Rapid Fire Presentations of Collaborators

Katia Verhamme, MD, PhD
Associate Professor of Use and Analysis of Observational Data
FEEDERNET (FEDERATED E-HEALTH BIG DATA FOR EVIDENCE RENOVATION NETWORK) PLATFORM IN KOREA

Seongwon Lee, Chungsoo Kim, Junhyuk Chang, Rae Woong Park
FeederNet (Federated E-Health Big Data for Evidence Renovation Network) platform in Korea

2022. 06. 24

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

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

* Presenting author: Chungsoo Kim, PharmD
FeederNet Project in Korea

FeederNet Platform

CDM-based Service

Distributed Analytic Platform

Bio-health Data Network

Nation-wide OMOP-CDM Data Network
- 1 national claim data of HIRA (from 2010 to 2020)
- 53 hospitals’ medical data of 72M Patients (including duplicates)
OMOP-CDM Data Network

Nation-wide CDM Data Network

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

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

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

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

Currently, 18 Hospitals joined the RFZ
OMOP-CDM Research in Korea

Analysis on FeederNet

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

Paper Publications

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

Paper search through Google scholar with a keyword, “OMOP-CDM” From 2017 to 2022 (n=880)

Papers after removing missing-year-paper (n=814)

Papers Screened (n=439)

Papers which published In non-journal (n=329)

Papers in non-English (n=46)

Papers written by Korean researcher as the first author (n=97)
New CDM project on Infectious Disease Surveillance

‘PHAROS’ Project

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

Infectious disease management using real-time clinical data

**Analytic apps for infectious disease**
- NLP module
- GIS module
- Bayesian statistical modeling
- Prediction modeling

**Utilizing the current DRN platforms**
- Collaboration with other CDM platform for infectious disease surveillance and research

**Promotion Strategy**
- Expert advisory group • Steering committee • Promotion and seminar • Expanding of institutions

FUTURE

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

- Change in infection symptoms
- Legal communicable diseases
- Precision monitoring using NLP

---

Data-driven infectious disease response system
Thank you
OMOP GENOMIC MAPPING CAPACITIES IN CONVERSION OF COMPREHENSIVE GENOMIC PROFILING RESULTS

Rogozhkina Maria, Odysseus
OMOP Genomic mapping capacities in conversion of comprehensive genomic profiling results

by ODYSSEUS DATA SERVICES

OHDSI OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS
ABL1 (ABL proto-oncogene 1, non-receptor tyrosine kinase) gene variant measurement

MET Gene Mutations measurement

Gene amplification

*All 5 codes for amplifications only

Postcoordination

ABL1 deletion

99.7% by 2 target concepts: measurement + value

MET amplification

MET Gene Amplification measurement

0.7% by 1 target concept: measurement

Precoordination

*All 5 codes for amplifications only
BRCA1 (BRCA1 DNA repair associated) gene variant measurement

JAK2 (Janus kinase 2) gene variant measurement

Abnormal nucleotide base sequence

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

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

1.6% by 1 target concept: measurement

99.9% by 2 target concepts: measurement + value

1 - data about position is missing in all target concepts
2 - 0.3% was mapped to Protein level
BONE MARROW BIOPSY | FISH | Cytogenetic Abnormality | AMP1Q21

Chromosome region 1q21 duplication* in Bone marrow by FISH

* customer approved variant of amplification storage
We found **46** duplicate concepts for DNA, **1276** for RNA and **224** for Protein variants in OMOP Genomic. All of them to be deduplicated.

A substantial number of codes are missing. **New codes** to be ingested in OMOP Genomics are essential to facilitate studies.

Concepts from non-Genomic vocabularies may be a valid target, but lacks of consistency and only a fraction of required targets is available.

Linking to other event tables such as Procedure and Specimen is required to properly represent source data. To efficiently run queries an easier way to link facts is required (direct fact modification).
OMOP MAPPING OF REAL-WORLD DATA FROM BRAZIL & PAKISTAN TOWARDS MANAGEMENT OF COVID-19 IN THE GLOBAL SOUTH

Sara Khalid, University of Oxford
OMOP Mapping of Real-World Data From Brazil & Pakistan Towards Management of COVID-19 In the Global South

Authors
Elzo Pereira Pinto Junior¹, Priscilla Normando¹, Renzo Flores-Ortiz¹, Muhammad Usman Afzal², Muhammad Asaad Jamil², Sergio Fernandez Bertolin³, Vinícius de Araújo Oliveira¹, Valentina Martufi¹, Edward Burn⁴, Maria Yury Ichihara¹, Maurício L. Barreto¹, Talita Duarte Salles³, Daniel Prieto-Alhambra⁴, Haroon Hafeez², Sara Khalid⁴

¹ The Center for Data and Knowledge Integration for Health (CIDACS – FIOCRUZ), Brazil
² Shaukat Khanum Memorial Cancer Hospital and Research (SKMHR&C), Pakistan
³ The Information System for Research in Primary Care (SIDIAP), Spain
⁴ Centre for Statistics in Medicine, University of Oxford, UK.
The COVID-19 pandemic highlighted need for rapid, reliable, representative evidence generation
The OHDSI COVID-19 Data Network

- Rapid
- Reliable
- Representative

<table>
<thead>
<tr>
<th>USA (8)</th>
<th>EUROPE (7)</th>
<th>ASIA-PACIFIC (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premier (National – Hospital Billing)</td>
<td>CPRD (UK – Electronic Health Records)</td>
<td>HIRA (South Korea – Administrative Claims)</td>
</tr>
<tr>
<td>HealthVerity (Claims linked to diagnostic testing)</td>
<td>SIDIAP (Spain – Electronic Health Records)</td>
<td>DCMC (South Korea – Electronic Health Records)</td>
</tr>
<tr>
<td>Optum EHR (National – Electronic Health Records)</td>
<td>SIDIAP-H (Spain – EHR hospital linkage)</td>
<td>Nanfang Hospital (China – Electronic Medical Records)</td>
</tr>
<tr>
<td>IQVIA Open Claims (National – Administrative Claims)</td>
<td>HM Hospitales (Spain – Hospital Billing)</td>
<td></td>
</tr>
<tr>
<td>Department of Veterans Affairs (National – Electronic Health Records)</td>
<td>ICPI (Netherlands – Electronic Health Records)</td>
<td></td>
</tr>
<tr>
<td>Stanford University (CA – Electronic Health Records)</td>
<td>LPD France (France – Electronic Health Records)</td>
<td></td>
</tr>
<tr>
<td>Tufts University (MA – Electronic Health Records)</td>
<td>Germany DA (Germany – Electronic Health Records)</td>
<td></td>
</tr>
<tr>
<td>Columbia University (NY – Electronic Health Records)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Together, OHDSI has studied (to date):
- >4.5m patients tested for SAR-COV-2
- >1.2m patients diagnosed or tested positive for COVID-19
- >250k hospitalized for COVID-19
The OHDSI COVID-19 Data Network
Data Partners – Pakistan

• Data source:
  – De-identified electronic health records
• Period:
  – 1994 – 2022 (ongoing)
• Unique records:
  – 8.3 million individuals
• Regional COVID-19 hub

Clinical activity snapshot (2021)

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

- **Data source/s:**
  - Brazil Ministry of Health Influenza Surveillance System (SIVEP-Gripe)
- **Period**
  - 2020 – 2022 (ongoing)
- **COVID-19 records:**
  - 2.6 million individuals
  - ~67,000 hospitalisations
  - ~27,000 deaths

**CIDACS-FIOCRUZ COVID-19 Data Platform**

- Brazil MoH Surveillance System
- Data source/s:
  - Brazil Ministry of Health Influenza Surveillance System (SIVEP-Gripe)
- Period
  - 2020 – 2022 (ongoing)
- COVID-19 records:
  - 2.6 million individuals
  - ~67,000 hospitalisations
  - ~27,000 deaths

**Bahia State Facts:**
- Population: 15 million
- 417 municipalities
Harmonisation to OMOP

ETL Process

1. Perform scan on source database using White Rabbit.
2. Perform column level mapping from source database to CDM (Tableau is a tool).
3. Export source data to CSV/Parquet after extraction of required variables / attributes.
4. Import CSV/Parquet in a staging database in PostgreSQL using DBeaver / Pentaho Transforms.
5. Concept Mapping (USEA and MIMIC4).
6. Apply inline preprocessing and transformation on staging database.
7. Load the data into CDM using PostgreSQL, concept procedures.

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

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

Pakistan

Brazil

The OHDSI Federated Data Network Model

Data Quality Assessment

Analysis Query

Open source analytical tools
(Chart builder, estimation, incidence rates, prediction)

Regional and international networks for health data governance and capacity building

Harmonisation to OMOP

ETL Process

1. Perform scan on source database using White Rabbit.
2. Perform column level mapping from source database to CDM (Tableau is a tool).
3. Export source data to CSV/Parquet after extraction of required variables / attributes.
4. Import CSV/Parquet in a staging database in PostgreSQL using DBeaver / Pentaho Transforms.
5. Concept Mapping (USEA and MIMIC4).
6. Apply inline preprocessing and transformation on staging database.
7. Load the data into CDM using PostgreSQL, concept procedures.

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

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

Pakistan

Brazil

The OHDSI Federated Data Network Model

Data Quality Assessment

Analysis Query

Open source analytical tools
(Chart builder, estimation, incidence rates, prediction)

Regional and international networks for health data governance and capacity building
COVID-19 cohorts – cases over time

<table>
<thead>
<tr>
<th>Tested population</th>
<th>COVID-19 diagnosis or positive test</th>
<th>Hospitalised with COVID-19</th>
<th>ICU admission with COVID-19</th>
<th>COVID-19 death</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1,312,832</td>
<td>752,699</td>
<td>34,699</td>
<td>17,041</td>
</tr>
</tbody>
</table>

**Brazil**

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

**Pakistan**

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Date of cohort entry</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2020</td>
<td>1000</td>
</tr>
<tr>
<td>Feb 2020</td>
<td>2000</td>
</tr>
<tr>
<td>Mar 2020</td>
<td>3000</td>
</tr>
<tr>
<td>Apr 2020</td>
<td>4000</td>
</tr>
<tr>
<td>May 2020</td>
<td>5000</td>
</tr>
</tbody>
</table>
```

**Distribution of cases over time (Jan/March 2020 – April 2022)**
COVID-19 cohorts – baseline characteristics

- COVID-19 outcomes were more severe in men, elderly, and those with co-morbidities

Distribution of cases by age and sex in each cohort
General population tested for COVID-19:

- Age: median (IQR) was 36 (25-75) and 38 (27-50) for Pakistan and Brazil
- Sex: 45.5% and 55% were female in Pakistan and Brazil
- Ethnicity/race: 1.2% Pakistan individuals had “Afghan” ethnicity. In Brazil, 52.3% had “Mixed” ethnicity.
Summary

• Two health databases covering 8.3 million people from Pakistan and 2.6 million people from Bahia, Brazil were analysed.

• 109,504 (Pakistan) and 921 (Brazil) medical concepts were harmonised to OMOP CDM.

• 341,505 (4.1%) people in the Pakistan dataset and 1,312,832 (49.2%) people in the Brazilian dataset tested for COVID-19 between 1\textsuperscript{st} Jan 2020 and 30 April 2022.

• In agreement with international findings, COVID-19 outcomes were more severe in men, elderly, and those with underlying health conditions.

• This proof-of-concept study demonstrates potential for OMOP-harmonised data from under-represented regions for global knowledge mobilisation and clinical translation for timely response to healthcare needs in pandemics and beyond.
What’s next

• COVID-19 – variant and vaccine surveillance study
• Communicable, NCDs
  – Cancer (OHDSI Oncology WG)
• Environment, equity, and artificial intelligence
  – Social deprivation dashboard (OHDSI GIS WG, OHDSI Equity WG)
• Data science ecosystem
  – Capacity building
  – Data re-use projects
  – Data governance
Acknowledgements

• This work was supported by funding from the Bill & Melinda Gates Foundation
OMOP Mapping of Real-World Data From Brazil & Pakistan Towards Management of COVID-19 In the Global South

Authors
Elzo Pereira Pinto Junior¹, Priscilla Normando¹, Renzo Flores-Ortiz¹, Muhammad Usman Afzal², Muhammad Asaad Jamil², Sergio Fernandez Bertolin³, Vinícius de Araújo Oliveira¹, Valentina Martufi¹, Edward Burn⁴, Maria Yury Ichihara¹, Maurício L. Barreto¹, Talita Duarte Salles³, Daniel Prieto-Alhambra⁴, Haroon Hafeez², Sara Khalid⁴

¹ The Center for Data and Knowledge Integration for Health (CIDACS – FIOCRUZ), Brazil
² Shaukat Khanum Memorial Cancer Hospital and Research (SKMHR&C), Pakistan
³ The Information System for Research in Primary Care (SIDIAP), Spain
⁴ Centre for Statistics in Medicine, University of Oxford, UK.
IMPACT OF RANDOM OVERSAMPLING AND RANDOM UNDERSAMPLING ON THE DEVELOPMENT AND VALIDATION OF PREDICTION MODELS USING OBSERVATIONAL HEALTH DATA

Cynthia Yang, Erasmus MC
Impact of random oversampling and random undersampling on the development and validation of prediction models using observational health data

Cynthia Yang
PhD Student at Erasmus MC
The class imbalance problem

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

- Random oversampling and random undersampling
- Imbalance ratio (IR) = the total number of patients without an outcome event (majority class) / the total number of patients with an outcome event (minority class)
- We vary IR = min(IR_{original}, x) with x ∈ {20, 10, 2, 1}
- 21 outcomes in depression (PLP framework paper)
- CCAE, MDCD, MDCR, IQVIA Germany
- XGBoost, lasso logistic regression
- Area under the receiver operating characteristic curve (AUROC)
Results
Conclusions

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

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

- For more information or questions please visit me at my poster #57.
REAL-WORLD EVIDENCE IS IN DEMAND!

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

Jamie Elvidge, National Institute for Health and Care Excellence (NICE)
Real-world evidence is in demand!

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

Jamie Elvidge
jamie.elvidge@nice.org.uk
Ravinder Claire, Shane Collins, Dalia Dawoud

NICE National Institute for Health and Care Excellence
About NICE

• The HTA agency for England
• Provides guidance to the healthcare system
• Key principles:
  • evidence-based medicine
  • opportunity cost
• Actively exploring ways to use RWE for decision making

NICE real-world evidence framework

Overview

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

Do NICE research recommendations ask for RWE studies?
There are calls for RWE from across NICE guidance
Lung cancer
Large retrospective analyses to predict determinants of lung cancer, characterise current practice, and predict likelihood of unnecessary surgery.

Ophthalmology
Routine treatment can be effective, but also unpleasant, costly and risky.
→ Observational studies to characterise treatment benefit and predict when the benefits cease.

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

https://www.nice.org.uk/about/what-we-do/science-policy-research/research-recommendations
WHY PREDICTING RISK CAN’T IDENTIFY ‘RISK FACTORS’
EMPIRICAL ASSESSMENT OF MODEL STABILITY IN MACHINE LEARNING ACROSS OBSERVATIONAL HEALTH DATABASES

Aniek Markus, Erasmus MC
WHY PREDICTING RISK CAN’T IDENTIFY ‘RISK FACTORS’

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

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

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

• Some researchers incorrectly interpret prediction models:
  1. Prediction models do not assess causality
  2. Also problematic to use models for ‘risk factor’ detection

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

Study population (3) \times Outcome phenotypes (3) \times Databases (7)

<table>
<thead>
<tr>
<th>General population def. 1</th>
<th>AMI def 1.</th>
<th>CCAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population def. 1</td>
<td>AMI def 2.</td>
<td>CCAE</td>
</tr>
<tr>
<td>General population def. 2</td>
<td>AMI def 3.</td>
<td>MDCD</td>
</tr>
<tr>
<td>General population def. 2</td>
<td>AMI def 3.</td>
<td>MDCD</td>
</tr>
<tr>
<td>General population def. 3</td>
<td>AMI def 3.</td>
<td>IQVIA EHR</td>
</tr>
</tbody>
</table>

Measure model stability for each outcome (9x)

+/− 63 models
Measuring model stability

Example model:

\[ \text{probability(acute myocardial infarction within 1 year)} = \sigma(+1.33 \times \text{age} - 0.2 \text{female} + 0.43 \text{obesity}) \]

1. How many variables are selected across models?

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

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

- Substantial variation in the selected variables
  - Different databases → different ‘risk factors’

- Sign of ‘risk factors’ can differ across models

- For ‘risk factor’ detection:
  - Investigate model robustness
  - Use other techniques (e.g. univariate analysis)
Take home message:

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

‘Risk factor’ = ‘variable associated to the outcome’
TRAJECTORYVIZ: INTERACTIVE VISUALIZATION OF TREATMENT TRAJECTORIES

Maarja Pajusalu, Institute of Computer Science, University of Tartu
TrajectoryViz: Interactive visualization of treatment trajectories

PRESENTER: Maarja Pajusalu
maarja.pajusalu@ut.ee

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

Create discrete sequences with R package Cohort2Trajectory

Create interactive visualisations with R package TrajectoryViz

Explore and find answers to your questions
Use Case 1:
Cervical Cancer patients

Sequences on a Sunburst Chart

PAP  HSIL  Colposcopy  6.53%

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

Did the patient follow suggested pattern for PAP tests before?
Use Case 2: Asthma patients

Sequences on a Sunburst Chart

- **21.3%**
  - 200 of 930

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

There was a subset of the patients who spent much more time taking the drugs.
Thank You!
ASSESSING TREATMENT EFFECT HETEROGENEITY USING THE RISKSTRATIFIEDESTIMATION R-PACKAGE

Alexandros Rekkas, Erasmus MC
Assessing treatment effect heterogeneity using the RiskStratifiedEstimation R-package

Alexandros Rekkas
DEFINING THE VALID ANALYTIC SPACE FOR QUANTITATIVE BIAS ANALYSIS IN PHARMACOEPIDEMIOLOGY

James Weaver, Janssen R&D
Defining the analytic space for valid QBA

OHDSI Europe Symposium
24 June 2022

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

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

Objective
• Evaluate QBA performance across large set of plausible scenarios

Methods
• Applied QBA across incidence x effect size x measurement error analytic space

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

Discussion
• QBA produces implausible or invalid estimates in many common comparative effect estimation scenarios
A PILOT STUDY TO EVALUATE THE FEASIBILITY OF USING OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS ANALYTICS TOOLS FOR SUPPORTING THE VALIDATION OF SAFETY SIGNALS

Ceyda Pekmez Kristiansen, Novo Nordisk
A pilot study to evaluate the feasibility of using OHDSI analytical tools for supporting safety surveillance

Ceyda Tugba Pekmez Kristiansen¹; Lasse Christensen¹; Michael Stellfeld¹; Atheline-Major Pedersen¹; Ditte Mølgaard-Nielsen¹; Mark White¹; Peter Jelnes¹

cypk@novonordisk.com

¹Novo Nordisk A/S, Vandtårnsvej 114, 2860 Søborg, Denmark
24.06.2022
Real world data (RWD) for supporting safety surveillance

- RWD supports the safety signal validation process, especially when the evidence from traditional safety data sources is scarce

- Gallbladder- or biliary tract-related events (including acute cholecystitis) are known risks for Victoza® (liraglutide) and Saxenda® (liraglutide)

- A known risk for liraglutide was chosen for the pilot study to evaluate the feasibility of implementing population level effect estimation into the safety surveillance process
Cohort definition – Truven Marketscan CCAE* 2020

Initial target cohort
Liraglutide users
ATC code: A10BJ02

Qualifying target cohort
New users**
≥ 1 exposures within the last year
7 days of latency
No prior outcome
n_subject = 34240

ATC code: A10BJ02
SGLT2 inhibitors and Sulfonylurease users

ATC codes: A10BK and A10BB

Initial comparator cohort
New users*
≥1 exposures within the last year
7 days of latency
No prior outcome
n_subject = 160644

SGLT2 and Sulfonylurease users

Outcome cohort
Acute cholecystitis condition
SNOMED code: 198809
n_subject = 778873

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

*Cohort definition
Truven Marketscan CCAE* 2020

IBM® Watson Health™. Commercial Claims and Encounters Database and Medicare Supplemental and Coordination of Benefits Database IBM Marketscan Research Databases User Guide. Certain data used in this study were supplied by International Business Machines Corporation. Any analysis, interpretation, or conclusion based on these data is solely that of the authors and not International Business Machines Corporation.
Propensity score matching

**Before matching**

<table>
<thead>
<tr>
<th>Age</th>
<th>Gender</th>
<th>BMI, Obesity</th>
<th>Parity</th>
<th>Diabetic complications (Retinopathy, Nephropathy, CV diseases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cases**

<table>
<thead>
<tr>
<th></th>
<th>Cases (n)</th>
<th>Prevalence (per 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>71</td>
<td>2.30</td>
</tr>
<tr>
<td>Comparator</td>
<td>83</td>
<td>1.23</td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td>1.57</td>
</tr>
</tbody>
</table>

**1:3 propensity score matching**
Population level risk estimation

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

Minimum detectable relative risk: 1.62 ± 0.17 (SE)
Conclusion – Next steps

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

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

• A new test case for another therapeutic area
  – Negative outcome controls
  – Data driven selection of covariates
Rapid Fire Presentations of Collaborators

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

24 June 2022

Julia Kurps, The Hyve
EUROPE: AN OCEAN OF DATA & A DESERT FOR ANALYSIS
Identify **relevant data** to answer your **research question**

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

...
ROLE OF THE EHDEN PORTAL IN THE STUDY WORKFLOW

Study Idea

What data is needed?

Where do I find the data?

Assemble the Team

Protocol

Is the Study Feasible?

EHDEN Portal
• Database Catalogue
• Network Dashboard

Publication

Dashboard

Results Collection

Execute Analysis

Define Analysis

EHDEN Evidence Hub

Atlas & Arachne

Arachne & Atlas
THE VISION OF THE EHDEN PORTAL – ONE-STOP-SHOP
EH DEN PO RTAL RE LEASE TO DAY

Database Catalogue

Network Dashboards

Academy

Available Q4 2023

Available Q4 2022

Available Q4 2022

Available Q4 2023
**GEOGRAPHICAL SPREAD DATA PARTNERS**

**DP Call #6**
56 applications in 17 countries

Final DP call
Oct/Nov 2022

**Applications (n=283)**
- Spain
- Italy
- UK
- Belgium
- France
- Portugal
- Germany
- The Netherlands
- Finland
- Switzerland
- Croatia
- Serbia
- Turkey
- Norway
- Hungary
- Greece
- Israel
- Czech Republic
- Austria
- Ireland
- Sweden
- Montenegro
- Luxembourg
- Estonia
- Denmark
- Bulgaria

**Awarded applications (n=140)**

Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more).
### DATABASE CATALOGUE – 67 DATA PARTNERS...AND COUNTING

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACI</td>
<td>Auria Clinical Informatics at Turku University Hospital District of Southwest Finland (HDSF)</td>
<td>Finland</td>
</tr>
<tr>
<td>ADW IMR</td>
<td>Anonymized Data Warehouse- IMR Innovative Medical Research S.A.</td>
<td>Greece</td>
</tr>
<tr>
<td>AP-HM</td>
<td>Health Data Warehouse of Assistance Publique - Hôpitaux de Marseille</td>
<td>France</td>
</tr>
<tr>
<td>ARCA</td>
<td>ARCA Cardiology Monasterio Foundation</td>
<td>Italy</td>
</tr>
</tbody>
</table>
COMPONENTS OF THE DATABASE CATALOGUE

Database Fingerprint - meta data -

- # of patients
- Gender distribution
- Age of first observation
- Year of birth

Database Dashboards - based on CDM data -

1. Database Description (20/20) 100%
2. Contact Details (10/10) 100%
3. Technical Details CDM (5/5) 100%
4. Data Governance and Ethics (9... 100%
5. Publications (2/2) 100%
All databases in **France**?

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

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

In which countries do we have **up to date data on COVID-19 vaccinations**?
NETWORK DASHBOARD
Network Dashboards - Overview

Total number of patients per country

Patients by Country

Database Types per Country
FILTER TO FIND RELEVANT DATA

Use case examples

- Geographical spread
  Identify all EHDEN Data Partners from France

- Range of Source data systems
  Identify Data Partners that can provide GP data in combination with hospital data
Distribution of age at first observation period
GETTING STARTED

Network Dashboard at a Glance

Getting Access to the Network Dashboard
- Enter credentials on the EHDEN portal (https://portal.ehdenn.eu)
- Select the Network Dashboard icon on the left panel

Understanding gender or year of birth or age coverage
1. Go to the tab ‘Demographics’
2. Gender distribution is shown in the top pie chart
3. Age at first observation is presented in tabular format, bar chart and boxplot
4. Year of birth is presented as bar chart.

Understanding what data are available for one or more data sources
1. Go to the tab ‘Data Domains’
2. The ‘average number of records per person’ indicates how many records per entity type are available. Use in conjunction with the above filters to select on country/database type or data source
3. The pie chart gives the total number of records per entity type

What is the source of the data?
1. Go to the tab ‘Data Provenance’
2. The pivot table ‘Visit Type Pivot’ gives a count per entity type of the respective concept types e.g. EHR problem list entry, primary or secondary condition, hospitalisation cost record etc.

What’s the available observation period?
1. Go to the tab ‘Observation Period’
2. Check the second graph – cumulative observation time
3. Click in the legend on a single item to select / deselect
4. Double click to select a single data source

Which data sources contain a particular concept?
1. Go to the tab ‘Concept Browser’
2. In the concept type, start typing part of the text string
3. Select the concept that you want to search for
4. Hit APPLY
5. In the pivot table you can see approx. total number of records and descendant number of records. Further down, you see a graphical representation

Nice-to-know
1. Knowing the last update per datasource
2. On the landing page of the network dashboard
3. Scroll to the bottom – meta data view
4. Cutoff date from the source data: source_release_date
5. CDM date: cdm_release_date

Total Number of datasources and patients in the network
- Use the country / database type / data source filter to restrict the selection

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

3. Maxime your graph
- When a graph is not readable: click on the icon and select ‘maximise graph’ from the pop-up
More Data Partners will be added
Submission deadline DP Call #6 just closed

Integration with Data Analytics and Study Management capabilities

Implementation of Production Instance of service desk for study requests
WELCOME – WE ARE OPEN

Register your account today at Portal.EHDEN.eu

After approval you can start exploring high quality real world data from > 44 M patients from 15 European countries
Welcome to the EHDEN Portal

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

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

- Database Catalogue: Provides metadata on the databases in the EHDEN data network
- Network Dashboards: Allows to analyse and compare aggregated data from the OMOP CDM databases in the
- EHDEN Academy: Our free, online and publicly available learning platform
- Publications: An overview of all deliverables and publications of EHDEN
WP4 – Building The Infrastructure

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

Study Management and Execution
- EHDEN Portal
- Database Catalogue
  - Arachne
  - Atlas
  - EHDEN Evidence Hub

Auxiliary
- EHDEN Academy
- Virtual Training Environment
  - Website
  - Forum
STUDY WORKFLOW IN EHDEN – TECHNICAL PERSPECTIVE

Study Eligibility

Study Idea

What data is needed?

Where do I find the data?

Assemble the Team

Study Feasibility

Protocol

Is the Study Feasible?

Study Execution

Define Analysis

Perform Analysis

Results Collection

Dashboard

Publication

Evidence: Results Collection, Analysis and Sharing
Lunch, Collaborator Showcase, and Early Investigator meetings

The lunch is made possible with the help of Synapse

The collaborator showcase is made possible with the help of Promptly

Early Investigators mentor meetings in the Queen’s Lounge