

# Health Data Science

## LARGE-SCALE PREDICTIVE MODEL DEVELOPMENT AND VALIDATION USING DATA STANDARDIZED TO THE OMOP COMMON DATA MODEL

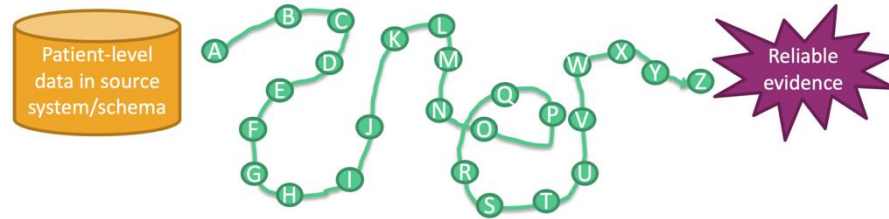


Peter R. Rijnbeek  
Professor of Medical Informatics  
Chair Department of Medical Informatics  
Erasmus MC, Rotterdam  
The Netherlands  
[www.healthdatascience.nl](http://www.healthdatascience.nl)

Erasmus MC  
University Medical Center Rotterdam



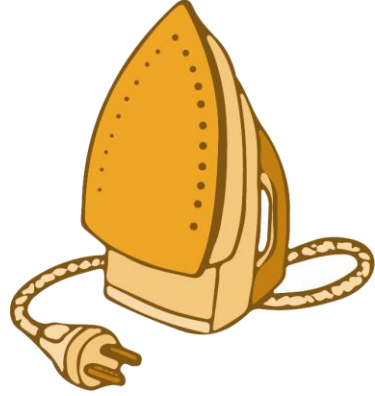
# Generating Reliable Evidence at Scale



*How can we generate reliable evidence at a large scale, i.e. on many data sources in Europe for many research questions?*

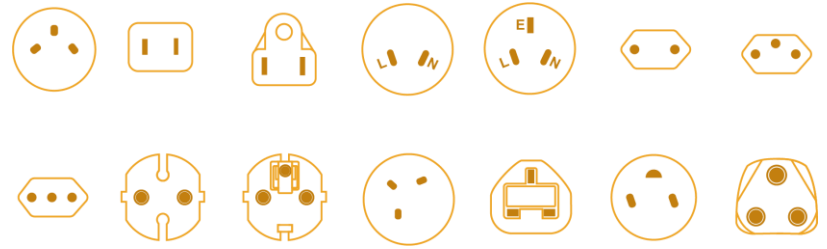
# The challenges of Real-World Data

Analytical method



Link to data

*The data...*



*What will it require?*



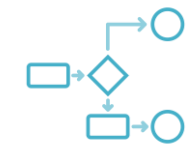
*Data interoperability*



*Strong community*

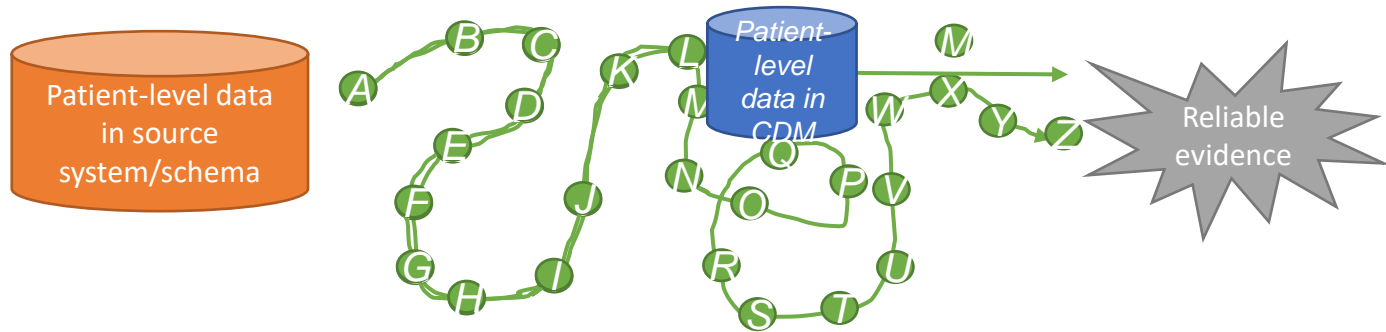
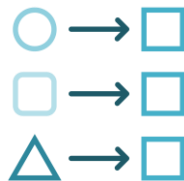


*Data network*



*Standardised analytics*

# Standardization to a common data model



We develop analytical pipelines that can be utilized by data partners across the world.



## Map of Collaborators

The OHDSI community brings together volunteers from around the world to establish open community data standards, develop open-source software, conduct methodological research, and apply scientific best practices to both answer public health questions and generate reliable clinical evidence.

Our community is ALWAYS seeking new collaborators. Do you want to focus on data standards or methodological research? Are you passionate about open-source development or clinical applications? Do you have data that you want to be part of global network studies? Do you want to be part of a global community that truly values the benefits of open science? Add a dot to the map below and JOIN THE JOURNEY!

### OHDSI By The Numbers

- 2,367 collaborators
- 74 countries
- 21 time zones
- 6 continents
- 1 community



# Large Active Community





# Oncology WG Structure

## Observational Cancer Research

CDM & Standardized Vocabularies

Research

Vocab & Development

2<sup>nd</sup> & 4<sup>th</sup> Thursday 1-2 pm EST

Genomic

2<sup>nd</sup> & 4<sup>th</sup> Tuesday 9-10 am EST

Outreach & Research

2<sup>nd</sup> & 4<sup>th</sup> Tuesday 3-4 pm EST

***[www.ohdsi.org](http://www.ohdsi.org)***



## *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*

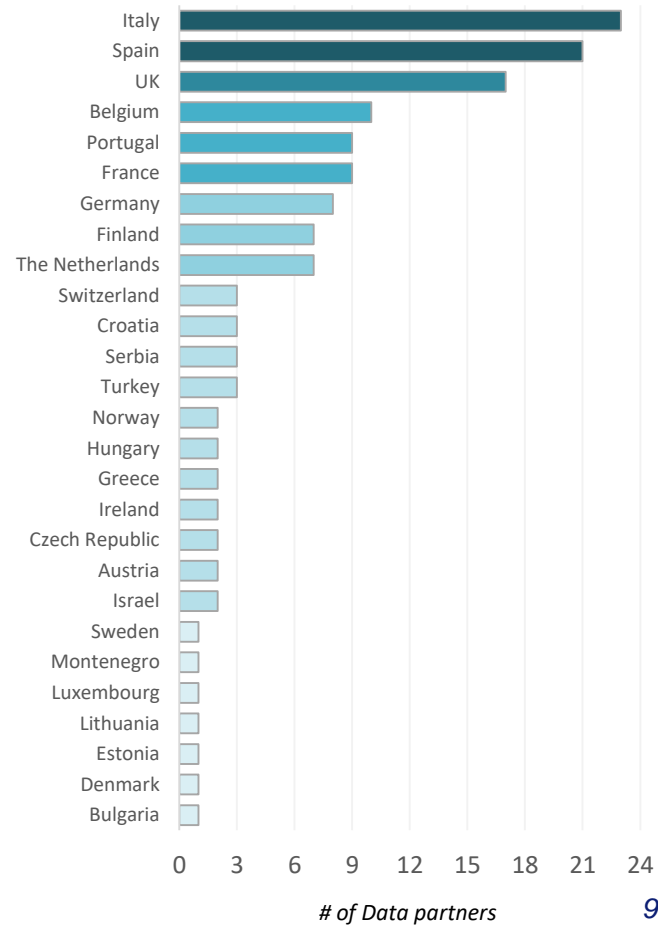
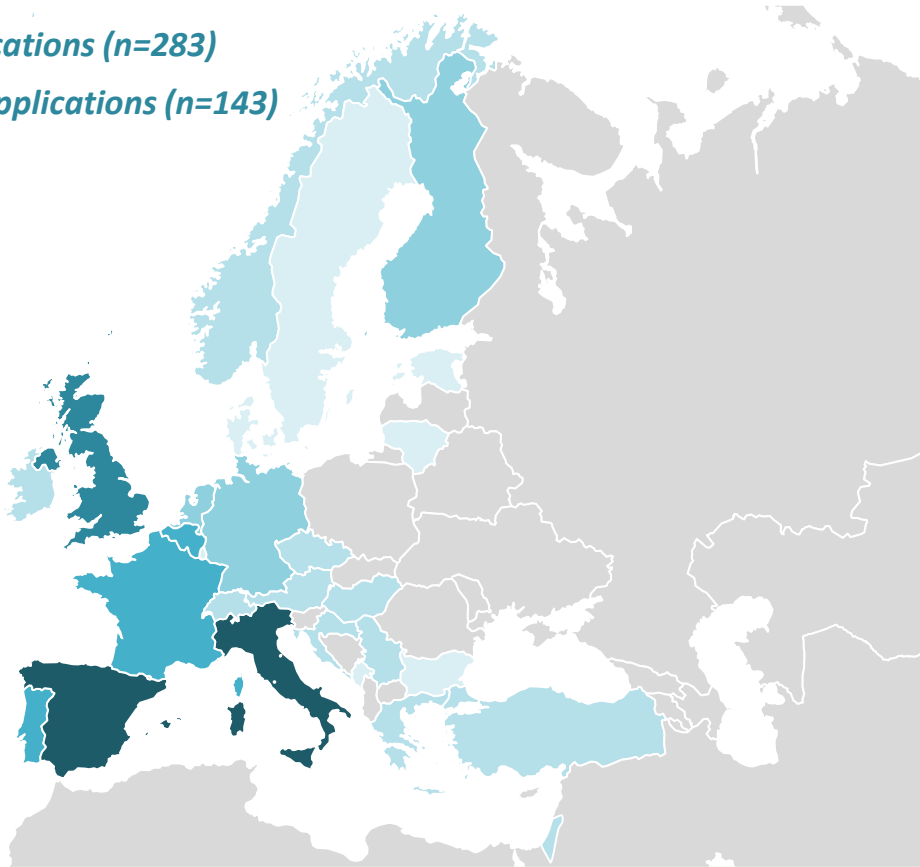




# EHDEN DATA NETWORK

Total applications (n=283)

Awarded applications (n=143)



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

## Initiation of DARWIN EU® Coordination Centre advances integration of real-world evidence into assessment of medicines in the EU [Share](#)

News 09/02/2022

EMA is initiating today the establishment of the Coordination Centre for the [Data Analysis and Real World Interrogation Network \(DARWIN EU®\)](#).

The role of the Coordination Centre is to develop and manage a network of real-world healthcare data sources across the EU and to conduct scientific studies requested by medicines regulators and, at a later stage, requested by other stakeholders.

The vision of DARWIN EU® is to give EMA and [national competent authorities](#) in EU Member States access to valid and trustworthy real-world evidence, for example on diseases, patient populations, and the use, safety and effectiveness of medicines, including vaccines, throughout the lifecycle of a [medicinal product](#).

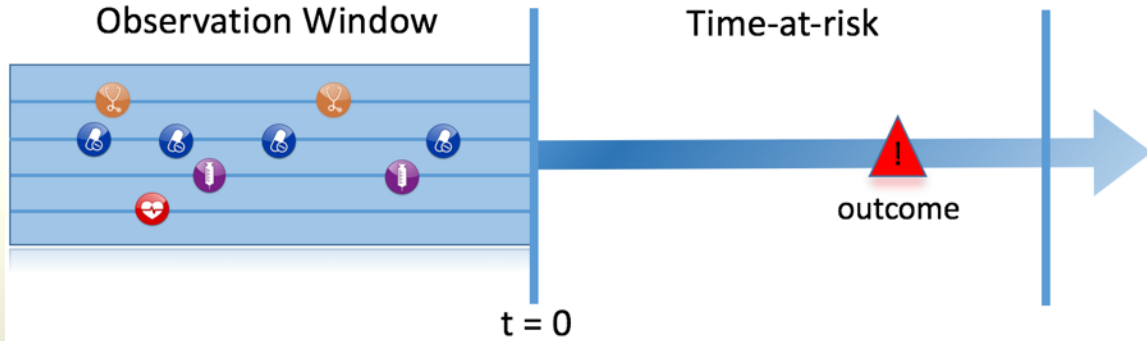
By supporting decision-making on the development, authorisation and surveillance of medicines, a wide range of stakeholders will benefit, from patients and healthcare professionals to [health technology assessment bodies](#) and the pharmaceutical industry. Additionally, DARWIN EU® will provide an invaluable resource to prepare for and respond to future healthcare crises and pandemics.

For example, the availability of timely and reliable real-world evidence can lead to [innovative medicines](#) becoming more quickly available to patients. Better evidence also supports more informed regulatory decision-making on the safe and effective use by patients of medicines on the market.

EMA will be working with Erasmus University Medical Center Rotterdam to establish the DARWIN EU® Coordination Centre. The contract was awarded to Erasmus University Medical Center Rotterdam following a call for tender for a service provider published in June 2021. The contractor will set up the necessary infrastructure of the Coordination Centre and establish the required business services. It will run scientific studies to answer research questions that come up during the evaluation of medicines in the EU and also maintain a catalogue of real-world data sources and metadata for use in medicine regulatory activities. All studies will be published in the EU catalogue of observational studies.



# 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.*

# Model Development Pipeline



**Dissemination** of study results should follow the minimum requirements as stated in the *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement*<sup>1</sup>.

- *Sharing of model development details*
- *Discrimination and Calibration*
- *Internal and external validation*
- *Etc.*

**Annals of Internal Medicine** RESEARCH AND REPORTING METHODS

## Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration

Karel G.M. Moons, PhD; Douglas G. Altman, DSc; Johannes B. Reitsma, MD, PhD; John P.A. Ioannidis, MD, DSc; Petra Macaskill, PhD; Ewout W. Steyerberg, PhD; Andrew J. Vickers, PhD; David F. Ransohoff, MD; and Gary S. Collins, PhD

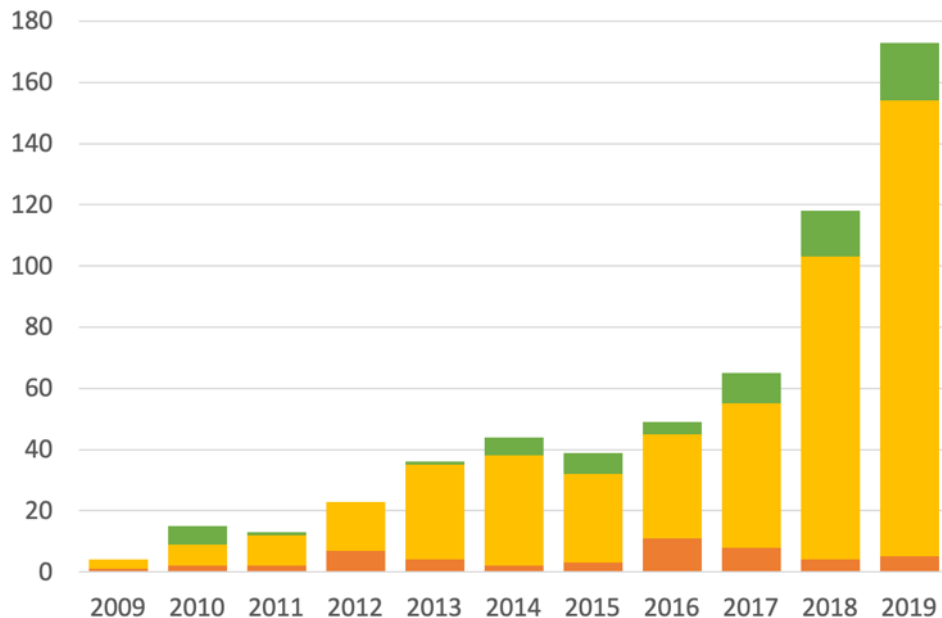
The TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) Statement includes a 22-item checklist, which aims to improve the reporting of studies developing, validating, or updating a prediction model, whether for diagnostic or prognostic purposes. The TRIPOD Statement aims to improve the transparency of the reporting of a prediction model study regardless of the study methods used. This explanation and elaboration document describes the rationale, clarifies the meaning of each item, and discusses why transparent reporting is important, with a view to assessing risk of bias and clinical usefulness of the prediction model. Each checklist item of the TRIPOD Statement is explained in detail and accom-

panied by published examples of good reporting. The document also provides a valuable reference of issues to consider when designing, conducting, and analyzing prediction model studies. To aid the editorial process and help peer reviewers and, ultimately, readers and systematic reviewers of prediction model studies, it is recommended that authors include a completed checklist in their submission. The TRIPOD checklist can also be downloaded from [www.tripod-statement.org](http://www.tripod-statement.org).

*Ann Intern Med* 2015;162:W1-W73. doi:10.7326/M14-0998 [www.annals.org](http://www.annals.org)  
For author affiliations, see end of text.  
For members of the TRIPOD Group, see the Appendix.

<sup>1</sup> Moons, KG et al. *Ann Intern Med*. 2015;162(1):W1-73

# Current status of predictive modelling:



> [J Am Med Inform Assoc. 2022 Apr 13;29\(5\):983-989. doi: 10.1093/jamia/ocac002.](#)

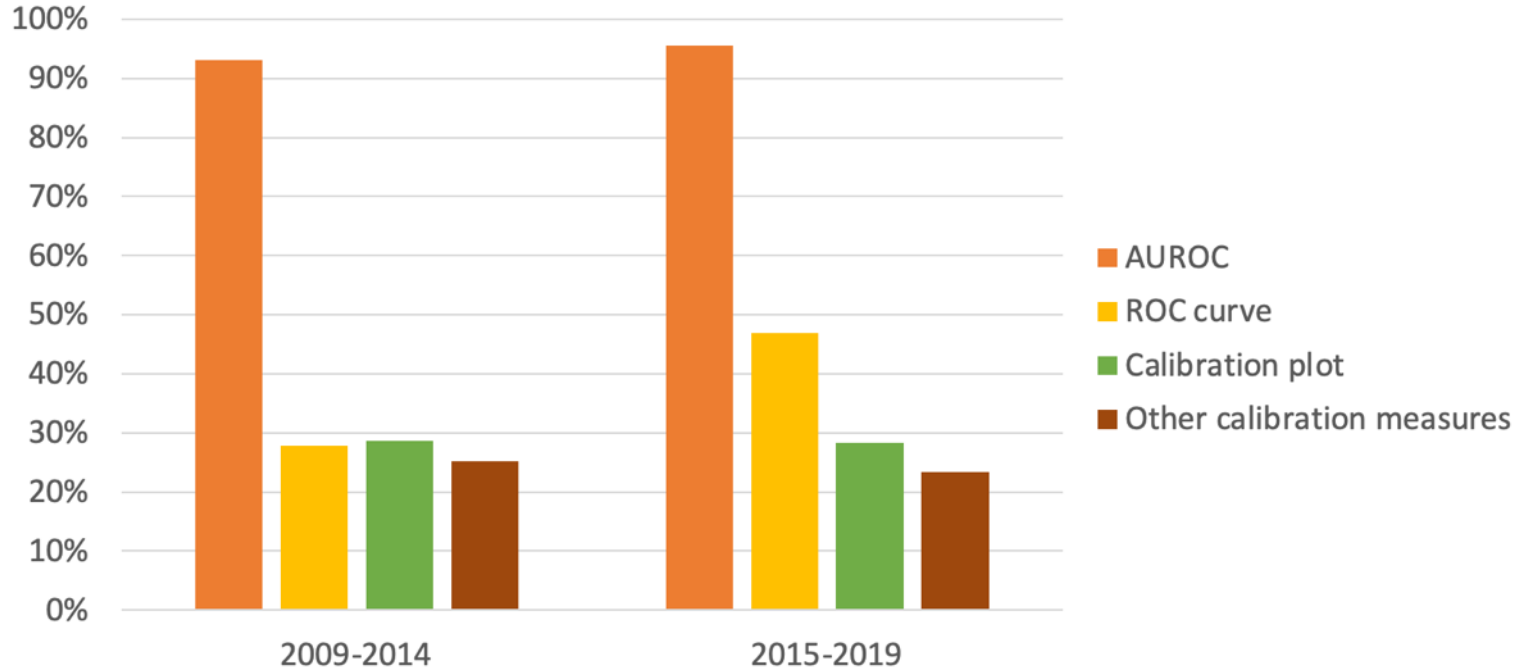
## Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review

Cynthia Yang<sup>1</sup>, Jan A Kors<sup>1</sup>, Solomon Ioannou<sup>1</sup>, Luis H John<sup>1</sup>, Aniek F Markus<sup>1</sup>, Alexandros Rekkas<sup>1</sup>, Maria A J de Ridder<sup>1</sup>, Tom M Seinen<sup>1</sup>, Ross D Williams<sup>1</sup>, Peter R Rijnbeek<sup>1</sup>

- External validation
- Internal validation
- No validation

***Review of 422 papers with 579 models***

# Current status of predictive modelling:



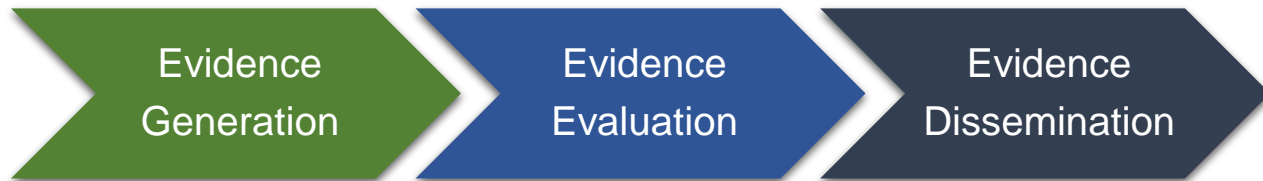
# What is needed?

## *Full transparency and reproducibility*

1. Standardised Health Data with respect to structure and terminology
2. Standardised Analytical Pipelines that enforce best modelling practices
3. Share models and allow extensive external validation across many databases
4. Disseminate all performance results

# Our mission for Patient-Level Prediction

The Observational Health Data Sciences and Informatics (OHDSI) and the European Health Data and Evidence Network (EHDEN) developed a systematic process to learn and evaluate large-scale patient-level prediction models using observational health data in a large data network



**OHDSI**

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

[www.ohdsi.org](http://www.ohdsi.org)



**EHDEN**

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

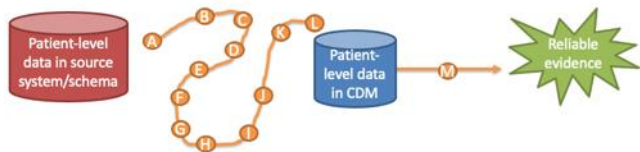
[www.ehden.eu](http://www.ehden.eu)


Erasmus MC





# Patient-Level Prediction Framework



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/ocy032>

Published: 27 April 2018 [Article history](#) ▼

 PDF  Split View  Cite  Permissions  Share ▼

## Abstract

### Objective

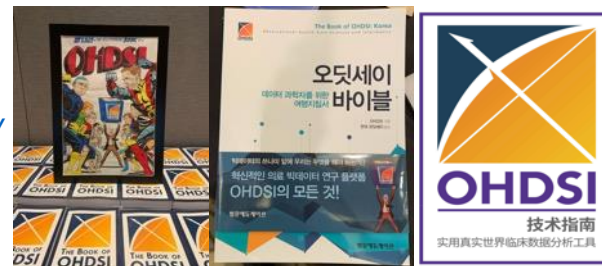
To develop a conceptual prediction model framework containing standardized steps and describe the corresponding open-source software developed to consistently implement the framework across computational environments and observational healthcare databases to enable model sharing and reproducibility.

*R*-package

[www.github.com/OHDSI/PatientLevelPrediction](https://www.github.com/OHDSI/PatientLevelPrediction)

*Book-of-OHDSI*

<https://book.ohdsi.org/>

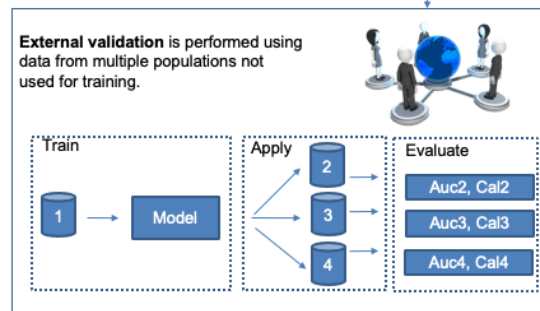
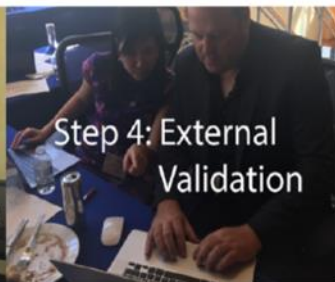
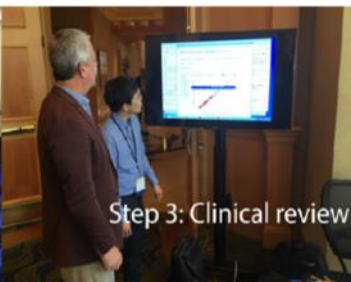
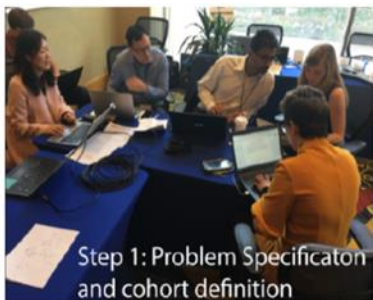


*EHDEN Academy* <https://academy.ehden.eu>

*Study Results*

<https://data.ohdsi.org>

# Prediction team in action



# Model Specification



ATLAS

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Characterizations
- Cohort Pathways
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Apache 2.0  
open source software  
provided by  
**OHDSI**

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

## Prediction Problem Settings

### Target Cohorts

+ Add Target Cohort

Show 10 entries

| Remove                   | Name   |
|--------------------------|--|
| <input type="checkbox"/> | [EHDEN RA] Female new users of methotrexate monotherapy used for PLP |
| <input type="checkbox"/> | [EHDEN RA] New users of methotrexate monotherapy used for PLP        |

Showing 1 to 2 of 2 entries

### Outcome Cohorts

Show 10 entries

| Remove                   | Name  |
|--------------------------|---|
| <input type="checkbox"/> | [EHDEN RA] Leukopenia events using diagnoses and measurements                           |
| <input type="checkbox"/> | [EHDEN RA] Pancytopenia events using diagnoses and measurements                         |
| <input type="checkbox"/> | [EHDEN RA] Pancytopenia or leukopenia events using diagnoses and measurements           |
| <input type="checkbox"/> | [EHDEN RA] Stroke (ischemic or hemorrhagic) events (any visit) (1)                      |
| <input type="checkbox"/> | [EHDEN RA] Opportunistic Infections (2)   |
| <input type="checkbox"/> | [EHDEN RA] Serious Infection events (2)   |
| <input type="checkbox"/> | [EHDEN RA] Serious Infection, opportunistic infections and other infections of interest |

## Model Settings

+ Add Model Settings

Show 10 entries

Filter:

| Remove                   | Model                           | Options                         |
|--------------------------|---------------------------------|---------------------------------|
| <input type="checkbox"/> | LassoLogisticRegressionSettings | (["variance":0.01,"seed":null]) |

Showing 1 to 1 of 1 entries

Previous

## Covariate Settings

+ Add Covariate Settings

Column visibility Copy CSV Show 10 entries

Filter:

| Remove                   | Options   |
|--------------------------|---|
| <input type="checkbox"/> | DemographicsGender, DemographicsAge, DemographicsAgeGroup, DemographicsIndexMonth, ConditionOccurrenceAnyTimePrior, DrugGroupEraLongTerm (+4 more covariate settings) |
| <input type="checkbox"/> | DemographicsGender, DemographicsAgeGroup, DemographicsIndexMonth, ConditionOccurrenceAnyTimePrior, CharlsonIndex, Dcsi (+1 more covariate settings)                   |

Showing 1 to 2 of 2 entries

Previous

## Population Settings

+ Add Population Settings

Column visibility Copy CSV Show 10 entries

Filter:

| Remove                   | Risk Window Start           | Risk Window End              | Washout Period | Include All Outcomes | Remove Subjects With Prior Outcome | Minimum Time At Risk |
|--------------------------|-----------------------------|------------------------------|----------------|----------------------|------------------------------------|----------------------|
| <input type="checkbox"/> | 1d from cohort start date   | 730d from cohort start date  | 365d           | true                 | true                               | 1d                   |
| <input type="checkbox"/> | 1d from cohort start date   | 90d from cohort start date   | 365d           | true                 | true                               | 1d                   |
| <input type="checkbox"/> | 365d from cohort start date | 1826d from cohort start date | 365d           | true                 | true                               | 1d                   |

Previous

# Generate R-Package and share with the world

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 132 Prediction Problem Settings 22 Analysis Settings 6

Column visibility Copy CSV Show 10 entries Filter:

|  | Target Cohort Name  | Outcome Cohort Name   | Model Name                      | Model Settings                | Covariate Settings               | Risk Window Start | Risk Window End |
|--|---|---|---------------------------------|-------------------------------|----------------------------------|-------------------|-----------------|
| Target Cohorts   |   |   |                                 |                               |                                  |                   |                 |
| [EHDEN RA] New users of methotrexate monotherapy used for PLP (66) | [EHDEN RA] New users of methotrexate monotherapy used for PLP | [EHDEN RA] Serious Infection, opportunistic infections and other infections of interest event (1) | LassoLogisticRegressionSettings | {"variance":0.01,"seed":null} | "attr_class":"covariateSettin... | 1                 | 730             |
| [EHDEN RA] Female new users of methotrexate monotherapy            |   |   |                                 |                               |                                  |                   |                 |
| Outcome Cohorts  |   |   |                                 |                               |                                  |                   |                 |
| [EHDEN RA] Leukopenia events using diagnoses and measurements (12) | [EHDEN RA] New users of methotrexate monotherapy used for PLP | [EHDEN RA] Serious Infection, opportunistic infections and other infections of interest event (1) | LassoLogisticRegressionSettings | {"variance":0.01,"seed":null} | "attr_class":"covariateSettin... | 1                 | 90              |
| [EHDEN RA] Opportunistic Infections (2) (12)                       |   |   |                                 |                               |                                  |                   |                 |
| Model Name   |   |   |                                 |                               |                                  |                   |                 |
| LassoLogisticRegressionSettings (132)                              | [EHDEN RA] New users of methotrexate monotherapy used for PLP | [EHDEN RA] Serious Infection, opportunistic infections and other infections of interest event (1) | LassoLogisticRegressionSettings | {"variance":0.01,"seed":null} | "attr_class":"covariateSettin... | 365               | 1826            |
| Risk Window  |   |   |                                 |                               |                                  |                   |                 |
| 1,00 / 144   | [EHDEN RA] New users  | [EHDEN RA] Serious Infection,   |                                 |                               |                                  |                   |                 |



# Share model performance

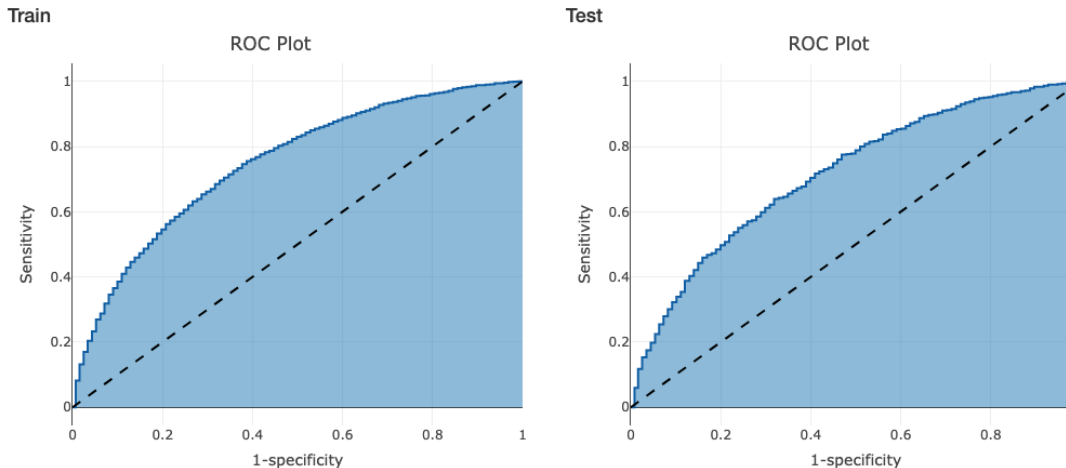


## Prediction Viewer

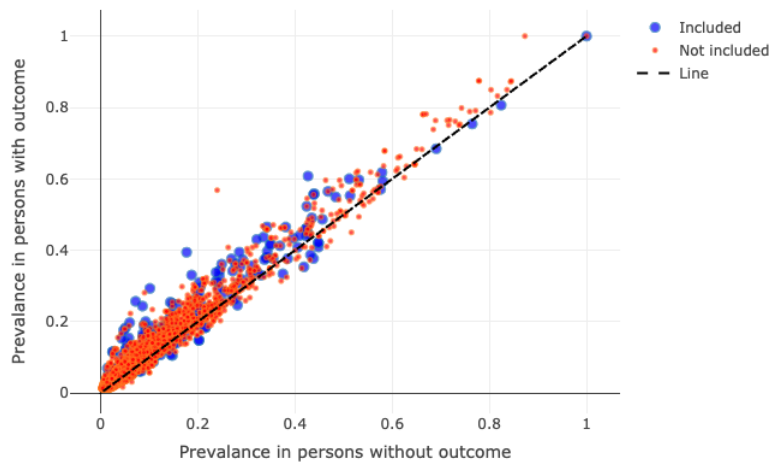
About Internal Validation External Validation

Evaluation Summary Characterization ROC Calibration Demographics Preference Box Plot Settings

Plot Table



**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).

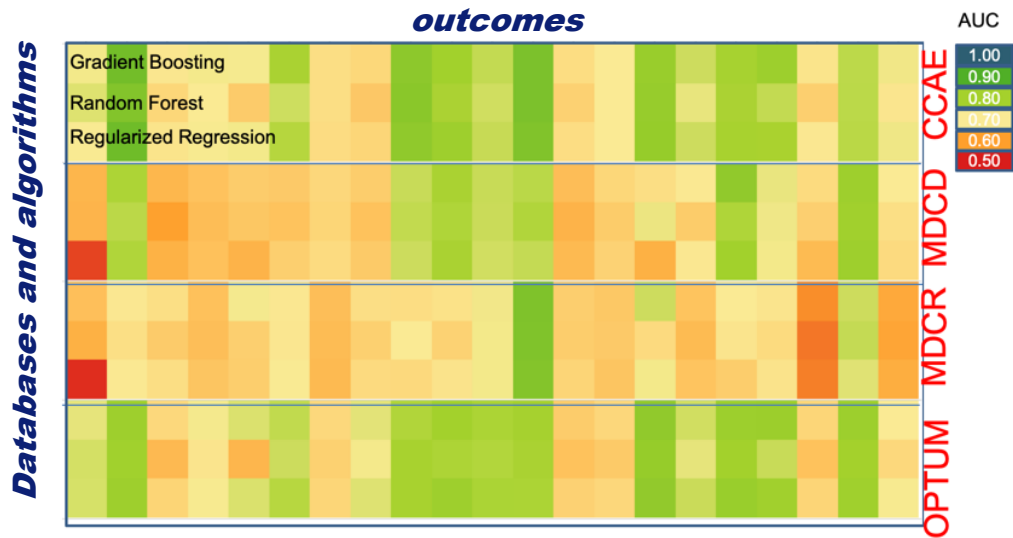


**Figure 1.** The variable scatter plot shows the mean covariate value for the people with the outcome against the mean covariate value for the people without the outcome. The meaning of the size and color of the dots depends on the settings on the left of the figure.

# Large scale validation and dissemination

The tool auto generates a word document containing all the model specifications, internal and external validation results, model details etc. etc. which serves as a kickstart for result dissemination.

Multiple interesting visualisation can be created:



*Development Database*

# Seek 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.

> BMC Med Res Methodol. 2022. Jan 30;22(1):35. doi: 10.1186/s12874-022-01505-z.

Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network

Ross D Williams # 1, Aniek F Markus # 1, Cynthia Yang 1, Talita Duarte-Salles 2, Scott L DuVal 3, Thomas Falconer 4, Jitendra Jonnagaddala 5, Chungsoo Kim 6, Yeunsook Rho 7, Andrew E Williams 8, Amanda Alberga Machado 9, Min Ho An 10, María Aragón 2, Carlos Areia 11, Edward Burn 2 12, Young Hwa Choi 13, Iannis Drakos 14, Maria Tereza Fernandes Abrahão 15, Sergio Fernández-Bertolín 2, George Hripscak 4, Benjamin Skov Kaas-Hansen 16 17, Prasanna L Kandukuri 18, Jan A Kors 1, Kristin Kostka 19, Siaw-Teng Liaw 5, Kristine E Lynch 3, Gerardo Machnicki 20, Michael E Matheny 21 22, Daniel Morales 23, Fredrik Nyberg 24, Rae Woong Park 25, Albert Prats-Urbe 12, Nicole Pratt 26, Gotham Rao 27, Christian G Reich 19, Marcela Rivera 28, Tom Seinen 1, Azza Shoabi 27, Matthew E Spohnitz 4, Ewout W Steyerberg 29 30, Marc A Suchard 31, Seng Chan You 25, Lin Zhang 32 33, Lili Zhou 18, Patrick B Ryan 27, Daniel Prieto-Alhambra 12, Jenna M Reips # 27, Peter R Rijnbeek # 34

Affiliations + expand

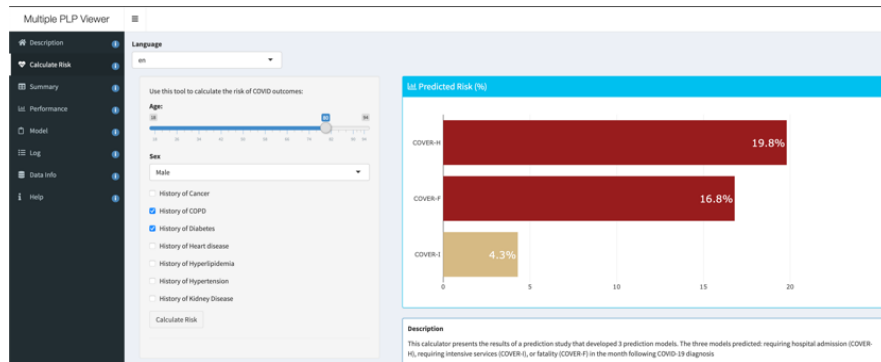
PMID: 35094685 PMCID: PMC8801189 DOI: 10.1186/s12874-022-01505-z

[Free PMC article](#)


## Abstract

**Background:** We investigated whether we could use influenza data to develop prediction models for COVID-19 to increase the speed at which prediction models can reliably be developed and validated early in a pandemic. We developed COVID-19 Estimated Risk (COVER) scores that quantify a patient's risk of hospital admission with pneumonia (COVER-H), hospitalization with pneumonia requiring intensive services or death (COVER-I), or fatality (COVER-F) in the 30-days following COVID-19 diagnosis using historical data from patients with influenza or flu-like symptoms and tested this in COVID-19 patients.

- **14 data sources from 6 countries**
- **Externally validated in 44,507 COVID cases from 5 data sources in South Korea, Spain, USA**



# Training prediction models for individual risk assessment of postoperative complications after surgery for colorectal cancer

[V. Lin](#) , [A. Tsouchnika](#), [E. Allakhverdiiev](#), [A. W. Rosen](#), [M. Gögenur](#), [J. S. R. Clausen](#), [K. B. Bräuner](#), [J. S. Walbech](#), [P. Rijnbeek](#), [I. Drakos](#) & [I. Gögenur](#)

[Techniques in Coloproctology](#) (2022) | [Cite this article](#)

108 Accesses | 1 Altmetric | [Metrics](#)

## Abstract

---

### Background

The occurrence of postoperative complications and anastomotic leakage are major drivers of mortality in the immediate phase after colorectal cancer surgery. We trained prediction models for calculating patients' individual risk of complications based only on preoperatively available data in a multidisciplinary team setting. Knowing prior to surgery the probability of developing a complication could aid in improving informed decision-making by surgeon and patient and individualize surgical treatment trajectories.

### Methods

All patients over 18 years of age undergoing any resection for colorectal cancer between January 1, 2014 and December 31, 2019 from the nationwide Danish Colorectal Cancer Group



# We need external validation at scale!

JMIR Med Inform. 2021 Apr; 9(4): e21547.  
Published online 2021 Apr 5. doi: 10.2196/21547  
PMCID: PMC8023380  
PMID: 33661754

## Implementation of the COVID-19 Vulnerability Index Across an International Network of Health Care Data Sets: Collaborative External Validation Study

Monitoring Editor: Christian Lovis

Reviewed by David Maslove, JianLi Wang, and Anoop Austin

Jenna M Repp, BSc, MSc, PhD,<sup>21</sup> Chungsoo Kim, PharmD,<sup>2</sup> Ross D Williams, MSc,<sup>3</sup> Aniek F Markus, BSc, MSc,<sup>3</sup> Cynthia Yang, BSc, MSc,<sup>3</sup> Talita Duarte-Salles, MPH, PhD,<sup>4</sup> Thomas Falconer, BSc, MSc,<sup>5</sup> Jitendra Jonnagaddala, MIS, PhD,<sup>6</sup> Andrew Williams, PhD,<sup>7</sup> Sergio Fernández-Bertolin, MSc,<sup>4</sup> Scott L DuVall, PhD,<sup>8</sup> Kristin Kostka, MPH,<sup>9</sup> Gowtham Rao, MD, PhD,<sup>1</sup> Azza Shoalbi, PhD,<sup>1</sup> Anna Ostropolets, MD,<sup>5</sup> Matthew E Spolnitz, MPH, MD,<sup>5</sup> Lin Zhang, PhD,<sup>10,11</sup> Paula Casajust, BSc, MSc,<sup>12</sup> Ewout W Steyerberg, MSc, PhD,<sup>13,14</sup> Fredrik Nyberg, MPH, MD, PhD,<sup>15</sup> Benjamin Skov Kaas-Hansen, MSc, MD,<sup>16,17</sup> Young Hwa Choi, MD, PhD,<sup>18</sup> Daniel Morales, PhD, MChB,<sup>19</sup> Siaw-Teng Liaw, PhD,<sup>6</sup> Maria Tereza Fernandes Abrahão, PhD,<sup>20</sup> Carlos Areia, MSc, PT,<sup>21</sup> Michael E Matheny, MD, MPH, MS,<sup>22</sup> Kristine E Lynch, PhD,<sup>6</sup> Maria Aragón, MSc,<sup>4</sup> Rae Woong Park, MD, PhD,<sup>23</sup> George Hripcaak, MD, MS,<sup>5</sup> Christian G Reich, MD, PhD,<sup>9</sup> Marc A Suchard, MD, PhD,<sup>24</sup> Seng Chan You, MD, MS,<sup>23</sup> Patrick B Ryan, PhD,<sup>1</sup> Daniel Prieto-Alhambra, MD, PhD,<sup>25</sup> and Peter R Rijnbeek, PhD<sup>3</sup>

Reps et al. BMC Medical Research Methodology (2021) 20:102  
https://doi.org/10.1186/s12874-020-00991-3  
BMC Medical Research Methodology

## RESEARCH ARTICLE

Open Access

### Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation

Jenna M. Repp<sup>1</sup>, Ross D. Williams<sup>2</sup>, Seng Chan You<sup>1</sup>, Thomas Falconer<sup>4</sup>, Evan Minty<sup>2</sup>, Alison Callahan<sup>1</sup>, Patrick B. Ryan<sup>1</sup>, Rae Woong Park<sup>23</sup>, Hong-Seok Lim<sup>1</sup> and Peter Rijnbeek<sup>2</sup>

**Abstract**  
**Background:** To demonstrate how the Observational Healthcare Data Science and Informatics (OHDSI) collaborative network and standardization can be utilized to scale-up external validation of patient-level prediction models by enabling validation across a large number of heterogeneous observational healthcare datasets.  
**Methods:** Five previously published prognostic models (ATRIA, CHADS<sub>2</sub>, CHADS<sub>2</sub>/VASC, Q-Stroke and Framingham) that predict future risk of stroke in patients with atrial fibrillation were replicated using the OHDSI frameworks. A network study was run that enabled the five models to be externally validated across nine observational healthcare datasets spanning three countries and five independent sites.  
**Results:** The five existing models were able to be integrated into the OHDSI framework for patient-level prediction and they obtained mean c-statistics ranging between 0.57–0.63 across the 6 databases with sufficient data to predict stroke within 1 year of initial atrial fibrillation diagnosis for females with atrial fibrillation. This was comparable with existing validation studies. The validation network study was run across nine datasets within 60 days once the models were replicated. An R package for the study was published at <https://github.com/OHDSI/StudyProtocolSandbox/tree/master/ExistingStrokeRiskExternalValidation>.  
(Continued on next page)

Comment on this paper

## Using the OHDSI network to develop and externally validate a patient-level prediction model for Heart Failure in Type II Diabetes Mellitus

Ross D. Williams, Jenna M. Repp, Jan A. Kors, Patrick B. Ryan, Ewout Steyerberg, Katia M. Verhamme, Peter R. Rijnbeek  
doi: <https://doi.org/10.1101/2021.04.06.21254966>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract Full Text Info/History Metrics Preview PDF

**Abstract**  
**Introduction** Heart Failure (HF) and Type 2 Diabetes Mellitus (T2DM) frequently coexist and exacerbate symptoms of each other. Treatments are available for T2DM that also provide beneficial treatment effects for HF. Guidelines recommend that patients with HF should be given Sodium-glucose co-transporter-2 inhibitors in preference to other second-line treatments for T2DM. Increasing personalization of treatment means that patients who have or are at risk of HF receive a customised treatment. We aimed to develop and externally validate prediction models to predict the 1-year risk of incident HF in T2DM patients starting second-line treatment.

Join the network!

# Acknowledgement

***This work would not have been possible without many contributors in the OHDSI community and the EHDEN project***



[www.ehden.eu](http://www.ehden.eu)



[@IMI\\_EHDEN](https://twitter.com/IMI_EHDEN)



[IMI\\_EHDEN](https://www.linkedin.com/company/IMI_EHDEN)



[github.com/EHDEN](https://github.com/EHDEN)



innovative  
medicines  
initiative



*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.*

Erasmus MC

