



What type of Evidence do we generate?

Peter Rijnbeek

Associate Professor Health Data Science
Erasmus MC, Rotterdam, The Netherlands



Classification by Scientific Tasks

Instead of defining health data science by its technical activities, e.g. management, processing, analysis, visualization, we should define the field by its scientific tasks:

1. Description -> Clinical Characterisation: **What happened to them?**
2. Prediction (inference) -> Patient-Level Prediction: **What will happen to me?**
3. Counterfactual Prediction (causal inference) -> Population-Level Effect Estimation: **What are the causal effects?**





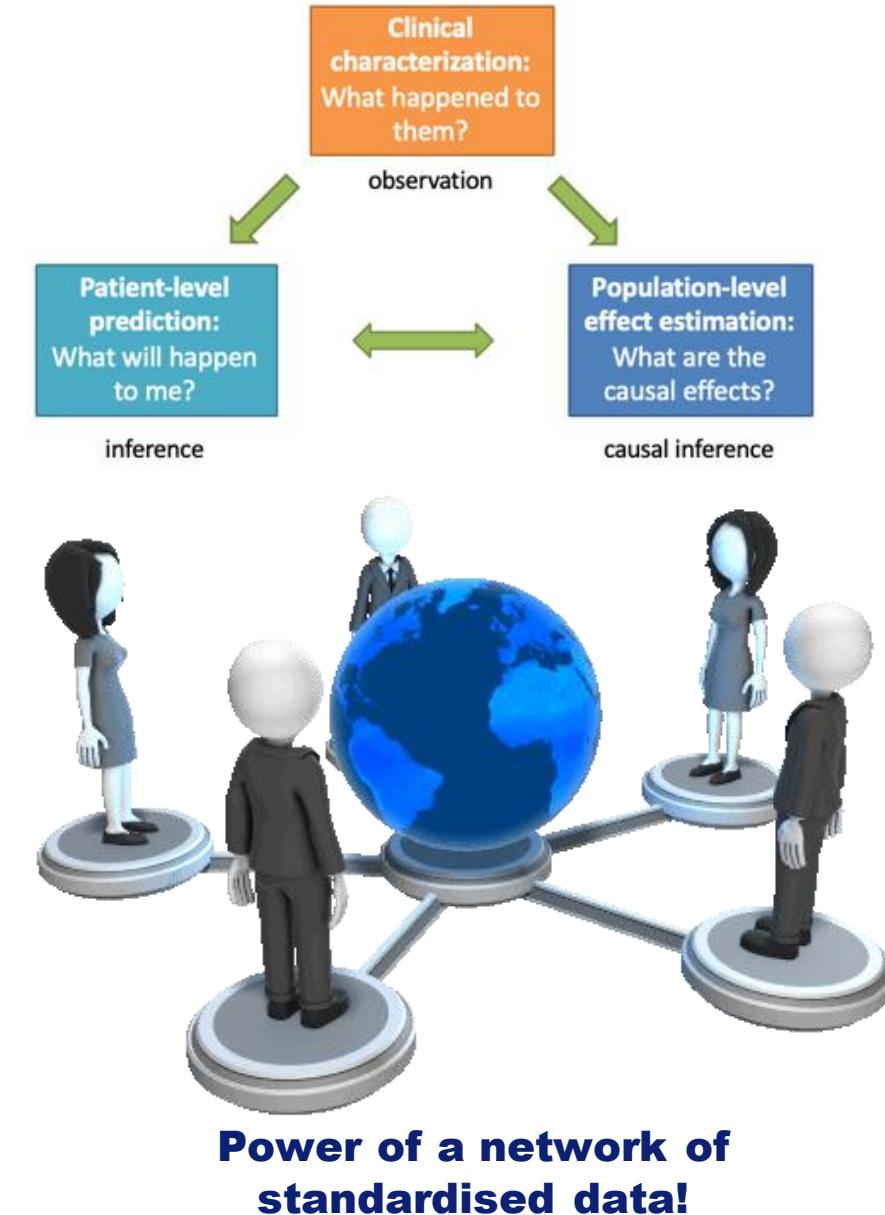
Questions asked across the patient journey

- **Clinical characterization**
 - Treatment Utilization: among patients with diabetes, which treatments are taken when
 - Natural history: Who has diabetes, and who takes metformin?
 - Quality improvement: What proportion of patients with diabetes experience complications?
- **Patient-level prediction**
 - Precision medicine: Given everything you know about me, now I started using metformin, what is the chance I will get lactic acidosis?
 - Disease interception: Given everything you know about me, what is the chance I will develop diabetes?
- **Population-level effect estimation**
 - Safety surveillance: Does metformin cause lactic acidosis?
 - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?



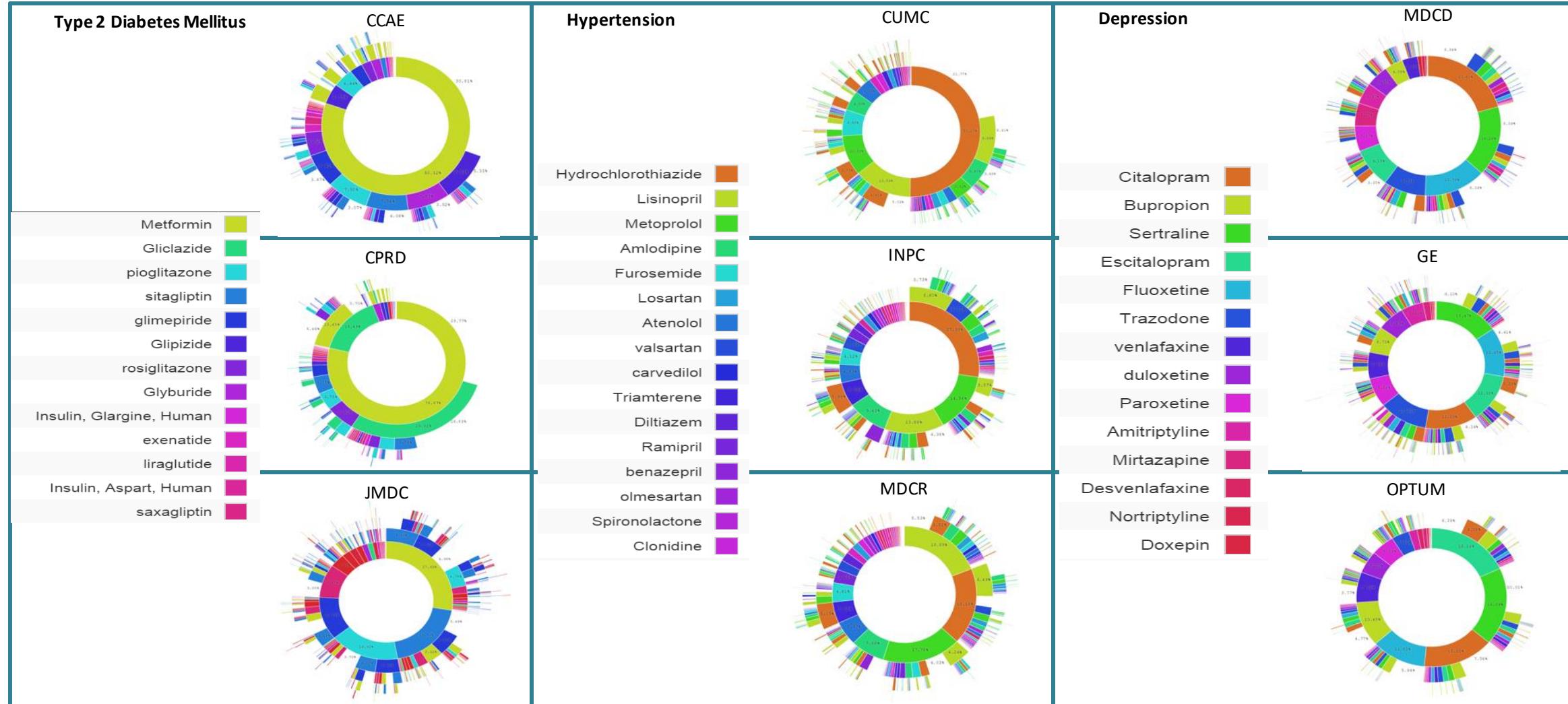
Goal of our work

- Develop transparent and fully reproducible analytical pipelines for all three scientific tasks
- Develop processes and tools to disseminate all the generated evidence
- Create an active community that collaboratively moves this field forward
- Train and educate all the stakeholders to maximally leverage the new paradigm





Clinical Characterization: Population-level heterogeneity across systems, and patient-level heterogeneity within systems

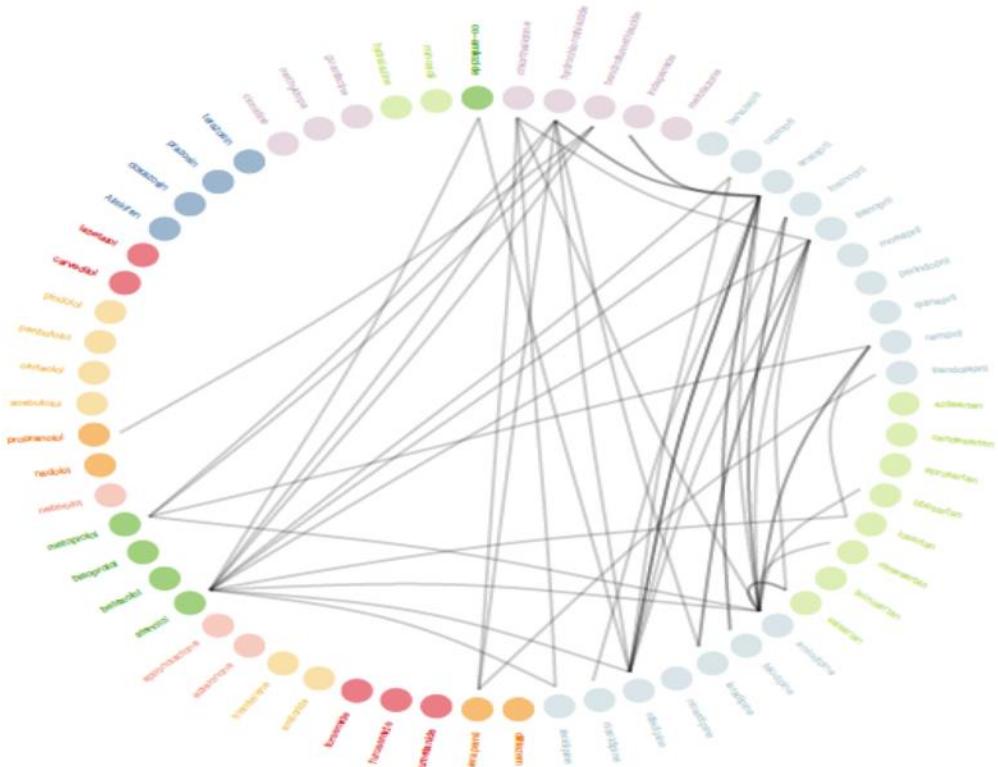




Population-Level Effect Estimation:

Large-Scale Evidence Generation and Evaluation in a Network of Databases (LEGEND)

Head-to-head HTN drug comparisons



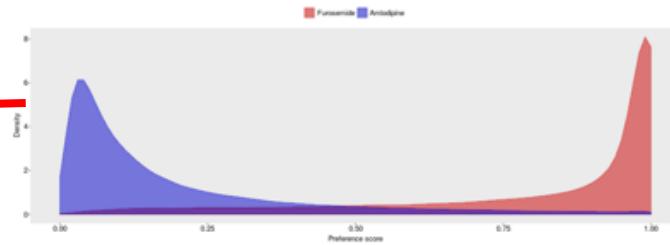
- Trials: 40
 - $N = 102 - [1148] - 33K$



Comparisons of hypertension treatments



Not all analyses are valid





LEGEND basic viewer

About Specific research questions

Indication

Hypertension

Exposure group

Drug major class

Include combination exposures

Target

ACE inhibitors

Comparator

Beta blockers

Outcome

Cardiovascular disease

Data source

- CCAE
- CUMC
- IMSG
- JMDC
- MDCC
- MDCR
- NHIS_NSC
- Optum
- Panther
- Meta-analysis

Analysis

- PS stratification, on-treatment
- PS stratification, intent-to-treat
- PS matching, on-treatment
- PS matching, intent-to-treat

Show 15 entries

Analysis	Data source	HR	LB	UB	P	Cal.HR	Cal.LB	Cal.UB	Cal.P
PS stratification, on-treatment	CCAE	0.72	0.68	0.76	0.00	0.73	0.66	0.82	0.00
PS stratification, on-treatment	CUMC	0.81	0.67	0.98	0.03	0.85	0.67	1.11	0.14
PS stratification, on-treatment	MDCC	0.78	0.70	0.87	0.00	0.78	0.70	0.88	0.00
PS stratification, on-treatment	MDCR	0.87	0.82	0.92	0.00	0.88	0.81	0.97	0.00
PS stratification, on-treatment	Meta-analysis	0.80	0.75	0.85	0.00	0.82	0.75	0.90	0.00
PS stratification, on-treatment	NHIS_NSC	1.25	0.88	1.78	0.21	1.26	0.90	1.79	0.22
PS stratification, on-treatment	Optum	0.81	0.78	0.84	0.00	0.82	0.74	0.92	0.00
PS stratification, on-treatment	Panther	0.78	0.74	0.82	0.00	0.81	0.72	0.93	0.00

Showing 1 to 8 of 8 entries

Previous 1 Next

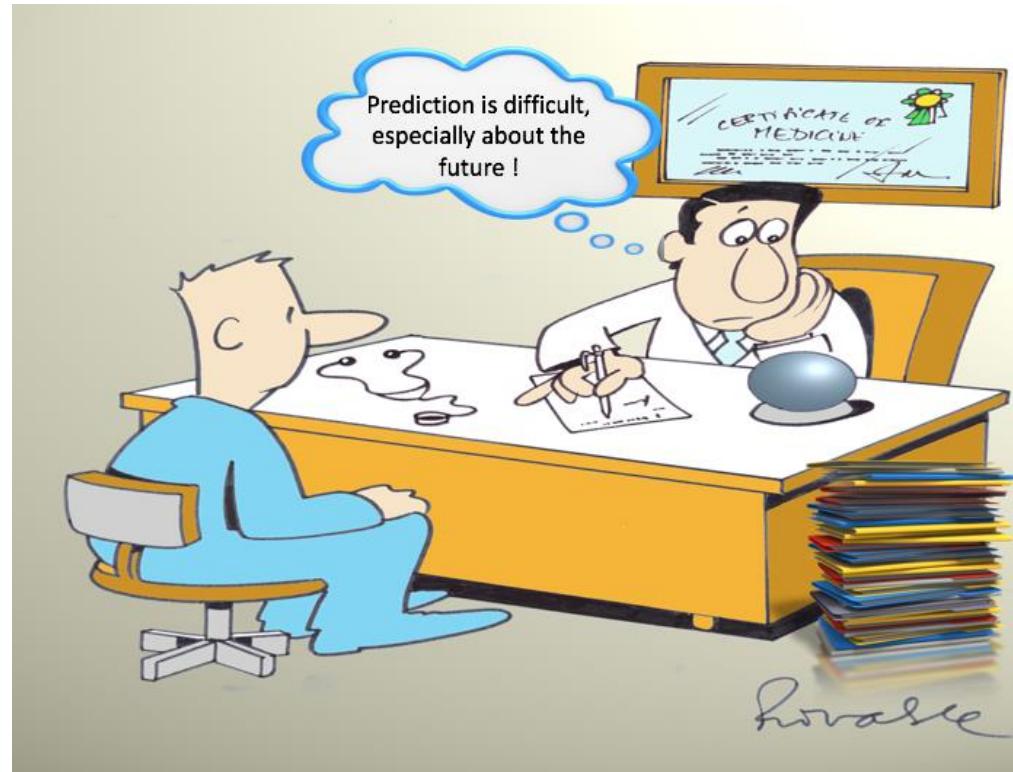
Power Attrition Population characteristics Propensity scores Covariate balance Systematic error Kaplan-Meier

Table 1a. Number of subjects, follow-up time (in years), number of outcome events, and event incidence rate (IR) per 1,000 patient years (PY) in the target (ACE inhibitors) and comparator (Beta blockers) group after stratification, as well as the minimum detectable relative risk (MDRR). Note that the IR does not account for any stratification.

Target subjects	Comparator subjects	Target years	Comparator years	Target events	Comparator events	Target IR (per 1,000 PY)	Comparator IR (per 1,000 PY)	MDRR
763,737	381,144	576,995	262,075	4,004	3,224	6.94	12.30	1.07

Table 1b. Time (days) at risk distribution expressed as minimum (min), 25th percentile (P25), median, 75th percentile (P75), and maximum (max) in the target (ACE inhibitors) and comparator (Beta blockers) cohort after stratification.

Cohort	Min	P10	P25	Median	P75	P90	Max
Target	1	29	37	116	528	718	6,218
Comparator	1	29	29	89	481	676	6,219

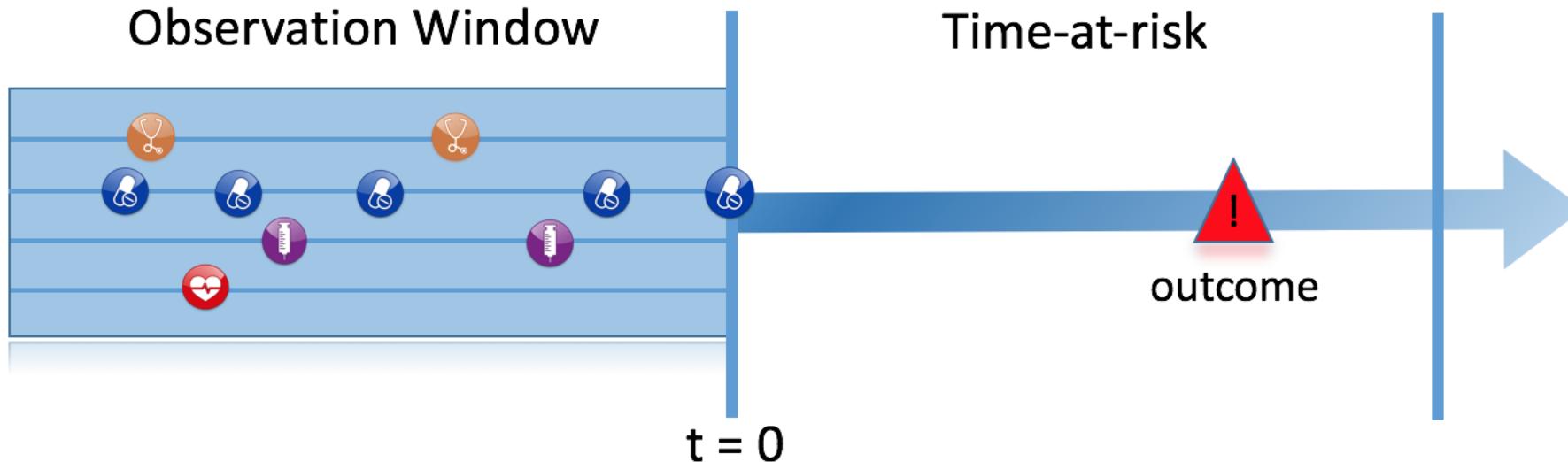


Journey of Patient-Level Prediction

An example of large-scale analysis enabled by data standardisation



Problem definition

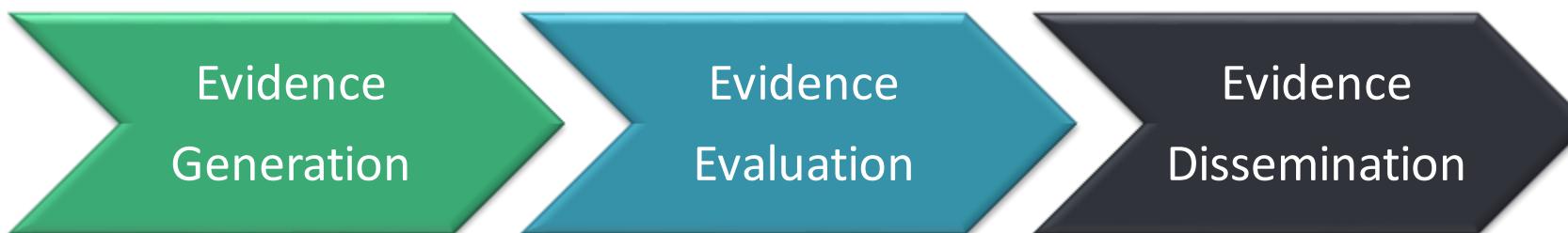


Among a target population (T), we aim to predict which patients at a defined moment in time ($t=0$) will experience some outcome (O) during a time-at-risk. Prediction is done using only information about the patients in an observation window prior to that moment in time.



OHDSI Mission for Patient-Level Prediction

OHDSI aims to develop a systematic process to learn and evaluate large-scale patient-level prediction models using observational health data in a data network





Patient-Level Prediction



A SCHOLARLY JOURNAL OF INFORMATICS IN HEALTH AND BIOMEDICINE

Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data

Jenna M Reps , Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek

Journal of the American Medical Informatics Association, Volume 25, Issue 8, August 2018,
Pages 969–975, <https://doi.org/10.1093/jamia/ocx032>

Published: 27 April 2018 Article history 

 PDF  Split View  Cite  Permissions  Share 

Abstract

Objective

To develop a conceptual prediction model framework containing standardized steps and describe the corresponding open-source software developed to consistently implement the framework across computational environments and observational healthcare databases to enable model sharing and reproducibility.

R-package

www.github.com/OHDSI/PatientLevelPrediction

- Vignettes
- Videos
- Online training material

Book-of-OHDSI

<https://ohdsi.github.io/TheBookOfOhdsi/>

Study Results

www.data.ohdsi.org



LARGE-SCALE PATIENT-LEVEL PREDICTION NOT THE FUTURE!



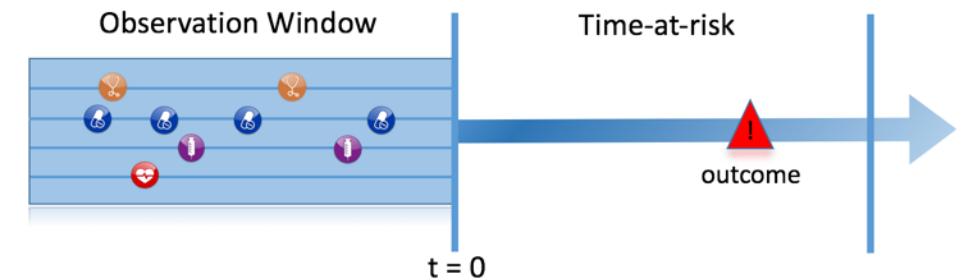
Patient-Level Prediction Team Work



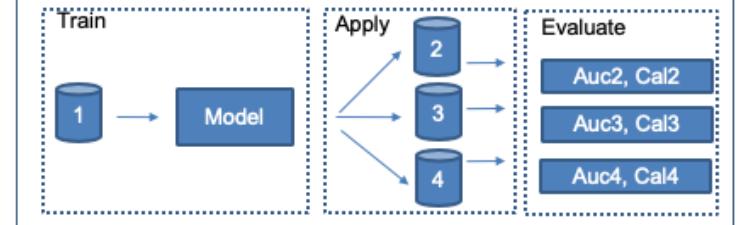
Step 4: External Validation

www.github.com/OHDSI/PatientLevelPrediction

Jenna M Reps, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek; Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data, Journal of the American Medical Informatics Association, Volume 25, Issue 8, 1 August 2018, Pages 969–975, <https://doi.org/10.1093/jamia/ocy032>



External validation is performed using data from multiple populations not used for training.





Model Specification

ATLAS

VIEW: All Prediction Problem Settings Analysis Settings Execution Settings Training Settings

Prediction Problem Settings

Target Cohorts

Show 10 entries Filter:

Remove Name

[PatientLevelPrediction Tutorial Europe 2019] T : patients who are newly diagnosed with Atrial fibrillation

Showing 1 to 1 of 1 entries Previous 1 Next

+ Add Target Cohort

Outcome Cohorts

Show 10 entries Filter:

Remove Name

[PatientLevelPrediction Tutorial Europe 2019] O : Ischemic stroke events

Showing 1 to 1 of 1 entries Previous 1 Next

+ Add Outcome Cohort

Analysis Settings

Model Settings

Show 10 entries Filter:

Remove Model Options

LassoLogisticRegressionSettings {"variance":0.01,"seed":null}

Showing 1 to 1 of 1 entries Previous 1 Next

+ Add Model Settings

Covariate Settings

Column visibility Copy CSV Show 10 entries Filter:

Remove Options

Chads2

+ Add Covariate Settings



Generate R-Package and share with the world

Review & Download Import Export

Review & Download

Review Full Study Specification

Please review the full study specification below and scroll down the page to download the study package.

Full Analysis List (3) Prediction Problem Settings (1) Analysis Settings (3)

Target Cohort Name	Outcome Cohort Name	Model Name	Model Settings	Covariate Settings	Risk Window Start
[PatientLevelPrediction Tutorial Europe 2019] T : patients who are newly diagnosed with Atrial fibrillation (3)	[PatientLevelPrediction Tutorial Europe 2019] O : Ischemic stroke events	LassoLogisticRegressionSettings	{"variance":0.01,"seed":null}	"attr_class":"covariateSetting...	1
[PatientLevelPrediction Tutorial Europe 2019] T : patients who are newly diagnosed with Atrial fibrillation	[PatientLevelPrediction Tutorial Europe 2019] O : Ischemic stroke events	LassoLogisticRegressionSettings	{"variance":0.01,"seed":null}	"attr_class":"covariateSetting...	1
[PatientLevelPrediction Tutorial Europe 2019] T : patients who are newly diagnosed with Atrial fibrillation	[PatientLevelPrediction Tutorial Europe 2019] O : Ischemic stroke events	LassoLogisticRegressionSettings	{"variance":0.01,"seed":null}	"attr_class":"covariateSetting...	1

Showing 1 to 3 of 3 entries

Previous 1 Next

Download Study Package

Please provide a name for the study package and click "download" to obtain the study package in ZIP format.

MyStudy

Please Note: The package name should consist of alphanumeric characters only. No spaces or special characters are permitted.

Download





Share model and performance

Prediction Viewer

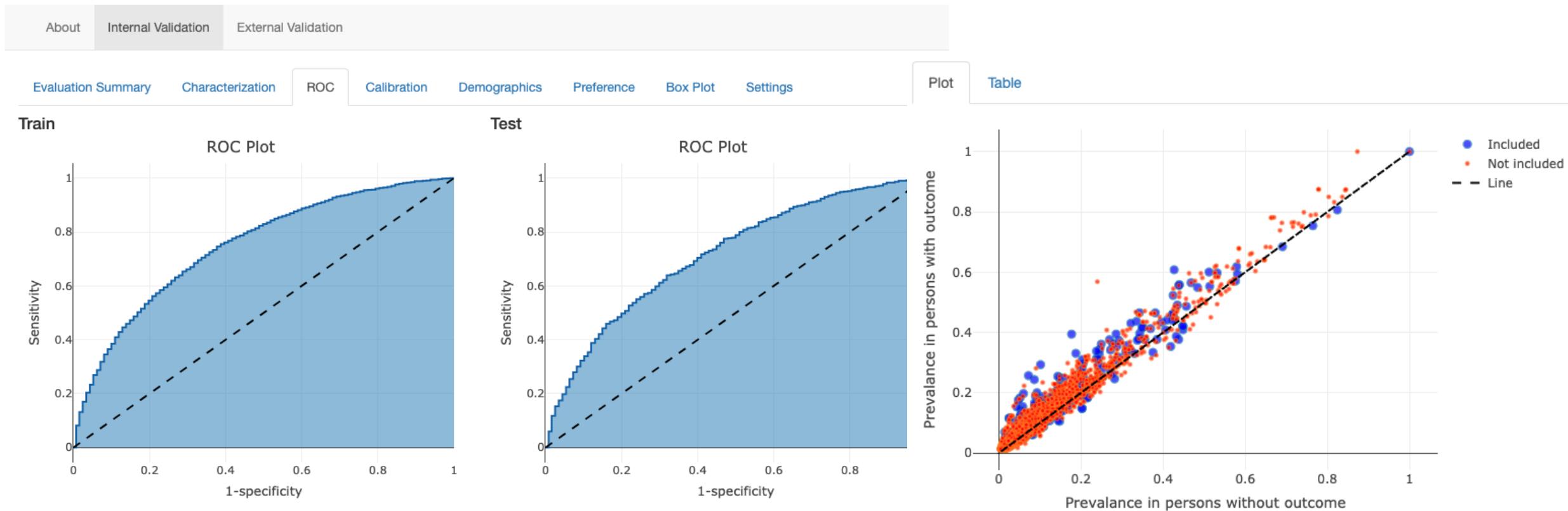


Figure 2. The Receiver Operating Characteristics (ROC) curve shows the ability of the model to discriminate between people with and without the outcome during the time at risk. It is a plot of sensitivity vs 1-specificity at every probability threshold. The higher the area under the ROC plot the higher the discriminative performance of the model. The diagonal refers to a model assigning a class at random (area under de ROC = 0.5).

Figure 1. The variable scatter plot shows the mean covariate value for the people with the outcome against the mean covariate value for the people without the outcome. The meaning of the size and color of the dots depends on the settings on the left of the figure.



EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Enabling Large-Scale Analysis of
Electronic Health Records in
Europe

The EHDEN Project
Peter Rijnbeek

Associate Professor Health Data Science
Erasmus MC, Rotterdam, The Netherlands





EHDEN: VISION AND MISSION



Vision

The European Health Data & Evidence Network (EHDEN) aspires to be the trusted observational research ecosystem to enable better health decisions, outcomes and care

Mission

Our 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 a common data model



EHDEN CONSORTIUM

Innovative Medicines Initiative Project



Start date: 1 Nov 2018
End date: 30 Apr 2024
Duration: 66 months



22 partners



Almost €29 million

Universities, public bodies and research organisations

Erasmus MC
Universitair Medisch Centrum Rotterdam

Uppsala Monitoring Centre

NICE
National Institute for Health and Care Excellence

TARTU ÜLIKOOL
UNIVERSITAS TARTUENSIS
1632

UNIVERSITY OF OXFORD

universidade de aveiro

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb

NOVARTIS

1632

BAYER

Celgene

Pfizer

Lilly

abbvie

AstraZeneca

SERVIER

SANOFI

ucb</p



EHDEN IS ABOUT ...

HARMONISATION

Harmonise more than 100 million anonymised health records to the OMOP common data model

OUTCOMES

Enabling outcomes driven healthcare at a European level

FEDERATION

Creation of an EU-wide architecture for federated analyses of real world data



EDUCATION

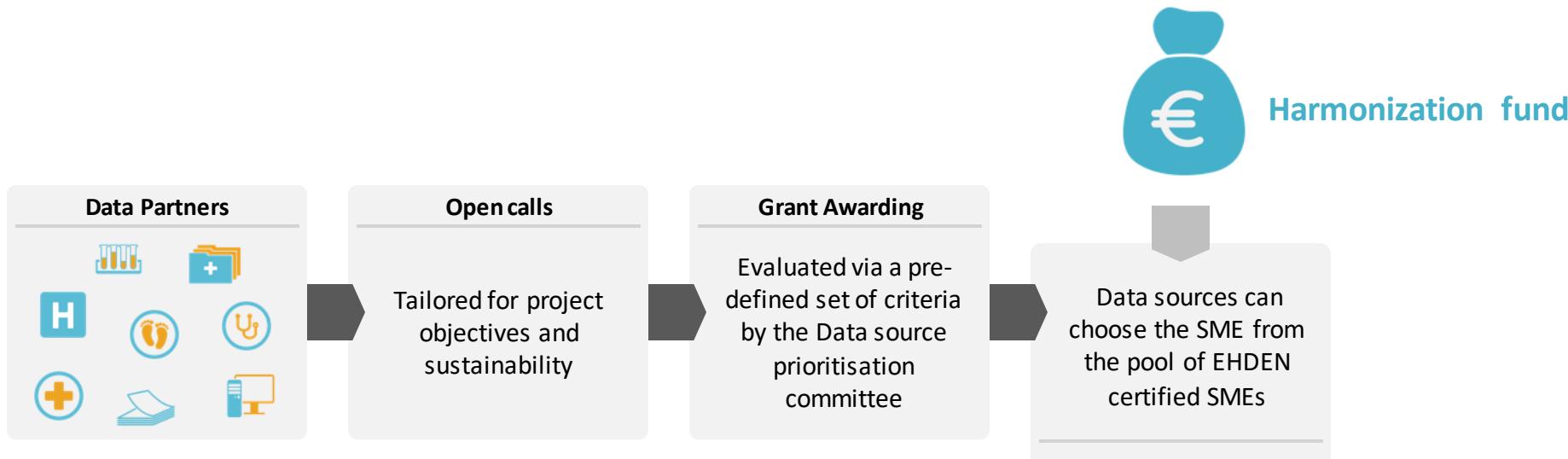
The establishment of an EHDEN Academy, webinars and face-to-face training sessions to train all stakeholders

COMMUNITY

Establish a self-sustaining open science collaboration in Europe, supporting academia, industry, regulators, payers, government, NGOs and others



CALL PROCESS FOR DATA PARTNERS AND SMALL TO MEDIUM-SIZED ENTERPRISE (SMEs)





PUBLIC RELEASE IN APRIL

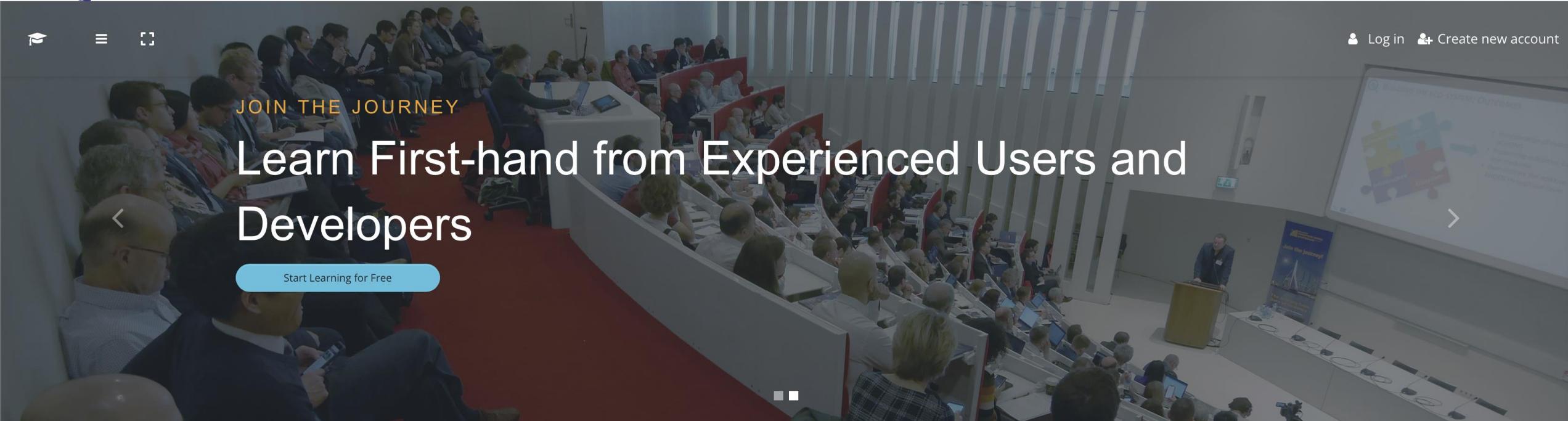


[Log in](#) [Create new account](#)

JOIN THE JOURNEY

Learn First-hand from Experienced Users and Developers

[Start Learning for Free](#)



Learning for Anyone, Anywhere

On-demand training and development programs developed by the OHDSI and EHDEN community.



Tool Ecosystem

Learn to use our latest tools for observational research in your organization.



Meet Like-minded People

Share and develop your ideas with an interdisciplinary group of learners.



Join the Journey

Prepare to participate in worldwide face-to-face events such as our Study-A-Thon.



Connect with Experts

Learn from experienced instructors that help to drive the development of our tools forward.



SME CERTIFICATION

Goal: to provide the SME all the skills to perform the data standardisation task to the OMOP-CDM and train them on the installation of the analytical infrastructure

Training

- 1) EHDEN Foundation: Introduction to IMI, EHDEN, OHDSI
- 2) OHDSI-IN-A-BOX Virtual Machine
- 3) OMOP CDM and Standardized Vocabularies
- 4) Extract, Transform and Load
- 5) Analytical Infrastructure



Certification

- Final certification will contain a two days face-to-face meeting at the Erasmus MC in Rotterdam with all SMEs in the current batch. Multiple persons per SME can participate.
- Final assessment will contain a mapping exercise and installation of the Analytical Infrastructure.

More course will be added in the EHDEN Academy in the future.

Course overview

All

Sort by Course name Card



Academy Support



EHDEN Foundation



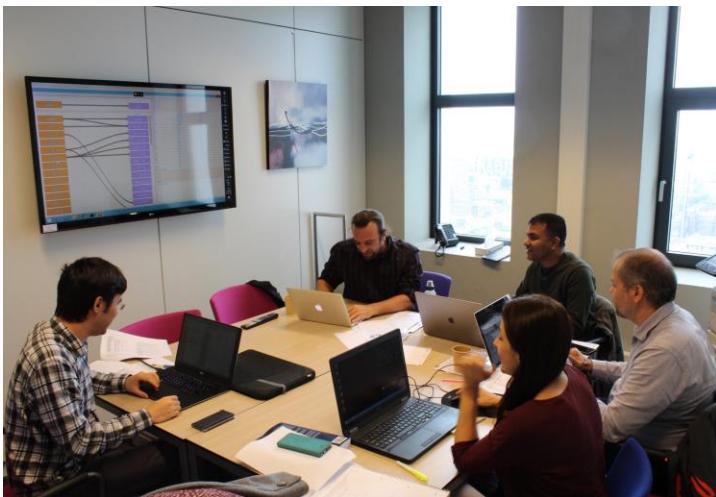
Extract, Transform and Load



OHDSI-In-A-Box Virtual Machine



SME PILOT CALL





SME OPEN CALL RESULTS – APRIL 2019

New Call in SME evaluation phase



34 SME profiles made



28 Eligible applications



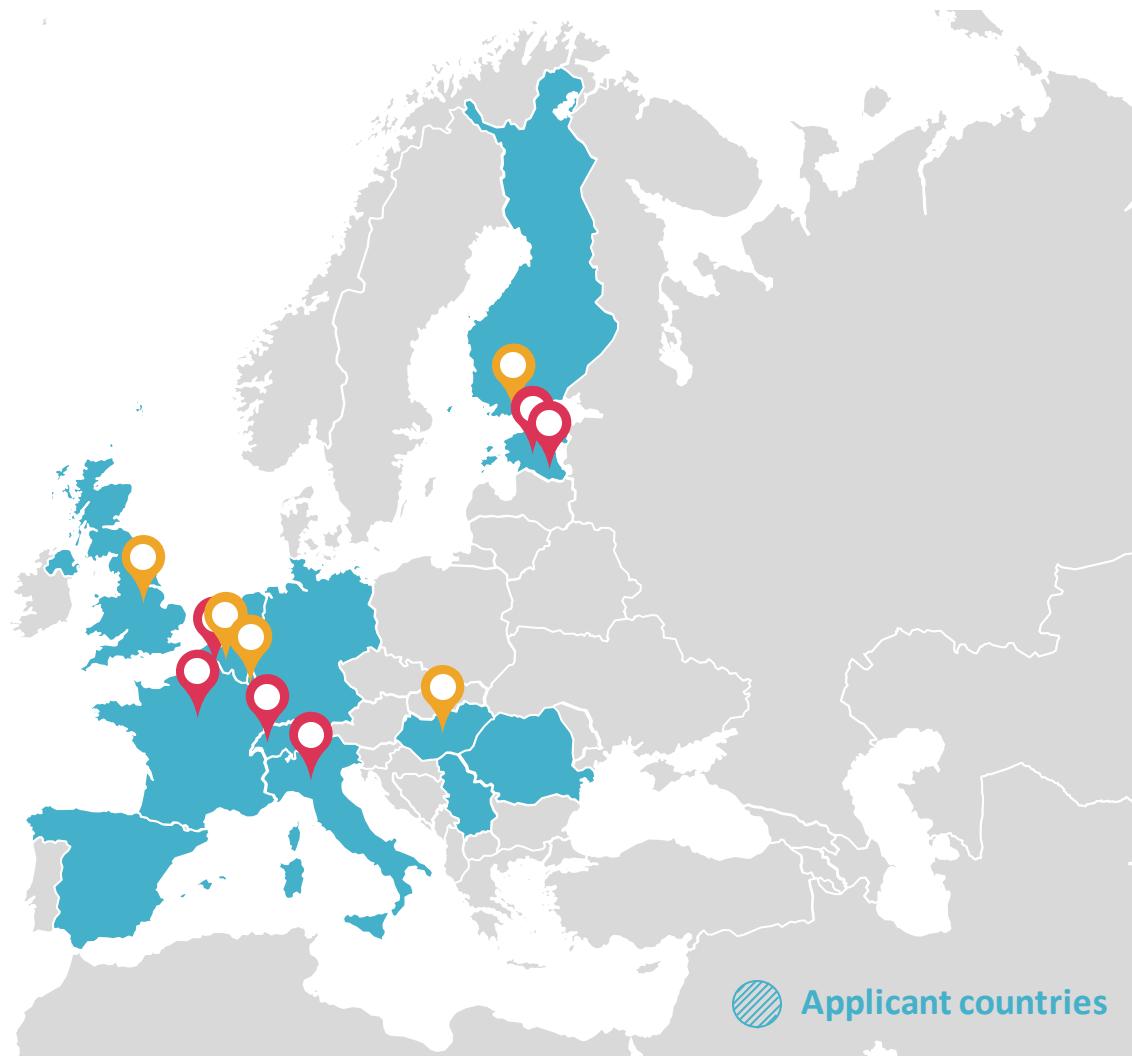
11 SMEs initially selected



Batch 1



Batch 2





OPEN CALL FOR DATA PARTNERS

- Different types of grants (max 100.000 Euro):
 - Create new Data Transformation and Analytical Infrastructure
 - Revise Existing Data Transformation and Analytical Infrastructure
 - Inspect Completed Data Transformation and Analytical Infrastructure
- Data Partners from EU Member States and H2020 countries can apply through online application portal.

For more information about the future Open Calls see the EHDEN website: www.ehdeneu.eu



DATA PARTNER PILOT CALL RESULTS



48 Data partner profiles made

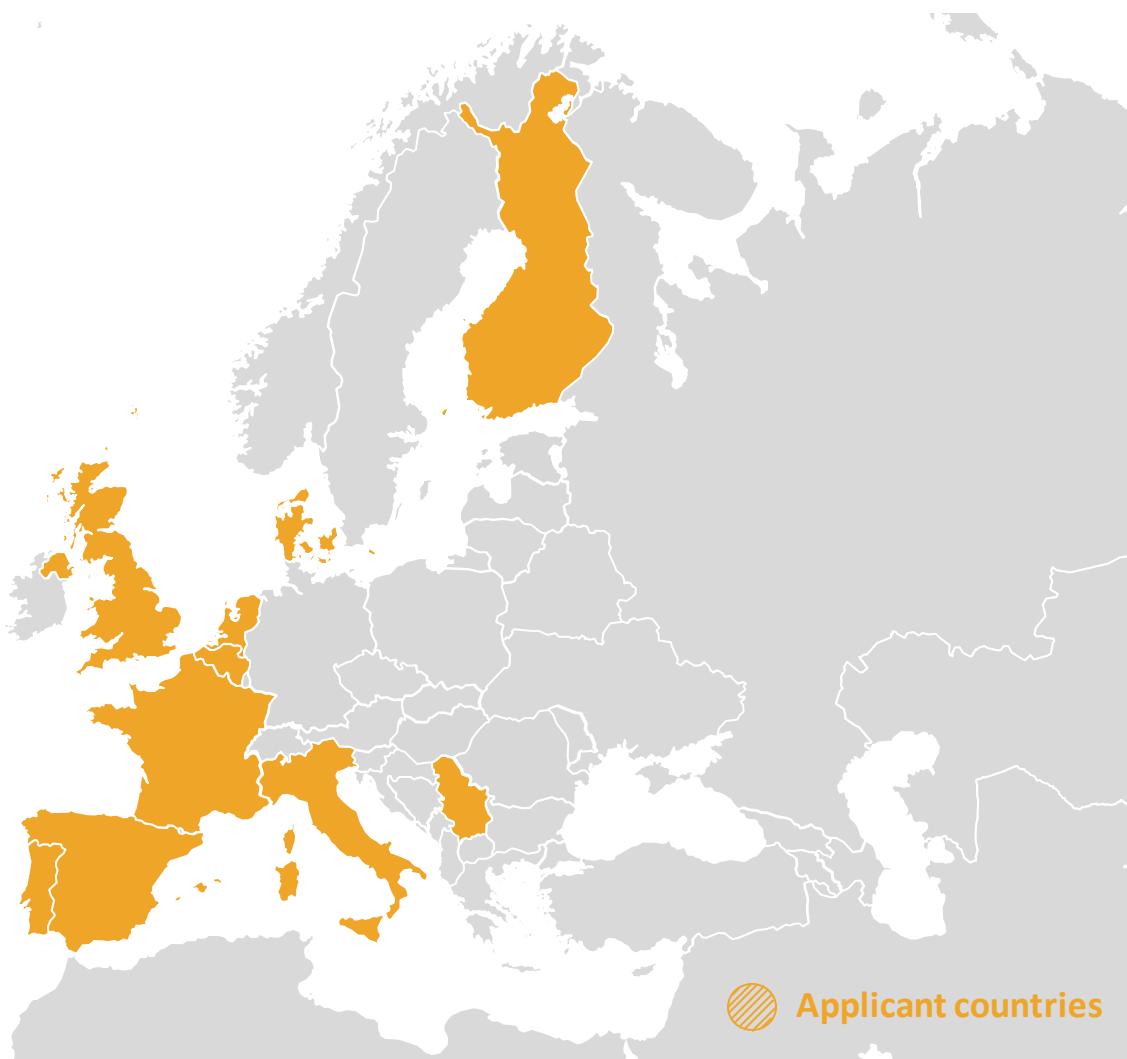


28 Submitted applications

20 Data Sources selected

>170 million Patient Records

Hospital, GP, Registries, etc.





THE POWER OF DISRUPTIVE OPEN SCIENCE: THE STUDY-A-THON CONCEPT

Why do we seem to accept that answering important clinical questions takes a lot of time?



We have an obligation to be disruptive and push hard to change the current paradigm!!

This requires a team effort, no one has all the necessary competences:
clinical knowledge, data source expertise, analytics, writing skills, etc.



Why not bring them together at a nice location and focus !!



OXFORD STUDY-A-THON

To compare the **risk** of post-operative complications and mortality between unicompartmental and total knee replacement.

THE LANCET
Rheumatology

Log in



ARTICLES | ONLINE FIRST

Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study

Edward Burn, MSc * • James Weaver, MSc * • Daniel Morales, PhD • Albert Prats-Uribe, MPH • Antonella Delmestri, PhD

Victoria Y Strauss, PhD • Ying He, PhD • Danielle E Robinson, PhD • Rafael Pinedo-Villanueva, PhD • Spyros Kolovos, PhD

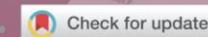
Talita Duarte-Salles, PhD • William Sproviero, PhD • Dahai Yu, PhD • Michel Van Speybroeck, MSc • Ross Williams, MSc •

Luis H John, MSc • Nigel Hughes, MSc • Anthony G Sena, BA • Ruth Costello, MSc • Belay Birlie, MSc • David Culliford, PhD

Caroline O'Leary, MSc • Henry Morgan, PhD • Theresa Burkard, MSc • Prof Daniel Prieto-Alhambra, PhD

Patrick Ryan, PhD † • Show less • Show footnotes

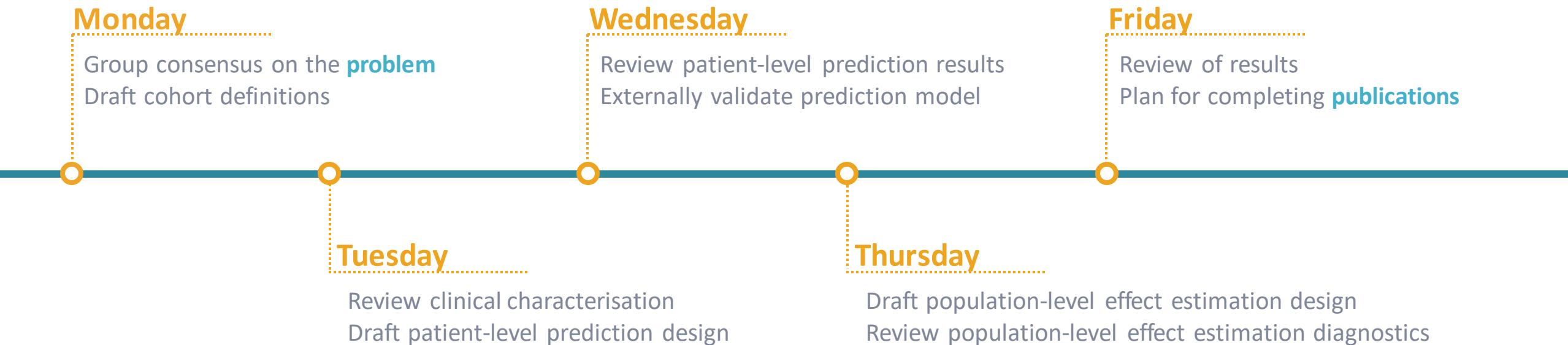
Published: November 07, 2019 • DOI: [https://doi.org/10.1016/S2665-9913\(19\)30075-X](https://doi.org/10.1016/S2665-9913(19)30075-X)





WE CAN DO THIS IN ONE WEEK (STUDY-A-THON)??

“To compare the **risk** of post-operative **complications** and **mortality** between unicompartmental *vs* total knee replacement.”





THE SECOND STUDY-A-THON IN BARCELONA



We will publish a video about this week soon!!

Different Location: Barcelona

Different Topic: Rheumatoid Arthritis

Different Team: RA Experts, Industry, Academia, Data Custodians

More datasources: 14

More countries: USA, Japan, Spain, The Netherlands, Estonia, UK, Germany, France, Belgium

Different approach:

Protocols were developed prior to the meeting and approved by governance board is applicable.

AIM: Submission of abstracts for European League Against Rheumatism (EULAR) and multiple publications



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

COVID-19 Study-A-Thon

ohdsi.org/covid-19-updates

Four day virtual study-a-thon
330 people from 30 nations
37 healthcare databases
Many studies run in parallel

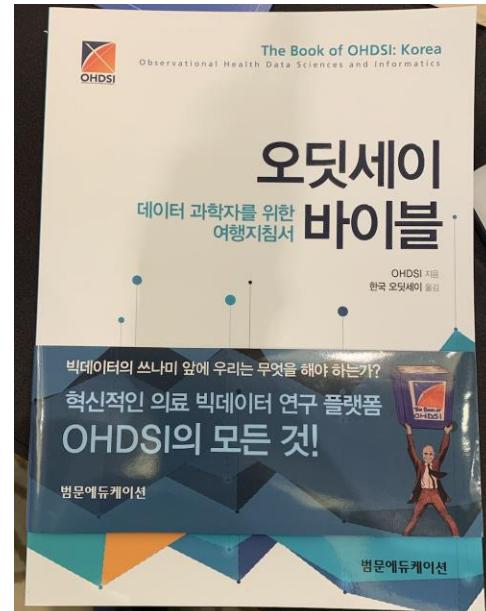
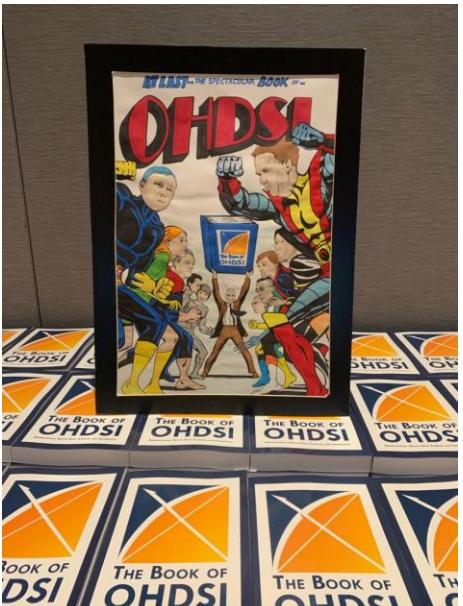




NEED MORE INFORMATION?



enquiries@ehden.eu



www.ohdsi.org



www.ehdeneu



@IMI_EHDEN



IMI_EHDEN



github.com/EHDEN



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

<https://book.ohdsi.org>