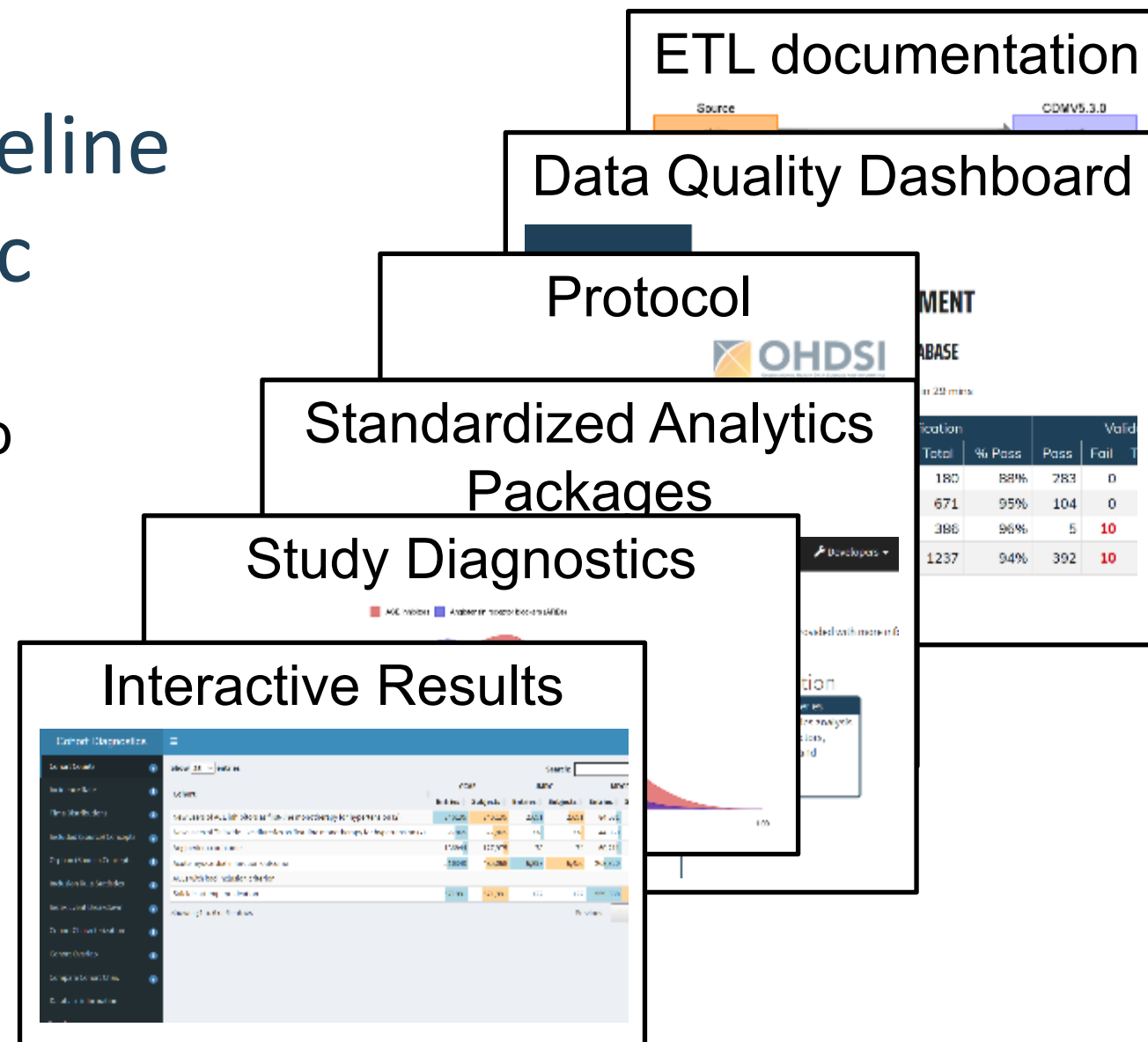
Clinical data





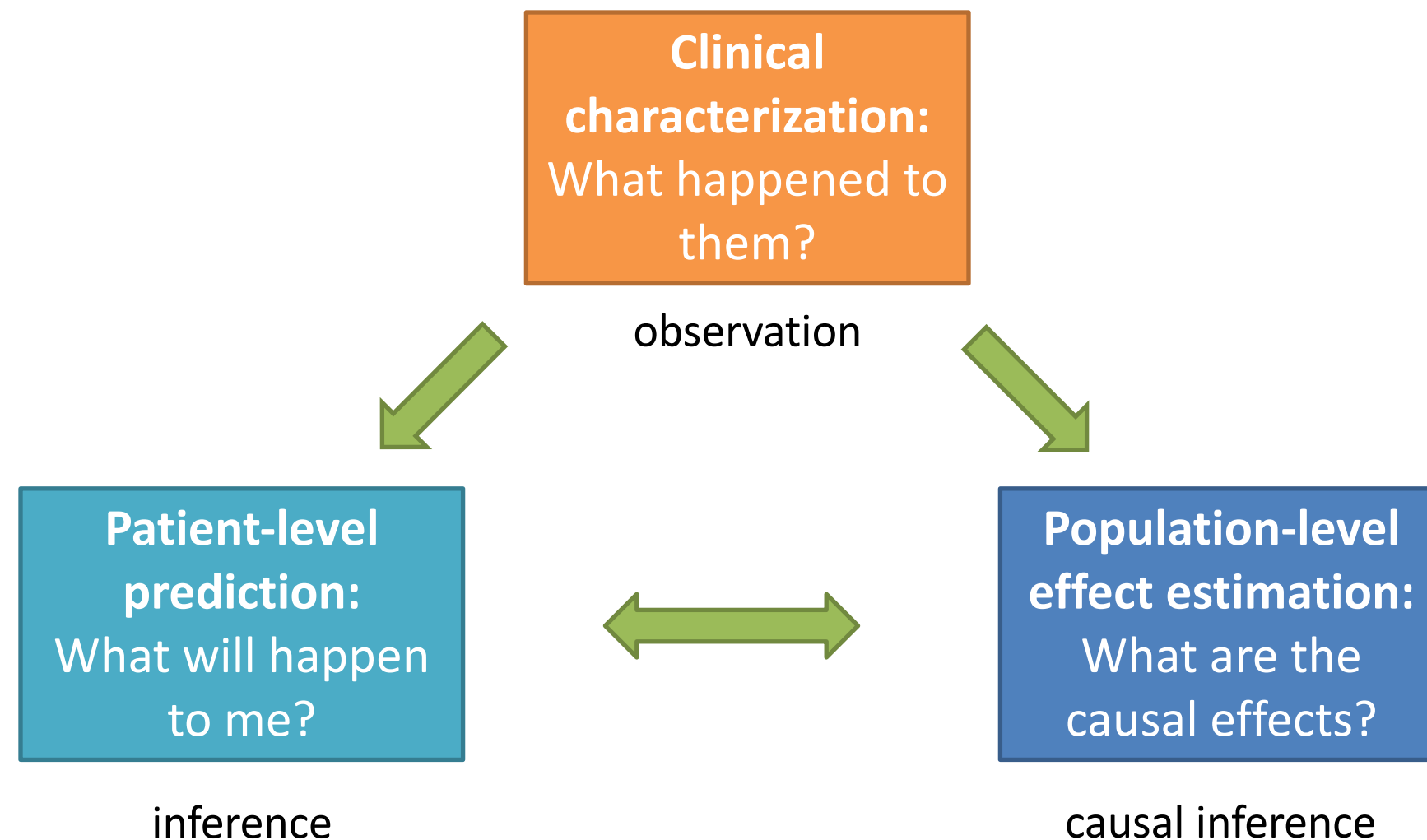
Driving agenda of full transparency

- 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

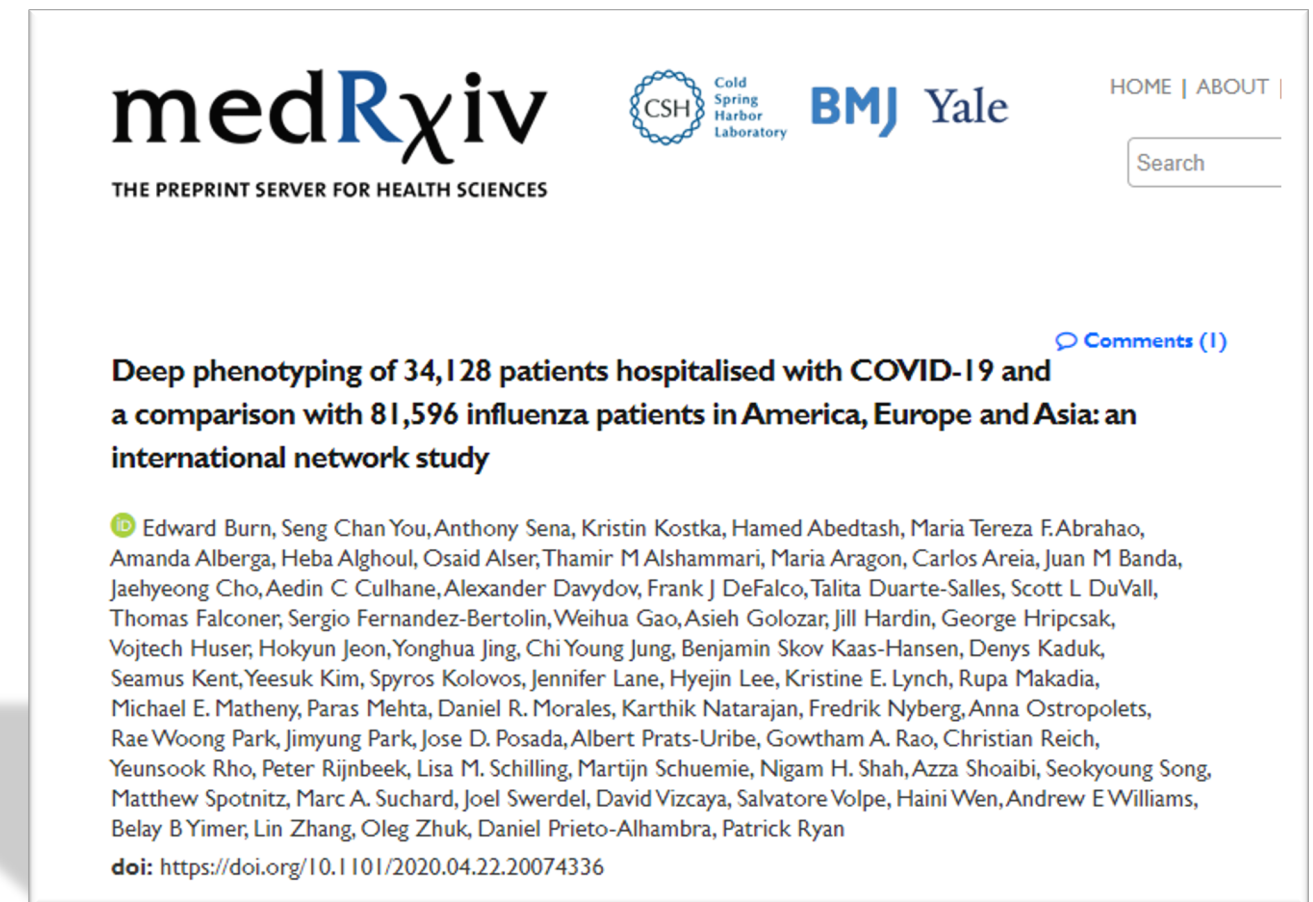




- Describe baseline characteristics for those hospitalized for COVID-19 as compared to those hospitalized for influenza

Findings:

- Patients hospitalized with COVID are systematically different from those hospitalized with flu
- COVID hospitalized patients, when compared 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

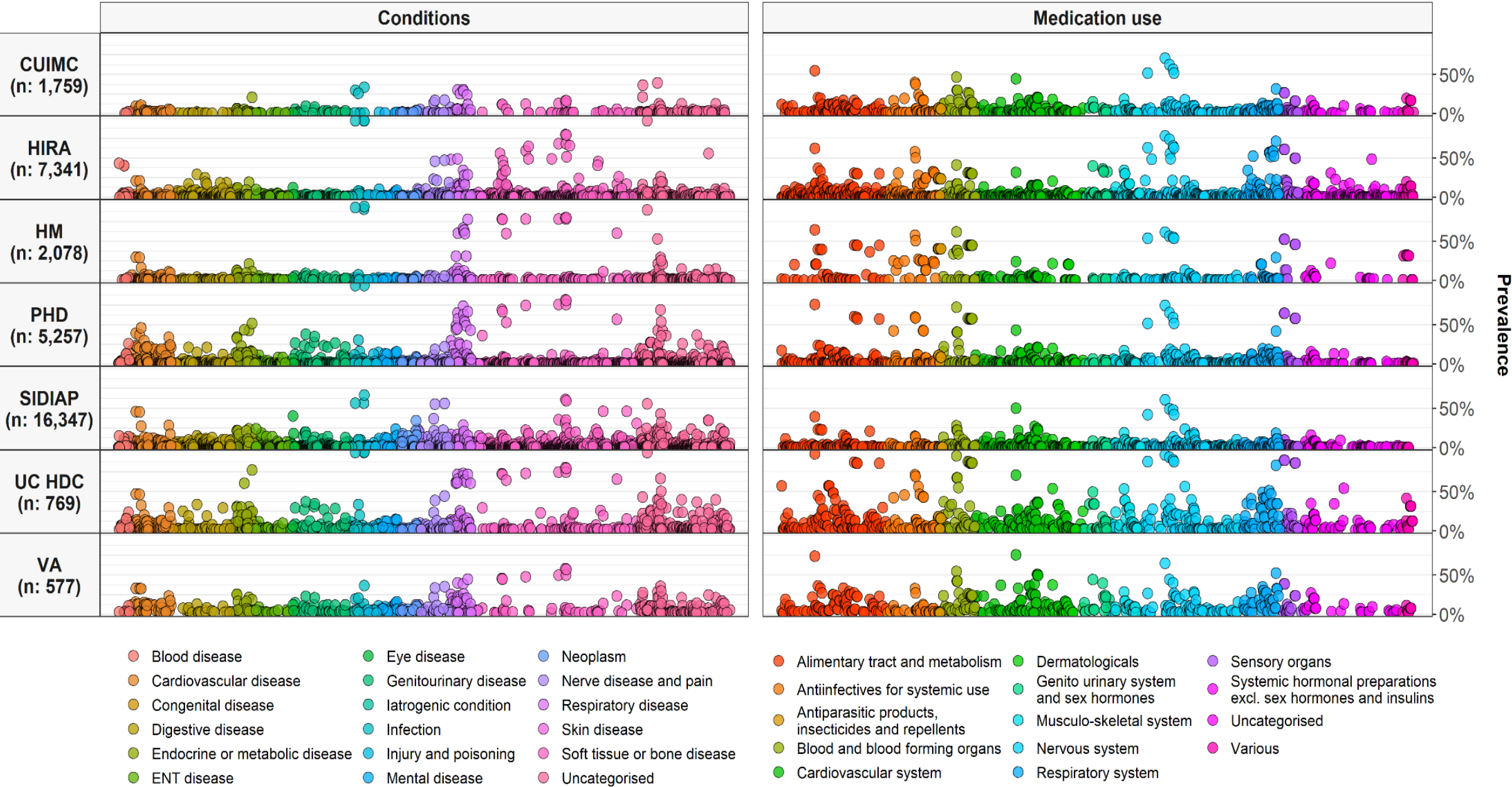


Accepted in *Nature-Communications*



Characterization

Disease Natural History of COVID-19



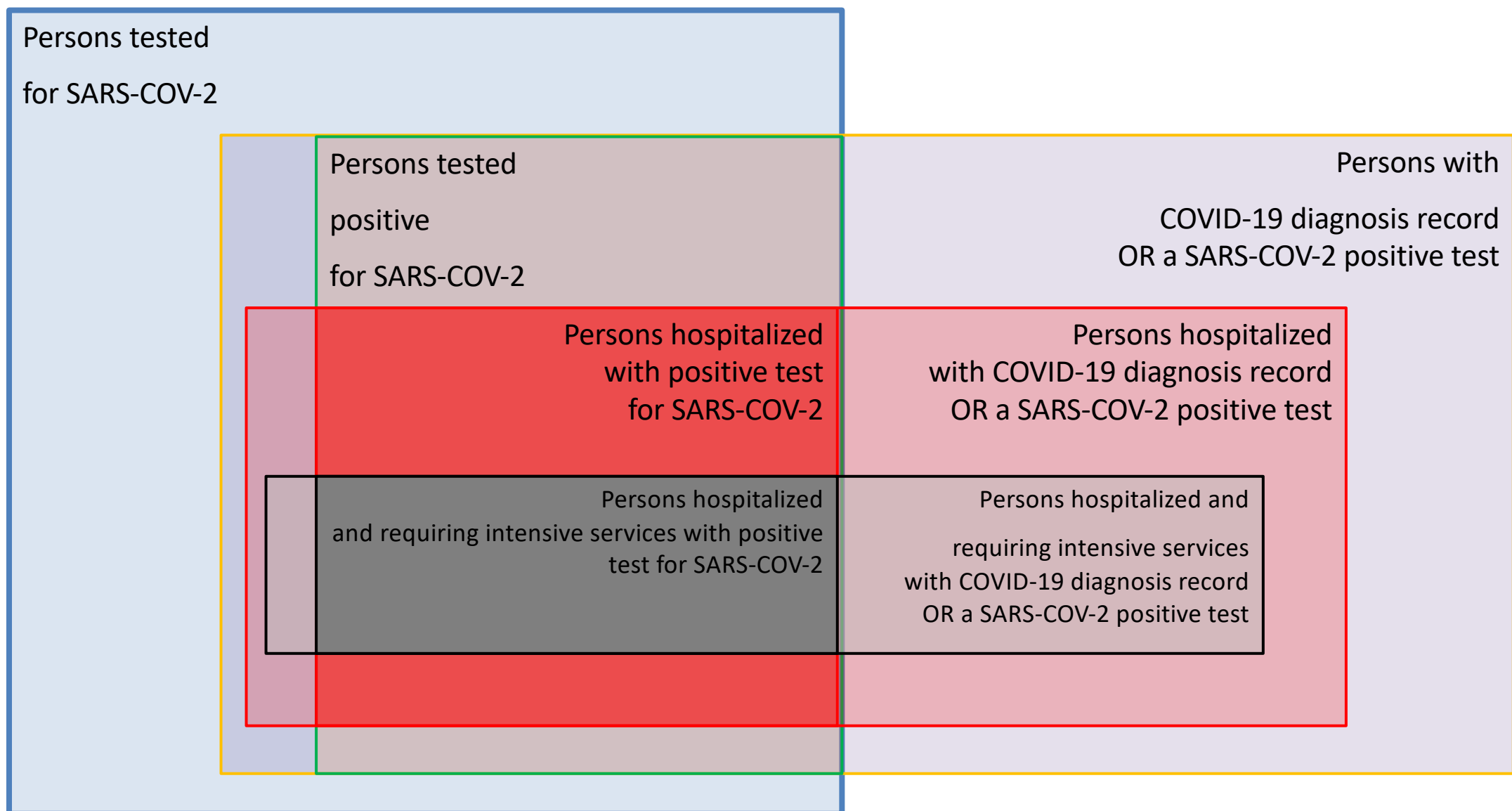


- CHARYBDIS – Characterizing Health Associated Risks, and your Baseline Disease in SARS-COV-2.
- Objectives
 - Describe the baseline demographic, clinical characteristics, treatments and outcomes among those tested for SARS-CoV-2 and/or diagnosed with COVID-19 overall and stratified by sex, age and specific 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



Characterization

CHARYBDIS target cohorts





Characterization

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 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

Stratification cohorts:

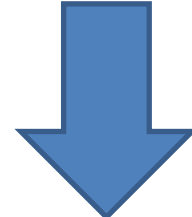
- Age: <18, >65
- Gender: Female/Male
- Race: Black/White
- Index month
- Hypertension
- Type 2 Diabetes
- Heart disease
- Obesity
- Asthma
- COPD
- Chronic kidney disease
- End stage renal disease
- Cancer
- Autoimmune conditions
- Dementia
- HIV
- Pregnant women



Characterization

CHARYBDIS time windows

Cohort
start date =
Index date



-365d to -1d

-30d to -1d

0d

0d to 30d

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)

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)



CHARYBDIS Results Viewer

Interactive application for exploring disease natural history:

- <https://data.ohdsi.org/Covid19CharacterizationCharybdis/>



Prediction

COVER: COVID risk prediction

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.



medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME | ABO Search

[Comment on this paper](#)

Seek COVER: Development and validation of a personalized risk calculator for COVID-19 outcomes in an international network

Ross D. Williams, Aniek F. Markus, Cynthia Yang, Talita Duarte Salles, Scott L. Duvall, Thomas Falconer, Jitendra Jonnagaddala, Chungsoo Kim, Yeunsook Rho, Andrew Williams, Amanda Alberga, Min Ho An, María Aragón, Carlos Areia, Edward Burn, Young Choi, Iannis Drakos, Maria Fernandes Abrahão, Sergio Fernández-Bertolín, George Hripcsak, Benjamin Kaas-Hansen, Prasanna Kandukuri, Jan A. Kors, Kristin Kostka, Siaw-Teng Liaw, Kristine E. Lynch, Michael E. Matheny, Gerardo Machnicki, Daniel Morales, Fredrik Nyberg, Rae Woong Park, Albert Prats-Urbe, Nicole Pratt, Gowtham Rao, Christian G. Reich, Marcela Rivera, Tom Seinen, Azza Shoaibi, Matthew E. Spotnitz, Ewout W. Steyerberg, Marc A. Suchard, Seng Chan You, Lin Zhang, Lili Zhou, Patrick B. Ryan, Daniel Prieto-Alhambra, Jenna M. Reps, Peter R. Rijnbeek

doi: <https://doi.org/10.1101/2020.05.26.20112649>

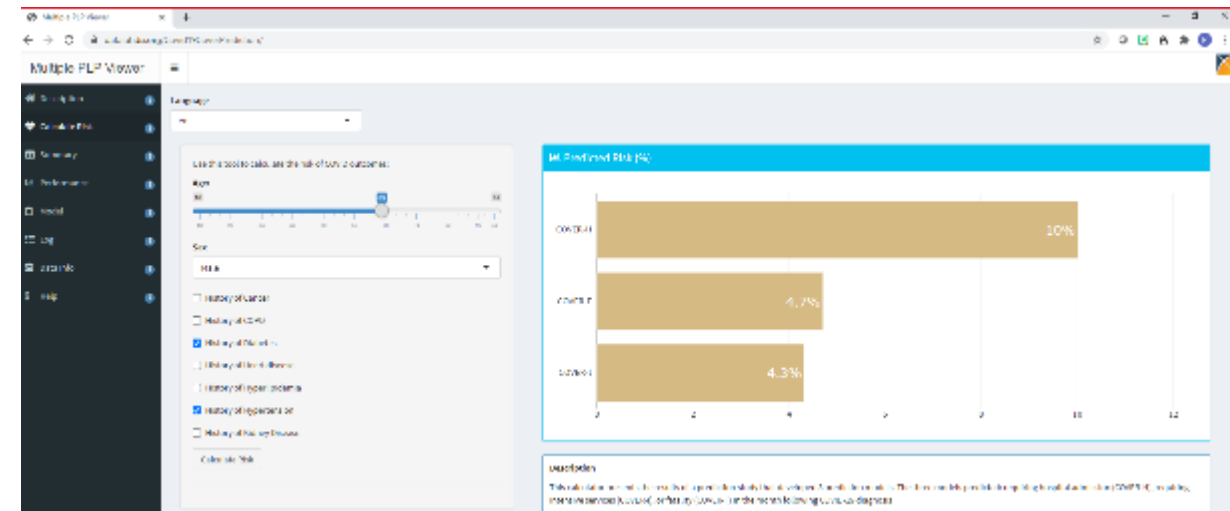
Under review at *PLOS Medicine*



Prediction

COVER: COVID risk prediction

- 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.



3. Indicadors

El Pla es basa en la mesura de **10 indicadors principals** que permeten una fotografia acurada de la realitat epidèmica a Catalunya.



En la interpretació dels indicadors s'aplicaran **factors de correcció** com: índex socioeconòmic complex, envelliment de la població o la densitat poblacional.

/Salut

Generalitat de Catalunya



Prediction

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