Scaling up the big data ecosystem across Europe

Prof Dani Prieto-Alhambra

Prof of Pharmaco- and Device Epidemiology
CSM, NDORMS, University of Oxford
AGENDA

• EHDEN – An Introduction
• The Oxford Study-A-Thon
• The Barcelona Study-A-Thon (if time allows... )
We need fully transparent and reproducible pipelines that enable large-scale federated analyses across Europe.
Why is this not current practice?

Analytical method

The data...

What will it require?

Data interoperability  Standardised analytics  Data network  Strong community
EHDEN will build on expertise and tools from prior IMI projects, such as EMIF, and will collaborate intensively with the global OHDSI community.
### Large-Scale Observational Research Is Feasible

**Characterizing treatment pathways at scale using the OHDSI network.**
George Hripcsak et al. - PNAS (2016)27:7329–7336

<table>
<thead>
<tr>
<th>T2 Diabetes Mellitus</th>
<th>Hypertension</th>
<th>Depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCAE</td>
<td>CUMC</td>
<td>MDCD</td>
</tr>
<tr>
<td>CPRD</td>
<td>INPC</td>
<td>GE</td>
</tr>
<tr>
<td>JMDC</td>
<td>MDCR</td>
<td>GE</td>
</tr>
<tr>
<td>OPTUM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data sources**
- 11 Data sources
- 4 Countries (incl Japan)
- > 250 million patients
"Comprehensive comparative effectiveness and safety of 1st line antihypertensive drug classes."

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 standardized to a common data model.
**EHDEN IS ABOUT ...**

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

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

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

**Outcomes**
Enabling outcomes driven healthcare at a European level

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

- Creation of an EU-wide federated network architecture
- **Privacy** by design
- **Data harmonisation** to the OMOP common data model
- Training & certification of SMEs

**Research & Outcomes**

- **Use cases** to evaluate the EHDEN federated network
- Collaboration on consistent methodologies
- Collaboration with the global OHDSI research network
- Incorporation of the ICHOM health outcome standards

**Education & Community**

- Establishment of an EHDEN Academy
- Expansion of the OHDSI network in Europe
- Collaboration on collective memory for research use cases
The EHDEN Federated Data Network

Local Governance

EMR
LIMS
Rx
Dx
Admin
...

Local Database
OMOP Database

Analysis query
Aggregated results

EHDEN will develop new tools and dashboards.

Many different open source tools (cohort builder, estimation, incidence rate, ....)

The EHDEN platform

ATLAS

EMR
LIMS
Rx
Dx
Admin
...
THE OMOP COMMON DATA MODEL

Patient-centric
Tabular
Extendable
Built for analytics
Relational design

Standardized clinical data
Person
- Observation period
- Specimen
- Death
- Visit occurrence
- Procedure occurrence
- Drug exposure
- Device exposure
- Condition occurrence
- Measurement
- Observation
- Note
- Note NLP
- Fact relationship

Standardized health system data
- Location
- Care site
- Provider
- Payer plan period
- Cost
- Cohort
- Cohort attribute
- Condition era
- Drug era
- Dose era

Standardized meta-data
- CDM source
- Concept
- Vocabulary
- Domain
- Concept class
- Concept relationship
- Relationship
- Concept synonym
- Concept ancestor
- Source-to-concept map
- Drug strength
- Cohort definition
- Attribute definition

Standardized derived elements

Standardized vocabularies

v 5.0.1
CALL PROCESS OVERVIEW

Data sources
- Tailored for project objectives and sustainability

Open calls
- Focusing on SMEs able to support mapping and sustainability

Grant awarding
- Evaluated via a pre-defined set of criteria by the Data source prioritisation committee

Data sources can choose the SME from the pool of EHDEN certified SMEs

SMEs are paid via grants from the harmonisation fund

Payments are milestone based

Supporting SMEs

Open calls

Training & Certification
- SME certification committee prioritizes SMEs for training and certification

Harmonisation fund

Workshop

Source Data Evaluation

Mapping Cycle

Audit

Mapping

Share of Mapping Process

Mapped data sources are encouraged to be active members of the EHDEN community, participating in research studies.
SME Pilot Call

Open Call
Submission of applications for the open call for SMEs from the 1st of April until the 1st of May (17h00) via the EHDEN website.

Evaluation
Following an eligibility check, applications were evaluated by the SME certification committee.

Training & Certification
Certification and training of selected SMEs in all necessary competencies.

34 SME profiles made
28 Eligible applications
11 SMEs initially selected
Data Partners Pilot Call

- Initial phase during which all interested data partners can review the call description, ask questions and comment on the call description (July 15 – August 15).

- Grant application portal open
  - Open call for data partners (September 1 – September 15)

- Evaluation
  - Evaluation of all applications by our committee of both internal and external experts.

- Agreement
  - Grant awarding and signing of grant agreement.

- SME linking
  - Identification and linking up with the SME of choice.

- Harmonisation
  - Initiation of the data harmonisation.

Selected Data Partners

- Applicant countries

20 Data Partners

>120 million patient records

(Preliminary results)
2020 Call Timelines

SME
The second open call for SMEs.

2020

Data partner
The second open call for Data partners.

SME
The third open call for SMEs.

Data partner
The third open call for Data partners.
Use Case 1 – Drug utilisation study
Validate the OMOP CDM in European data sources comparing results from source versus OMOP data

Use Case 2 – Drug and device safety study
Test and further develop existing (OHDSI) population-level estimation and population-level prediction packages in EU Data sources

Use Case 3 – HTA study
Assess whether data commonly used for HTA purposes can easily be measured and analysed using the OMOP CDM
“To compare the **risk** of post-operative **complications** and **mortality** between unicompartmental vs total knee replacement.”

---

**Monday**
- Group consensus on the **problem**
- Draft cohort definitions

**Tuesday**
- Review clinical characterisation
- Draft patient-level prediction design

**Wednesday**
- Review patient-level prediction results
- Externally validate prediction model

**Thursday**
- Draft population-level effect estimation design
- Review population-level effect estimation diagnostics

**Friday**
- Review of results
- Plan for completing **publications**

(EHDEN 1st Study-a-thon, Oxford, December 2018)
Aim
To develop an e-learning environment to train all stakeholders on the use of the tools and processes that are being adopted within EHDEN and OHDSI.

Collaboration
Course development on the OMOP Common Data Model and the rich set of OHDSI tools that are being developed in collaboration with the OHDSI community.

Infrastructure
The EHDEN Academy is developed in Moodle and is hosted in the Amazon AWS cloud. We use virtual machines for assignments.
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.
The Oxford Study - Athon
OUR JOURNEY WITH OHDSI AND EHDEN TO REAL-WORLD USEFUL EVIDENCE

Important clinical question

Data partners standardized to OMOP CDM: Iqvia Janssen

Standardized analysis tools from OHDSI

Valuable clinical answers disseminated to medical decision-makers

Reliable evidence
**Hands-on knowledge of 1+ data source/s,**
including its structure and content, the provenance of the underlying population and data capture process, data quality issues and temporal variability, so that you can responsibly use the data to generate reliable evidence and recognize its limitations.

**Direct knowledge of the diagnosis, treatment, and management of severe knee OA,**
including healthcare delivery, natural history and patient prognosis.

**Hands-on knowledge in designing studies and executing statistical methods to generate aggregate summary statistics from patient-level data.** Different expertise required for clinical characterization, patient-level prediction, and population-level effect estimation.
EXPERTISE REQUIRED

• Clinical knowledge in knee OA/arthroplasty?
• UK electronic health records (THIN)?
• US claims (MarketScan, Optum, PharMetrics)?
• OHDSI tools?
• R programming?
• Literature review?
• Publication writing?

Who has all of these prerequisites?
“To compare the risk of post-operative complications (infection, venous thromboembolism, mortality) and long-term implant revision between unicompartmental vs total knee replacement.”
WHAT WE KNEW BEFORE WE STARTED

• N = 60 quality studies

• From 1998 to 2018 (20 years of research!!)

• Reduced risk of VTE w UKR

Fig 3 | Forest plot comparing risk of venous thromboembolism after unicompartmental (UKA) versus total knee replacement (TKA). Also appears in the supplementary material as supplementary figure 5. M-H=Mantel-Haenszel test.
**What we knew before we started (2)**

- Possibly a reduction in post-operative mortality ...
- [although little data available on this]

<table>
<thead>
<tr>
<th>Study</th>
<th>No of events</th>
<th>Total no of procedures</th>
<th>Risk ratio M-H Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Beard 2017</td>
<td>0</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>Costa 2011</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Newman 1996</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Sun 2010</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>314</td>
<td>300</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: Not applicable
Test for overall effect: Not applicable

<table>
<thead>
<tr>
<th>Group 2</th>
<th>No of events</th>
<th>Total no of procedures</th>
<th>Risk ratio M-H Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bolognesi 2013</td>
<td>0</td>
<td>3098</td>
<td>455</td>
</tr>
<tr>
<td>Courtney 2018</td>
<td>0</td>
<td>1351</td>
<td>24</td>
</tr>
<tr>
<td>Drager 2016</td>
<td>0</td>
<td>1340</td>
<td>0</td>
</tr>
<tr>
<td>Duchman 2014</td>
<td>0</td>
<td>1583</td>
<td>3</td>
</tr>
<tr>
<td>Hunt 2014</td>
<td>33</td>
<td>40 428</td>
<td>1,183</td>
</tr>
<tr>
<td>Liddle 2015</td>
<td>7</td>
<td>25 358</td>
<td>90</td>
</tr>
<tr>
<td>Subtotal</td>
<td>75 163</td>
<td>69 6356</td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>40</td>
<td>1755</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: $\chi^2=1.08$, $p=0.04$, $df=4$, $P=0.13$; $I^2=21%$
Test for overall effect: $z=5.02$, $P<0.001$

<table>
<thead>
<tr>
<th>Group 3</th>
<th>No of events</th>
<th>Total no of procedures</th>
<th>Risk ratio M-H Random (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang 2003</td>
<td>0</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Subtotal</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: Not applicable
Test for overall effect: Not applicable

**Figure 4** | Forest plot comparing risk of early mortality (at 45 days) after unicompartmental (UKA) versus total knee replacement (TKA). Also appears in the supplementary material as supplementary figure 7. M-H = Mantel-Haenszel test.
• ... BUT

• An increase (around double) in long-term revision risk
What we knew before we started (5)

- Caveats with quality of these 60 papers (and 20y) of data (mostly observational and from different sources)
  - NIHR UK-funded
    - 1 surgical RCT (TOPKAT)
    - 1 observational study (UTMOST)
“We need less research, better research, And research done for the right reasons”

i.e., can we do in a week a study what has taken so far 20+ years, 60+ papers and loads of cash?
• Can we ‘predict’ the TOPKAT results (on complications) before they publish?

• And more:
  • can we report on the results of UKR (vs TKR) in the older, more complex patients, excluded from TOPKAT?
  • Can we predict who is likely to have a post-operative complication following knee replacement surgery
WE CAN DO THIS IN ONE WEEK (STUDY-A-THON)??

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

**Monday**
- Group consensus on the problem
- Draft cohort definitions

**Tuesday**
- Review clinical characterisation
- Draft patient-level prediction design

**Wednesday**
- Review patient-level prediction results
- Externally validate prediction model

**Thursday**
- Draft population-level effect estimation design
- Review population-level effect estimation diagnostics

**Friday**
- Review of results
- Plan for completing publications

(EHDEN 1st Study-a-thon, Oxford, December 2018)
“To compare the risk of post-operative complications (infection, revision, and venous thrombo-embolism) and mortality between uni-compartmental vs total knee replacement.”
LET’S START COLLABORATING!

• Open the shared group notes: Link

• Ground rules:
  • During group exercises, take all your notes here together
  • During breakout exercises, assign one person in your team to make sure notes are recorded so other groups can learn from our experience
• Patient-Level Prediction: Link

• Population-Level Effect Estimation: Link
LET’S START LEARNING ATLAS!

- Public version from OHDSI (v2.6, simulated data), go to: http://ohdsi.org/web/ATLAS

- Private version from IQVIA (v2.4, THIN data), go to: https://training.atlasplus.imshealth.com
SO WHAT DID WE LEARN (BY FRIDAY!!)

• Population-level effect estimation:
  http://data.ohdsi.org/UkaTkaSafetyEffectiveness/

• Patient-level prediction
  http://data.ohdsi.org/TKROutcomesExplorer/
• VTE
  • RR 0.49 [0.20 to 1.17] (20 y)
  • vs HR 0.62 [0.36-0.96] (1 week)

• Long-term revision
  • RR 1.68 [1.07 to 2.64] (20y)
  • vs HR 1.51 to 2.16 (1 week)
• Small improvement in pain/function with UKR in TOPKAT
• = Small reduction in opioid/s use in Study-a-thon
• No power for safety in TOPKAT
• Findings compatible w 20y of data in Study-a-thon
PREDICTION... PREDICTING POST-OP MORTALITY
THE MODEL

Model Table

<table>
<thead>
<tr>
<th>Covariate Name</th>
<th>Value</th>
<th>Outcome Mean</th>
<th>Non-outcome Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 index month: 1</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>2 Charlson index - Romano adaptation</td>
<td>0</td>
<td>3.81</td>
<td>2.16</td>
</tr>
<tr>
<td>3 Diabetes Comorbidity Severity Index (DCS)</td>
<td>0</td>
<td>3.71</td>
<td>1.96</td>
</tr>
<tr>
<td>4 CHADS2</td>
<td>0</td>
<td>2.54</td>
<td>1.57</td>
</tr>
<tr>
<td>5 CHADS2VASc</td>
<td>0</td>
<td>4.18</td>
<td>2.94</td>
</tr>
<tr>
<td>6 visit_occurrence concept count during day -1095 through -1 concept_count relative to index</td>
<td>0</td>
<td>71.99</td>
<td>50.13</td>
</tr>
<tr>
<td>7 visit_occurrence concept count during day -365 through -1 concept_count relative to index</td>
<td>0</td>
<td>32.46</td>
<td>24.26</td>
</tr>
<tr>
<td>8 index month: 2</td>
<td>0</td>
<td>0.1</td>
<td>0.08</td>
</tr>
<tr>
<td>9 index month: 3</td>
<td>0</td>
<td>0.1</td>
<td>0.09</td>
</tr>
<tr>
<td>10 index month: 4</td>
<td>0</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>
PREDICTION... PREDICTING POST-OP MORTALITY

DISCRIMINATION

CALIBRATION
• **Conference/s**
  • 1 Podium ppt at IOF/ESCEO Paris’19 (>3,000 researchers)
  • 1 Podium ppt and Press Release at EULAR Madrid ‘19 (>15,000 rheumatologists)
  • 1 Podium ppt and 1 Spotlight poster at ICPE’19 (>1,500 epidemiologists)

• **Scientific journal manuscripts**
  • PLE = Major review at Lancet Rheumatology
  • PLP = Submitted to JAMA Surgery
AND WHAT DID I LEARN (BY FRIDAY TOO!!)
The Oxford Study-Athon

“Why had we not joined the journey earlier?”

Finished in 1430

Finished in the 1600s
EULAR recommendations for the management of rheumatoid arthritis – 2019 Update

BACKGROUND

We knew this was coming...
After 7 years of silence ...