Rapid Fire Presentations of Collaborators

Katia Verhamme, MD, PhD

Associate Professor of Use and Analysis of Observational Data

FEEDERNET (FEDERATED E-HEALTH BIG DATA FOR EVIDENCE RENOVATION NETWORK) PLATFORM IN KOREA

Seongwon Lee, Chungsoo Kim, Junhyuk Chang, Rae Woong Park

FeederNet (Federated E-Health Big Data for Evidence Renovation Network) platform in Korea

2022. 06. 24

Seongwon Lee, Chungsoo Kim, Junhyuk Chang, Rae Woong Park

Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea
Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea

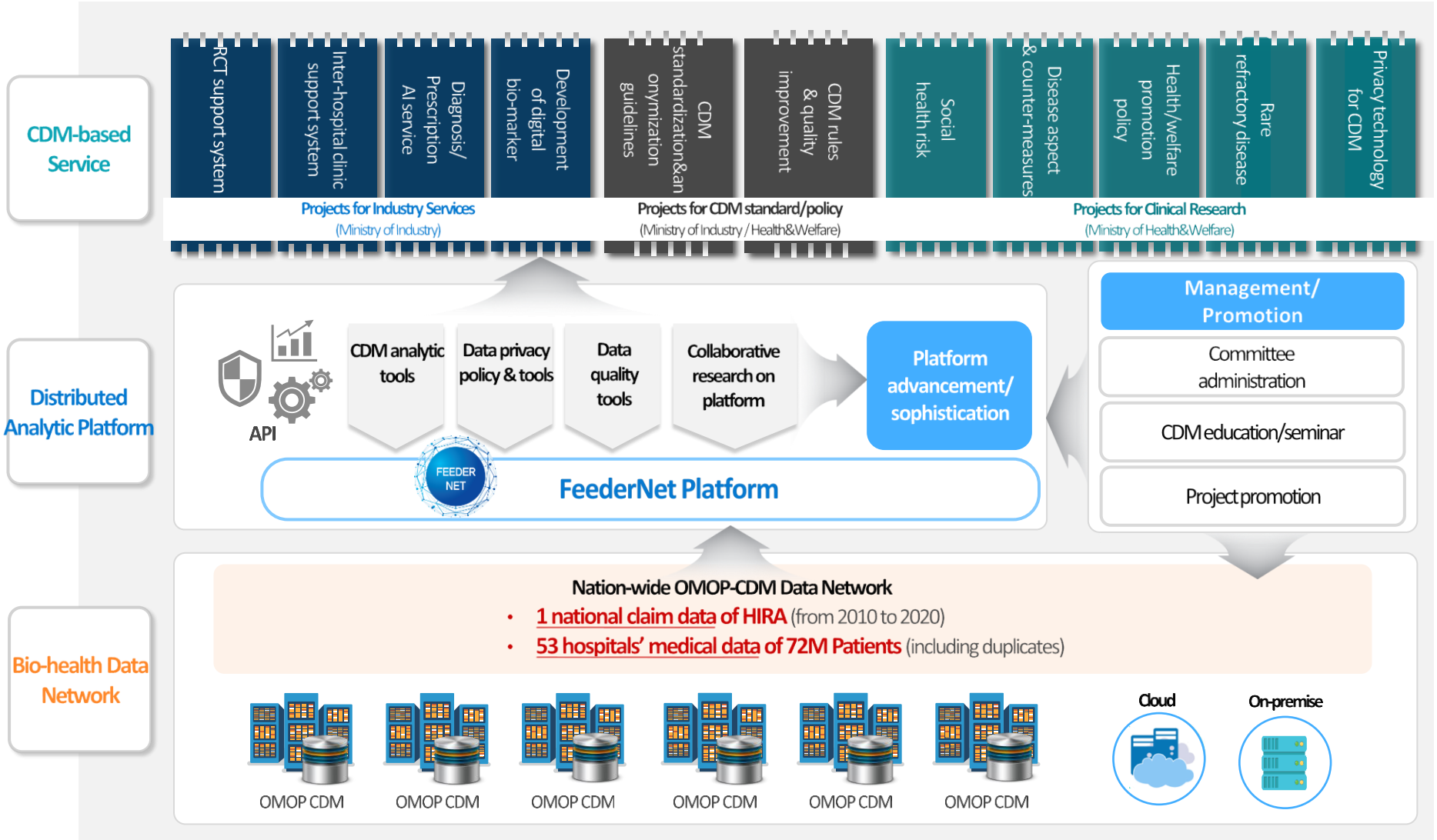
* Presenting author: Chungsoo Kim, Pharm D

아주대학교 AJOU UNIVERSITY

OHDSI EUROPE OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OHDSI OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

FeederNet Project in Korea

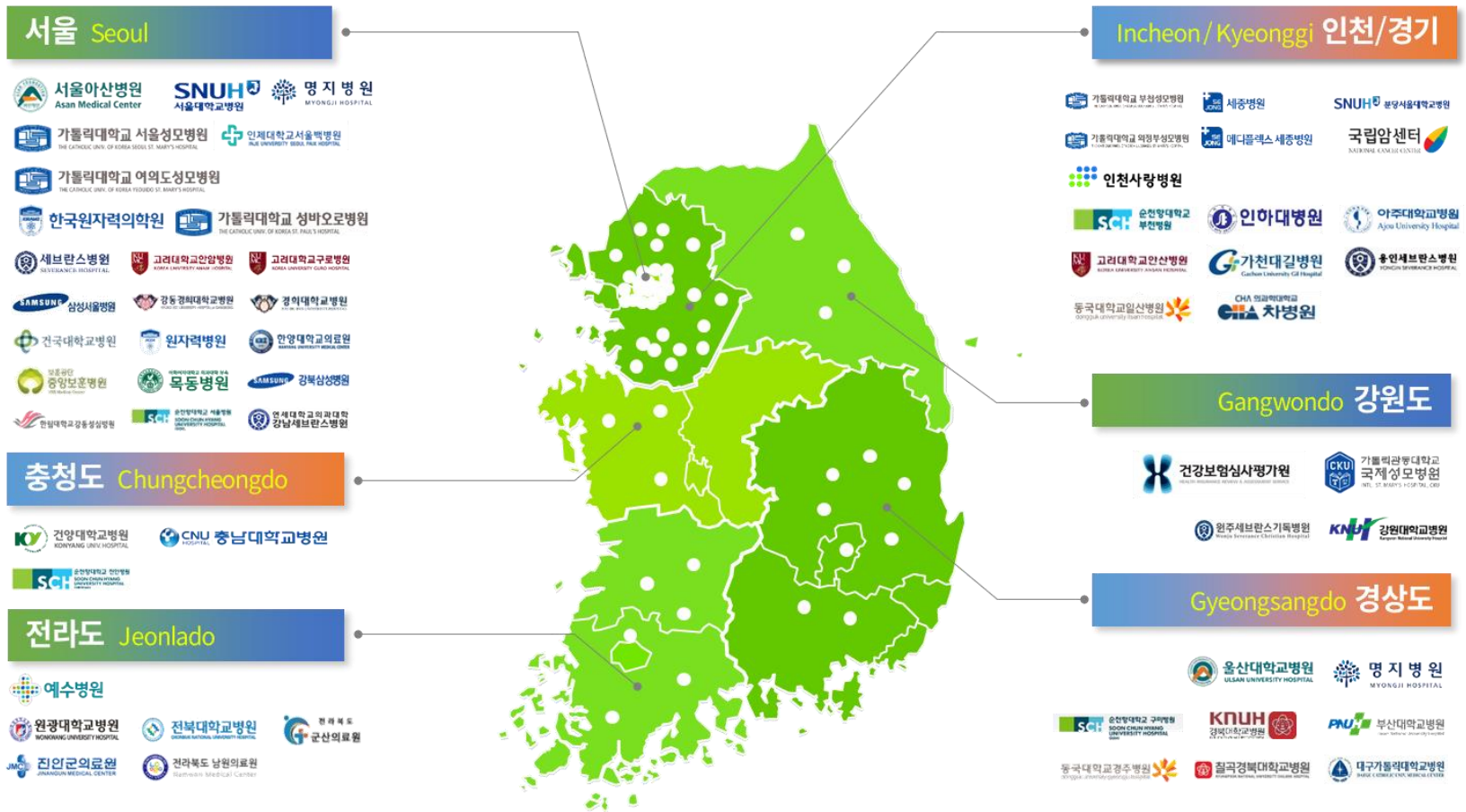


OMOP-CDM Data Network

Nation-wide CDM Data Network

- **53** hospitals (including 72% of all tertiary hospitals in Korea)
- **71,987,327** patients (including duplicates)

※ National claim data (2010-2020) of **All Korean** was also converted to OMOP-CDM

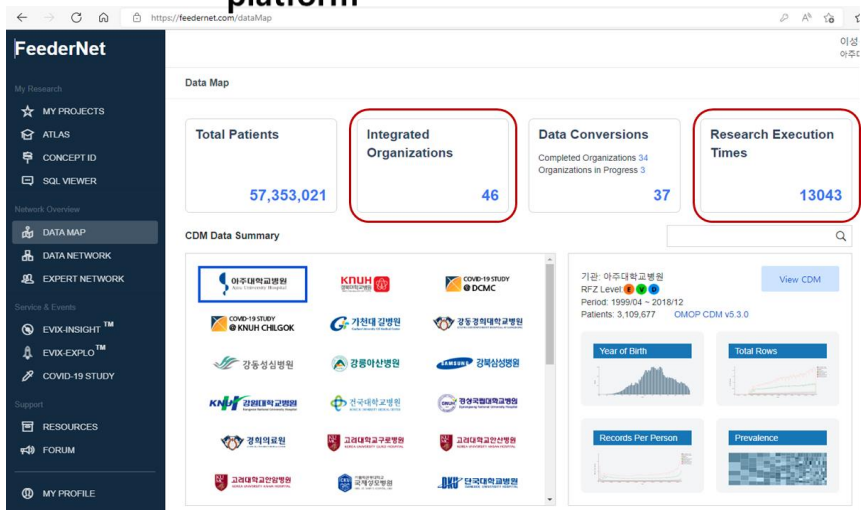


FeederNet Platform

FeederNet Platform

- **46** CDM DBs of hospitals have been integrated with FeederNet platform
- FeederNet Central + FeederNet Node
- Since May 2019, total**13,043** analyses have been executed

Main page of FeederNet platform

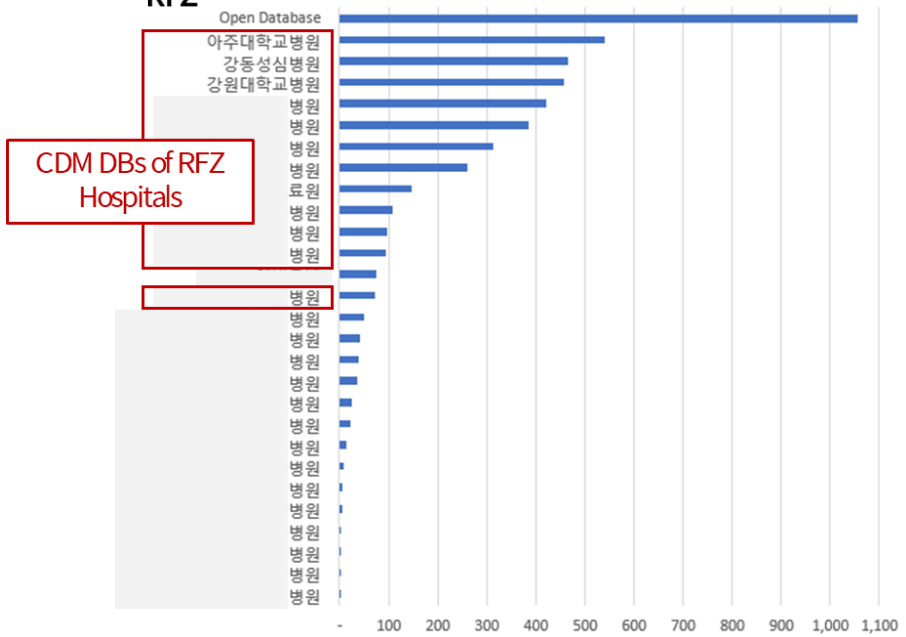


Research Free Zone

- Mutual Cooperative MOU for promoting joint research
- Clauses of RFZ
  1. Grant **same authorities on CDM DB analysis** to all researchers in RFZ hospitals
  2. Approve **single IRB** among RFZ hospitals

Currently, **18** Hospitals joined the RFZ

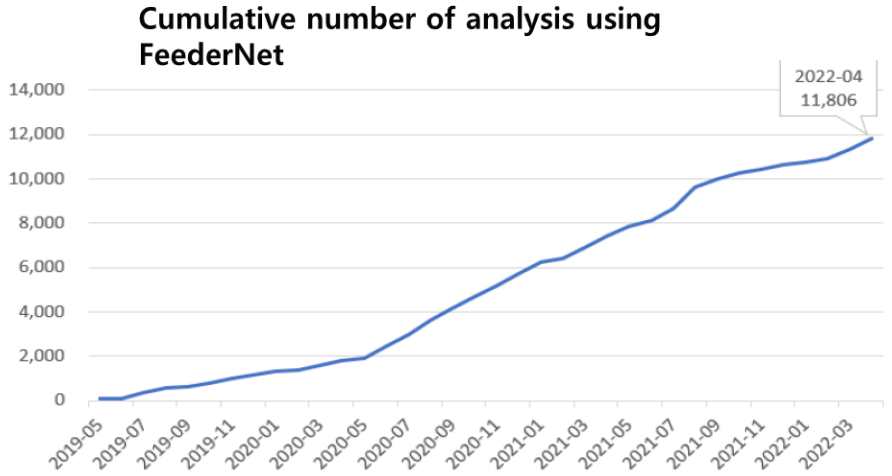
Comparison of CDM usage between RFZ and non-RFZ



OMOP-CDM Research in Korea

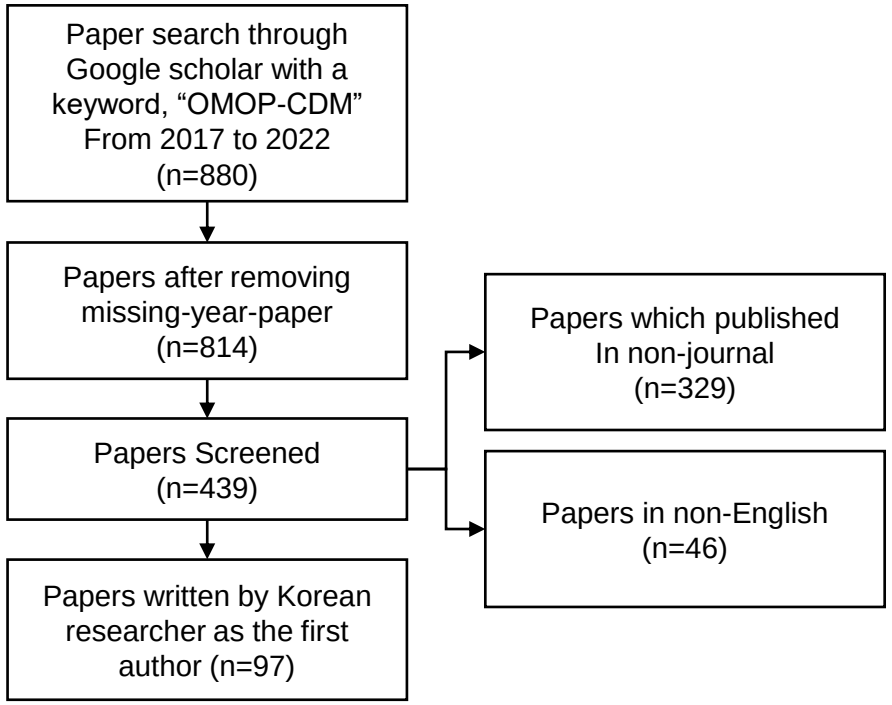
Analysis on FeederNet

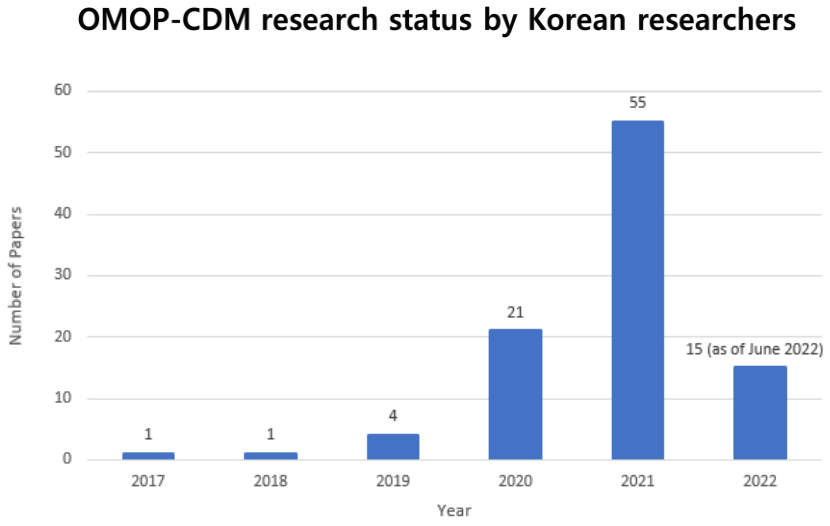
- Number of analyses continues to increase
- From June 2020, about 500 analyses have been being conducted every month



Paper Publications

- **Total 97** papers of Korean researcher as a first author have been published since 2017 and **55** were published **in only 2021**

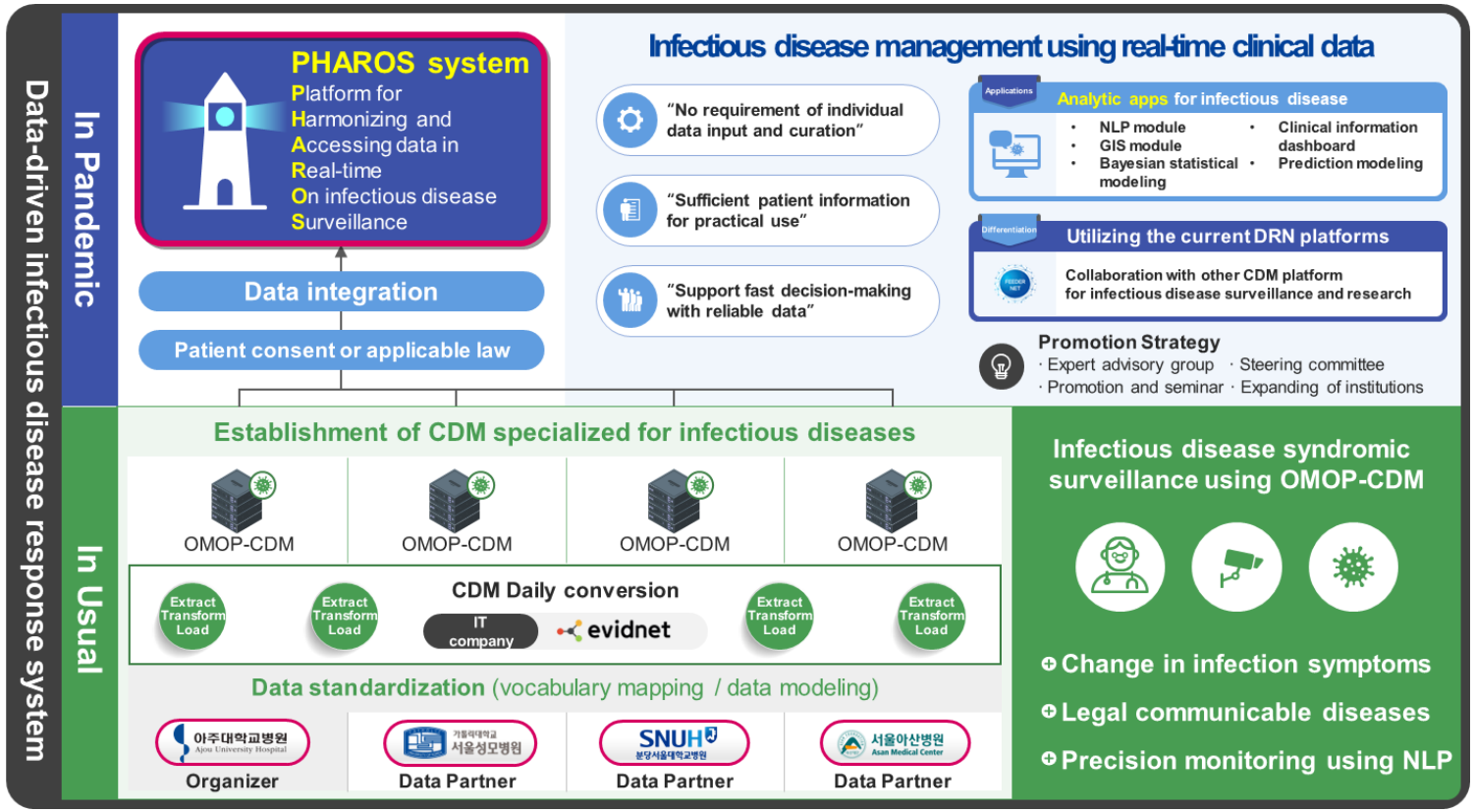




New CDM project on Infectious Disease Surveillance

'PHAROS' Project

- PHAROS (Platform for Harmonizing and Accessing data in Real-time On Infectious disease Surveillance)
  - **Purpose** : Establishment of **infectious disease surveillance** system using **real-time CDM** data of multi centers
  - **Uniqueness**
    1. **Daily automatic CDM ETL**
    2. **Patient-centered Integration** of CDMs from multi-centers



FUTURE

The same framework can be used for **pragmatic clinical trials (PCT)**

Thank you

OMOP GENOMIC MAPPING CAPACITIES IN CONVERSION OF COMPREHENSIVE GENOMIC PROFILING RESULTS

Rogozhkina Maria, Odysseus

OMOP Genomic mapping capacities in conversion of comprehensive genomic profiling results

by

ODYSSEUS DATA SERVICES

EUROPE

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Representation of Copy Number aberrations

***Post*coordination**

ABL1 (ABL proto-oncogene 1, non-receptor tyrosine kinase) gene variant measurement

Gene deletion

ABL1|deletion

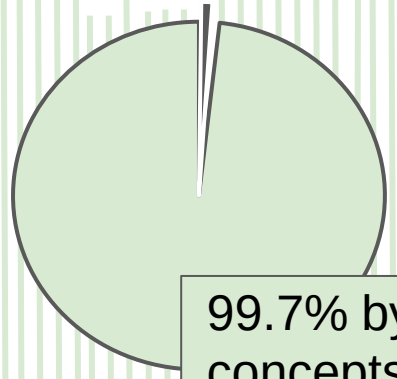
MET Gene Mutation measurement

Gene amplification

MET|amplification

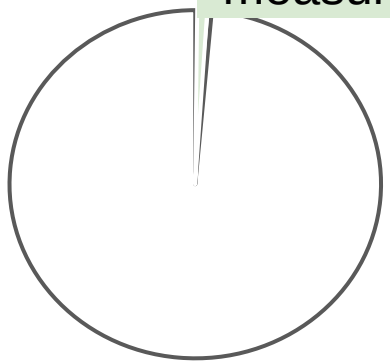
***Pre*coordination**

MET Gene Amplification measurement



99.7% by 2 target concepts: measurement + value

0.7*% by 1 target concept: measurement



**All 5 codes for amplifications only*

Representation of Single Nucleotide Variants

***Post*coordination**

JAK2 (Janus kinase 2) gene variant measurement

Abnormal nucleotide base sequence

BRCA1 (BRCA1 DNA repair associated) gene variant measurement

Abnormal nucleotide base sequence

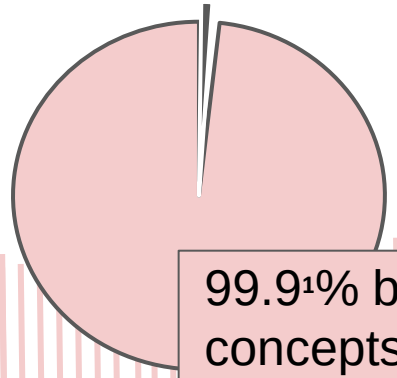
JAK2 p.L830= c.2490G>A

BRCA1 p.K1183R c.3548A>G

***Pre*coordination**

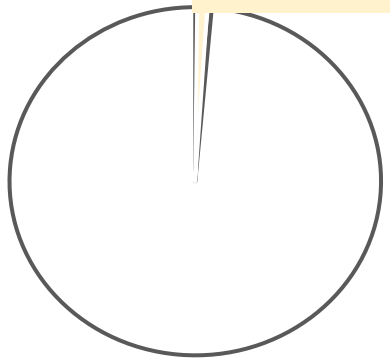
JAK2 transcript: Substitution in position 2490 of G replaced by A measurement

BRCA1 transcript: Substitution in position 3548 of A replaced by G measurement



99.9[1]% by 2 target concepts: measurement + value

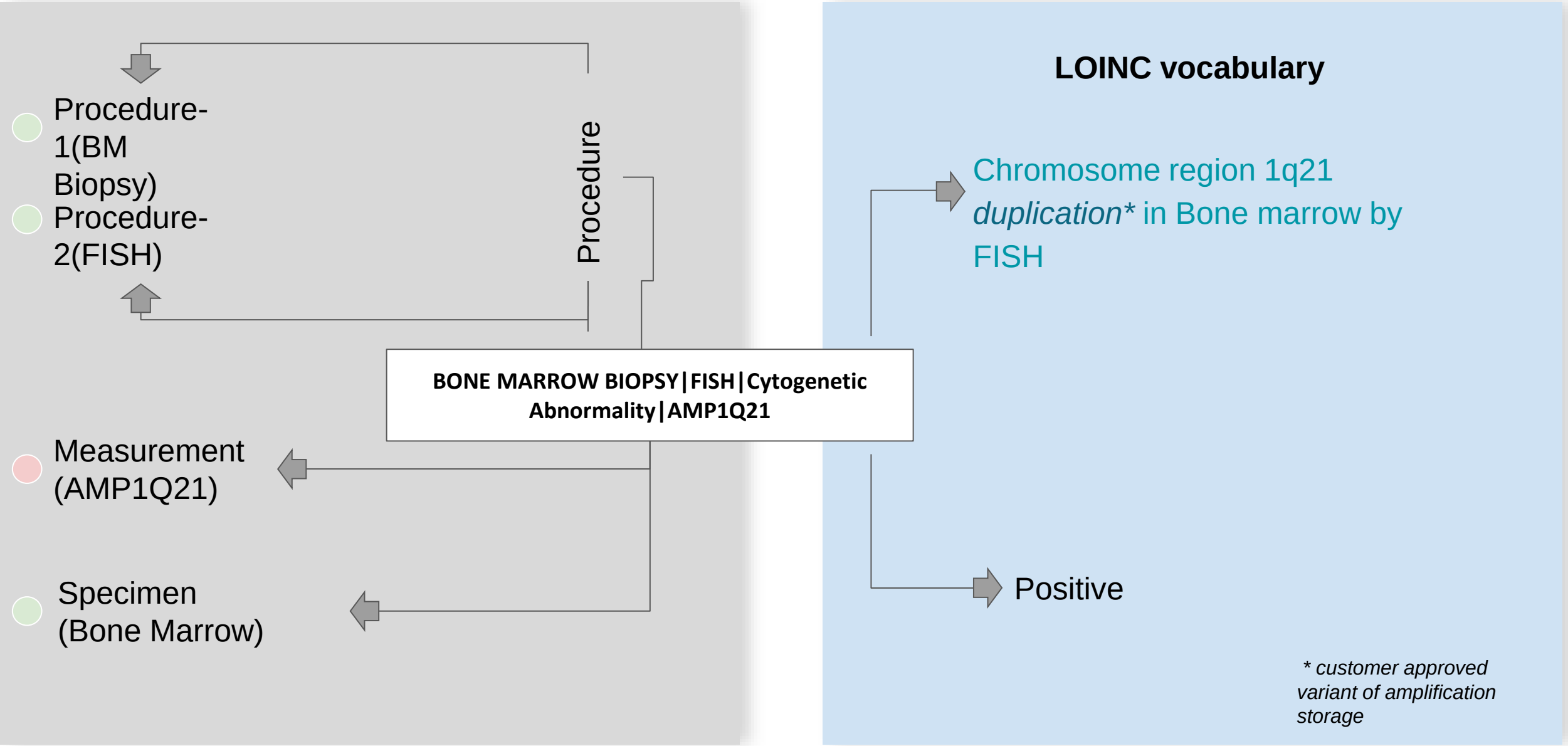
1.6[2]% by 1 target concept: measurement



[1]- *data about position is missing in all target concepts*

[2] - *0.3% was mapped to Protein level*

Using existing vocabularies as a fallback



Conclusion

- We found *46* duplicate concepts for DNA, *1276* for RNA and *224* for Protein variants in OMOP Genomic. All of them **to be deduplicated**.
- A substantial number of codes are missing**. New codes** to be ingested in OMOP Genomics are essential to facilitate studies.
- Concepts from non-Genomic vocabularies may be a valid target, but lacks of consistency and only a fraction of required targets is available.
- Linking to other event tables such as Procedure and Specimen is required to properly represent source data. To efficiently run queries an easier way to **link facts** is required (**direct** fact modification)

OMOP MAPPING OF REAL-WORLD DATA FROM BRAZIL & PAKISTAN TOWARDS MANAGEMENT OF COVID-19 IN THE GLOBAL SOUTH

Sara Khalid, University of Oxford

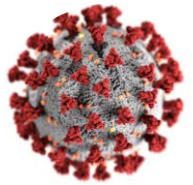
OMOP Mapping of Real-World Data From Brazil & Pakistan Towards Management of COVID-19 In the Global South

Authors

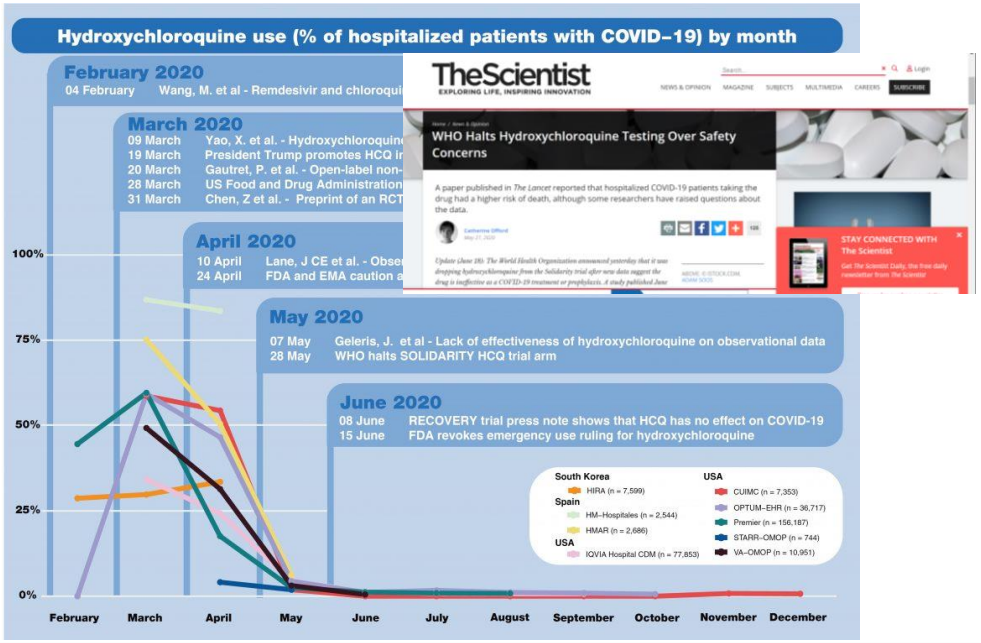
Elzo Pereira Pinto Junior[1], Priscilla Normando[1], Renzo Flores-Ortiz[1], Muhammad Usman Afzal[2], Muhammad Asaad Jamil[2], Sergio Fernandez Bertolin[3], Vinícius de Araújo Oliveira[1], Valentina Martufi[1], Edward Burn[4], Maria Yury Ichihara[1], Maurício L. Barreto[1], Talita Duarte Salles[3], Daniel Prieto-Alhambra[4], Haroon Hafeez[2], Sara Khalid[4]

[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



Planetary Health Informatics Group | sara.khalid@ndorms.ox.ac.uk
Centre for Statistics in Medicine | University of Oxford | www.csm.ox.ac.uk | @CSMOxford

The OHDSI COVID-19 Data Network

- Rapid
- Reliable
- **Representative**

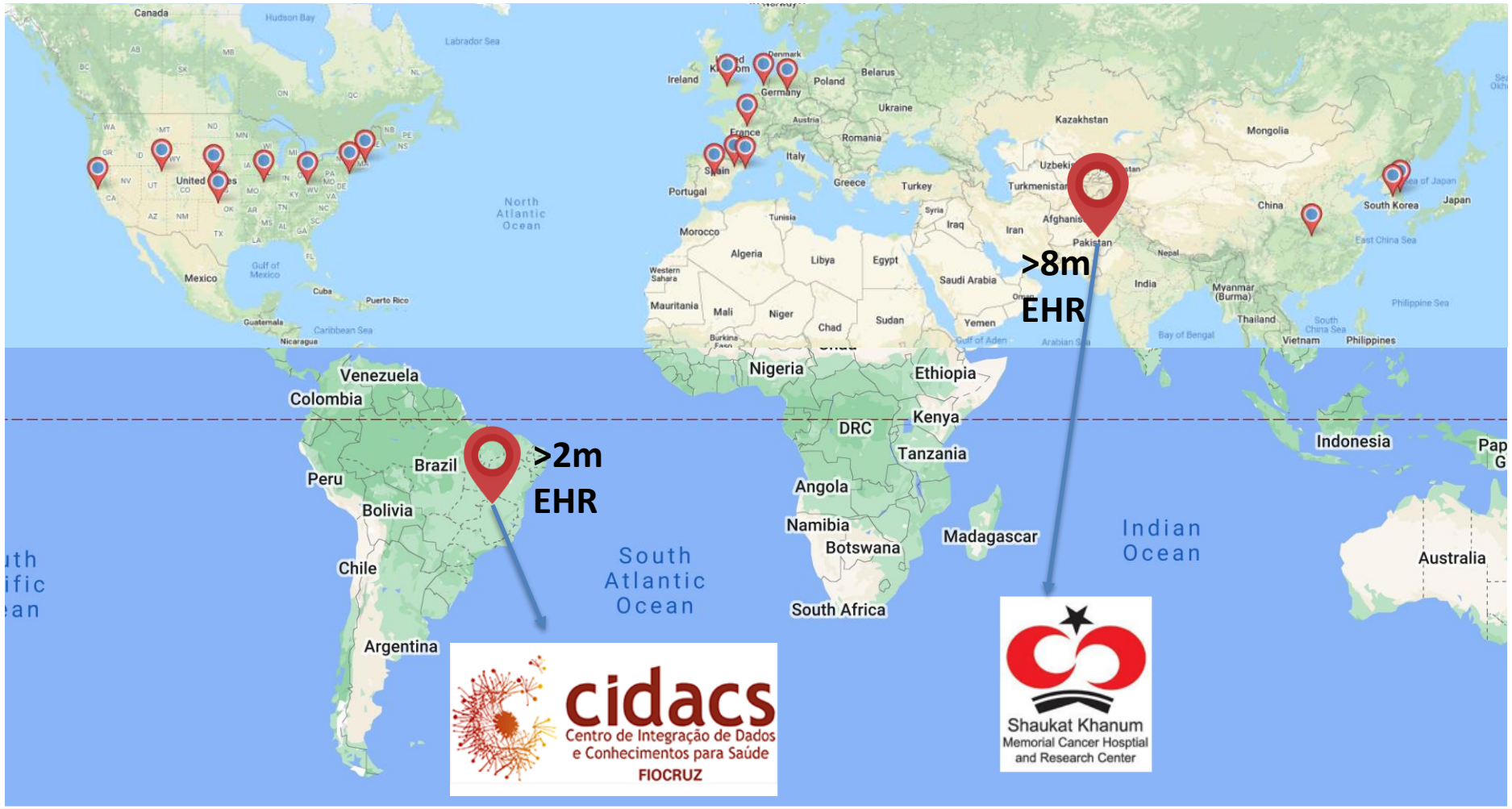
USA (8)	EUROPE (7)	ASIA-PACIFIC (3)
Premier (National – Hospital Billing)	CPRD (UK – Electronic Health Records)	HIRA (South Korea – Administrative Claims)
HealthVerity (Claims linked to diagnostic testing)	SIDIAP (Spain – Electronic Health Records)	DCMC (South Korea – Electronic Health Records)
Optum 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)	
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)		

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

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

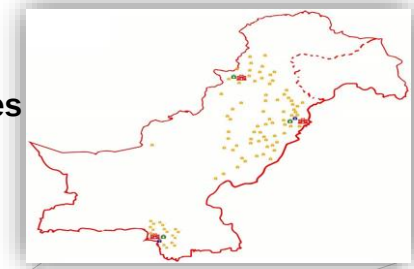


Data Partners – Pakistan

- 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

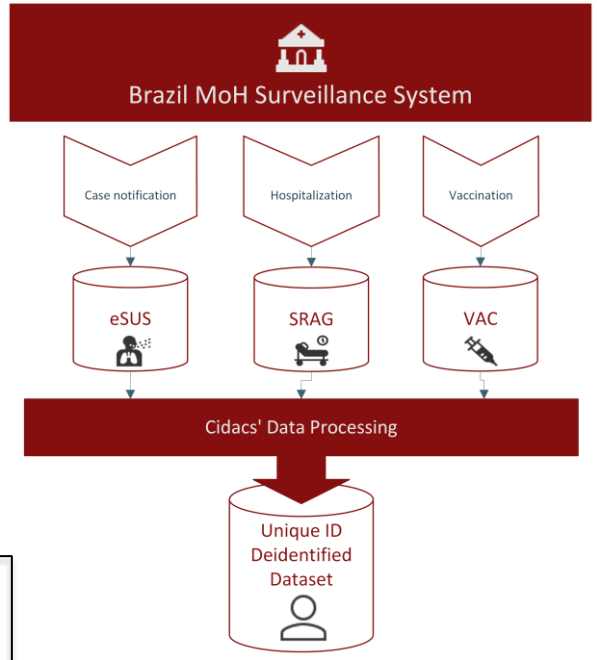
Clinical activity snapshot (2021)

12,018	210,667	14,840	63,725	19,128	77,809	198,393	6,277,572
New Registrations	OPD Visits	Admissions	Chemotherapy Sessions	Surgical Procedures	Radiotherapy Sessions	Imaging Studies	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



Harmonisation to OMOP

ETL Process

Pakistan

1 - Perform scan on source database using White Rabbit.

2 - Perform table/ column level mapping from source database to CDM (Rabbit in a Hat).

3 - Export source data to CSV files after extraction of required variables / attributes.

4 - Import CSV files in a staging database in PostgreSQL using DBeaver/ Pentaho Transformations.

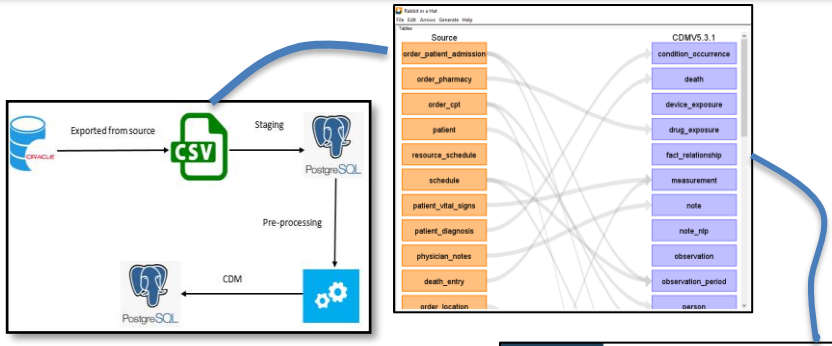
5 - Concept Mapping (USAGI and Athena)

6 - Apply data preprocessing and transformation on staging database.

7 – Load the data into CDM using PostgreSQL stored procedures.

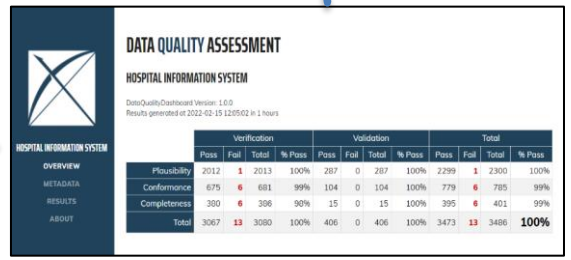
Software used during ETL:

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

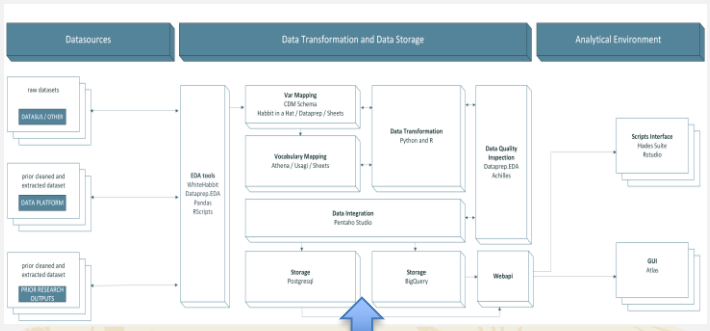


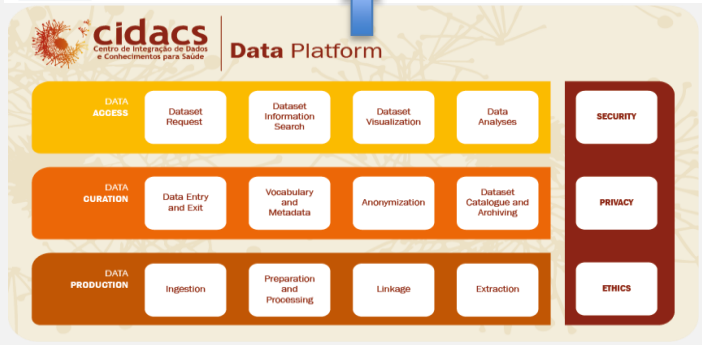
Mapping in Numbers

- >100K (source) to 108K (CDM) concepts
- >33M measurements
- >2M procedures
- >600K observations
- <1% missing matching concepts



Brazil

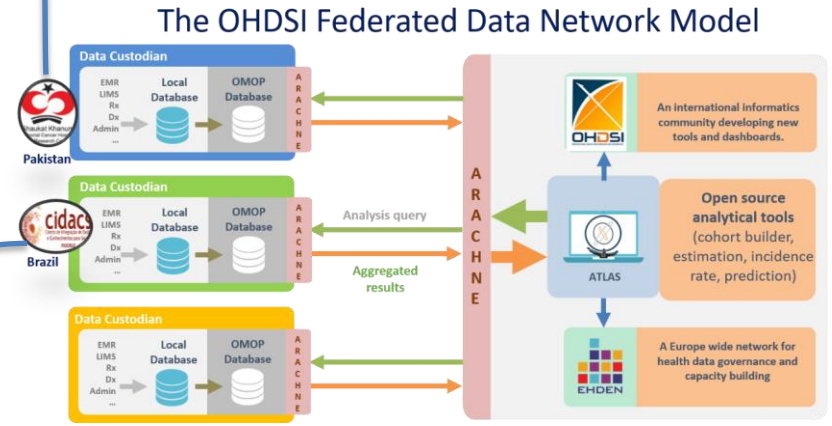




Barreto, Maurício Lima, et al. Pop Data Science 2019

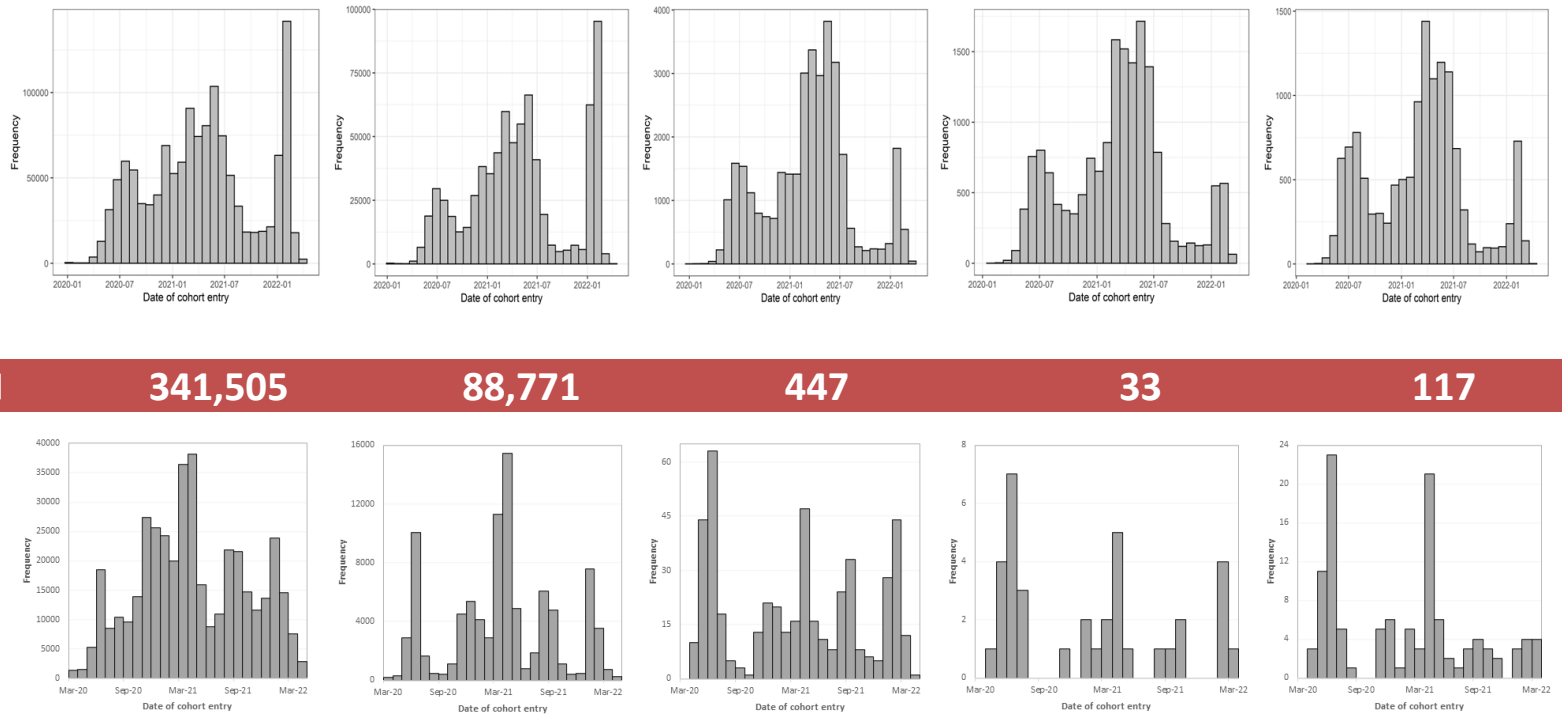
Mapping Complete

Mapping In Progress



COVID-19 cohorts – cases over time

	Tested population	COVID-19 diagnosis or positive test	Hospitalised with COVID-19	ICU admission with COVID-19	COVID-19 death
Brazil					
N	1,312,832	752,699	34,699	17,041	13,877
Pakistan					
N	341,505	88,771	447	33	117

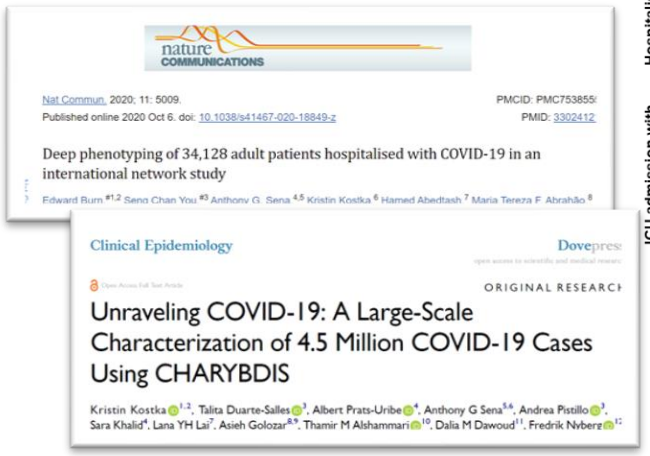


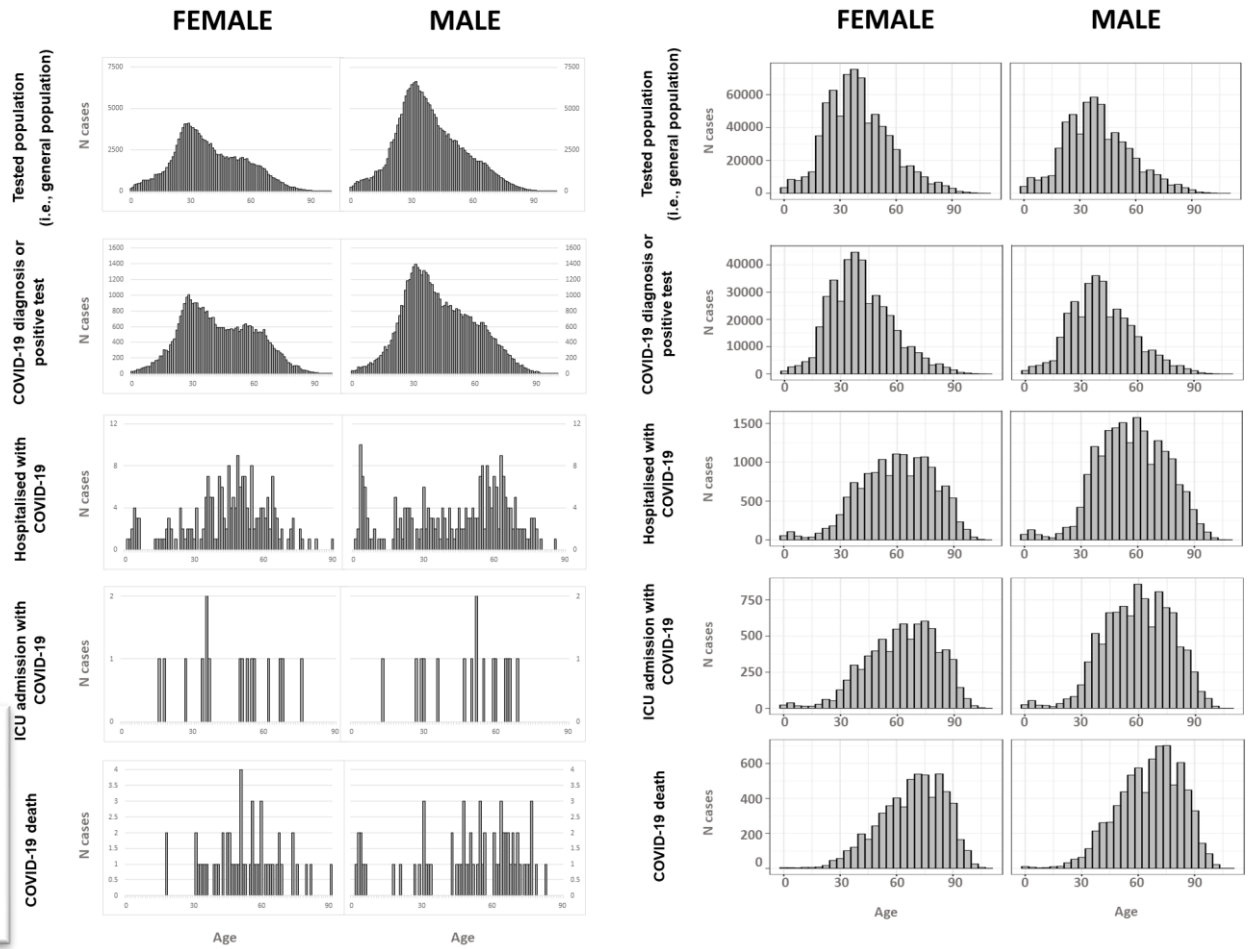
Distribution of cases over time (Jan/March 2020 – April 2022)

In Review. *JAMIA* 2022

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

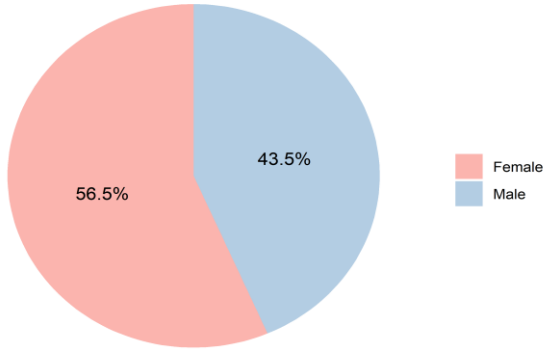
In Review. *JAMIA* 2022

COVID-19 cohorts – baseline characteristics

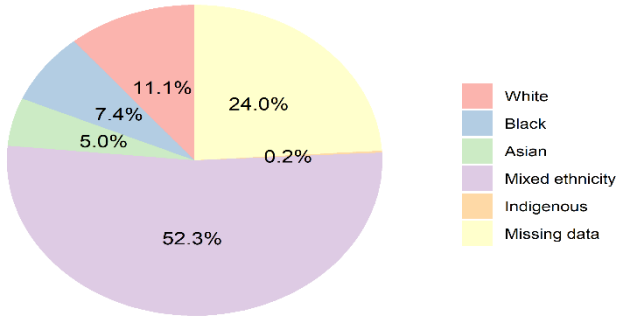
Brazil

N = 1,312,832

Sex

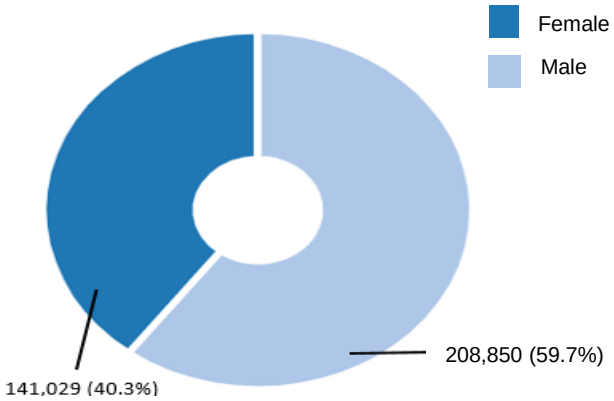


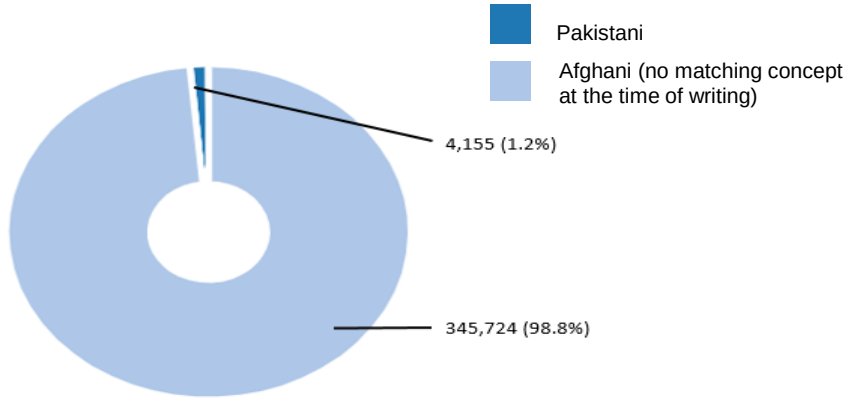
Ethnicity



Pakistan

N = 341,505





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.

In Review. *JAMIA* 2022

Summary

- 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

Acknowledgements

- This work was supported by funding from the Bill & Melinda Gates Foundation

Planetary Health Informatics Group | sara.khalid@ndorms.ox.ac.uk
Centre for Statistics in Medicine | University of Oxford | www.csm.ox.ac.uk | @CSMOxford

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

Authors

Elzo Pereira Pinto Junior[1], Priscilla Normando[1], Renzo Flores-Ortiz[1], Muhammad Usman Afzal[2], Muhammad Asaad Jamil[2], Sergio Fernandez Bertolin[3], Vinícius de Araújo Oliveira[1], Valentina Martufi[1], Edward Burn[4], Maria Yury Ichihara[1], Maurício L. Barreto[1], Talita Duarte Salles[3], Daniel Prieto-Alhambra[4], Haroon Hafeez[2], Sara Khalid[4]

[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.

IMPACT OF RANDOM OVERSAMPLING AND RANDOM UNDERSAMPLING ON THE DEVELOPMENT AND VALIDATION OF PREDICTION MODELS USING OBSERVATIONAL HEALTH DATA

Cynthia Yang, Erasmus MC

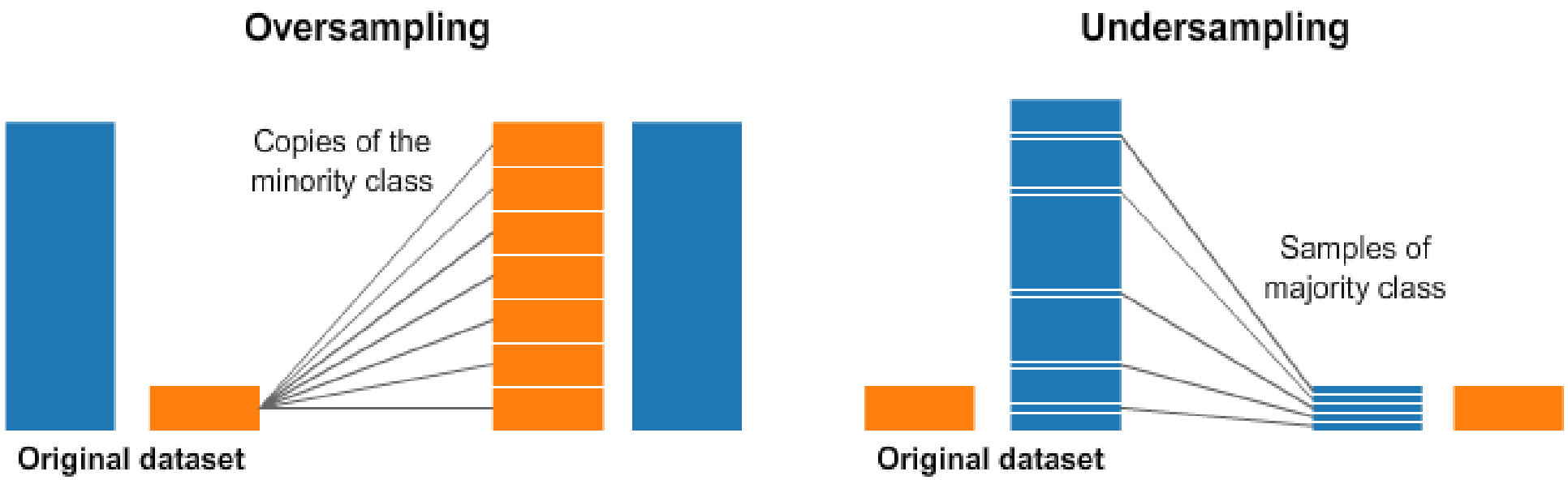
Impact of random oversampling and random undersampling on the development and validation of prediction models using observational health data

Cynthia Yang
PhD Student at Erasmus MC

Erasmus MC
University Medical Center Rotterdam

The class imbalance problem

- Class imbalance: a small proportion of patients in a study population (minority class) experiences a certain outcome of interest.

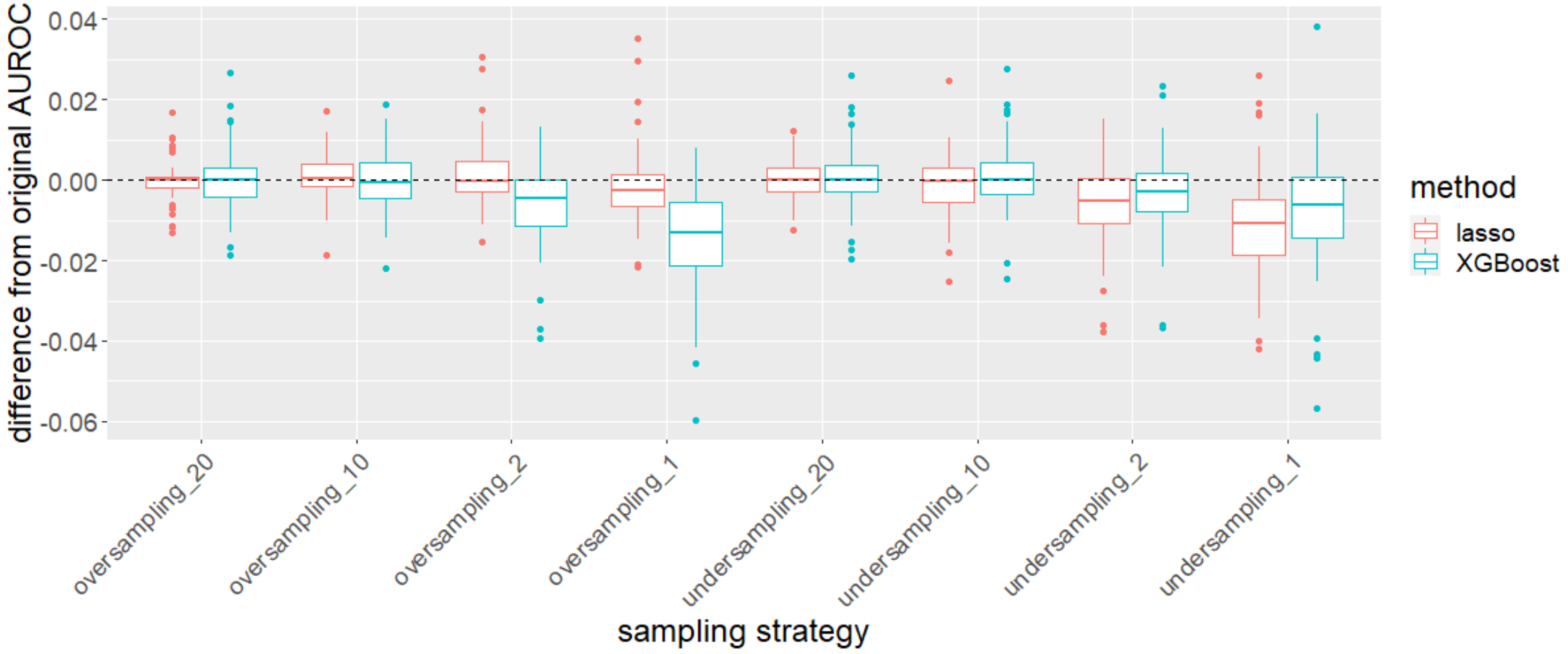


Methods

- Random oversampling and random undersampling
- Imbalance ratio (IR) = the total number of patients **without** an outcome event (majority class) / the total number of patients **with** an outcome event (minority class)
- We vary IR = min($IR_{original}$, x) with $x \in \{20, 10, 2, 1\}$

- 21 outcomes in depression (PLP framework paper)
- CCAE, MDCD, MDCR, IQVIA Germany
- XGBoost, lasso logistic regression
- Area under the receiver operating characteristic curve (AUROC)

Results



Erasmus MC

Conclusions

- The impact of random sampling on the AUROC is limited.
- Random sampling strategies on average do not improve the AUROC.

- Next steps: model calibration, random forest, external validation.

- For more information or questions please visit me at my poster #57.

REAL-WORLD EVIDENCE IS IN DEMAND!

A SUMMARY OF ‘LIVE’ REQUESTS FOR RWE STUDIES PUBLISHED BY A EUROPEAN HEALTH TECHNOLOGY ASSESSMENT (HTA) AGENCY

Jamie Elvidge, National Institute for Health and Care Excellence

(NICE)

Real-world evidence is in demand!

Live requests for real-world evidence (RWE) studies from a health technology assessment (HTA) agency

Jamie Elvidge

jamie.elvidge@nice.org.uk

Ravinder Claire, Shane Collins, Dalia Dawoud

NICE National Institute for Health and Care Excellence

About NICE

- The HTA agency for England
- Provides guidance to the healthcare system
- Key principles:
  - **evidence-based medicine**
  - **opportunity cost**

- Actively exploring ways to use RWE for decision making

NICE

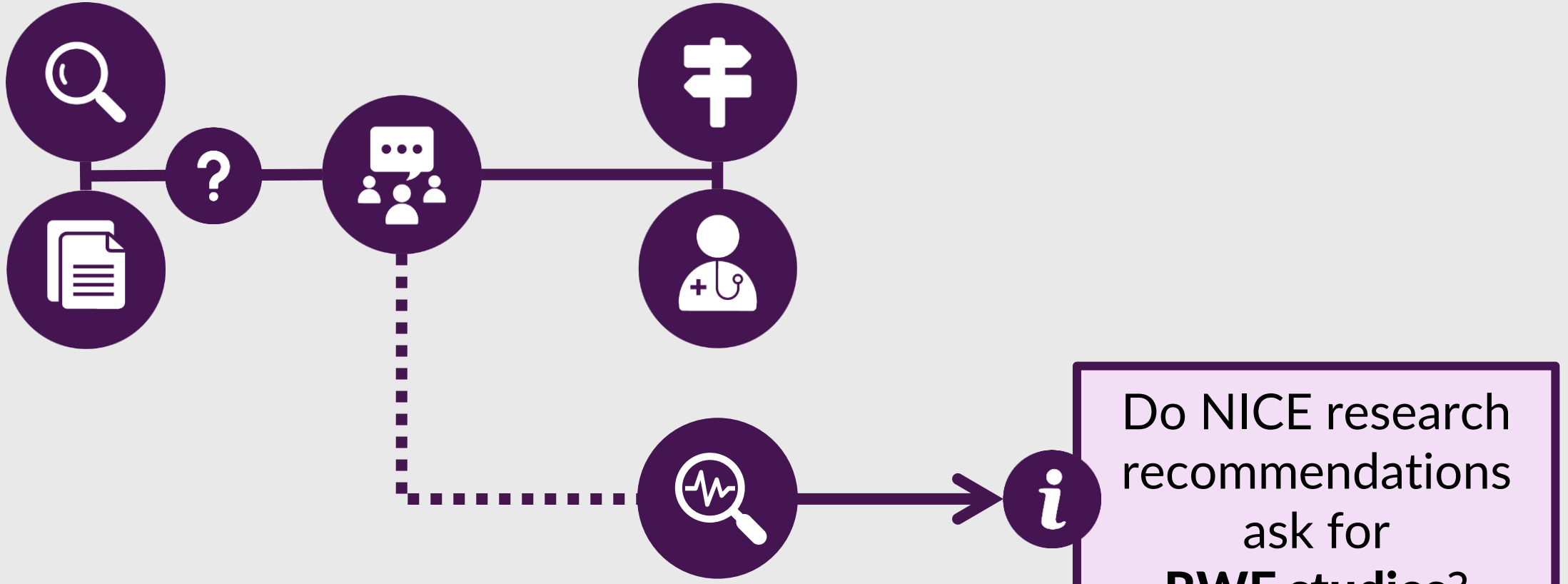
NICE real-world evidence framework

Overview

Key messages

- The NICE Strategy 2021 to 2026 states our ambition to use real-world data to resolve gaps in knowledge and drive forward access to innovations for patients. Real-world data is essential to enabling rapid, robust, and responsive technology evaluations and dynamic, living guidelines.
- We developed the Real-World Evidence Framework to help deliver on this ambition. It does this by:
  - Identifying when real-world data can be used to reduce uncertainties and improve

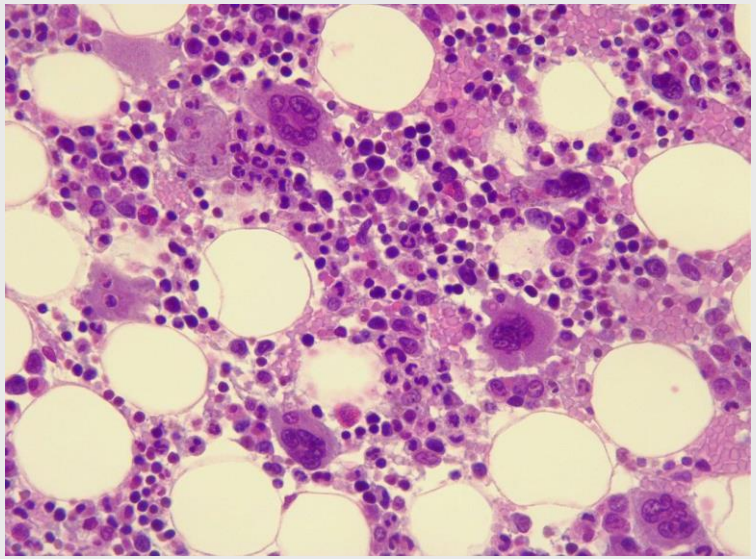
NICE recommends research to fill evidence gaps identified during guidance development



NICE

There are calls for RWE from across NICE guidance

NICE



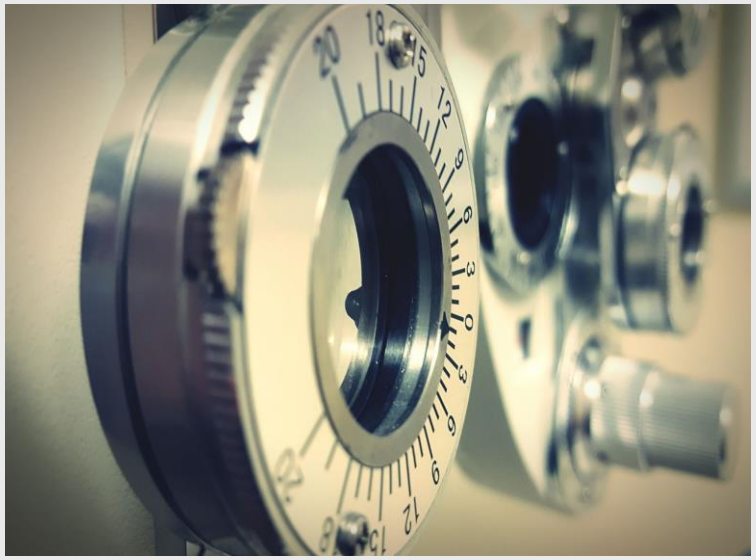
Ophthalmology

Routine treatment can be effective, but also unpleasant, costly and risky.

→ Observational studies to characterise treatment benefit and predict when the benefits cease.

Lung cancer

Large retrospective analyses to predict determinants of lung cancer, characterise current practice, and predict likelihood of unnecessary surgery.



Delirium

Observational studies to characterise prevalence of delirium in long-term care settings, and explore whether delirium is predictive of adverse outcomes and death.

https://www.nice.org.uk/about/what-we-do/science-policy-research/research-recommendations

WHY PREDICTING RISK CAN'T IDENTIFY 'RISK FACTORS'
EMPIRICAL ASSESSMENT OF MODEL STABILITY IN MACHINE LEARNING ACROSS OBSERVATIONAL HEALTH DATABASES

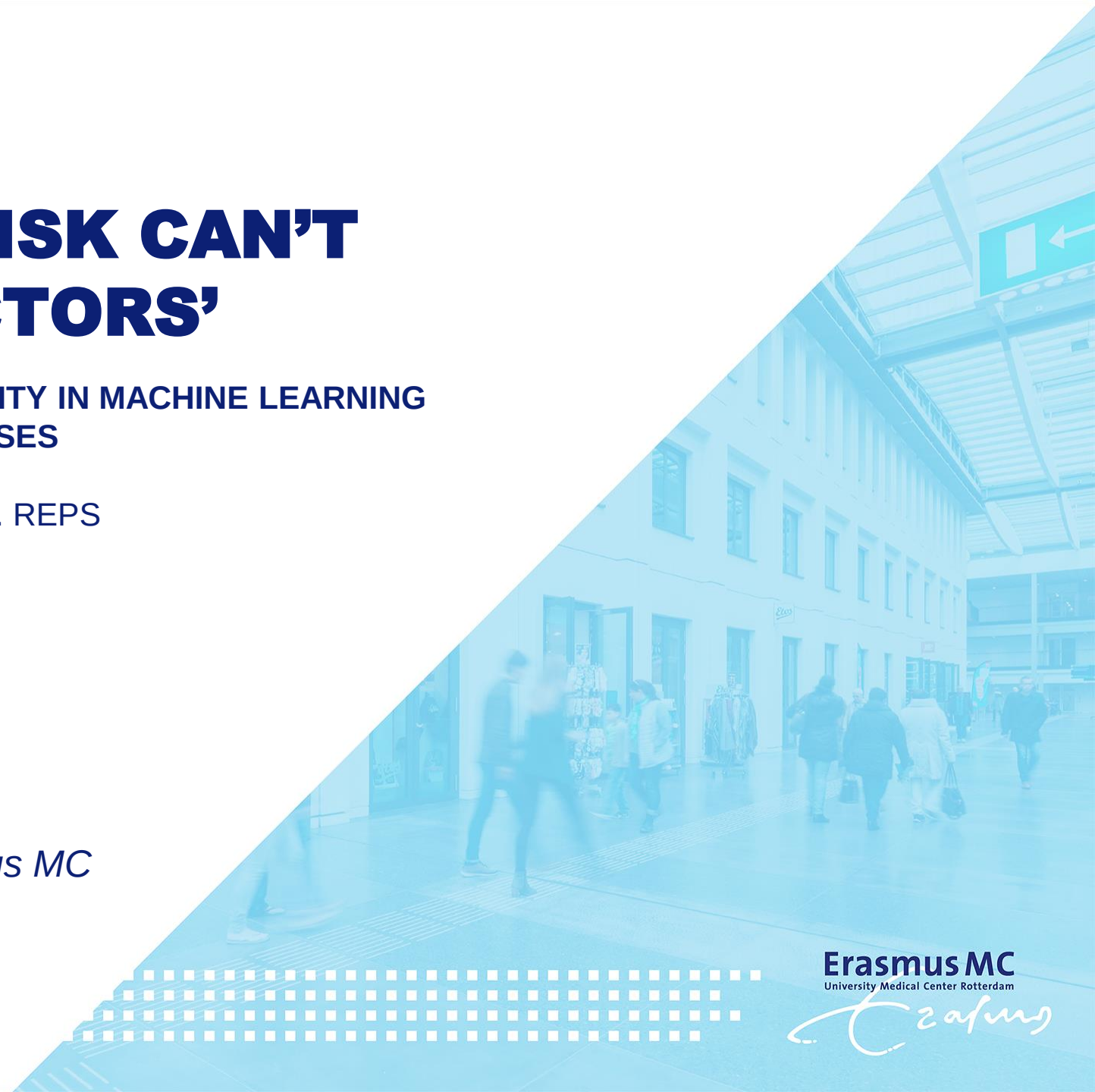
Aniek Markus, Erasmus MC

WHY PREDICTING RISK CAN'T IDENTIFY 'RISK FACTORS'

EMPIRICAL ASSESSMENT OF MODEL STABILITY IN MACHINE LEARNING ACROSS OBSERVATIONAL HEALTH DATABASES

CO-AUTHORS: PETER R. RIJNBEEK, JENNA M. REPS

Aniek Markus
PhD Student
Department of Medical Informatics, Erasmus MC
a.markus@erasmusmc.nl



Erasmus MC University Medical Center Rotterdam

Health Data Science

OHDSI Observational Health Data Sciences and Informatics

EHDEN European Health Data & Evidence Network

Motivation

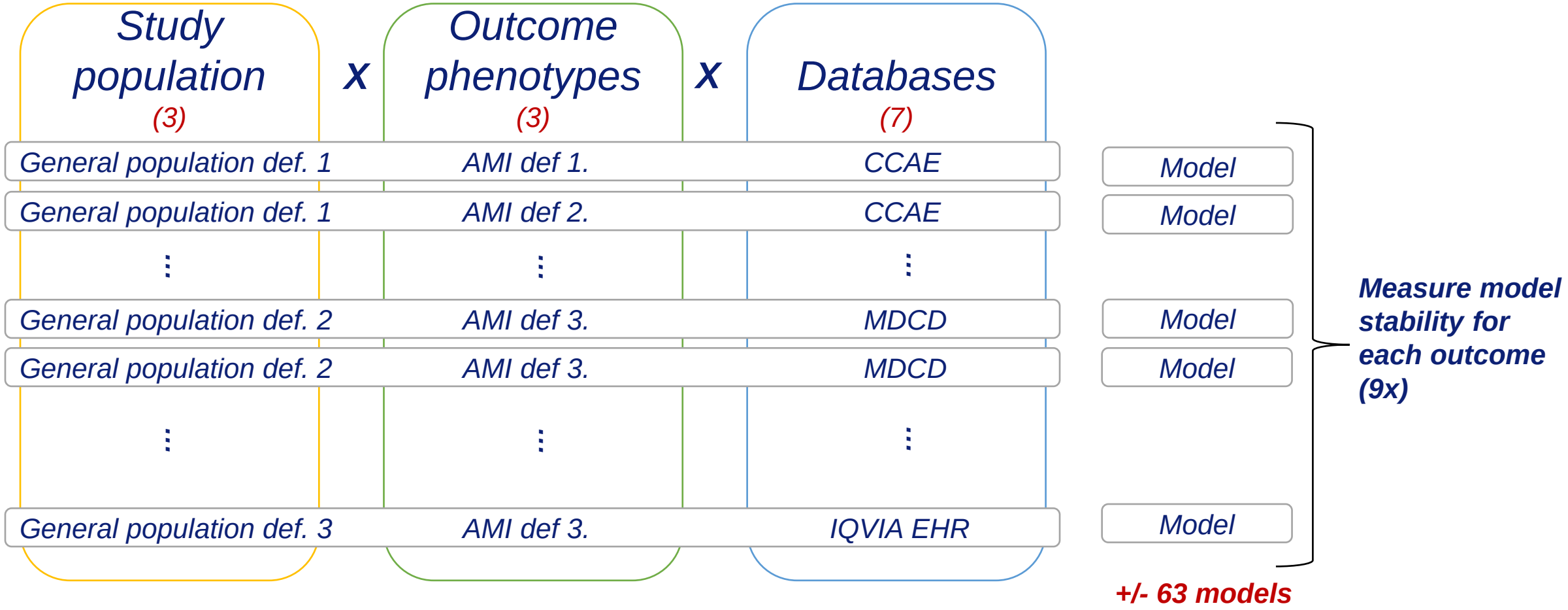
'Variable associated to the outcome'

- Some researchers incorrectly interpret prediction models:
  1. Prediction models do not assess causality
  2. **Also problematic to use models for 'risk factor' detection**

- Why?
  - Procedures, medical conditions, drugs often *co-occur*
  - LASSO logistic regression might ignore some of them

Erasmus MC

Study design



Erasmus MC

Measuring model stability

Example model:

$$probability(acute\ myocardial\ infarction\ within\ 1\ year) = \sigma(+1.33 * age - 0.2\ female + 0.43\ obesity)$$

1. How many variables are selected across models?

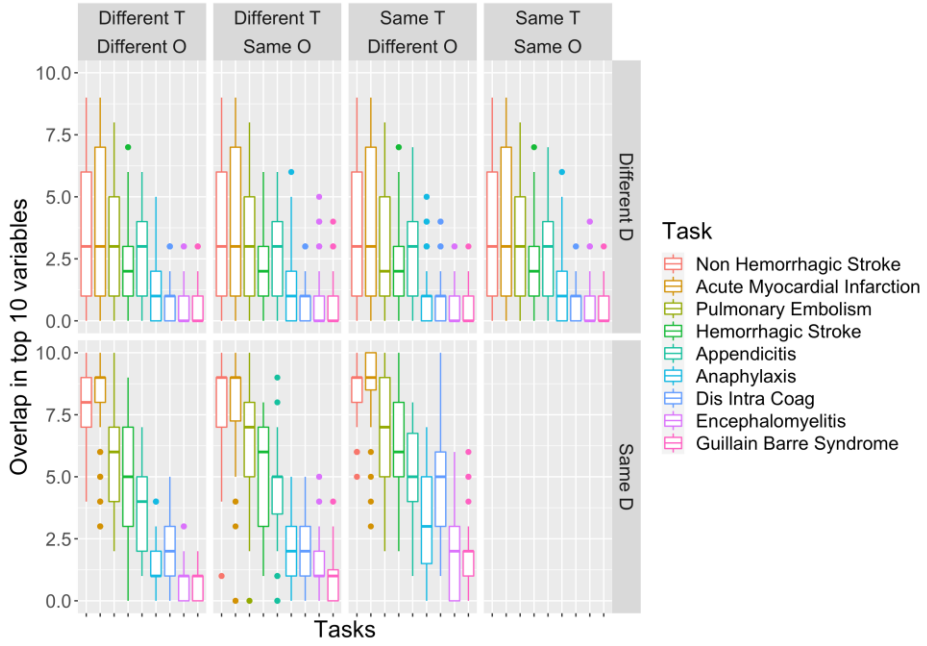
2. Are the same or different variables included across models?

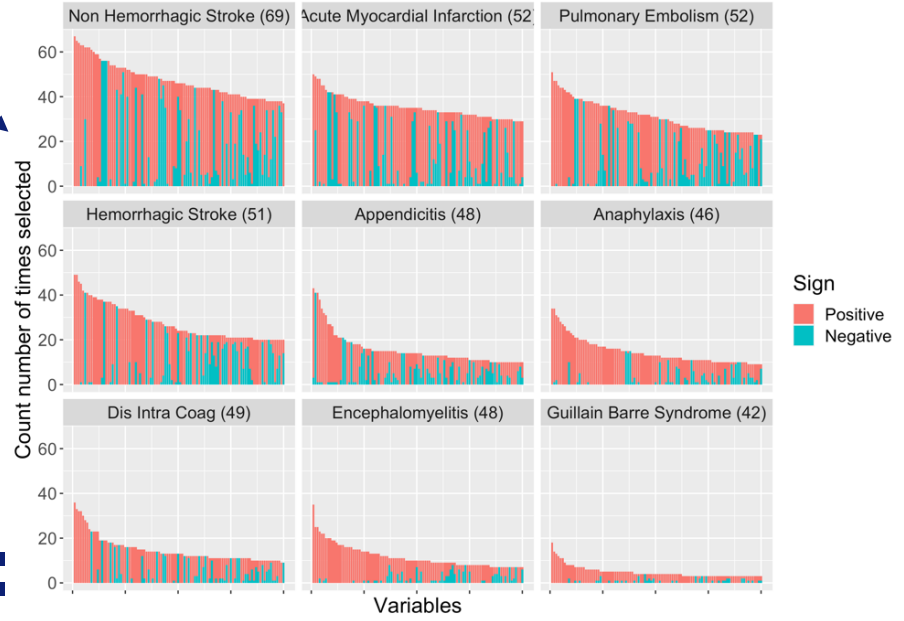
3. Is the direction of the effect (+/-) of variables the same across models?

Erasmus MC

Findings

- Substantial variation in the selected variables
  - Different databases → different ‘risk factors’
- Sign of ‘risk factors’ can differ across models
- For ‘risk factor’ detection:
  - Investigate model robustness
  - Use other techniques (e.g. univariate analysis)





Erasmus MC

Take home message:

Be careful interpreting prediction models as the identified ‘risk factors’ appear to depend on study design choices.

‘Risk factor’ = ‘variable associated to the outcome’

TRAJECTORYVIZ: INTERACTIVE VISUALIZATION OF TREATMENT TRAJECTORIES

Maarja Pajusalu, Institute of Computer Science, University of Tartu

TrajectoryViz: Interactive visualization of treatment trajectories

PRESENTER: Maarja Pajusalu
maarja.pajusalu@ut.ee

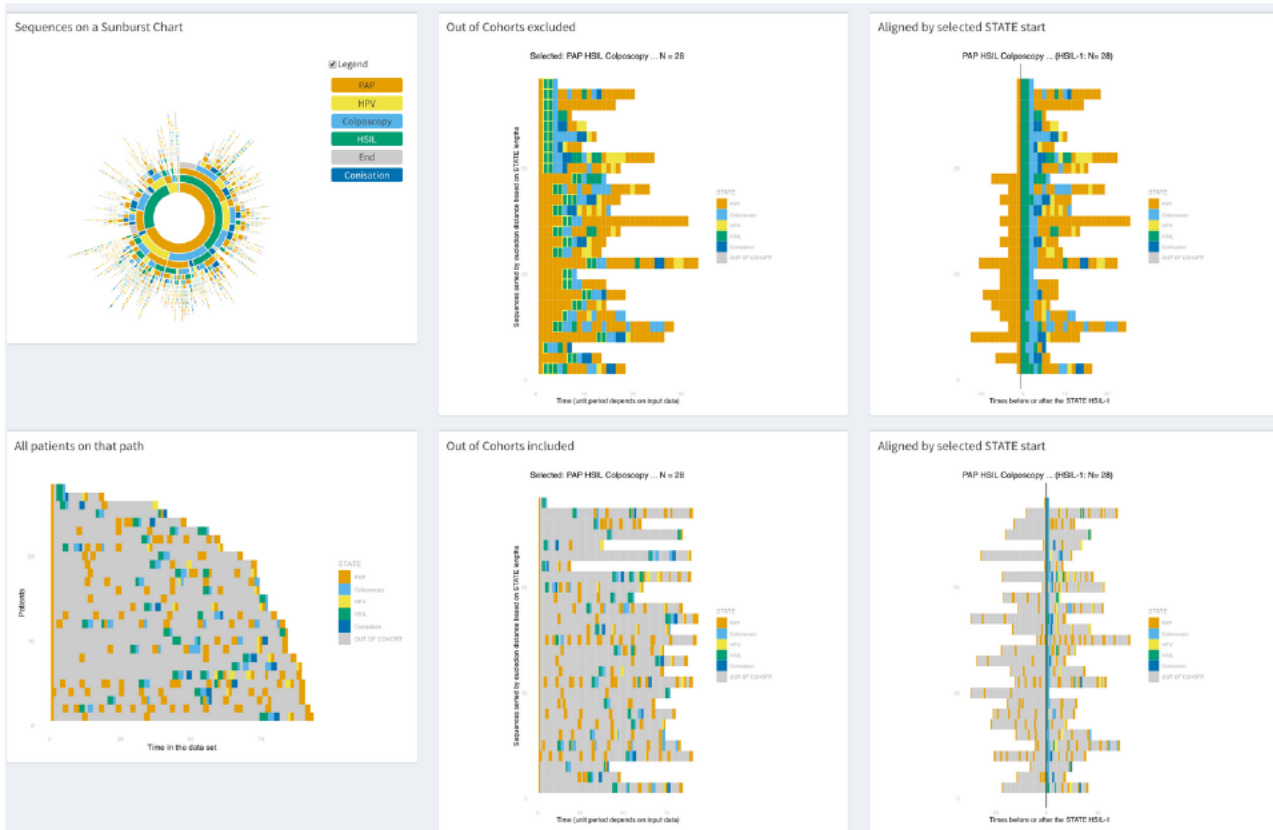
Co-authors:
Marek Oja, Sirli Tamm, Markus Haug, Raivo Kolde
Institute of Computer Science
University of Tartu, Estonia

Create cohorts in **ATLAS**

Create discrete sequences with R package **Cohort2Trajectory**

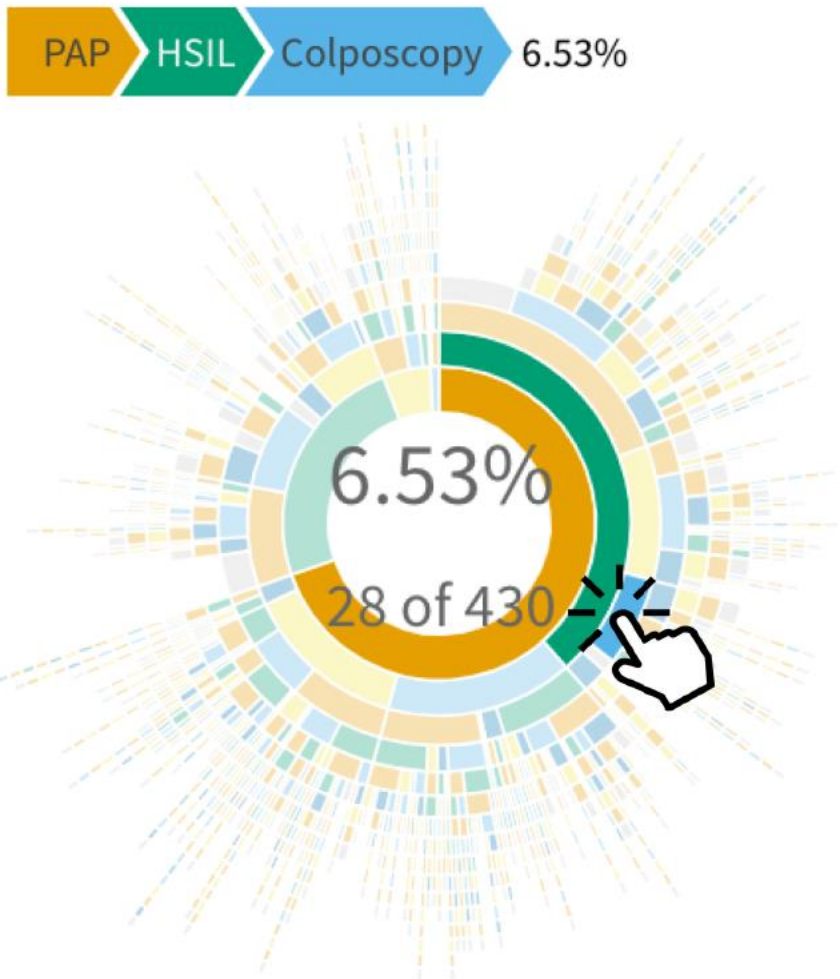
Create interactive visualisations with R package **TrajectoryViz**

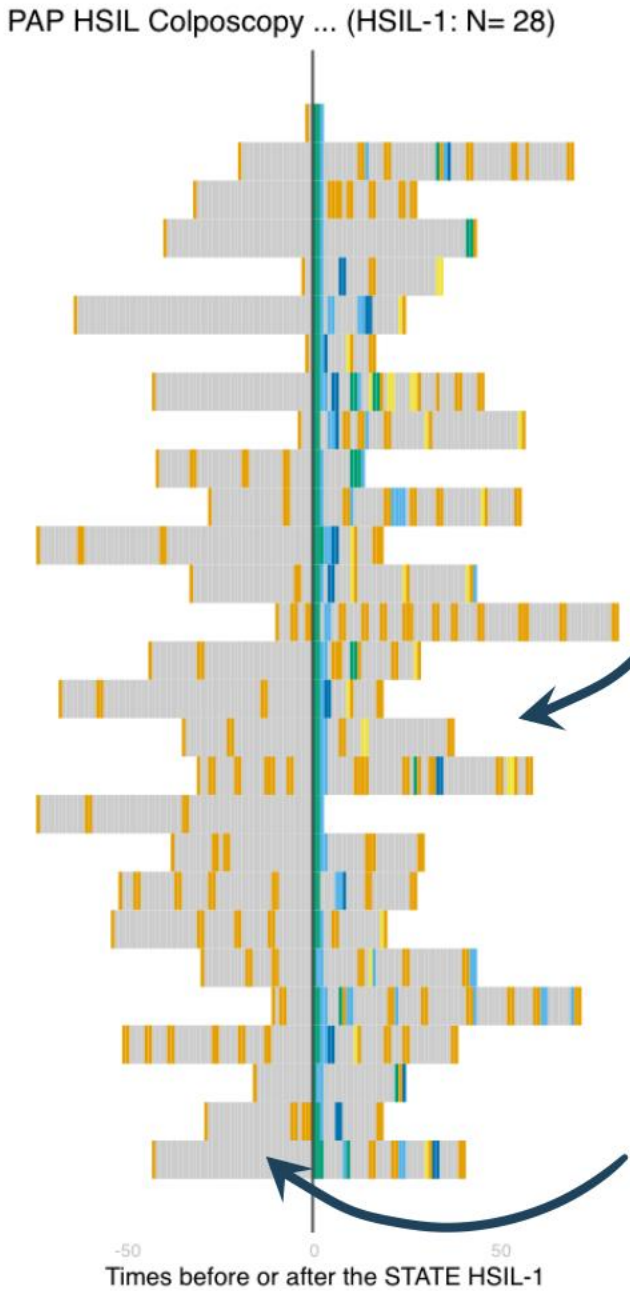
Explore and find **answers** to Your **questions**



Use Case 1:
Cervical Cancer patients

Sequences on a Sunburst Chart



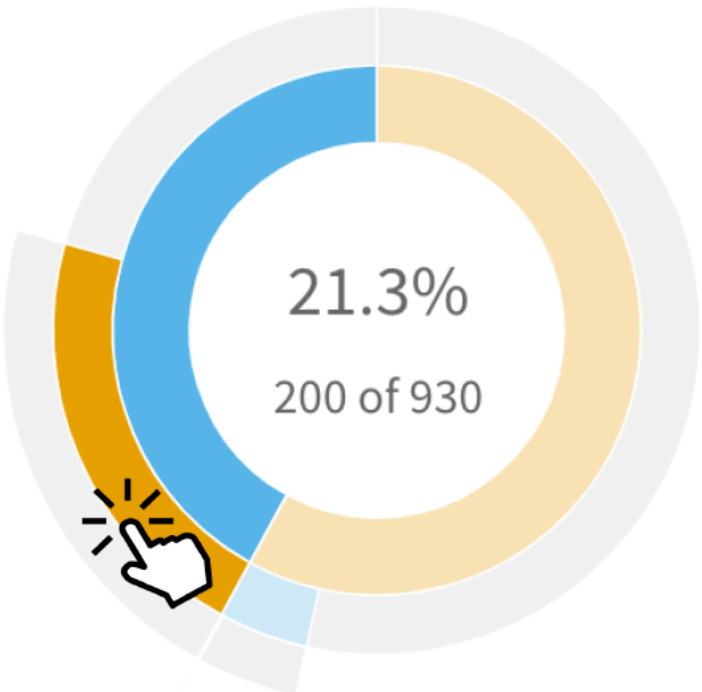


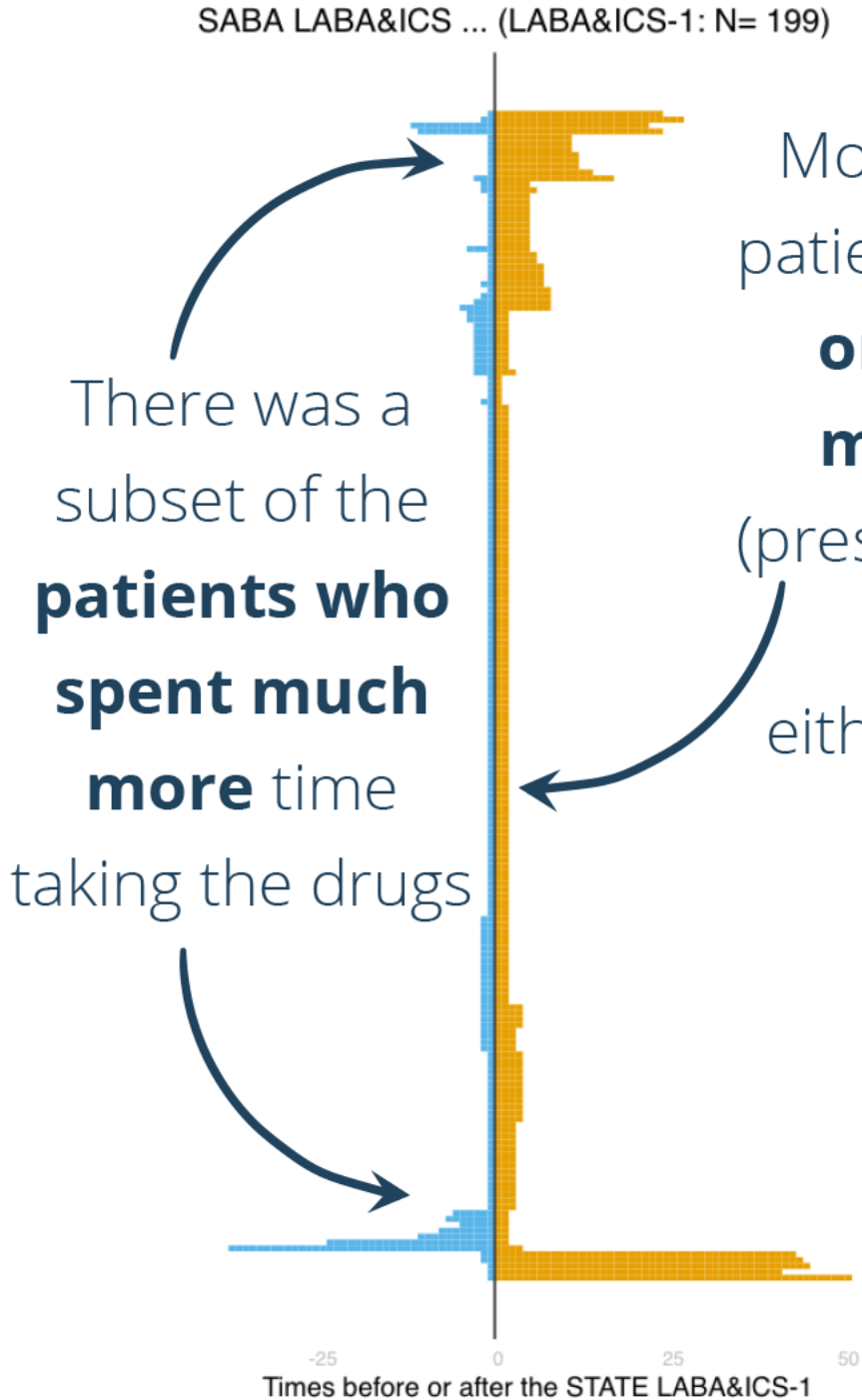
Did the patient follow suggested **pattern** for PAP tests **after selected** diagnosis?

Did the patient follow suggested **pattern** for PAP tests **before**?

Use Case 2:
Asthma patients

Sequences on a Sunburst Chart





There was a subset of the **patients who spent much more** time taking the drugs

Most of the patients spent **only 1-2 months** (prescriptions) on either of the drugs

UNIVERSITY OF TARTU

Thank You!

unitartu

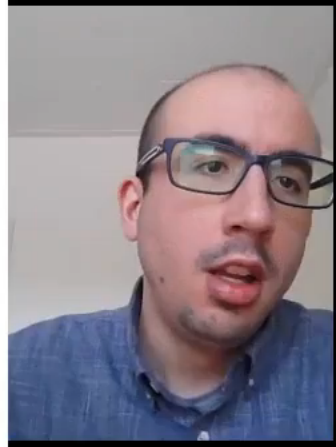
tartuuniversity

ASSESSING TREATMENT EFFECT HETEROGENEITY USING THE RISKSTRATIFIEDESTIMATION R-PACKAGE

Alexandros Rekkas, Erasmus MC

Assessing treatment effect heterogeneity using the RiskStratifiedEstimation R-package

Alexandros Rekkas



Erasmus MC
University Medical Center Rotterdam

Health Data Science

DEFINING THE VALID ANALYTIC SPACE FOR QUANTITATIVE BIAS ANALYSIS IN PHARMACOEPIDEMIOLOGY

James Weaver, Janssen R&D

OHDSI EUROPE
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Defining the analytic space for valid QBA

OHDSI Europe Symposium
24 June 2022

James Weaver [jweave17@its.jnj.com]
Global Epidemiology, Janssen R&D
NDORMS, University of Oxford

Background

- Bias from outcome misclassification often ignored in casual estimation
- QBA: a proposed solution

Objective

- Evaluate QBA performance across large set of plausible scenarios

Methods

- Applied QBA across incidence x effect size x measurement error analytic space

Results

- Small specificity change has large impact on effect estimate
- Limited impact of sensitivity at low incidence

Discussion

- QBA produces implausible or invalid estimates in many common comparative effect estimation scenarios

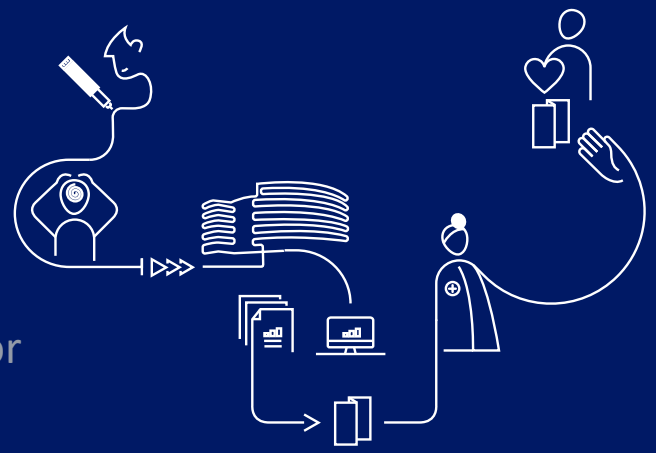
NDORMS
NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSCULOSKELETAL SCIENCES
UNIVERSITY OF OXFORD

janssen
PHARMACEUTICAL COMPANIES OF Johnson & Johnson

A PILOT STUDY TO EVALUATE THE FEASIBILITY OF USING OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS ANALYTICS TOOLS FOR SUPPORTING THE VALIDATION OF SAFETY SIGNALS

Ceyda Pekmez Kristiansen, Novo Nordisk

A pilot study to evaluate the feasibility of using OHDSI analytical tools for supporting safety surveillance

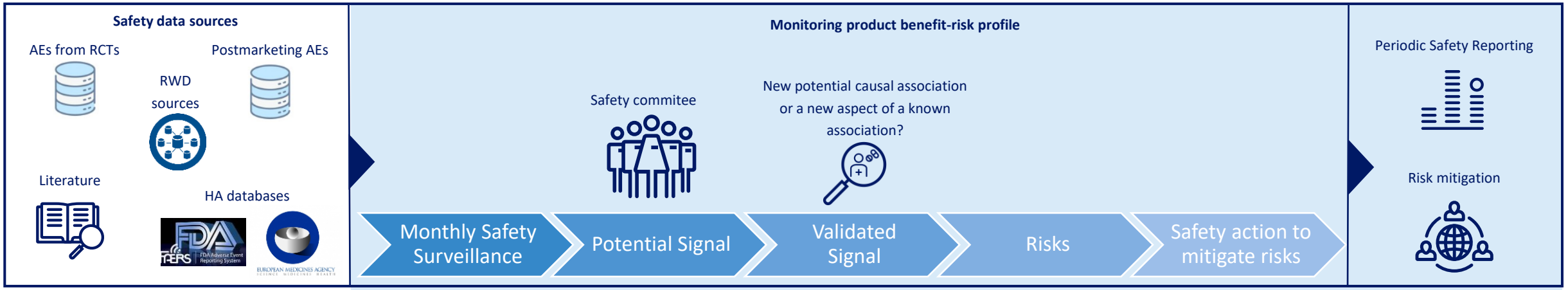


Ceyda Tugba Pekmez Kristiansen[1]; Lasse Christensen[1]; Michael Stellfeld[1]; Atheline-Major Pedersen[1]; Ditte Mølgaard-Nielsen[1]; Mark White[1]; Peter Jelnes[1]

cypk@novonordisk.com

[1]Novo Nordisk A/S, Vandtårnsvej 114, 2860 Søborg, Denmark

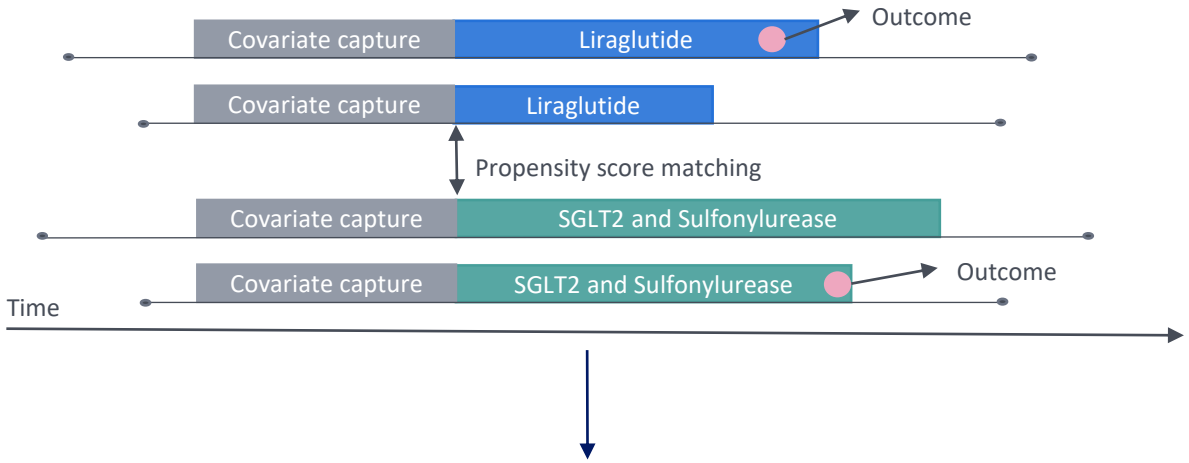
24.06.2022

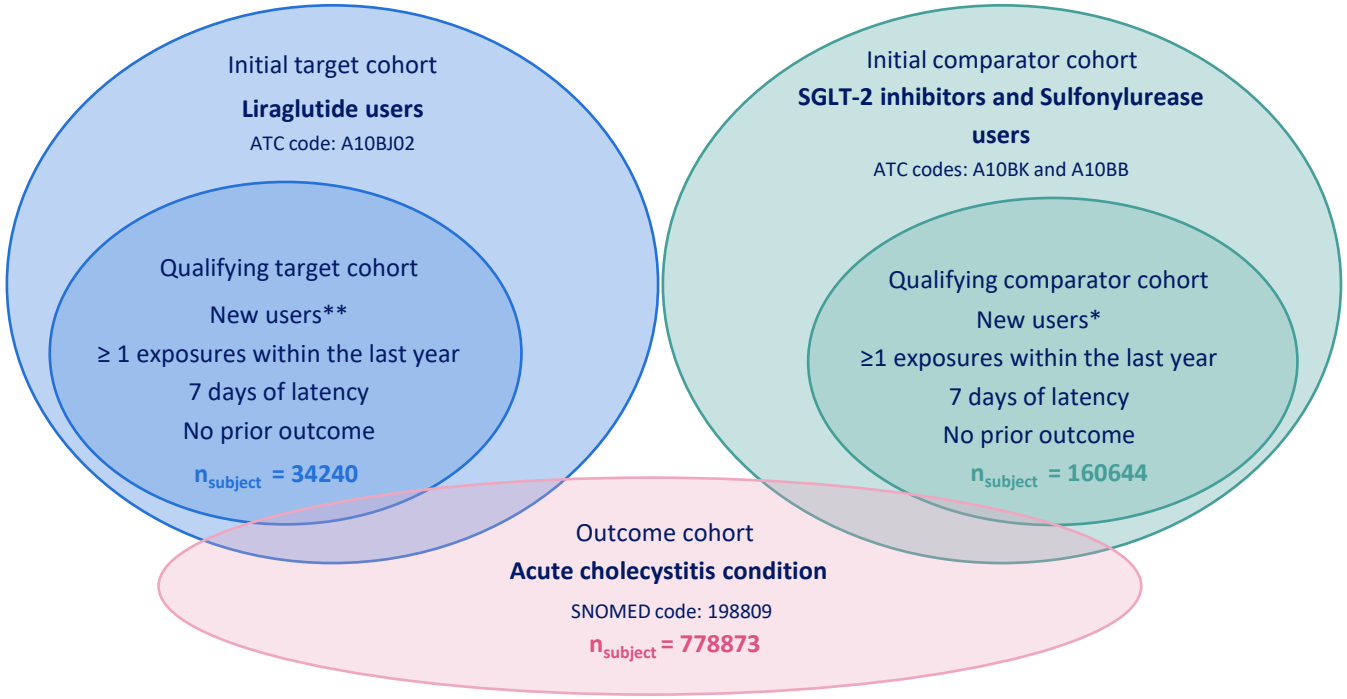


Real world data (RWD) for supporting safety surveillance

- RWD supports the safety signal validation process, especially when the evidence from traditional safety data sources is scarce
- Gallbladder- or biliary tract-related events (including acute cholecystitis) are known risks for Victoza® (liraglutide) and Saxenda® (liraglutide)
- A known risk for liraglutide was chosen for the pilot study to evaluate the feasibility of implementing population level effect estimation into the safety surveillance process

Cohort definition – Truven Marketscan CCAE* 2020





HADES Health Analytics Data-to-Evidence Suite

CohortMethod
New-user cohort studies using large-scale regression for propensity and outcome models.
Learn more...

SqlRender
Generate SQL on the fly for the various SQL dialects.
Learn more...

IBM® Watson Health™. Commercial Claims and Encounters Database and Medicare Supplemental and Coordination of Benefits Database IBM Marketscan Research Databases User Guide. Certain data used in this study were supplied by International Business Machines

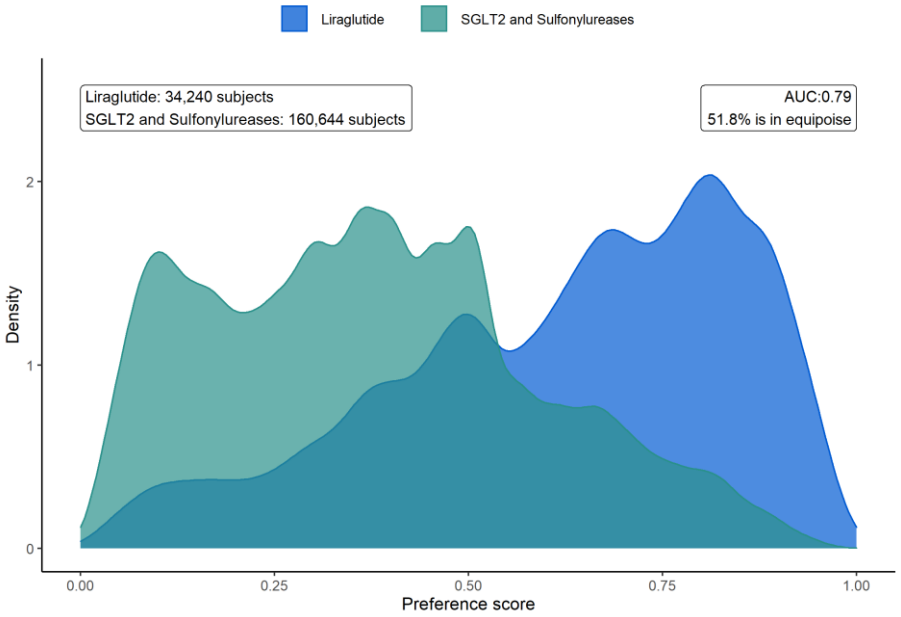
Corporation. Any analysis, interpretation, or conclusion based on these data is solely that of the authors and not International Business Machines Corporation.

**Commercial Claims and Encounters*

***First time use of drug within last 1 years of observation period*

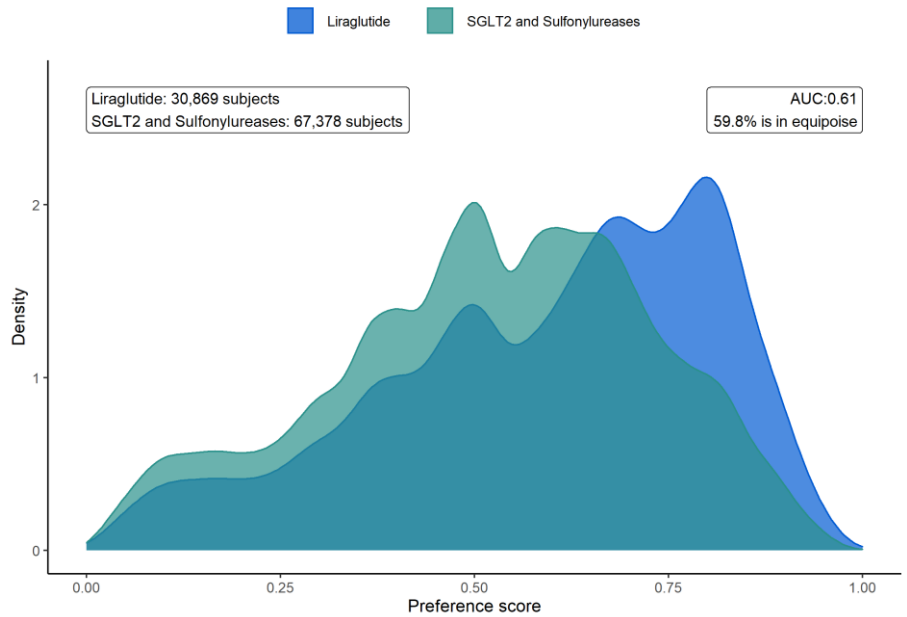
Propensity score matching

Before matching



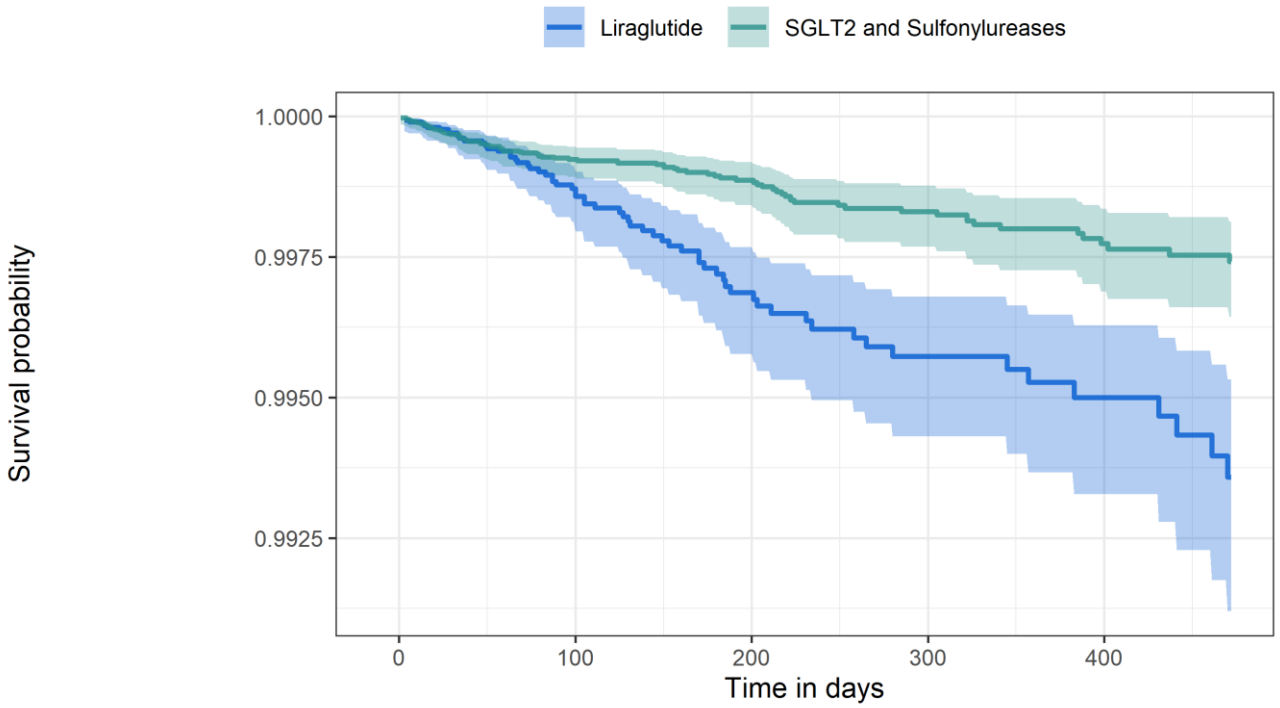
Age
Gender
BMI, Obesity
Parity
Diabetic complications (Retinopathy, Nephropathy, CV diseases)

1:3 propensity score matching



	Cases (n)	Prevalence (per 1000)
Target	71	2.30
Comparator	83	1.23
Total	154	1.57

Population level risk estimation



Number at risk					
Liraglutide	30,869	14,771	8,383	5,327	3,420
SGLT2 and Sulfonylureases	67,378	32,826	20,374	13,380	9,227

Cox-proportional Hazard ratio: 2.26
CI: [1.70 – 3.03]

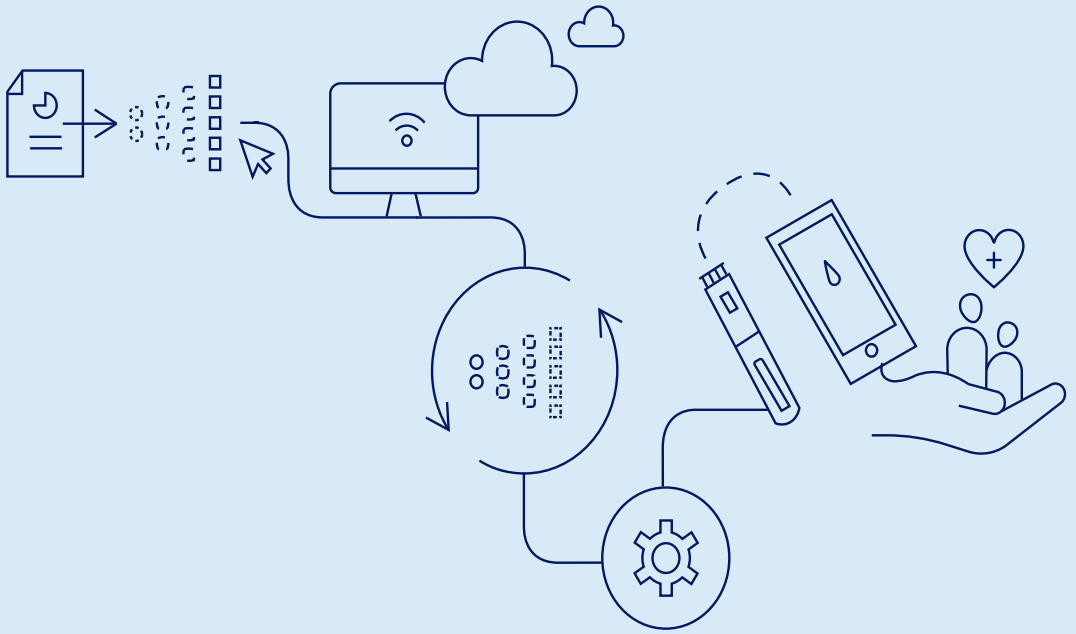
Minimum detectable relative risk:
1.62 ± 0.17 (SE)

Conclusion – Next steps

- OHDSI analytics tools have promising potential for utilising RWD sources to support the validation of safety signals

- The result supports a known risk of acute cholecystitis for liraglutide on a RWD source

- A new test case for another therapeutic area
  - Negative outcome controls
  - Data driven selection of covariates



Rapid Fire Presentations of Collaborators

THANK YOU!

EHDEN

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

- Findable, standardised data at scale through the EHDEN Database Catalogue -

24 June 2022

Julia Kurps, The Hyve

imi innovative medicines initiative

efpia

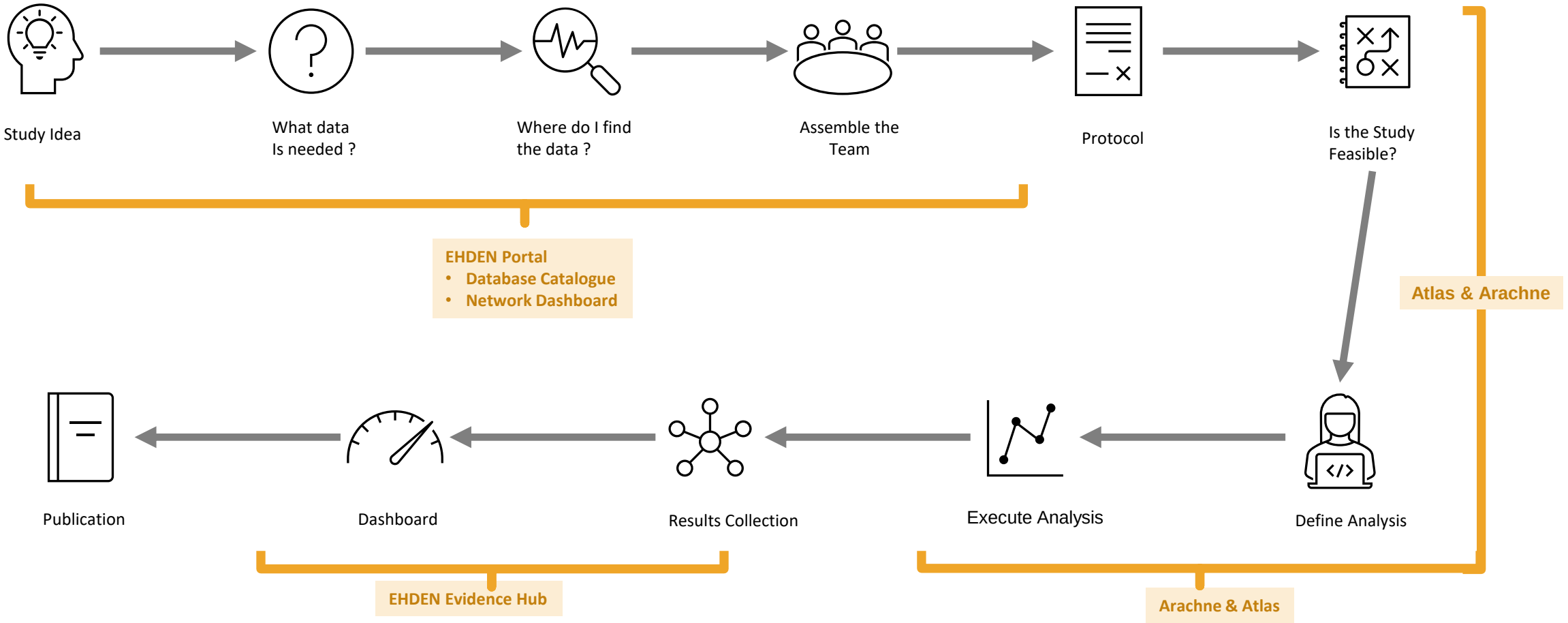
Europe: an Ocean of Data & a Desert For Analysis

OPEN UP – KEY RESEARCH TOOL IN EHDEN

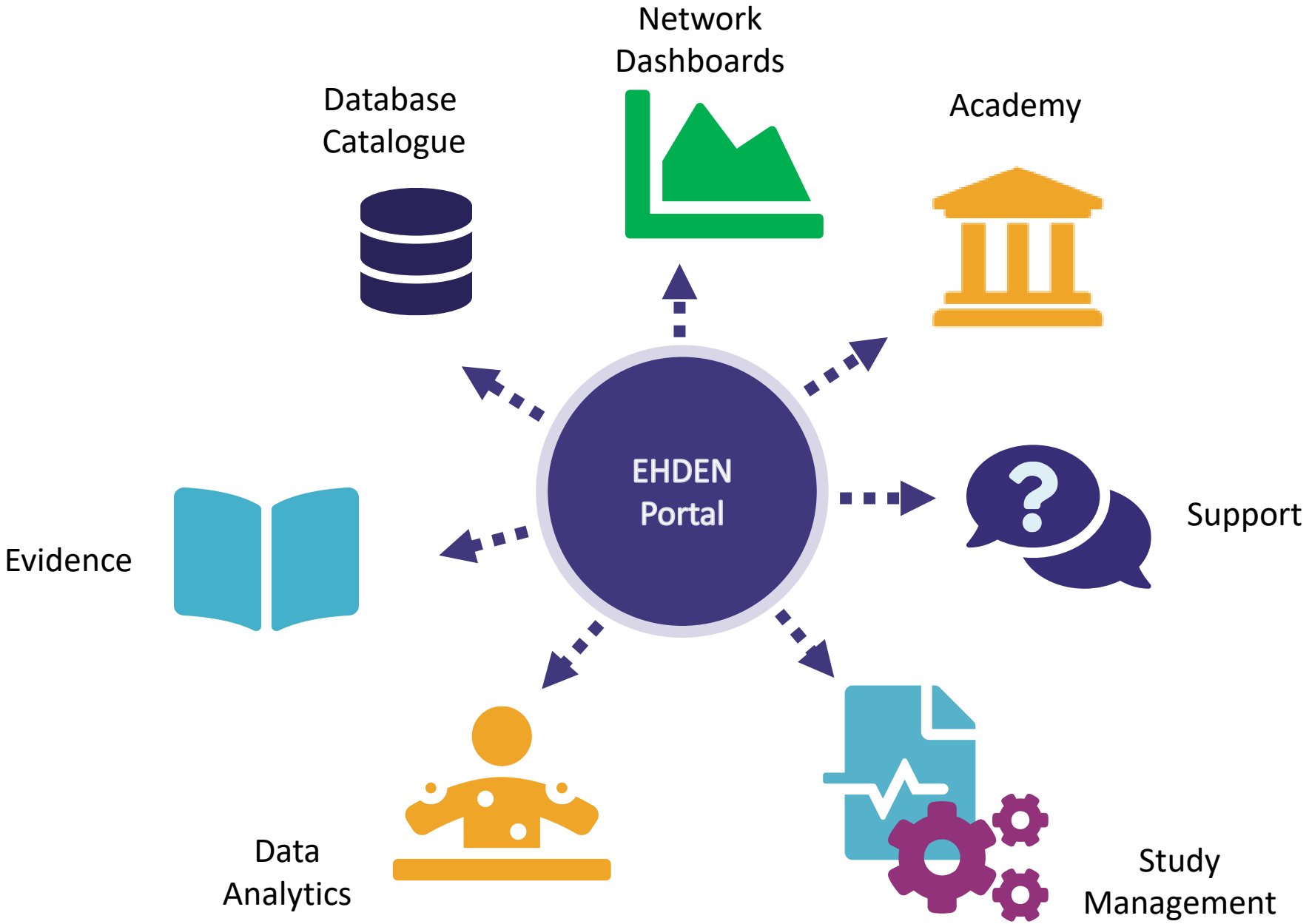
Identify **relevant data** to answer your **research question**

What data is available?
From how many patients?
Can I analyse this data for my research?
…

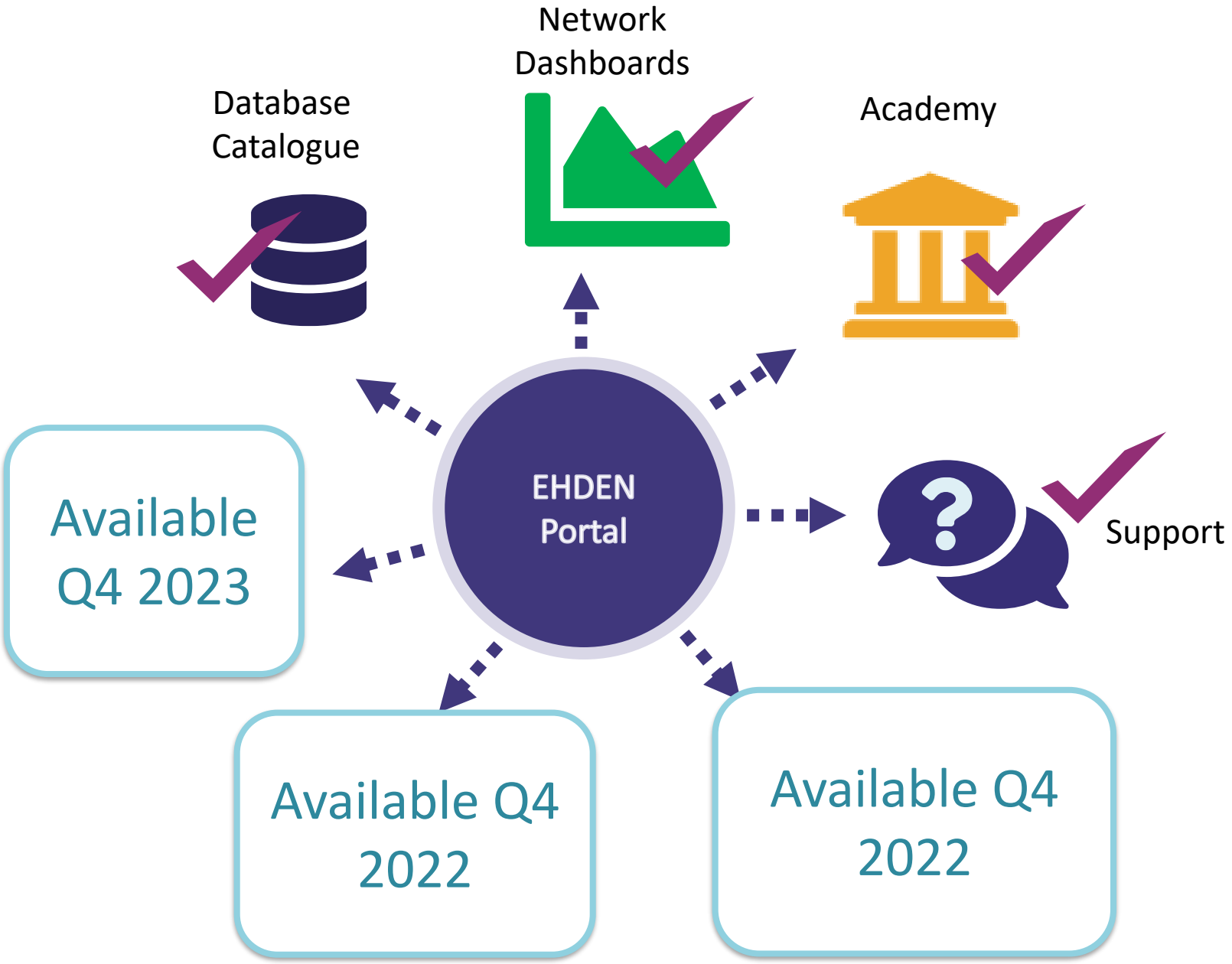
ROLE OF THE EHDEN PORTAL IN THE STUDY WORKFLOW



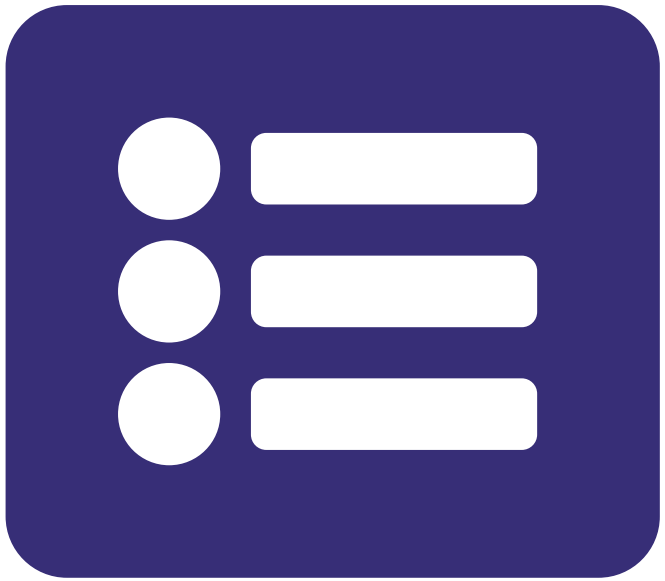
THE VISION OF THE EHDEN PORTAL – ONE-STOP-SHOP



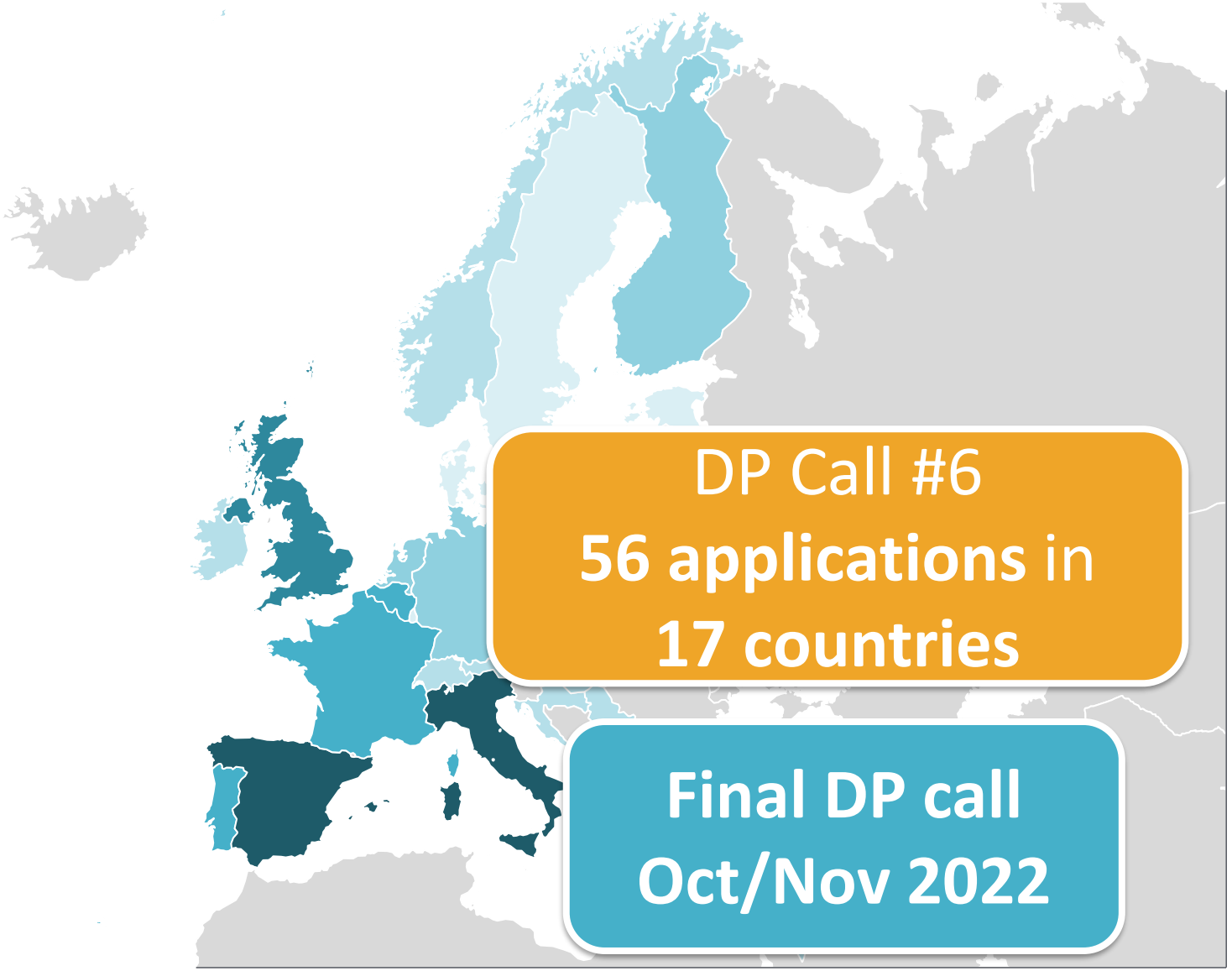
EHDEN Portal Release Today



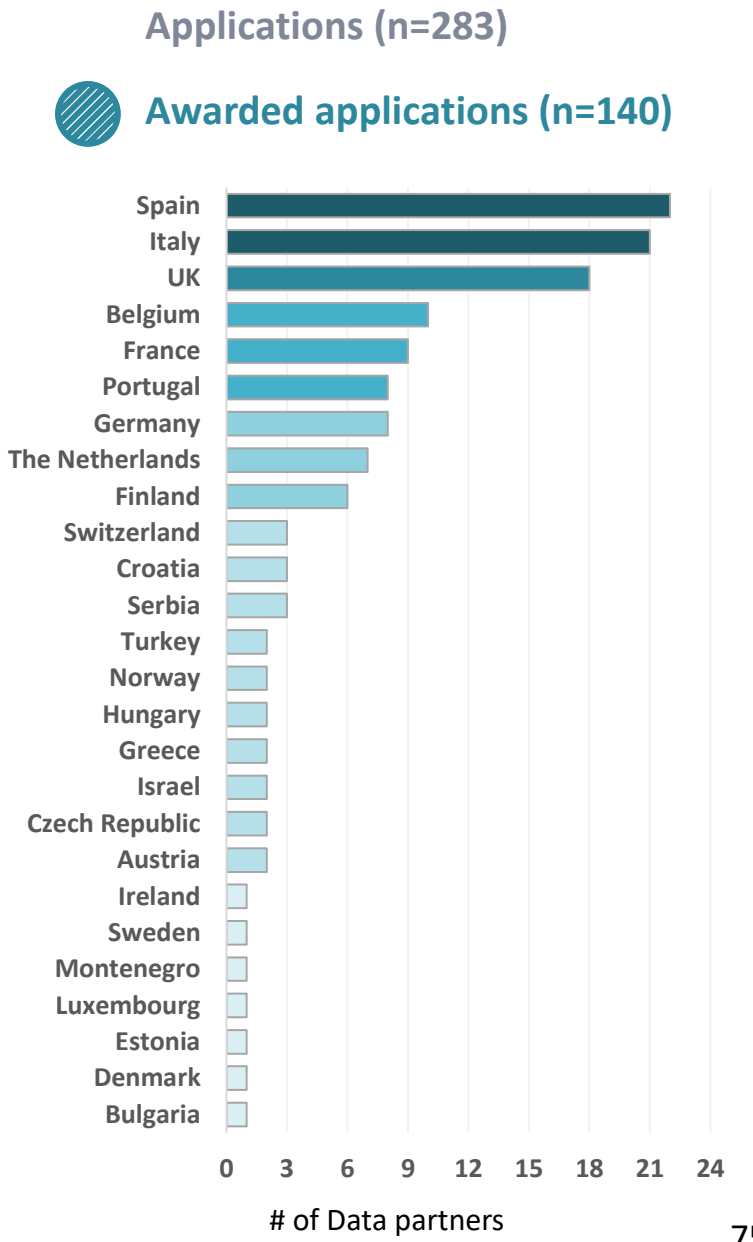
DATABASE CATALOGUE



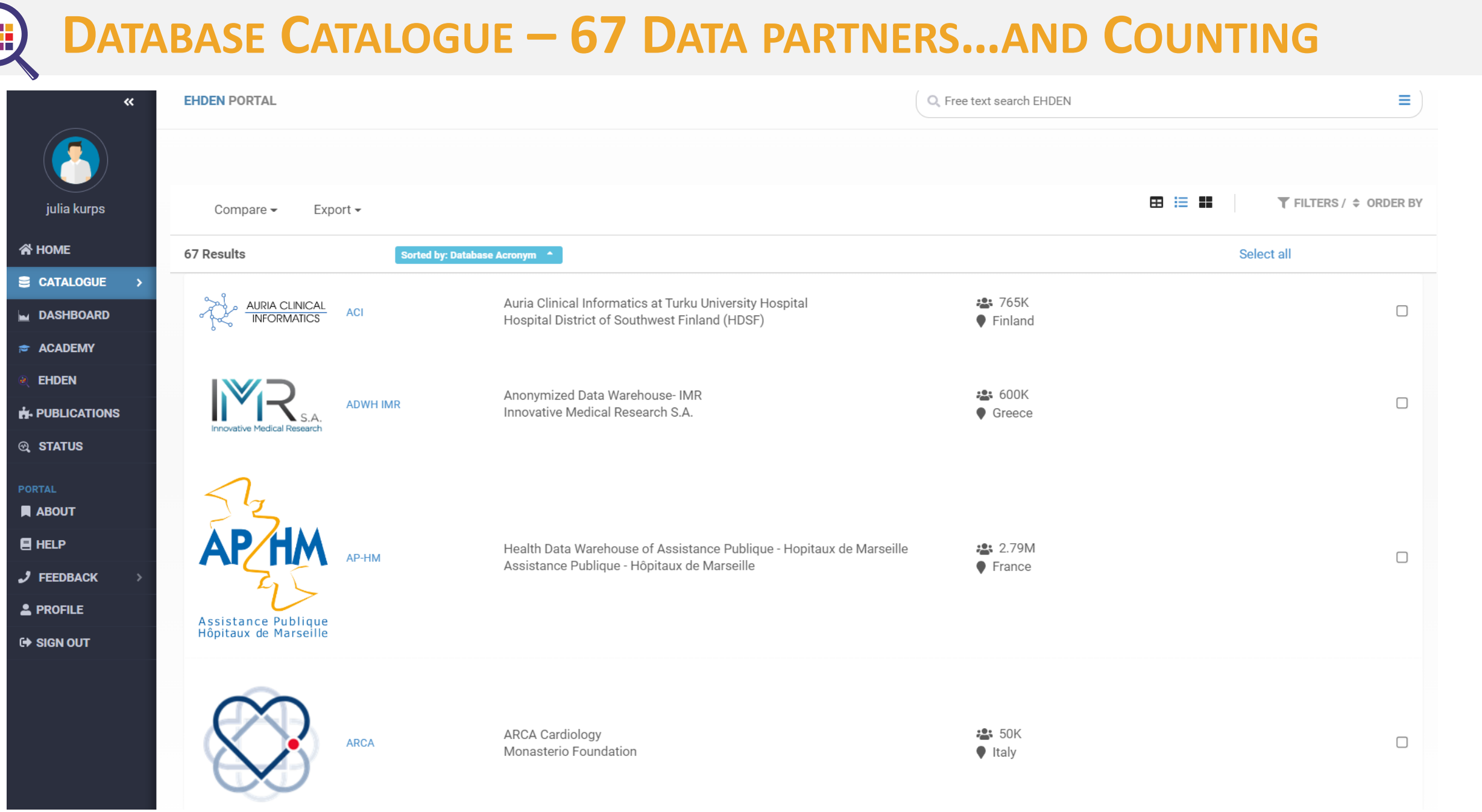
Geographical Spread Data Partners



Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)



DATABASE CATALOGUE – 67 DATA PARTNERS…AND COUNTING



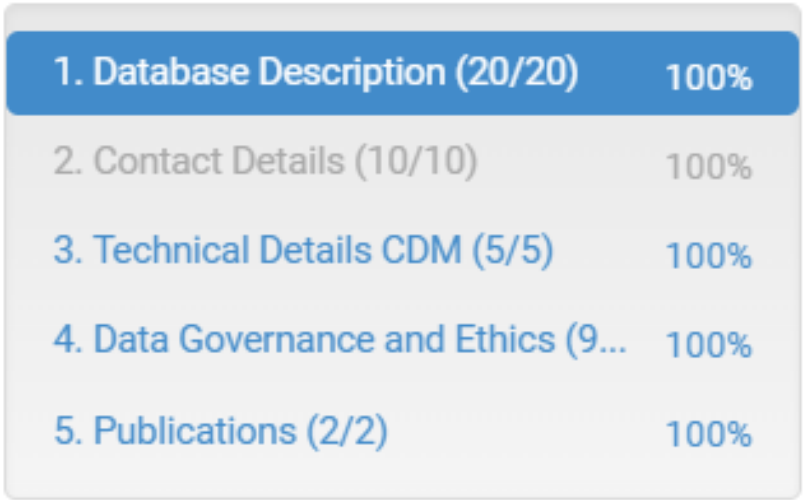
Components Of the Database catalogue

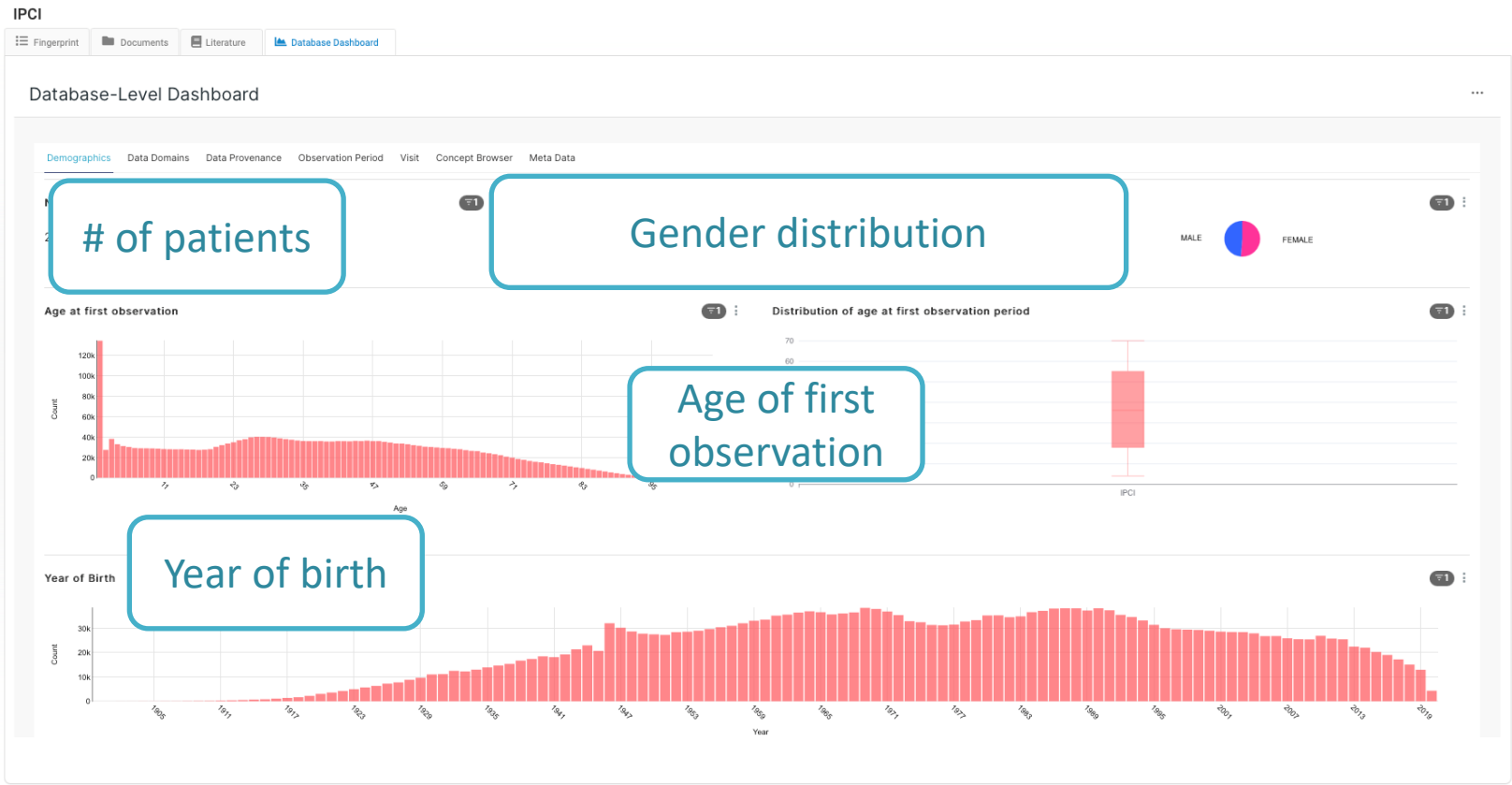
Database Fingerprint

- meta data -

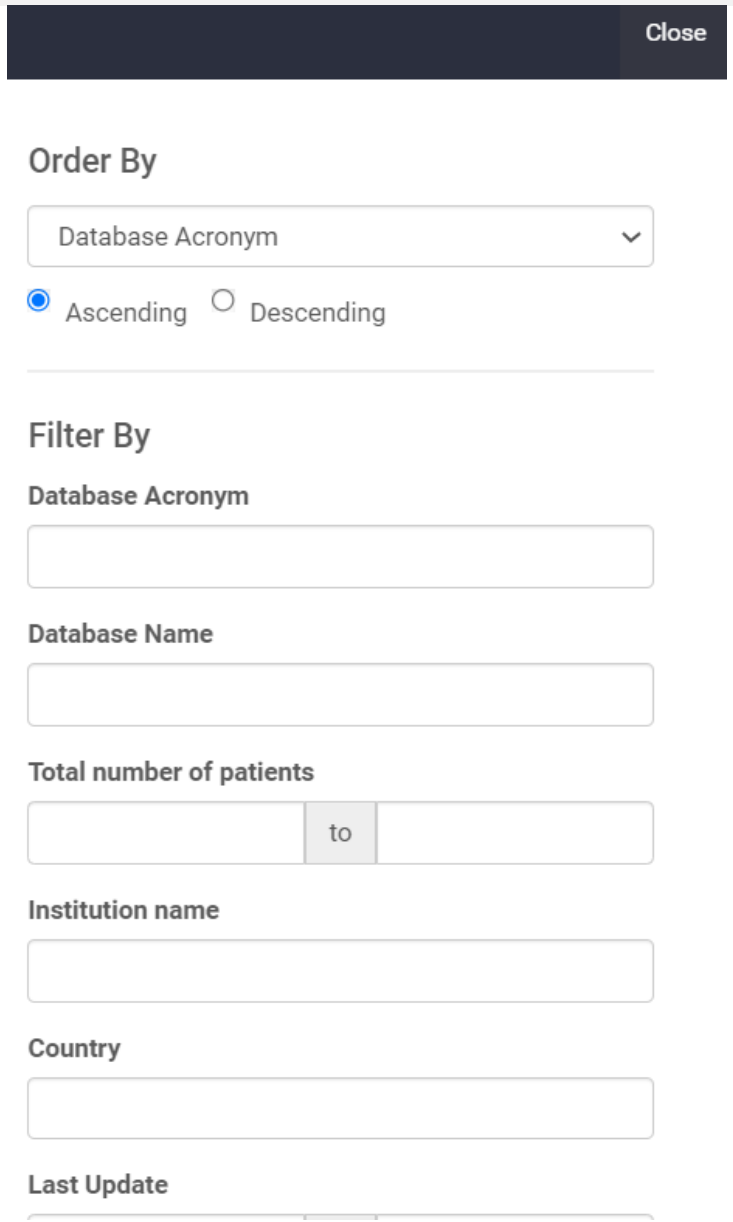
Database Dashboards

- based on CDM data -





CUSTOMISE SEARCH



All databases in **France**?

Which databases have occurrences of **myocardial infarction**?

Which databases have **GP data** in combination with **hospital data**?

In which countries do we have **up to date data** on **COVID-19 vaccinations**?

NETWORK DASHBOARD

NETWORK DASHBOARDS - OVERVIEW

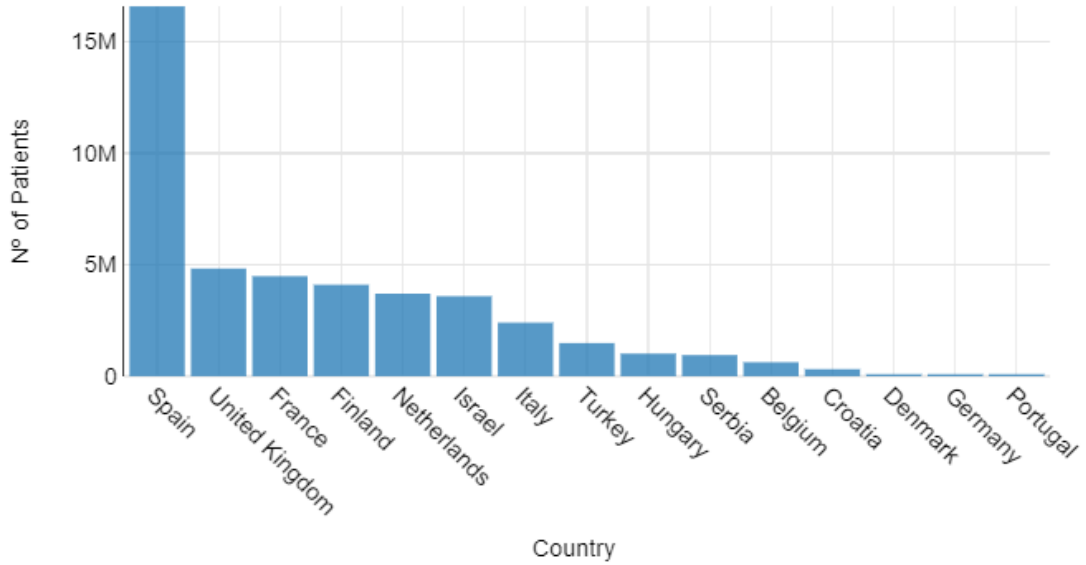
Countries	Data Sources	Patients
15	35	44.2M

Datasource Types

Hospital | Hospital + Lab Results | Hosp | 1/4

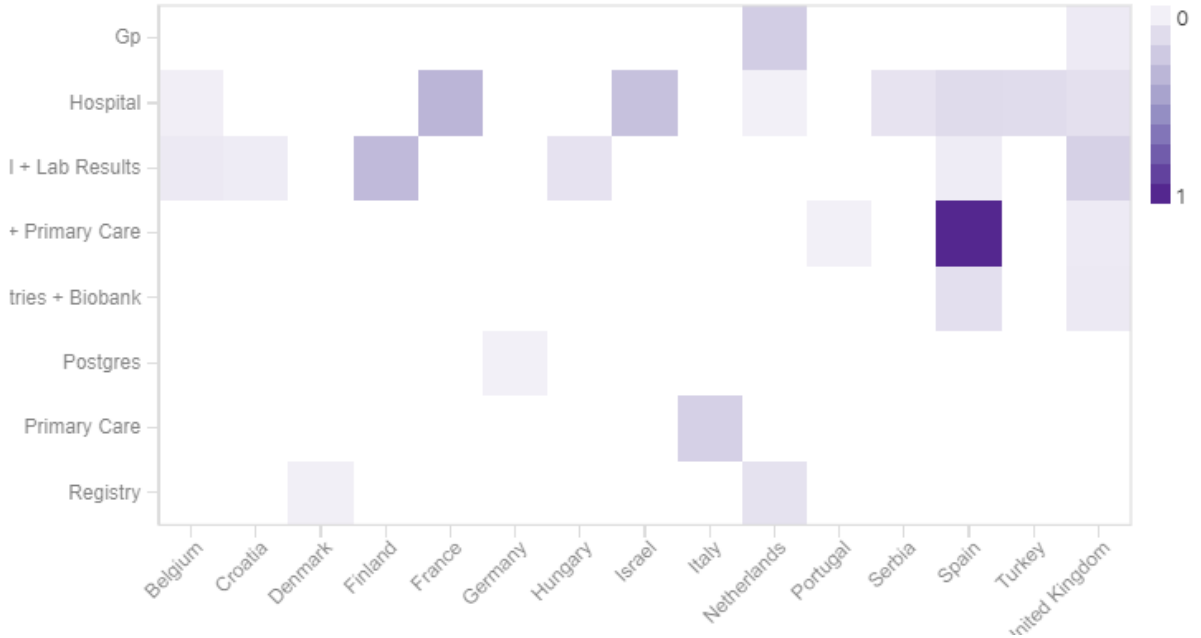
Total number of patients per country

Patients by Country

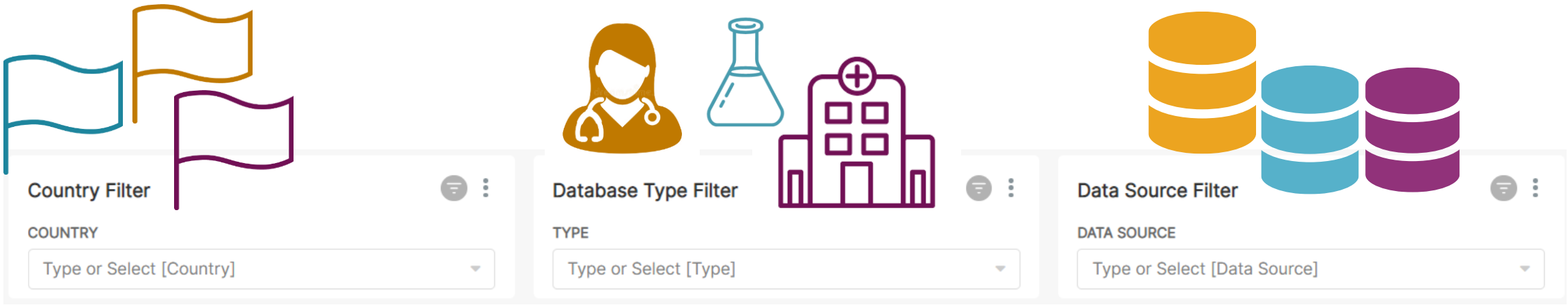


Number of patients (color) per database type for each country

Database Types per Country



Filter to Find Relevant Data



Use case examples

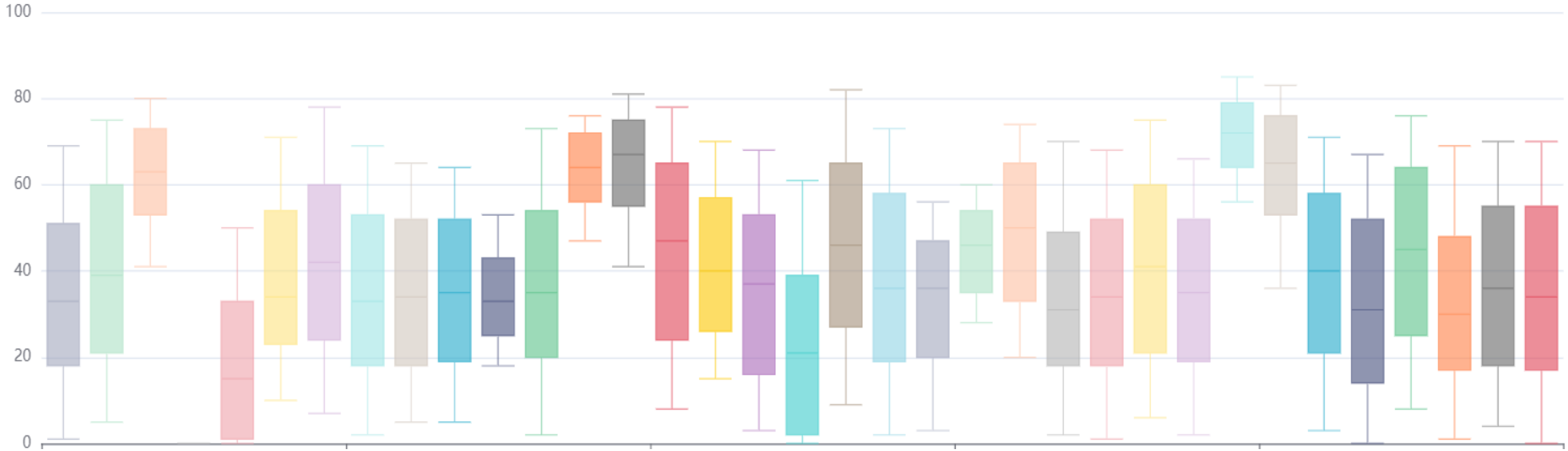
- Geographical spread
  Identify all EHDEN Data Partners from *France*

- Range of Source data systems
  Identify Data Partners that can provide *GP data in combination with hospital data*

NETWORK DASHBOARDS – BROWSE – EXPLORE - DISCOVER

Overview | Demographics | Data Domains | Data Provenance | Observation Period | Visit | Concept Browser | About

Distribution of age at first observation period

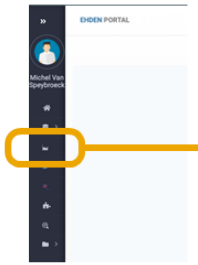


GETTING STARTED

EHDEN Network Dashboard at a Glance

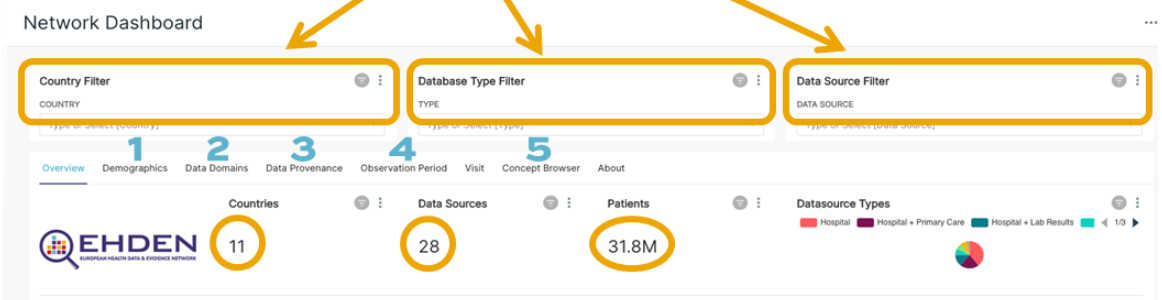
Getting Access to the Network Dashboard

- Enter credentials on the EHDEN portal (https://portal.ehden.eu)
- Select the **Network Dashboard** icon on the left panel



Total Number of datasources and patients in the network

- Use the country / database type / data source filter to restrict the selection



1 Understanding gender or year of birth or age coverage

- Go to the tab '**Demographics**'
- Gender distribution is shown in the top pie chart
- Age at first observation is presented in tabular format , bar chart and boxplot
- Year of birth is presented as bar chart.

2 Understanding what data are available for one or more data sources

- Go to the tab '**Data Domains**'
- The '**Average number of records per person**' indicates how many records per entity type are available. Use in conjunction with the above filters to select on country/database type or data source
- The pie chart gives the total number of records per entity type

3 What is the source of the data?

- Go to the tab '**Data Provenance**'
- The pivot table 'Visit Type Pivot' gives a count per entity type of the respective concept types e.g. EHR problem list entry, primary or secondary condition, hospitalization cost record etc.

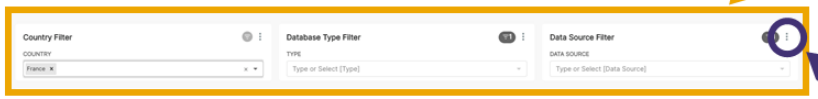
4 What's the available observation period?

- Go to the tab '**Observation Period**'
- Check the second graph – cumulative observation time
- Click in the legend on a single item to select / deselect
- Double click to select a single data source

5 Which data sources do contain a particular concept ?

Go to the tab '**Concept Browser**'

- In the concept type, start typing part of the text string
- Select the concept that you want to search for
- Hit APPLY
- In the pivot table you can see approx. total number of records and descendant number of records. Further down, you see a graphical representation



Nice-to-know

1. Knowing the last update per datasource

- On the landing page of the network dashboard
- Scroll to the bottom – **meta data view**
- Cutoff date from the source data: source_release_date
- CDM date: cdm_release_date

2. Filtering results and other tricks

- **Option A**: top filters for selection on country / database type or data source
- **Option B**: When a legend is displayed, individual items can be selected / unselected or double click – select a single item from another list

3. Maxime your graph

- When a graph is not readable: click on the **icon** and select '**maximise graph**' from the pop-up

AND THAT'S JUST THE BEGINNING…

More Data Partners will be added

Submission deadline DP Call #6 just closed

Integration with Data Analytics and Study Management capabilities

Data Analytics

Study Management

Implementation of Production Instance of service desk for study requests

WELCOME – WE ARE OPEN

Register your account today at

Portal.EHDEN.eu

After approval
you can start exploring
high quality
real world data from
> 44 M patients
from **15 European countries**

Welcome to the EHDEN Portal

The European Health Data & Evidence Network (EHDEN) project aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care. Its mission is to provide a new paradigm for the discovery and analysis of health data in Europe, by building a large-scale, federated network of data sources standardised to the OMOP common data model.

The EHDEN Portal provides an entry point to the growing list of tools in our ecosystem. **Of these tools, Database Catalogue (including Dashboards) and EHDEN Academy are available for general use, while Arachne, Atlas and the Service Desk are still in development and only available to EHDEN Consortium members.** Click on the icons below for more information. Your feedback is highly appreciated.

Database Catalogue

Provides metadata on the databases in the EHDEN data network

Network Dashboards

Allows to analyse and compare aggregated data from the OMOP CDM databases in the

EHDEN Academy

Our free, online and publicly available learning platform

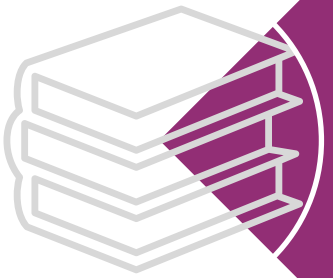
Publications

An overview of all deliverables and publications of EHDEN

WP4 – Building The Infrastructure

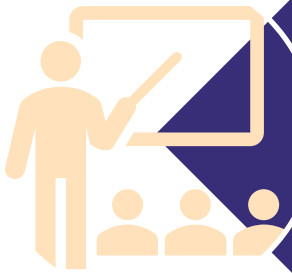
Mapping

- White Rabbit / Rabbit in a Hat
- Usagi
- Data Quality Dashboard



Study Management and Execution

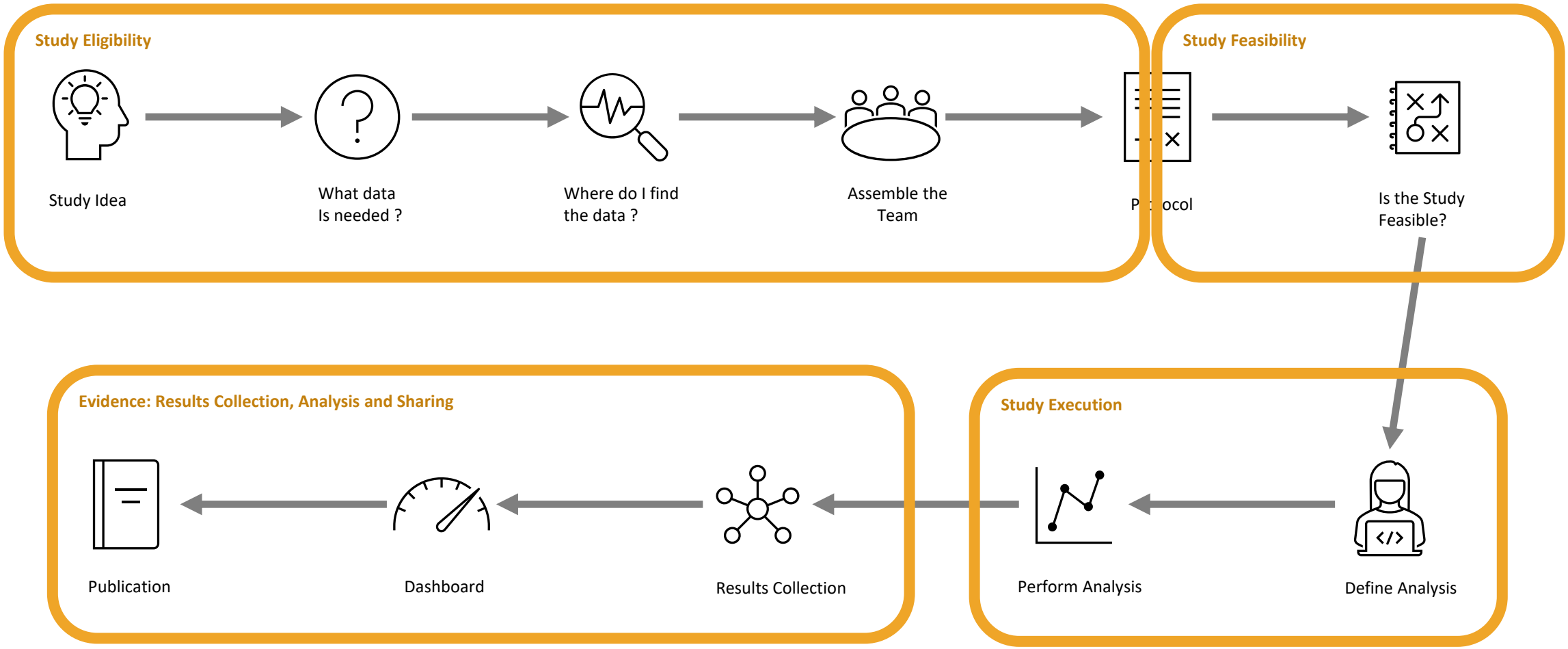
- EHDEN Portal
- Database Catalogue
- Arachne
- Atlas
- EHDEN Evidence Hub



Auxilliary

- EHDEN Academy
- Virtual Training Environment
- Website
- Forum

Study Workflow in EHDEN – Technical Perspective



Lunch, Collaborator Showcase, and Early Investigator meetings

The lunch is made possible with the help of Synapse

SYNAPSE
RESEARCH MANAGEMENT PARTNERS

The collaborator showcase is made possible with the help of Promptly

PROMPTLY
Health Analytics

Early Investigators mentor meetings in the Queen's Lounge