Welcome to the European OHDSI Journey

Prof. Dr. Ir. Peter R. Rijnbeek
Professor of Medical Informatics
Chair Department of Medical Informatics
Erasmus MC, The Netherlands
Thank you for your support!
Welcome on the SS Rotterdam

The SS Rotterdam – ‘La Grande Dame’
Launched on 14 December 1956 here in Rotterdam
Originally served as a transatlantic line to connect Rotterdam with New York
Open Science Community

Driving Scalability of Reliable Evidence Generation
Changing the Paradigm

Interoperability

Community

Common Analytics
Objectives of OHDSI Europe

• Enable the generation of reliable evidence from European health data: promote the adoption of the OMOP-CDM and analytics.

• Focus on European Challenges and Opportunities.

• Community building
  – Point of contact for all stakeholders
  – Organization of European OHDSI Symposia
  – Training of stakeholders
  – Stimulate national and international collaborations in Europe
National nodes

- Belgium
- Germany
- Greece
- Italy
- Luxemburg
- Netherlands
- Portugal
- Spain
- United Kingdom
First Annual OHDSI Symposium, March 23th 2018

- 200 participants
- 24 countries
- 40 posters
- 5 software demos
- 2 full day tutorials
Second Annual OHDSI Symposium, March 29th 2019

- 250 participants
- 27 countries
- 35 posters
- 8 software demos
- 5 full day tutorials
Third Annual OHDSI Symposium, June 24th 2022

- 350 participants
- 80 posters
- 4 software demos
- 2 days with tutorials and workshops
Meeting Goals Fourth OHDSI Symposium

- Educate and train the community through workshop (Saturday) and multiple Workgroup Meetings (Sunday) (180 participants)

- Facilitate meetings for multiple projects and initiatives that use the OMOP CDM to generate evidence for patient care

- 350 Participants, 100+ submissions for collaborator showcase.

We hope you will learn a lot and fill your notebook with valuable information!
Thanks to all faculty!!
Breakdown of Participants: 28 Countries

The Netherlands
Belgium
Germany
Portugal
Greece
Denmark
Finland
Estonia
Czechia
Lithuania
Ukraine
Luxembourg
Hungary
Latvia
United Kingdom
Spain
Italy
United States
Norway
Switzerland
France
India
Sweden
Hong Kong
Saudi Arabia
Israel
Republic of Korea
China
Breakdown of Participants: Stakeholders

- Academia
- Technology
- Health System
- Government
- Pharmaceutical
Relationship with OHDSI

- New to OHDSI: 30%
- Participate OHDSI meetings: 39%
- Database conversion: 30%
- OMOP CDM instance: 28%
- Use OHDSI tools & methods: 37%
- Participate OHDSI Forum: 19%
- Participate OHDSI studies: 19%
- Visited OHDSI Europe Symp: 30%
- Visited OHDSI Global Symp: 21%

N=350
<table>
<thead>
<tr>
<th>Time</th>
<th>Title</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:10 – 9:40</td>
<td><strong>Journey of OHDSI: Where have we been and where can we go together?</strong> Speaker: Patrick Ryan, PhD, Janssen Research and Development, Department of Biomedical Informatics, Columbia University Medical Center</td>
<td>Theatre</td>
</tr>
<tr>
<td>9:40 – 11:00</td>
<td><strong>European Initiatives Using the OMOP CDM</strong> Moderator: Renske Los, PhD, Assistant Professor of Medical Informatics, Department of Medical Informatics, Erasmus MC Multiple presentations of European Projects and Initiatives</td>
<td>Theatre</td>
</tr>
<tr>
<td>11:00 – 11:30</td>
<td><strong>Coffee Break</strong></td>
<td>Queen’s Lounge</td>
</tr>
<tr>
<td>11:30 – 12:45</td>
<td><strong>Collaborator Showcase: Rapid fire presentations</strong> Moderator: Katia Verhamme, MD, Associate Professor of Use and Analysis of Observational Data, Department of Medical Informatics, Erasmus MC, Rotterdam.</td>
<td>Theatre</td>
</tr>
<tr>
<td>12:45 – 13:45</td>
<td><strong>Lunch</strong></td>
<td>La Fontaine &amp; Odyssee Room</td>
</tr>
<tr>
<td>Time</td>
<td>Title</td>
<td>Location</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
</tbody>
</table>
| 13:00 – 14:30 | **OHDSI Collaborator Showcase**  
Poster presentations and open-source software demonstrations from OHDSI collaborators:  
- Observational data standards and management  
- Open-source analytics development  
- National nodes | La Fontaine & Odyssee Room, Queen’s Lounge   |
| 14:30 – 16:00 | **OHDSI Collaborator Showcase**  
- Clinical applications  
- Methodological research | La Fontaine & Odyssee Room, Queen’s Lounge   |
| 16:00 – 16:30 | **Real-World Evidence in use in Medicines Regulation**  
Speaker: Andrej Segec, European Medicines Agency | Theatre   |
| 16:35 – 17:45 | **Data Analysis and Real World Interrogation Network (DARWIN EU®)**  
Multiple Speakers from the DARWIN EU® Coordination Centre  
Q&A Session | Theatre   |
| 17:45 – 18:00 | **Closure**                          | Theatre   |
| 18:00 – 19:30 | **Networking Reception**                   | Queen’s Lounge   |
Journey of OHDSI: Where have we been and where we can go together?

Patrick Ryan, PhD
Johnson & Johnson
Columbia University Irving Medical Center
OHDSI’s mission

To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care
To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.
Our Journey
Where The OHDSI Community Has Been
And Where We Are Going
2022 edition

OMOP Common Data Model

Concept

Concept_relationship

Concept_ancestor

Vocabulary

Source_to_concept_map

Relationship

Concept_synonym

Drug_strength

Standardized vocabularies

Domain

Concept_class

Condition_era

Drug_era

Dose_era

Results schema

Cohort

Cohort_definition

Standardized derived elements

Cost

Payer_plan_period

Standardized health economics

Standardized health system

Location

Care_site

Provider

Standardized clinical data

Person

Visit_occurrence

Visit_detail

Condition_occurrence

Drug_exposure

Procedure_occurrence

Device_exposure

Measurement

Observation

Note

Note_NLP

Episode

Specimen

Fact_relationship

Observation_period

Death

Standardized derived elements

Standardized metadata

CDM_source

Metadata
OHDSI vocabularies

OHDSI Vocabularies By The Numbers
• 10,218,572 concepts
• 3,549,524 standard concepts
• 789,207 classification concepts
• 81,243,356 concept relationships
• 85,241,004 ancestral relationships
• 3,268,183 concept synonyms

1 Shared Resource to Enable Data Standards
OHDSI data partners

OHDSI Data By The Numbers

- 453 data sources
  - 374 EHRs
  - 34 registries
  - 30 administrative claims
- 41 countries
- 928 million unique patient records
  (12% of world’s population)
Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation

Alexandra Pacuraru, Kelly Plueschke, Patricia McGettigan, Daniel R Morales, Jim Slattery, Dagmar Vogl, Thomas Goedecke, Xavier Kurz, Alison Cave

Results A total of 34 EHDs were selected after applying key criteria relevant for regulatory purposes. The most represented regions were Northern, Central and Western Europe. The most frequent types of data source were electronic medical records (44.1%) and record linkage systems (29.4%). The median number of patients registered in the 34 data sources was 5 million (range 0.07–15 million) while the median time covered by a database was 18.5 years. Paediatric patients were included in 32 databases (94%). Completeness of information on drug exposure was variable. Published validation studies were found for only 17 databases (50%). Some level of access exists for 25 databases (73.5%), and 23 databases (67.6%) can be linked through a personal identification number to other databases with parent–child linkage possible in 7 (21%) databases. Eight databases (23.5%) were already transformed or were in the process of being transformed into a common data model that could facilitate multidatabase studies.

Conclusion A few European databases meet minimal regulatory requirements and are readily available to be used in a regulatory context. Accessibility and validity information of the included information needs to be improved. This study confirmed the fragmentation, heterogeneity and lack of transparency existing in many European EHDs.
European data standardization from EHDEN

A federated network of Data Partners

The EHDEN project aims to collaborate with diverse institutions, data sources and data custodians across the EU, with a goal of harmonising source data to the OMOP common data model locally, within a federated network.

Following the 7 open calls to date we have organised, we currently have 187 Data Partners from 29 different countries which are mapping their data to the OMOP common data model. This includes several EHDEN project partners who have also mapped their data to the OMOP CDM for use in the federated network.

https://www.ehden.eu/datapartners/
OHDSI workgroups

- APAC (Asia-Pacific)
  - Current Participants: 269
  - Lead: Mui Van Zandt

- ATLAS/WebAPI
  - Current Participants: 226
  - Lead: Anthony Sena

- Clinical Trials
  - Current Participants: 252
  - Leads: Mike Hamidi, Lin Zhen

- Common Data Model
  - Current Participants: 596
  - Lead: Clair Blacketer

- Data Quality Dashboard Development
  - Current Participants: 260
  - Lead: Clair Blacketer

- Early-Stage Researchers
  - Current Participants: 214
  - Leads: Faizaah Arshad, Ross Williams

- Medical Imaging
  - Current Participants: 114
  - Leads: Paul Nagy, Seng Chan You

- Natural Language Processing
  - Current Participants: 379
  - Lead: Hua Xa

- Oncology
  - Current Participants: 241
  - Lead: Aseih Golozar

- Education
  - Current Participants: 116
  - Lead: Nigel Hughes

- Eye Care & Vision Research
  - Current Participants: 40
  - Leads: Sally Baxter, Kerry Goetz

- FHIR and OMOP
  - Current Participants: 214
  - Leads: Jon Duke, Christian Reich, Dana Stephenson

- Open-Source Community
  - Current Participants: 118
  - Leads: Adam Black, Paul Nagy

- Patient-Level Prediction
  - Current Participants: 355
  - Leads: Jenna Reps, Ross Williams

- Phenotype Development & Evaluation
  - Current Participants: 249
  - Lead: Gowtham Rao

- Geographic Information System (GIS)
  - Current Participants: 122
  - Leads: Robert Miller, Andrew Williams

- HADES (Health Analytics Data-to-Evidence Suite)
  - Current Participants: 262
  - Lead: Martijn Schuemie

- Health Equity
  - Current Participants: 201
  - Lead: Jake Gillberg

- Population-Level Effect Estimation
  - Current Participants: 355
  - Leads: Martijn Schuemie, Marc Suchard

- Psychiatry
  - Current Participants: 115
  - Leads: Dmitry Dymshyts, Andrew Williams

- Registry
  - Current Participants: 115
  - Lead: Tina Parciak

- Healthcare Systems
  - Current Participants: 430
  - Lead: Melanie Philcfsky

- Latin America
  - Current Participants: 48
  - Lead: Jose Posada

- Medical Devices
  - Current Participants: 130
  - Leads: Vojtech Huser, Asiyah Lin

- Steering Group
  - Current Participants: 70
  - Lead: Patrick Ryan

- Surgery and Perioperative Medicine
  - Current Participants: 37
  - Lead: Evan Minty

- Vaccine Vocabulary
  - Current Participants: 76
  - Lead: Adam Black
### OHDSI regional chapters

<table>
<thead>
<tr>
<th>Region</th>
<th>Current Participants</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>66</td>
<td>Nega Gebreyesus</td>
</tr>
<tr>
<td>Australia</td>
<td>74</td>
<td>Nicole Pratt</td>
</tr>
<tr>
<td>China</td>
<td>228</td>
<td>Hua Xu</td>
</tr>
<tr>
<td>Europe</td>
<td>321</td>
<td>Peter Rijnbeek</td>
</tr>
<tr>
<td>India</td>
<td>new</td>
<td>Swetha Kiranmayi Jakuva</td>
</tr>
<tr>
<td>Japan</td>
<td>49</td>
<td>Tatsuo Hiramatsu</td>
</tr>
<tr>
<td>Korea</td>
<td>55</td>
<td>Seng Chan You</td>
</tr>
<tr>
<td>Singapore</td>
<td>58</td>
<td>Mengling Feng</td>
</tr>
<tr>
<td>Taiwan</td>
<td>71</td>
<td>Jason Hsu</td>
</tr>
</tbody>
</table>

And now National Nodes!
OHDSI + large community initiatives
OHDSI publications

Publications & Cumulative Citations

Summary

581
PubMed Manuscripts

3434
PubMed Authors

https://dash.ohdsi.org/
2023 publication in data standards

Short communication

Representing and utilizing clinical textual data for real world studies: An OHDSI approach

Vipina K. Keloth a, Juan M. Banda b, Michael Gurley c, Paul M. Heider d, Georgina Kennedy e, Hongfang Liu f, Feifan Liu g, Timothy Miller h, Karthik Natarajan i, Olga V Patterson j, k, l, Yifan Peng m, Kalpana Raja n, Ruth M. Reeves n, o, Masoud Rouhizadeh o, p, Jianlin Shi h, k, r, Xiaoyan Wang q, Yanshan Wang r, Wei-Qi Wei s, Andrew E. Williams u, Rui Zhang v, Rimma Belenkaya w, Christian Reich x, Clair Blacketer y, z, Patrick Ryan h, l, George Hripcsak i, Noémie Elhadad h, k, Hua Xu a, z

Fig. 2. An overview of the workflow for transforming clinical text in the NOTE table.
2023 publication in methodological research

**Drug Safety**
https://doi.org/10.1007/s40264-023-01324-1

**ORIGINAL RESEARCH ARTICLE**

**Serially Combining Epidemiological Designs Does Not Improve Overall Signal Detection in Vaccine Safety Surveillance**

Faaizah Arshad¹,² · Martijn J. Schuemie¹,²,³ · Fan Bu¹,² · Evan P. Minty⁴ · Thamir M. Alshammari⁵ · Lana Y. H. Lai⁶ · Talita Duarte-Salles⁷ · Stephen Fortin³ · Fredrik Nyberg⁸ · Patrick B. Ryan²,³ · George Hripcsak²,⁹,¹⁰ · Daniel Prieto-Alhambra¹¹,¹² · Marc A. Suchard¹,²,¹³,¹⁴

**Table 1**

<table>
<thead>
<tr>
<th>Design</th>
<th>Type 1 Error</th>
<th>Type 2 Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>Group 2</td>
<td>0.10</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Graph 1**

Type I and II Errors Without Empirical Calibration for all Databases
A standardized framework for risk-based assessment of treatment effect heterogeneity in observational healthcare databases

Alexandros Rekkas,1,3, David van Klaveren,2,3, Patrick B. Ryan,4,5, Ewout W. Steyerberg,6, David M. Kent,6 and Peter R. Rijnbeek1

Treatment effects are often anticipated to vary across groups of patients with different baseline risk. The Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement focused on baseline risk as a robust predictor of treatment effect and provided guidance on risk-based assessment of treatment effect heterogeneity in a randomized controlled trial. The aim of this study is to extend this approach to the observational setting using a standardized scalable framework. The proposed framework consists of five steps: (1) definition of the research aim, i.e., the population, the treatment, the comparator and the outcome(s) of interest; (2) identification of relevant databases; (3) development of a prediction model for the outcome(s) of interest; (4) estimation of relative and absolute treatment effect within strata of predicted risk, after adjusting for observed confounding; (5) presentation of the results. We demonstrate our framework by evaluating heterogeneity of the effect of thiazide or thiazide-like diuretics versus angiotensin-converting enzyme inhibitors on three efficacy and nine safety outcomes across three observational databases. We provide a publicly available R software package for applying this framework to any database mapped to the Observational Medical Outcomes Partnership Common Data Model. In our demonstration, patients at low risk of acute myocardial infarction receive negligible absolute benefits for all three efficacy outcomes, though they are more pronounced in the highest risk group, especially for acute myocardial infarction. Our framework allows for the evaluation of differential treatment effects across risk strata, which offers the opportunity to consider the benefit-harm trade-off between alternative treatments.

npj Digital Medicine (2023)6:58; https://doi.org/10.1038/s41746-023-00794-y
2023 publication in clinical applications

Part of THE LANCET Discovery Science

Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study


65 co-authors
26 databases in 11 countries
492 million patient records

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>Data provenance</th>
<th>Dates covered</th>
<th>Total persons</th>
<th>% F</th>
<th>Ages covered mean</th>
<th>Patients with COVID-19 %</th>
<th>Pre-national background</th>
<th>Measurements with values</th>
</tr>
</thead>
<tbody>
<tr>
<td>APHM</td>
<td>FR</td>
<td>EHR</td>
<td>1998/01/2017/07</td>
<td>2,466,265</td>
<td>51.9</td>
<td>40 (21-60)</td>
<td>11,435</td>
<td>673,031</td>
<td>N</td>
</tr>
<tr>
<td>CPMD_AURUK</td>
<td>UK</td>
<td>GP</td>
<td>1995/01/2017/03</td>
<td>39,379,547</td>
<td>51.8</td>
<td>31 (18-84)</td>
<td>587,886</td>
<td>14,949,032</td>
<td>Y</td>
</tr>
<tr>
<td>CL_AUC</td>
<td>US</td>
<td>EHR</td>
<td>2013/01/2017/04</td>
<td>4,799,392</td>
<td>54.0</td>
<td>39 (22-95)</td>
<td>72,649</td>
<td>8,095,597</td>
<td>Y</td>
</tr>
<tr>
<td>COMM</td>
<td>ES</td>
<td>EHR</td>
<td>2015/01/2017/01</td>
<td>6,808,476</td>
<td>51.8</td>
<td>38 (21-66)</td>
<td>28,044</td>
<td>72,573</td>
<td>Y</td>
</tr>
<tr>
<td>FINAI</td>
<td>ES</td>
<td>EHR</td>
<td>2015/01/2017/08</td>
<td>7,292,395</td>
<td>54.7</td>
<td>41 (33-78)</td>
<td>7,139</td>
<td>78</td>
<td>N</td>
</tr>
<tr>
<td>HIC</td>
<td>SC</td>
<td>EHR</td>
<td>2005/01/2017/12</td>
<td>1,254,646</td>
<td>50.4</td>
<td>45 (18-66)</td>
<td>11,813</td>
<td>881,235</td>
<td>Y</td>
</tr>
<tr>
<td>IBM_CCAE</td>
<td>US</td>
<td>Claims</td>
<td>2000/01/2017/03</td>
<td>159,440,276</td>
<td>51.1</td>
<td>31 (17-66)</td>
<td>983,689</td>
<td>21,483,191</td>
<td>N</td>
</tr>
<tr>
<td>IBM_MOCD</td>
<td>US</td>
<td>Claims</td>
<td>2005/01/2017/02</td>
<td>21,066,807</td>
<td>56.2</td>
<td>24 (3-88)</td>
<td>196,997</td>
<td>11,870,595</td>
<td>Y</td>
</tr>
<tr>
<td>IBM_MCRD</td>
<td>US</td>
<td>Claims</td>
<td>2005/01/2017/02</td>
<td>10,356,249</td>
<td>55.3</td>
<td>73 (26-77)</td>
<td>41,547</td>
<td>1,461,963</td>
<td>Y</td>
</tr>
<tr>
<td>IMAIS</td>
<td>ES</td>
<td>EHR</td>
<td>1994/01/2017/06</td>
<td>974,524</td>
<td>47.4</td>
<td>38 (23-54)</td>
<td>9,139</td>
<td>148,012</td>
<td>Y</td>
</tr>
<tr>
<td>IPC</td>
<td>NL</td>
<td>GP</td>
<td>2006/01/2017/06</td>
<td>2,529,355</td>
<td>51.2</td>
<td>37 (18-85)</td>
<td>91,759</td>
<td>1,329,674</td>
<td>Y</td>
</tr>
<tr>
<td>IQVIA_FRANCE_DA</td>
<td>FR</td>
<td>GP</td>
<td>2005/01/2017/05</td>
<td>5,472,012</td>
<td>51.2</td>
<td>38 (21-68)</td>
<td>2,785</td>
<td>1,394,917</td>
<td>Y</td>
</tr>
<tr>
<td>IQVIA_GERMANY_DA</td>
<td>DE</td>
<td>GP</td>
<td>2006/01/2017/03</td>
<td>5,730,723</td>
<td>55.7</td>
<td>44 (15-82)</td>
<td>45,508</td>
<td>7,048,331</td>
<td>Y</td>
</tr>
<tr>
<td>IQVIA_OPENCLAIMS</td>
<td>US</td>
<td>Claims</td>
<td>2006/01/2017/03</td>
<td>102,000,000</td>
<td>62.4</td>
<td>34 (14-92)</td>
<td>17,480,443</td>
<td>306,000,000</td>
<td>N</td>
</tr>
<tr>
<td>IQVIA_PHARMACISTS</td>
<td>US</td>
<td>Claims</td>
<td>2013/01/2017/09</td>
<td>166,422,594</td>
<td>50.6</td>
<td>38 (19-99)</td>
<td>1,593,578</td>
<td>46,947,246</td>
<td>Y</td>
</tr>
<tr>
<td>IU</td>
<td>TR</td>
<td>EHR</td>
<td>2015/01/2017/01</td>
<td>899,515</td>
<td>53.0</td>
<td>35 (18-82)</td>
<td>619</td>
<td>619</td>
<td>Y</td>
</tr>
<tr>
<td>JAMC</td>
<td>JP</td>
<td>Claims</td>
<td>2005/01/2017/03</td>
<td>12,541,088</td>
<td>48.8</td>
<td>32 (19-49)</td>
<td>17,564</td>
<td>6,680,196</td>
<td>N</td>
</tr>
<tr>
<td>MHD</td>
<td>BE</td>
<td>EHR</td>
<td>2015/01/2017/12</td>
<td>372,131</td>
<td>50.9</td>
<td>37 (21-79)</td>
<td>203</td>
<td>73,754</td>
<td>N</td>
</tr>
<tr>
<td>OPTUM_EHR</td>
<td>US</td>
<td>EHR</td>
<td>2005/01/2017/02</td>
<td>90,454,775</td>
<td>53.3</td>
<td>37 (19-96)</td>
<td>693,334</td>
<td>41,281,647</td>
<td>Y</td>
</tr>
<tr>
<td>OPTUM_SES</td>
<td>US</td>
<td>Claims</td>
<td>2000/01/2017/05</td>
<td>90,283,592</td>
<td>50.5</td>
<td>36 (19-92)</td>
<td>899,969</td>
<td>21,722,611</td>
<td>Y</td>
</tr>
<tr>
<td>SDIAP</td>
<td>ES</td>
<td>GP</td>
<td>2003/01/2017/03</td>
<td>8,022,374</td>
<td>50.1</td>
<td>31 (17-54)</td>
<td>495,739</td>
<td>5,934,449</td>
<td>Y</td>
</tr>
<tr>
<td>STARR</td>
<td>US</td>
<td>EHR</td>
<td>2008/01/2017/04</td>
<td>3,475,673</td>
<td>51.3</td>
<td>36 (18-84)</td>
<td>31,928</td>
<td>1,118,549</td>
<td>Y</td>
</tr>
<tr>
<td>U_OF_TARTU</td>
<td>EU</td>
<td>EHR</td>
<td>2015/01/2017/12</td>
<td>353,557</td>
<td>52.3</td>
<td>32 (19-57)</td>
<td>84,957</td>
<td>378,842</td>
<td>Y</td>
</tr>
<tr>
<td>UCS</td>
<td>ES</td>
<td>EHR</td>
<td>2015/01/2017/03</td>
<td>823,962</td>
<td>54.1</td>
<td>51 (35-67)</td>
<td>16,764</td>
<td>49,643</td>
<td>Y</td>
</tr>
<tr>
<td>UCHW</td>
<td>US</td>
<td>EHR</td>
<td>2012/01/2017/05</td>
<td>316,119</td>
<td>53.8</td>
<td>37 (22-54)</td>
<td>61,037</td>
<td>240,831</td>
<td>Y</td>
</tr>
<tr>
<td>UK_BIOBANK</td>
<td>UK</td>
<td>EHR + Registry</td>
<td>2015/01/2017/02</td>
<td>502,504</td>
<td>54.4</td>
<td>58 (40-69)</td>
<td>1,717</td>
<td>458,889</td>
<td>Y</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23,840,986</td>
<td>492,730,913</td>
<td>Y</td>
</tr>
</tbody>
</table>

APHM = Health Data Warehouse of Assistance Publique – Hôpitaux de Marseille, BE = Belgium, CPMD_AURUK = Clinical Practice Research Datalink AURUK, CL_AUC = University of Colorado Anschutz Medical Campus, CL_ICM = Columbia University Irving Medical Center, DE = Germany, ES = Spain, EHR = Electronic Health Record, ESI = Spain, F = Female, FINAI = Fundación para la Investigación e Innovación Sanitaria en Atención Primaria COVID19, FR = France, GP = General Practice, HC = Health Informatics Centre, IBM_CCAE = IBM MarketScan® Commercial Claims and Encounters Database, IBM_MOCD = IBM MarketScan® Multi-State Medicaid Database, IBM_MOCD = IBM MarketScan® Medicaid Supplemental and Coordination of Benefits Database, IMAIS = Pasc de Salut Mar Barcelona Information System, IPC = Integrated Primary Care Information, IQVIA_OPENCLAIMS = IQVIA Elixion Open Claims, IQVIA_PHARMACISTS = IQVIA Pharmacists, KJ = Istanbul Faculty of Medicine, University of Istanbul, JP = Japan, MHD = Mediam Health Data, N = Netherland, OPTUM_EHR = Optum® De-identified Electronic Health Record Dataset, OPTUM_SES = Optum De-identified Chromatic® Data Mart Database – Socio-Economic Status, RS = Russia, SC = Scotland, SDIAP = The Information System for Research on Primary Care, STAR = Swedish Medical Research Data Repository-CMOP, TR = Turkey, U_OF_TARTU = University of Tartu (U_OF_TARTU), UCS = University Clinical Center of Serbia, UCHW = University of California Health Data Warehouse, UK = United Kingdom, UK_BIOBANK = UK Biobank, US = United States, Y = Yes, COVID-19 only subset, COVID-19 = Controls.

Table 1: Database characteristics.
### YouTube Summary

- **549 Videos Published**
- **164K+ Hours Watched**

### Ehden Learning Summary

- **25 Number of Courses**
- **3397 Course Completions**

---

**OMOP Common Data Model and Standardized Vocabularies**

**15- September-2019**

Christian Reich, MD, PhD, Mui van Zandt
Erica A. Voss, MPH, PMP, Hamed Abedtash, PharmD, PhD
Dmitry Dymshyts, MD, Melanie Philofsky, RN, MS
Don Torok, MS

**2019 Tutorials - OMOP Common Data Model and Standardized Vocabularies (Full Tutorial)**

https://dash.ohdsi.org/
Save Our Sisyphus Challenge

The OHDSI mission is to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. The 2023 Save Our Sisyphus (SOS) Challenge attempted to fulfill that mission, but not through one study at a time.

Our community simultaneously collaborated on four studies, each of which were designed, implemented, executed and will ultimately be disseminated by members of the OHDSI global community.

As you can see on the right, there were two weekly tutorials taught by different members of the community. These focused on two of the four studies (see below), but they also serve as educational tools for people who plan to lead or participate in a network study in the future.
System characteristics:

• Standardized procedures with defined inputs and outputs
• Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
• Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
• Pre-specified and objective decision thresholds for go/no go criteria
• Measurable operating characteristics of system performance
Engineering open science systems that build trust into the real-world evidence generation and dissemination process

‘System’ required elements:
- Required phenotypes
- Analysis specifications
- Decision thresholds

Week 1: Study initiation

Week 0: Research questions
- VEGF→ESRD
- FQ→AA
- MS biologics→PML
- risankizumab→CVA

Final unblinded results
Interface for exploration

VEGF→ESRD
FQ→AA
MS biologics→PML
risankizumab→CVA
Engineering open science systems that build trust into the real-world evidence generation and dissemination process

‘System’ required elements:
- Required phenotypes
- Analysis specifications
- Decision thresholds

Week 2: Data diagnostics

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation
Research question
Database diagnostics
Pass
Fail
STOP

Phenotype development and evaluation
Cohort definitions
Cohort diagnostics
Pass
Fail
STOP

Analysis reliability evaluation
Analysis design choices
Study diagnostics
Pass
Fail
STOP

Final unblinded results
Interface for exploration
Engineering open science systems that build trust into the real-world evidence generation and dissemination process

'Distributed data network, standardized to common data model'

'Network coordination'

Data quality evaluation

Pass

Fail

Research question

Database diagnostics

Cohort definitions

Cohort diagnostics

Analysis reliability evaluation

Analysis design choices

Study diagnostics

Final unblinded results

Interface for exploration

Week 3-4: Phenotype development and evaluation

Required elements:
- Required phenotypes
- Analysis specifications
- Decision thresholds

'System' required elements:
Engineering open science systems that build trust into the real-world evidence generation and dissemination process

- Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Data quality evaluation
- Research question
- Database diagnostics

Phenotype development and evaluation
- Cohort definitions
- Cohort diagnostics

Analysis reliability evaluation
- Analysis design choices
- Study diagnostics

STOP

Network coordination

Week 5: Analysis design

Only possible because of standardized analytics developed across our community

Final unblinded results

Interface for exploration
Engineering open science systems that build trust into the real-world evidence generation and dissemination process

‘System’ required elements:
- Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
- Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration

Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
- Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration

Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
- Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration

Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
-Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration

Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
- Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration

Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
- Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration

Required phenotypes
- Analysis specifications
- Decision thresholds

Distributed data network, standardized to common data model

Network coordination

Data quality evaluation

Research question → Database diagnostics → Phenotype development and evaluation → Analysis reliability evaluation → Final unblinded results

- Pass
- Fail
- STOP

Week 6-8: Analysis execution, study diagnostics, evidence synthesis

Interface for exploration
Future directions of OHDSI

George Hripcsak MD MS
Director, Columbia University OHDSI Coordinating Center
Professor and Chair of Biomedical Informatics
Columbia University Irving Medical Center
NewYork-Presbyterian Hospital

Patrick Ryan PhD
Johnson & Johnson
Columbia University Irving Medical Center
Foundational elements to enable reliable evidence generation

Clinical characterization: What happened to them?

Patient-level prediction: What will happen to me?

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

Evidence use cases:

Foundational pillars:

- Standardized vocabularies
- Standardized data network
- Standardized open-source tools
Pillar #1: Standardized vocabularies

• Opportunity: Increase transparency and maturity with vocabulary development and evaluation process

• Proposed solutions:
  – Conduct landscape assessment to understand community needs
  – Develop code of conduct and developer guidelines
  – Disseminate vocabulary process and end-user documentation and roadmap
  – Establish centralized development infrastructure
  – Create standardized test development
  – Build vocabulary version release distribution service
Landscape assessment

Part I: 188 responses from 144 institutions | Part II: Vocabularies use in 60 data sources

FINDINGS

- 87% of the community feels confident about Vocabularies’ integrity
  (to be used as a benchmark)

- Most used vocabularies: SNOMED, ICD 9/10 (US and int. versions), MedDRA, ICDO3, ATC, RxNorm / RxE, ICD10PCS, ICD9Proc, CPT4, LOINC, CVX, HCPCS, UCM, NDC, NAACCR, Cancer Modifier

- Most of the community updates their data annually or semi-annually
Roadmap

Release planning

This page provides you with the planned maintenance and improvement activities around the CHDS Standardized Vocabularies. This is to be treated as a forecast. Below you can find the content of each release and an overview of the planned improvement activities (detailed content to be posted separately).

Roadmap 2023 Q1 - 2024 Q2:

More information: Vocabulary-v5.0 GitHub Wiki bit.ly/43q8yc6
Community contribution Part I: currently supported use cases

**ADD NEW CONTENT**
- Does not impact the rest of the community
- Can be incorporated quickly (no review)
  - Adding *non-standard* concepts to existing vocabularies
  - Adding synonyms to existing concepts
  - Adding new mappings
  - Adding *non-standard* vocabularies

**MODIFY EXISTING CONTENT**
- Impacts community as already is used by others
- Requires review of the Vocab Team & other WG's/broader community
  - Modifying existing mappings
  - Modifying concept attributes (e.g., domain)
  - Promoting non-standard concepts to standard

More info: [bit.ly/42qQscr](bit.ly/42qQscr)
Pillar #2: Standardized data network

• Opportunity: Increase transparency and maturity of OHDSI data network

• Proposed solutions:
  – Create OHDSI data network catalog to encourage network studies across interested partners and promote data quality practices
  – Generate OHDSI network concept prevalence data and make accessible for ATLAS users to enable more generalizable phenotype development
  – Promote database diagnostics by having data partners share limited subset of ACHILLES to allow for users to identify databases that satisfy study criteria
<table>
<thead>
<tr>
<th>databased</th>
<th>D1: risankizumab vs. other biologics for psoriasis and risk of ischemic stroke</th>
<th>A1: aflibercept vs. bevacizumab for blinding diseases with esrd outcome</th>
<th>B1: fluoroquinolone vs. cephalosporin for urinary tract infection and risk of aortic aneurysm</th>
<th>C2: biologics vs disease modifying treatments for multiple sclerosis and risk of PML</th>
</tr>
</thead>
<tbody>
<tr>
<td>truven_ccae_2324_20230201</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>US_PharMetrics_Plus_20230330</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>optum_ehr_2247_20221205</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>US_OPEN_CLAIMS_20230313</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Japan_HIS_20220120</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>jmdc_2325_20230126</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>truven_mdc_2322_20230127</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pmxb_202203_20220831</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>JHM_OMOP_20230406</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>US_Hospital_20230130</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CUIMC_20221214</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Japan_Claims_20230215</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>VA-OMOP_20230411</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>optum_extended_cod_2323_20230201</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>optum_extended_se_2327_20230204</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>truven_mdc_2359_20230215</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LPD_Italy_20221126</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>France_LPD_20230118</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>premier_3326_20230125</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UK_IMRD_THIN_20221230</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

1-20 of 34 rows
Pillar #3: Standardized open-source tools

- **Opportunity:** Increase adoption and ease-of-use of HADES packages and other OHDSI open-source analytic tools

- **Proposed solutions:**
  - Create central infrastructure to enable testing of all OHDSI tools against each of the supporting database platforms
  - Establish referent benchmark study that all organizations can execute to demonstrate that OHDSI toolstack is operational
  - Improve documentation and educational materials to promote adoption of OHDSI tools
  - Encourage greater community support of open-source development activities: we need more help to maintain our solutions!
OHDSI HADES package releases in 2023 alone...

- jreps
  - Characterization 0.0.5 has been released.
- Chris_Knoll
  - CirceR 1.3.0 has been released.
- jreps
  - ShinyModules v1.0.2 has been released.
  - ShinyAppBuilder v1.1.1 has been released.
- schuemie
  - Martijn Schuemie
  - BigKnn 1.0.2 has been released.
- Gowtham_Rao
  - CohortExplorer v0.0.11 has been updated.
- schuemie
  - Martijn Schuemie
  - ROhdslWebApi 1.3.2 has been released.
- Adam_Black
  - Andromeda 0.6.3 has been released on Github and CRAN.
- jpegilbert
  - Jamie Gilbert
  - ResultModelManager v0.4.0 has been released.
- anthonysena
  - CohortGenerator v0.8.0 was released a few weeks ago.
- jpegilbert
  - Jamie Gilbert
  - CohortDiagnostics version 3.2.1 has been released.
- schuemie
  - Martijn Schuemie
  - SelfControlledCaseSeries 4.2.0 has been released.
- msuchard
  - Marc Suchard
  - BrokenAdaptiveRidge v1.0.0 has been released.
- msuchard
  - Marc Suchard
  - Cyclops v3.3.1 has been released and is now on CRAN.
- schuemie
  - Martijn Schuemie
  - CohortMethod 5.0.0 has been released.
- fanbu
  - EvidenceSynthesis 0.5.0 has been released.
- jswerdel
  - Joel N. Swerdel
  - PhEvaluato 2.2.8 has been released.
- Frank
  - Frank DeFalco
  - Achilles 1.7.2 has been released.

https://forums.ohdsi.org/t/hades-development-announcements/12293
A question to ask yourself throughout the day

If you had:

• an open community of international, multi-disciplinary, cross-stakeholder collaborators
• an open community data standard used by >400 databases around the world
• established and evaluated scientific best practices that can ensure reliable evidence
• a suite of open source tools capable of supporting the entire journey from data to evidence

What would you do?
European Initiatives Using the OMOP CDM

Moderator: Renske Los, PhD, Assistant Professor of Medical Informatics
European Initiatives Using the OMOP CDM

1. European Health Data and Evidence Network: building a sustainable ecosystem for generating reliable evidence in Europe
   Carlos Diaz, Synapse

2. Harmonizing rare cancer data within EURACAN
   Dr. Maaike van Swieten, IKNL

3. HONEUR: Building a federated network in haematology
   Michel van Speybroeck, Janssen Pharmaceutica

4. PIONEER and OPTIMA, two EU-IMI funded big data projects led by the European Association of Urology
   Monique Roobol, Professor Decision Making in Urology, Erasmus MC
European Health Data and Evidence Network: building a sustainable ecosystem for generating reliable evidence in Europe

Carlos Diaz
Synapse
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Carlos Diaz, SYNAPSE
On behalf of the EHDEN consortium
EHDEN IS ABOUT...

Providing 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 (OMOP), significantly speeding up the generation of reliable evidence.
**EHDEN BASICS**

Start date: 1 Nov 2018  
End date: 30 Oct 2024  
Duration: 70 months

25 partners

~€30 million

---

### Universities, public bodies and research organisations

- Erasmus MC
- Academic coordinator
- Uppsala Monitoring Centre
- universidade de aveiro
- NICE
- National Institute for Health and Care Excellence
- University of Oxford

### Small & Mid-sized companies

- SYNAPSE
- ODYSSEUS DATA SERVICES INC
- The Hyve

### Other organisations

- EFPIA & Associated partners
- EFPIA Lead
- Janssen
- Abbvie
- Celgene
- Pfizer
- Boehringer Ingelheim
- Janssen
- Abbvie
- Roche
- Servier
- AstraZeneca
- Sanofi
- UCB
- Novartis
- Lundbeck
EHDEN BASICS

Data and Network

Methods research

Training and education

Tools and infrastructure

Evidence Generation
DATA PARTNER NETWORK [AFTER 7 CALLS]

~850 million records being mapped to OMOP CDM in 29 European countries

https://www.ehden.eu/datapartners/
SME Network [After 4 Calls]

https://www.ehden.eu/business-directory/

Certified SMEs (n=64)

Applications (n=143)

<table>
<thead>
<tr>
<th>Country</th>
<th># of SMEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>7</td>
</tr>
<tr>
<td>Spain</td>
<td>6</td>
</tr>
<tr>
<td>Italy</td>
<td>5</td>
</tr>
<tr>
<td>France</td>
<td>5</td>
</tr>
<tr>
<td>Greece</td>
<td>5</td>
</tr>
<tr>
<td>Belgium</td>
<td>4</td>
</tr>
<tr>
<td>Portugal</td>
<td>4</td>
</tr>
<tr>
<td>Hungary</td>
<td>4</td>
</tr>
<tr>
<td>Germany</td>
<td>4</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>3</td>
</tr>
<tr>
<td>Switzerland</td>
<td>3</td>
</tr>
<tr>
<td>Finland</td>
<td>3</td>
</tr>
<tr>
<td>Estonia</td>
<td>3</td>
</tr>
<tr>
<td>Ukraine</td>
<td>3</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
</tr>
<tr>
<td>Serbia</td>
<td>2</td>
</tr>
<tr>
<td>Poland</td>
<td>2</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>2</td>
</tr>
<tr>
<td>Ireland</td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1</td>
</tr>
<tr>
<td>Austria</td>
<td>1</td>
</tr>
</tbody>
</table>

# of SMEs
METHODS: MANY PUBLICATIONS. SOME EXAMPLES...

TreatmentPatterns: An R package to facilitate the standardized development and analysis of treatment patterns across disease domains

**METHODS**

**Methods:**

A number of publications. Some examples...

**A. Markus**

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Peter R. Rijneveld

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Jenna M. Reps

Janssen Research and Development, Barren, New Jersey, United States

Why predicting risk can’t identify ‘risk factors’: empirical assessment of model stability in machine learning across observational health databases

**A. Markus**

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Peter R. Rijneveld

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Jenna M. Reps

Janssen Research and Development, Barren, New Jersey, United States

**Abstract**

Background: Identifying clinical outcomes remains a fundamental goal of medicine. The prediction of adverse health outcomes using electronic health data is an active area of research.

Objective: To develop a framework for identifying temporal clinical event sequences from observational medical outcomes partnerships.

Methods and materials: A 4-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. The package is used to identify adverse health outcomes using Electronic Health Data.


Using Iterative Pairwise External Validation to Contextualize Prediction Model Performance: A Use Case Predicting 1-Year Heart Failure Risk in Patients with Diabetes Across Five Data Sources

**A. Markus**

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Peter R. Rijneveld

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Jenna M. Reps

Janssen Research and Development, Barren, New Jersey, United States

**Abstract**

Objective: To develop a framework for identifying temporal clinical event sequences from observational medical outcomes partnerships.

Methods and materials: A 4-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. The package is used to identify adverse health outcomes using Electronic Health Data.


**Frequentist**

Logistic regression models for patient-level prediction based on massive observational data: Do we need all data?

**A. Markus**

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Peter R. Rijneveld

Department of Medical Informatics, Erasmus University Medical Center Rotterdam, The Netherlands

Jenna M. Reps

Janssen Research and Development, Barren, New Jersey, United States

**Abstract**

Objective: To develop a framework for identifying temporal clinical event sequences from observational medical outcomes partnerships.

Methods and materials: A 4-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. The package is used to identify adverse health outcomes using Electronic Health Data.


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

Cynthia Yang, Jan Alexandrinos Rekkas, Peter R Rijneveld

**Abstract**

Objective: To develop a framework for identifying temporal clinical event sequences from observational medical outcomes partnerships.

Methods and materials: A 4-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. The package is used to identify adverse health outcomes using Electronic Health Data.


Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: A multinational real-world cohort analysis

Cynthia Yang, Ross D Williams, Joel N Swedo, Joao Rafeij Almeida, Emily B Brower, Edward Burn, Lorentz Coenen, Carine Claudel, Twu Tulsana, Peter R Rijneveld, Wald Fakhoury, Ant J Aman, Lambe Kullena, Henry Morgan, Bruns, Carmen G Torre, K. Daniel Prieto-Alhambra

**Abstract**

Objective: To develop a framework for identifying temporal clinical event sequences from observational medical outcomes partnerships.

Methods and materials: A 4-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. The package is used to identify adverse health outcomes using Electronic Health Data.


90–Day all-cause mortality can be predicted following a total knee replacement: an international network study to develop and validate a prediction model

Ross D Williams, Jenna M Reps, OHSIE/OEHD Knee Arthroplasty Group;
TRAINING: EHDEN ACADEMY – FREE ONLINE TRAINING FOR ALL

19 courses covering:

- EHDEN Academy
- SKILLS
- TOOLS
- METHODS

and a course for non-experts

> 4,000 students

Also used as a training program for certification of SMEs to support Data Partners mapping their data to the OMOP CDM
TOOLS: THE EHDEN PORTAL – ONE-STOP-SHOP LAUNCHED 24 JUNE 2022

https://www.ehden.eu/ehden-portal/

Free Enrolment!

<table>
<thead>
<tr>
<th># databases in Database Catalogue</th>
<th># databases in Network Dashboard</th>
<th># countries</th>
<th># patients</th>
<th># users</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>35</td>
<td>15</td>
<td>~ 44 Mio</td>
<td>~ 400</td>
</tr>
<tr>
<td>118</td>
<td>90</td>
<td>23</td>
<td>~ 149 Mio</td>
<td>~ 900</td>
</tr>
<tr>
<td>76%</td>
<td>157%</td>
<td>53%</td>
<td>238%</td>
<td>125%</td>
</tr>
<tr>
<td>Total number at end of project</td>
<td>187</td>
<td>187</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>

Network Dashboards

Continuously updated
EVIDENCE GENERATION EXAMPLE: LONG-COVID STUDY-A-THON

13 Data Partners

- GOLD
- AURUM

ADDITIONAL DPs (online)  Pharmetrics+, IQVIA (USA), HSD (Italy), Ajou University (S Korea)

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-10:00</td>
<td>Kick-off</td>
<td>Café break</td>
<td>Café break</td>
<td>Café break</td>
<td>Café break</td>
</tr>
<tr>
<td>10:00-11:00</td>
<td>5:00-5:30 Missouri (data session)</td>
<td>5:00-5:30 Missouri (data session)</td>
<td>5:00-5:30 Missouri (data session)</td>
<td>5:00-5:30 Missouri (data session)</td>
<td>5:00-5:30 Missouri (data session)</td>
</tr>
<tr>
<td>11:00-13:00</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
</tr>
<tr>
<td>13:00-14:00</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>14:00-18:00</td>
<td>Welcome and introductions</td>
<td>Manuscript discussion</td>
<td>All progress update</td>
<td>All progress update</td>
<td>All progress update</td>
</tr>
<tr>
<td>15:30-16:00</td>
<td>Literature review</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
</tr>
<tr>
<td>16:00-17:00</td>
<td>Introductions of the databases</td>
<td>All updates and summary plan for the first days</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
<td>Breakout session 1-3 groups</td>
</tr>
</tbody>
</table>

Team 1: Work Package
- Incidence
- Characteristics and Outcomes
- Trajectories

Team 2: Expert contributors
- Long COVID Experts
- Data Scientists
- Pharmacists
- Epidemiologists

Team 3: Technical experts
- Programming
- Statistics
- Data Management

Team 4: Site coordinators
- GOLD
- AURUM

GOLD: Pharmetrics+, IQVIA (USA), HSD (Italy), Ajou University (S Korea)

AURUM: Long COVID Study-A-Thon
EVIDENCE GENERATION: EXAMPLE OUTPUTS

Long COVID characterisation

Health conditions: individual cohorts

Table of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
</table>
| BGAN     | Baseline Group A
| CRUGold  | Control Group B

Preliminary clusters identified using Latent Class Analyses across 4 databases

SUNBURST PLOTS

Treatments after COVID-19 diagnosis in the NL and the UK

CPRD GOLD

I PCI

Antibiotics

Anti-inflammatory drugs

ACEI

ARB

Metformin

HMG CoA reductase inhibitors

H2 receptor antagonists

 Others

Corixosteroids

Antibiotics_agents

Antibiotics

Antifungal_agents

ACEI

ARB

Metformin

HMG CoA reductase inhibitors

H2 receptor antagonists

Others
BMI in COVID-19: JCEM pick of the week!!
EVIDENCE GENERATION: EXAMPLES OF UPCOMING STUDIES

• Heavy menstrual bleeding
  - To describe the incidence and prevalence of women diagnosed with HMB
  - To characterise women with a diagnosis of HMB
  - To describe the treatment pathways of multiple therapeutic options for HMB
  - To estimate the frequency of guideline-compliant treatment for HMB, and to characterise women with guideline-compliant versus non-guideline compliant treatment

• Alopecia Areata study-a-thon
  - To identify & characterise a cohort of European patients diagnosed with AA

• Paediatrics research
  - To identify data sources suitable for paediatric studies – characterise and generate background rates of selected phenotypes in paediatric and adult subjects

• EHDEN Mega-Study
  - To describe the incidence and prevalence of medicines with suggested shortages between 2015 and 2023
  - To characterise incident users of medicines with suggested shortages between 2015 and 2023
Future of EHDEN

- Need for high-quality RWE increasing and recognised by all stakeholders.

- Project’s legacy needs to be sustained and further developed.

- IMI consortium construct creates huge opportunities to fuel and build this system but enabling it for external collaborations requires a different focus of resources.

- Acting on sustainability now is the best way to ensure smooth and progressive transition to a project after-life.

- Legal entity in the Netherlands established in late 2021 (EHDEN Foundation)

- Consistent with the EHDEN project's vision, workplan and results

- Flexible and scalable by design to prepare for the future
Vision

The EHDEN Foundation aspires to be the trusted key actor in Europe to facilitate and accelerate the generation of high-quality real-world evidence to improve healthcare of patients.

Values

Innovative
Professional
Inclusive
Independent

Mission

Our mission is to operationalize a new paradigm for the discovery and analysis of health data, building on a large-scale federated network of data sources standardised to the OMOP common data model.
**EHREN FOUNDATION BLUEPRINT**

## Governance Structure

### Research
- Disease-specific research programmes
- Regulatory research programmes
- One-off studies
- ....

### Support
- EHDEN Academy
- Vocabularies
- Community, Data Network
- Tools & platforms
- Certification

---

**Management, communication, legal, etc.**
<table>
<thead>
<tr>
<th><strong>EH DEN F OUndATION – INITIAL PILOT SERVICES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advisory</strong></td>
</tr>
<tr>
<td>Guidance or advice, e.g., preliminary study exploration, or methodological</td>
</tr>
<tr>
<td><strong>Data Landscaping</strong></td>
</tr>
<tr>
<td>Provisional look at Data Partner network datasets, e.g., variable detection</td>
</tr>
<tr>
<td><strong>Feasibility</strong></td>
</tr>
<tr>
<td>Feasibility study/studies on likelihood of full study implementation</td>
</tr>
<tr>
<td><strong>Simple Studies</strong></td>
</tr>
<tr>
<td>Rapid, quick to deploy studies, dashboards and/or recurrent</td>
</tr>
<tr>
<td><strong>Complex Studies</strong></td>
</tr>
<tr>
<td>More time consuming, complicated (methodologically) and with customisations</td>
</tr>
<tr>
<td><strong>Study-a-thon</strong></td>
</tr>
<tr>
<td>Deployment of study-a-thons as in-house study/training events</td>
</tr>
<tr>
<td><strong>Research Programmes</strong></td>
</tr>
<tr>
<td>Public Private Partnerships, multi-year research programmes in specific TAs</td>
</tr>
</tbody>
</table>
CURRENT STATUS

• Initiation of the first Research Programme on Neuroscience
• Discussing EHDEN value streams for transition to Foundation (training, certification, studies, tools...)
• Managing considerable interest from industry for all services
• Piloting procedures and workflows for prioritised services
• Setting up core team, support team, consultants, etc.
• Establishing MoU (as precedent to Collaboration Agreements) with interested Data Partners
• Participation in Horizon Europe Calls
• EHDEN, in liaison with OHDSI, has been a driving force contributing to the European uptake of OMOP CDM
• Impressive array of results on all pillars: data network, methods, training, tools and evidence generation
• EHDEN Foundation aims to leverage the network created for real-world implementation and ultimate patient benefit

www.ehden.eu
@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.
LOCAL BEERS
AUTHENTIC ITALIAN PIZZAS
HEATED BENCHES IN GARDEN
INDEPENDENT FREE HOUSE
OXFORD LIVING WAGE EMPLOYERS
COMMITTED TO NET ZERO BY 2023
ALL WELCOME!
Harmonizing rare cancer data within EURACAN

Dr. Maaike van Swieten
IKNL
Harmonizing rare cancer data within EURACAN

Dr. Maaike van Swieten
Clinical Data Scientist (IKNL)
Outline

• EURACAN Introduction
• Data Harmonization Strategy
• Challenges in Data Harmonization
• Overcoming Challenges
• Putting it into perspective
• Conclusion & Next Steps
EURACAN (European Reference Network for Rare Adult Solid Cancers) is a collaborative initiative aimed at improving the diagnosis, treatment, and research of rare adult solid cancers across Europe.

- Low incidence (less than 6 per 100,000 people)
- Diverse clinical manifestations
- 25% of all cancer diagnoses

» Lack of data interoperability
Two EURACAN projects: Blueberry & IDEA4RC

Blueberry
• Blueprint for a sustainable, effective, scalable EURACAN registry
• 7 data partners
• OMOP CDM

IDEA4RC
• An intelligent data ecosystem for rare cancers
• 25 partners from 12 European countries
• FHIR and OMOP CDM
From health data to registry data

EHR → Data extraction and AI → Expert registration → Registry → ETL → OMOP-CDM

Blueberry
Challenges in Data Harmonization

- Missing data elements
- Data heterogeneity
- Data quality
Population registries vs Clinical registries

- Patient specific data
- Risk factors and co-morbidities
- Disease specific data
- Treatment related information
- Follow up information
Different data coding systems

Tumor grade

1. FNCLCC score
2. Tumor grade

- Differentiation grade
- Mitotic count
- Necrosis
- Differentiation grade
Data that cannot be converted to OMOP CDM

*mpnst: malignant peripheral nerve sheath tumor.
Overcoming challenges

- Collaborative efforts with the **OHDSI community**
  - Model implementation validation
  - Vocabulary extensions

- Collaborative efforts **within the network**
  - Mapping & Data Quality
  - Model implementation validation
  - Expert knowledge from clinicians and data experts

- Collaborative effort with **EHDEN**
  - Ontology platform for semantic mapping
Putting it into Perspective

- Governance and Legal Framework
- Implementation of a federated learning network
- Use case implementation
- Valorization and Financial Sustainability
From health data to registry data
From registry data to OHDSI network

Each data partner runs the same script locally and sends back the results
From registry data to federated learning network

OMOP-CDM

Node

OMOP-CDM

Node

Server

Choose analysis

Analysis results

Analysis script

User Interface

Researchers

‘Grouped’ analysis of different data sources as if data is stored locally
Sarcoma use cases

1. **Simple use case**
   > Distribution of patients

2. **Clinically relevant use case**
   > RWE of STRASSII trial

3. **Sustainability relevant use case**
   > Improve SARCULATOR app
Conclusion & Next Steps

- 4 out 7 data sources converted
- First successful study-a-thon in May 2023
- Finalize OMOP conversion for remaining partners
- Implement federated learning network
- Execute the use cases
Acknowledgements

m.vanswieten@iknl.nl
National Nodes

- Belgium
- Germany
- Greece
- Italy
- Luxemburg
- Netherlands
- Portugal
- Spain
- United Kingdom
HONEUR: Building a federated network in haematology

Michel van Speybroeck
Janssen Pharmaceutica
A Federated Data Network for the study of Haematological Diseases

Michel Van Speybroeck | Janssen Pharmaceutica
What is HONEUR?

A federated data network of real world datasets in haematological diseases in Europe

OMOP Common Data Model

Hospital

Admin

Regional

Registries & cohorts

Academia

Discovery

Post-Authorisation

HTA/Outcomes

1° Care

2° Care

Biobanks
Working with Disease Specific Datasets

Focus on deep clinical data:

- Specific variables that are often not available as structured data:
  - Regimens across different lines of treatment
  - Response / Minimal Residual Disease
  - Staging
  - Information on lesions (#, size, ..)
  - Cytogenetic risk factors
  - Reason for stopping a regimen

- Important variables can often be in text -> manual curation
Scaling of OMOP’ing in haematology

- The mapping is coordinated by one single team with predefined conventions
- Predefined variables of interest
- DQD extended to check on conventions
- Data profiler to get insight into mapped data

- Analyze Data Source
  - white rabbit
- Define Mapping Logic
  - rabbit in a hat
- Develop Mapping
  - unagi
  - github
- Test Mappings
  - Execute the mapping in a production environment
- Acceptance Testing / QA
  - Achilles
- Evaluate Infrastructure
HONEUR Architecture – the central infrastructure

- Metadata and high-level aggregate data about
- Definition of studies, selection of data partners, exchange of scripts and results
- Univariate stats on data across participating sites
- Availability of variables across different data partners
HONEUR Architecture – local infrastructure

OMOP instance for HONEUR

Script execution:
- As docker containers
- R scripts
- Python scripts

Local dashboards (Posit, Streamlit, Dash, Notebook)

Configuration: API, database connection, central account
Cornerstones of our approach

- **Transparency:** Data Partner(s) must be able to see what they are asked/agreed to execute.
- **Reliability:** We must ensure that the shared code, is also what is executed.
- **Traceability:** Results can be tied back to a script and to a particular data-cut.
- **Ease-of-use:** Execute with the 'click of a button' to avoid ad-hoc adaptations.
From feasibility to result – technical flow

Identify relevant Data sources

Perform a feasibility Query

Formulate Research Study

Aside the technical flow, the governance aspect includes an HONEUR board, forum and study Working group
Analysing your own data – Disease Trajectory Analyser (DiTrAn)

DiTrAn allows the clinical expert, non-technical user to see and explore her/his own data

On basis of a custom clinical event model
OMOP data are re-mapped to this event model
Descriptive stats, treatment analysis, time-to-event analysis
Flexible cohort definitions
Export of results (pdf, html, csv)
Cohort Creation in DiTrAn

- Integration with Atlas – import of cohorts
- From the clinical event model: any subgraph can be converted into a cohort
- Interactive cohort creation on list of selected attributes
What else...

- eCRF Lite solution for data enhancement
- Federated statistics
- Federated Machine Learning
- Application of synthetic data in the ETL process (underway)
- Integration of genomics data (underway)
- Integration of image analysis (just starting)
- Underlying platform - Feder8 setup across disease areas
Where are we now?

55,000 patients

21 partners
Acknowledgements & Links

- The HONEUR core team: Kristina Bardenheuer, Solenn Salaun, Romy Goossens, Wout Vekemans, Ido Lapidot, Eric Schoenamsgruber, Cristian Benza

- The technical team: Peter Moorthamer, Shawn Gyina, Flavio Camarrone, Lars Halvorsen, Freija Descamps, Rehan Sonmez

- Our data partners

More information:
PIONEER and OPTIMA, two EU-IMI funded big data projects led by the European Association of Urology

Monique Roobol
Professor Decision Making in Urology, Erasmus MC
PIONEER and OPTIMA, two EU-IMI funded big data projects led by the European Association of Urology

MJ. Roobol, Erasmus MC, Rotterdam
BIG DATA PLATFORM

THE EUROPEAN NETWORK OF EXCELLENCE FOR BIG DATA IN PROSTATE CANCER
Together we can ensure each individual patient receives the right treatment for them at the right time.

Clinical benefit of determining patients’ genetic risk profile

Tumour- & patient-specific variables that affect prognosis

What differentiates patients with lethal vs. non-lethal disease

Best therapeutic window & approach for recurrent prostate cancer

Impacts of comorbidities & life expectancy on patient outcomes

Upfront chemotherapy: who benefits & real-life side effects

DATA SOURCES
- Hospitals
- Pharma
- Research Institutes
- Biobanks/OMICS
- Biotech companies

DATA ELEMENTS
- Demographics
- Epidemiological
- Diagnosis & monitoring
- Imaging & lab results
- Treatments
- Outcomes
- Quality of life
- Genomics

DATA SOURCES

DATA anonymised

Data uploaded (centralised model)

Tools set up, data stays with owner (federated model)

OMICS

Data harmonisation

Quality evaluation

Transformation to common data model (OMOP)

Data platform and analytical tools on Atlas and Jupyter

RESEARCH QUESTIONS

PRIORITISED DATA SOURCES

BIG DATA ANALYTICS

KNOWLEDGE GAPS

DATA SOURCES

BIG DATA PROCESSING

PIOENEER OUTCOMES

Answers to the most important prostate cancer questions

Invididualised evidence based medicine

Improved decision-making & optimise care for prostate cancer patients and their families
PIioneer Data sources and access overview

95 Data sources identified
23 Datasets ready in platform
4 Datasets being OMOP mapped
1 Datasets in priority pipeline

Dataset onboarding to the PIONEER Platform
(number of prostate cancer patients per dataset shown)

July 2018
MarketScan CCAE 654,672
CPRD 58,603
OPTUM 651,765

Mar 2020
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #

Mar 2021
MarketScan MDCR 608,330
CPRD 58,603
OPTUM 651,765

July 2021
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #

Feb 2022
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #

June 2022
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #

Sept 2022
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #

Jan 2023
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #
CPRD 58,603
MarketScan MDCR 608,330
ERSPC Rotterdam 42,376 #
PRIAS: AS 7,302
Active Biotech 400 #

OMOP mapping ongoing

Scottish EHR 39,000
Erasmus MC PCa EHR 3,500
Janssen Prostate Cancer Registry
CUHK 2,000 #

Netherlands Cancer Registry 295,000
Stanford 7,158
TUD MRI Fusion 2,500 #
DIAMOND UK 2,024 #

PRIAS: AS 7,302
Estonian Biobank 200,000
SIDIAP 55,000
UniSR Milan 5,435
Malmö Diet & Cancer 1,585 #

Martini-Klinik 30,000 #
Stanford 7,158
TUD MRI Fusion 2,500 #
DIAMOND UK 2,024 #

Malmö Preventative Medicine 2,625 #
Malmö Diet & Cancer 1,585 #
Bordeaux University 66.6 million
Freeburg University 1000 #

CPRD 58,603
MarketScan MDCR 608,330
Estonian Biobank 200,000
SIDIAP 55,000
UniSR Milan 5,435
Malmö Preventative Medicine 2,625 #
Malmö Diet & Cancer 1,585 #
Bordeaux University 66.6 million
Freeburg University 1000 #

CPRD 58,603
MarketScan MDCR 608,330
Estonian Biobank 200,000
SIDIAP 55,000
UniSR Milan 5,435
Malmö Preventative Medicine 2,625 #
Malmö Diet & Cancer 1,585 #
Bordeaux University 66.6 million
Freeburg University 1000 #

CPRD 58,603
MarketScan MDCR 608,330
Estonian Biobank 200,000
SIDIAP 55,000
UniSR Milan 5,435
Malmö Preventative Medicine 2,625 #
Malmö Diet & Cancer 1,585 #
Bordeaux University 66.6 million
Freeburg University 1000 #

CPRD 58,603
MarketScan MDCR 608,330
Estonian Biobank 200,000
SIDIAP 55,000
UniSR Milan 5,435
Malmö Preventative Medicine 2,625 #
Malmö Diet & Cancer 1,585 #
Bordeaux University 66.6 million
Freeburg University 1000 #

CPRD 58,603
MarketScan MDCR 608,330
Estonian Biobank 200,000
SIDIAP 55,000
UniSR Milan 5,435
Malmö Preventative Medicine 2,625 #
Malmö Diet & Cancer 1,585 #
Bordeaux University 66.6 million
Freeburg University 1000 #

Central datasets:

- CPRD 58,603
- MarketScan CCAE 654,672
- MarketScan MDCR 608,330
- Pharmetrics 926,699
Studyathon 1

March 8th to 12th 2021

Collaborative effort:

245 participants including patients
20 countries & 5 time zones
5 days

8 datasets
Initial cohort > 100,000,000
**Objective:** To describe the *clinical characteristics and long-term outcomes* of PCa patients *on conservative management* by using an international large network of real-world data.

**Design, setting & participants:**

- Initial cohort > 100,000,000 adult subjects included in 8 databases
- From this 527,311 newly diagnosed PCa cases were identified
- **123,146** patients who did not receive curative or palliative treatment within six months from diagnosis were further selected
**Results**: The most common comorbidities were hypertension (35-73%), obesity (9.2-54%) and type 2 diabetes (11-28%). The rate of PCa-related symptomatic progression ranged between 2.6-6.2%.
Studyathon 1 - A

Results: The probability of being free from both palliative and curative treatments decreased during follow-up.
Conclusions:

• During the first year of follow up, up to 6% of men who were managed with conservative approach experienced PCa-related symptomatic progression, and up to 25% and 14% experienced a hospitalisation and ED events.

• The probability of receiving curative and palliative intent therapies decreased according to time elapsed between diagnosis and patient assessment.

• Older age and selected comorbidities identified patients at higher risk of adverse outcomes.

• Paper accepted (awaiting publication)
Objective: To describe clinical baseline characteristics and outcomes of PCa patients on deferred palliative management by using an international network of real-world data.

Design, setting & participants:

• Descriptive study of patients with a diagnosis of PCa on delayed management from 6 databases

• 17,659 men, with a median age range of 68-75 and Charlson Comorbidity index of 5-8 were included
Results: The most common comorbidities were hypertension (80%-93%), type 2 diabetes (29%-41%) and asthma/chronic obstructive pulmonary disease (COPD) (28%-37%). Obesity was more prevalent among younger patients (19%-41%).
Results & limitations:

- Emergency department visits and hospitalisations were highest across databases in the first year of follow-up (19-33% and 21-48% respectively).
- There were no significant differences in symptomatic progression across databases.
Conclusions:

- Men on delayed palliative management were in their mid-seventies, with the three most common comorbidities being hypertension, type 2 DM and asthma/COPD, regardless of age groups.
- ED visits, hospitalisations and symptomatic progression were similar across databases and were highest within the first year of follow-up.
- This study reflects the potential of PIONEER as an international federated network of databases that may be used to harness big data in PCa research.
- Manuscript drafted
Studyathon 2  October 31st to November 4th, 2022

Research question:
• Which specific patient will benefit the most according to the different treatment schemes in metastatic hormone-sensitive prostate cancer?
• Two parts – characterisation & prediction

35 participants including patients
10 countries
5 days hybrid event

16 datasets
Study package successful completed in 10 datasets
Current status & next steps

Status:
1. Study protocol published
2. Final study package run on 2 data sets
3. Protocol accepted for publication in International Journal of Surgery Protocols (accepted on June 24th)

Next steps:
1. Review of study package results
2. Afterwards outreach to other data owners
3. Write manuscript(s) depending on depth of results
Studyathon 3  
June 13th and 16th 2023.

53 participants including patients  
9 countries  
4 days hybrid event

20 datasets identified  
2 test R packages successfully run on 5 datasets

Research goal:
• An observational health data analysis on the adverse events of systemic treatment in patients with metastatic hormone-sensitive prostate cancer (mHSPC)

• Systemic treatment: Androgen deprivation therapy (ADT); Taxane chemotherapy; and Androgen receptor Signalling Inhibitor (ARSI)
Current status & next steps

Status:
1. Study package ready and data collected from 5 data sets
2. 1st version of the protocol under finalisation
3. 2 shiny apps:
   - cohort diagnostics & characterisation
   - incidence rates analysis (new development)

Next step:
1. Review of results and outreach to 15 more data sets
2. Characterisation analysis and incidence rates analysis
3. 4 publications planned so far
Online search Tool

Diagnostic and prognostic factors are:

- Multiple diagnostic and prognostic factors or biomarkers are available for prostate cancer
- A diagnostic biomarker or factor allows for early detection of prostate cancer
- A prognostic biomarker or factor is a clinical or biological tool that provides information on the likely course of the disease
- However, due to limited data the use of these factors is not routinely implemented in clinical practice

Identification of prostate cancer diagnostic & prognostic factors:

- Online search tool for diagnostic & prognostic factors for prostate cancer
- Which biomarkers have been reported in the literature?
- Identification and selection of biomarkers
- Assessment of the quality of the evidence for the identified biomarkers

Outcome: Online search tool which summarises systematic review findings to facilitate development of new diagnostic & prognostic factors or for setting up future studies, including clinical trials

Studies included:

- Quantitative studies in English
- Published between Jan 2016 & 2020
- Patient population: adult men (≥18) diagnosed with localized, locally advanced, metastatic, or non-metastatic castration-resistant prostate cancer
- Included studies must report on an outcome which is part of the PIONEER core outcome sets
- Interventions considered are treatments supported by the 2019 EAU guidelines

www.prostate-pioneer.eu

@ProstatePioneer
Optimal treatment for patients with solid tumours in Europe through artificial intelligence

2022 - 2027

Funded by OPTIMA
A strong alliance of public & private partners

39 Partners across 9 countries in different fields.

- IMI experienced researchers / SMEs
- Medical societies
- Opinion leaders in oncology
- Leading guideline authors
- Patient organisations
- Committed industrial partners
- Members of the EMA Steering Group
- Healthcare AI Experts
- Implementation Scientist
Objectives

1. To establish a **data catalogue**.

2. To develop a secure and **interoperable platform**.

3. To develop a scalable and regularly updated **guideline decision-support toolset**.

4. To drive new **knowledge generation**.

5. To ensure the **sustainability** of OPTIMA's platform.
Research areas of interest for OPTIMA

- Prostate Cancer
- Breast Cancer
- Lung Cancer
9 Work packages

OPTIMA is divided into 9 work packages, which all work closely together. WP 1, 2 and 8 in particular work across all packages.
• Knowledge gaps  
• Guidelines  
• Key recommendations  
• Care pathways

WP3  
Guideline-based decision tool

WP4  
Platform technical and infrastructural requirement analysis

WP5  
Platform implementation and evaluation

WP6  
Data gathering and access

WP7  
AI knowledge based implementation

Improved decision making at an individual level

Interaction with electronic health record

2022 – 2027
WP 6
Identified 11 data assets with completed / planned mapping to CIOMS core data management

WP 5
Identified list of user personas and organised the platform development in waves accordingly: IT and data scientists, care providers and patients, AI-aware physicians, external users

WP 7
First draft of an ML OPs pipeline and set-up of the first software for the pipeline with WP4

WP 2
Initial Data Protection Impact Assessment finalized

WP 2
Initial Report on the Legal and Ethical Framework of the Project

WP 3
Development of Research Questions Surveys and subsequent prioritization

WP 4
Requirement and Specification Analysis Document

WP 4
Initial Data management plan finalized

WP 6
Signed memorandum of understanding with EHELEN to leverage on efforts to generate analytical packages

WP 3
Prototype of decision support tool built and tested

WP 3
Assessments of Clinical Practice Guidelines for Prostate Cancer, Breast Cancer and Lung Cancer and identified care pathways

What we have achieved so far.

Contact: communication@optima-oncology.eu
Thank you for your attention!

OPTIMA is funded through the IMI2 Joint Undertaking and is listed under grant agreement No. 101034347. IMI2 receives support from the European Union’s Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA). IMI supports collaborative research projects and builds networks of industrial and academic experts in order to boost pharmaceutical innovation in Europe.
European Initiatives Using the OMOP CDM - discussion

Carlos Diaz, Maaike van Swieten, Michel van Speybroeck, Monique Roobol
Coffee break! Next session starts here @ 11:30!

This coffee break is sponsored by Promptly Health.