OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

European Symposium Review

OHDSI Community Call
June 28, 2022 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
July 5	NO MEETING
July 12	New Adopter Introductions and Q&A
July 19	Workgroup Updates
July 26	CDM Update Process

Upcoming OHDSI Community Calls

Date	Topic
July 5	NO MEETING
July 12	New Adopter Introductions and Q&A
July 19	Workgroup Updates
July 26	CDM Update Process

July 12: New Adopters & Community Members

Our July 12 Community Call will be focused on new adopters of the OMOP CDM or new members of the OHDSI community.

We are welcoming people to introduce themselves, share why they have joined the community and what impact they hope to make, and also ask a question to the broader community (if you wish). If you would like to take part in this event, please fill out this form to help us plan the session: ***https://bit.ly/3A7JNkV***

Form in chat and on community calls page

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

A Record-Setting Submission Year!

Thank you to everybody who submitted brief reports to join our #OHDSI2022 Collaborator Showcase. We had a record amount (more than 130!) of submissions for poster presentations, software demos and oral presentations for the 2022 OHDSI Symposium, which will be held Oct. 14-16 in Bethesda, Md.

The scientific committee meets this week to begin the process of reviewing all submissions, and **selected presenters will be notified by August 1.**

OHDSI Shoutouts!

Any shoutouts from the community? Please share and help promote and celebrate OHDSI work!

Have a study published? Please send to sachson@ohdsi.org so we can share during this call and on our social channels.
Let's work together to promote the collaborative work happening in OHDSI!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

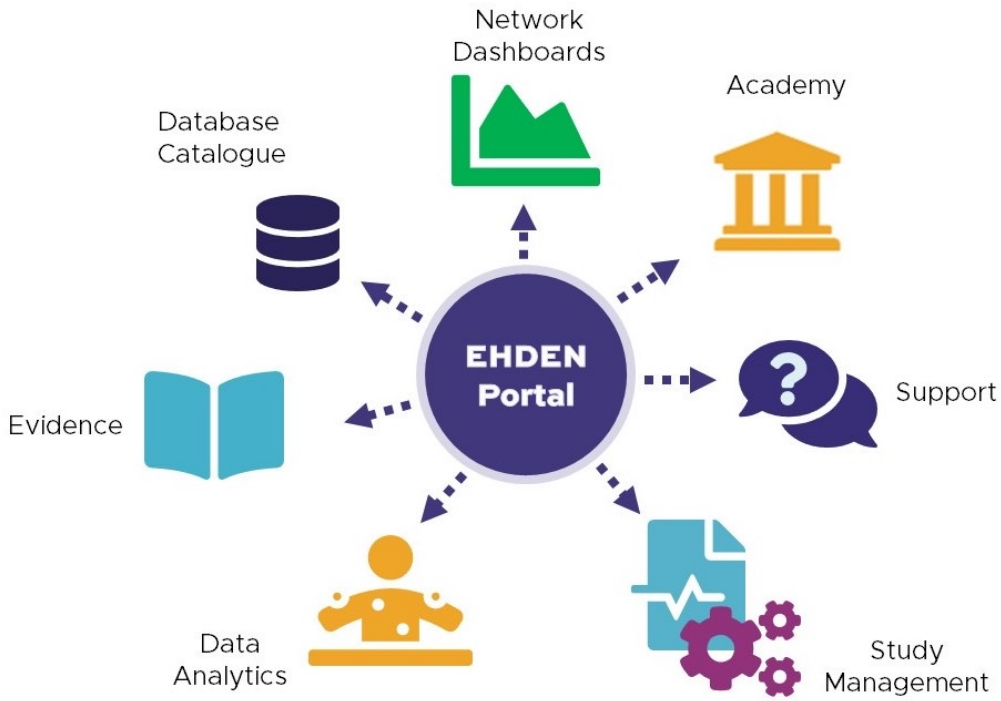
Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model Vocabulary Subgroup
Tuesday	3 pm	OMOP CDM Ongology Outreach/Research Subgroup
Wednesday	11 am	Open-Source Community
Wednesday	12 pm	FHIR and OMOP Terminologies Subgroup
Wednesday	7 pm	Medical Imaging
Thursday	10 am	Data Quality Dashboard
Thursday	12 pm	FHIR and OMOP Oncology Subgroup
Friday	9 am	GIS – Geographic Information Systems
Tueesday	10 am	Common Data Model

www.ohdsi.org/upcoming-working-group-calls

EHDEN Portal/Data Catalogue

The EHDEN Portal, which provides free access to the research community, was launched at the 2022 OHDSI European Symposium. The Portal includes a Data Partner Catalogue (140 partners, >500M anonymous patient records) and Feasibility Dashboards that support data discoverability (findable under Findable, Accessible, Interoperable and Reusable (FAIR) principles).



www.ohdsi.org/ohdsi-news-updates/

OHDSI EHR Data Survey

MPhilofsky Melanie Philofsky 6h

Hello friends with EHR data,

One of the Healthcare Systems group's objectives this year is "To provide support for transforming source EHR data to the CDM". Currently, we provide support through answering questions on the forums and during our regularly scheduled work group meetings. Another product we would like to provide to the community is a central repository of different OMOP sites, their underlying EHR system, and attributes. This will allow new OHDSI collaborators to find and reach out to sites with similar infrastructure, EHR systems, and/or research goals. Participating in this survey does NOT commit you to being a mentor, providing your ETL script, or even answering your email. However, we hope you embrace the spirit of our open source community and contribute to the cause. We all learn as we OMOP our data. I've been very active in the OHDSI community and digging deep into EHR data for 8 years, and I still learn something new every day. But I think all persons in any field of science continue to learn because science is continually evolving. Here's the link 1 to the google form.

Reply

Job Openings

Professor **Peter Rijnbeek** announced an opening for an epidemiologist to work with his team at Erasmus MC.

This position will be responsible for all aspects of observational research including protocol writing, input in the statistical analysis plan, study execution, interpretation of results and report/manuscript writing.

The application deadline is July 8, 2022.

Epidemiologist

Published	Deadline	Location
9 Jun	7 Jul	Rotterdam

JOB DESCRIPTION

This research will be performed in close collaboration with the Observational Health Data Sciences and Informatics (OHDSI) initiative, which is a global, multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics , and the EU-sponsored European Health Data and Evidence Network (EHDEN) which develop frameworks to generate reliable real-world evidence.

In your function as Epidemiologist you will be responsible for all aspects of observational research including protocol writing, input in the statistical analysis plan, study execution, interpretation of results and report/manuscript writing.

Job Openings

Odysseus Data Services recently announced two openings, one for an epidemiologist and one for a data scientist.

Check out the links on the community calls page or reach out to a member of the Odysseus team to learn more!

Odysseus Data Services (Odysseus) has an exciting opening for an **Epidemiologist**. This role will be responsible for supporting the development, maintaining, and troubleshooting of the cutting-edge distributed solutions in the Real-World Evidence (RWE) area, utilized by the researchers in Pharmaceutical, Healthcare and Payer industries. Odysseus is looking for a self-driven individual who can hit the ground running, quick learner and wants to be a part of our dynamic global team.

Responsibilities

- Lead and contribute to the design of observational database analysis, including authoring protocol, reviewing and providing relevant epidemiological and project-specific comments to statistical analysis plans and analysis output
- Participate in the design and development of standardized analytic tools to generate reliable and reproducible evidence in a network of observational data
- Contribute to the execution of observational database analyses using standardized analytical tools and writing statistical packages
- Contribute to the dissemination of scientific information through technical reports and publications in peer-reviewed literature.
- Work closely with healthcare and pharmaceutical customers to identify their needs
- Contribute to the development of complex phenotypes using advanced analytic approaches (i.e. machine learning, incorporating unstructured data sources using NLP, etc.)

Qualifications

- Graduate degree (MS, PhD, MD, etc) in epidemiology, biostatistics, pharmacy, public health or related clinical discipline plus two years' experience in observational research. PhD preferred
- Experience in designing and conducting healthcare studies and in development and applications of advanced analytics solutions
- Strong epidemiology and biostatistics background
- Experience using OHDSI tools and analytical methods is a big plus

Odysseus Data Services (Odysseus) has an exciting opening for a **Healthcare/Clinical Data Scientist**. This role will be responsible for supporting the development, maintaining, and troubleshooting of the cutting-edge distributed solutions in the Real-World Evidence (RWE) area, utilized by the researchers in Pharmaceutical, Healthcare and Payer industries. Odysseus is looking for a self-driven individual who can hit the ground running, quick learner and wants to be a part of our dynamic global team.

Responsibilities

- Lead and contribute to the design, development and documentation of standardized analytic tools that will be executed against a network of observational data
- Lead the execution of observational database analyses using standardized analytical tools and writing statistical packages
- Provide technical support for the data and analysis infrastructure and scientific support
- Contribute to writing of protocols and statistical analysis plans, methods development, conduct of simulation studies and statistical/mathematical modeling studies
- Lead and contribute to the development of complex phenotypes using advanced analytic approaches (i.e. machine learning, incorporating unstructured data sources using NLP, etc
- Contribute to the dissemination of scientific information through technical reports and publications in peer-reviewed literature.
- Lead and contribute to the development of novel analytic tools and techniques to leverage the EHR data for rapid, reliable and reproducible evidence generation

2022 OHDSI Symposium

Registration is OPEN for #OHDSI2022!





The 2022 OHDSI Symposium will be held Oct. 14-16 at the Bethesda North Marriott Hotel & Conference Center.

www.ohdsi.org/ohdsi2022symposium

An Introductory Journey From Data To Evidence

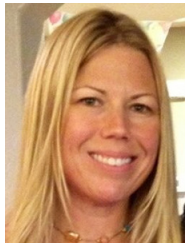
OHDSI2022 Tutorial • Saturday, Oct. 15 • Bethesda, Md.

The OHDSI Journey: Where Are We Going?

Patrick Ryan

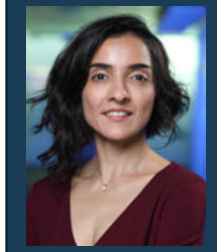
OMOP Common Data Model and Vocabulary

Clair Blacketer



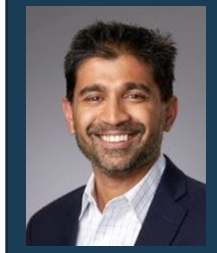
ETL – A Source Database Into OMOP CDM

Melanie Philofsky



Creating Cohort Definitions

Asieh Golozar



Phenotype Evaluations

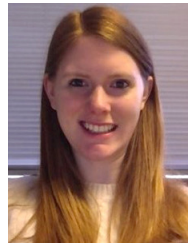
Gowtham Rao

Characterization

Kristin Kostka

Estimation

Martijn Schuemie



Prediction

Jenna Reps

The OHDSI Journey: Where Do We Go From Here?

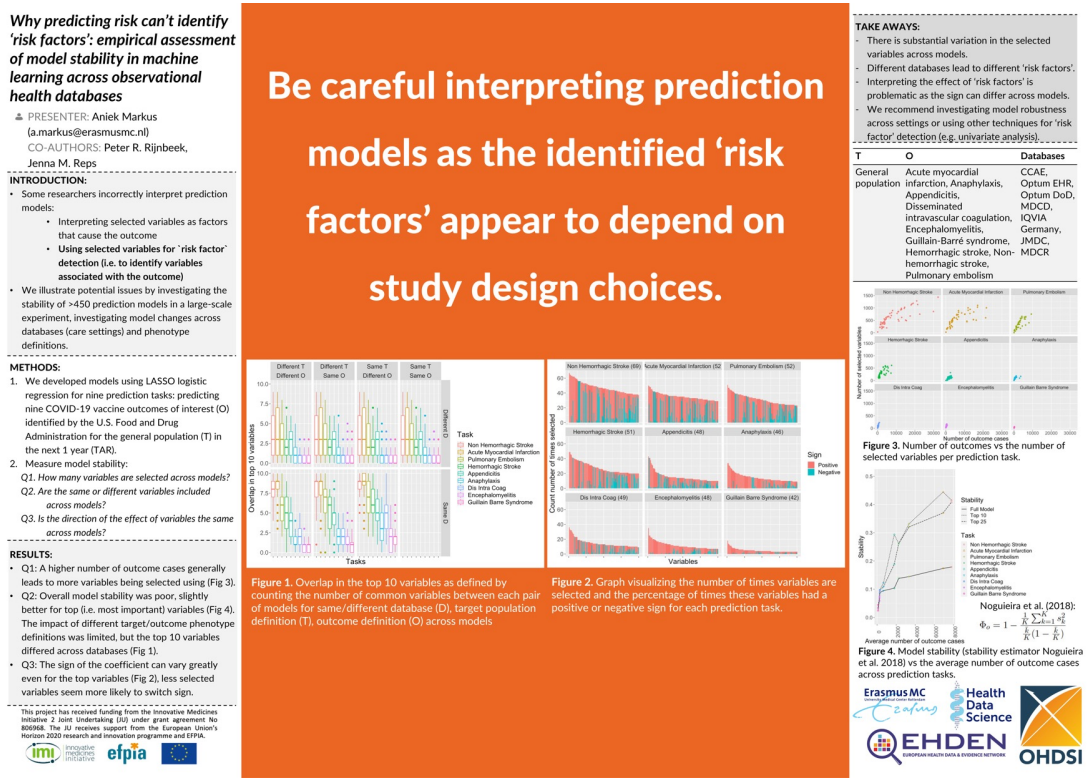
George Hripcsak

Workgroup Activities

Saturday, Oct. 15, and Sunday, Oct. 16

Saturday, Oct 15					
Start Time (ET)	End Time (ET)				
800	900	Tutorial	HADES Hack-a-thon: Part 1	Oncology WG	FHIR-OMOP: Terminologies Subgroup, Part 1
900	1000				
1000	1100				FHIR-OMOP: Increasing the Value of Data Through a Rich Set of Attributes
1100	1200				
1200	1300		Lunch	Lunch	Lunch
1300	1400		Methods Research (PLE/PLP)	Oncology WG (continued)	FHIR-OMOP: Data Model Harmonization Subgroup
1400	1500				
1500	1600			Natural Language Processing	FHIR-OMOP: Oncology Subgroup
1600	1700				
1700	1800				FHIR-OMOP: Terminologies Subgroup, Part 2
1800	1900				
Sunday, Oct 16					
800	900	All-Hands Workgroup Meeting			
900	1000				
1000	1100				
1100	1200				
1200	1300	Lunch		Lunch	Lunch
1300	1400	Phenotype Evaluation	HADES Hack-a-thon: Part 2	Education	CDM and Data Quality
1400	1500				
1500	1600			Health Equity	
1600	1700				

#OHDSISocialShowcase This Week



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

PRESENTER: Aniek Markus (a.markus@erasmusmc.nl)
CO-AUTHORS: Peter R. Rijnbeek, Jenna M. Reps

INTRODUCTION:

- Some researchers incorrectly interpret prediction models:
  - Interpreting selected variables as factors that cause the outcome
  - **Using selected variables for `risk factor` detection (i.e. to identify variables associated with the outcome)**
- We illustrate potential issues by investigating the stability of >450 prediction models in a large-scale experiment, investigating model changes across databases (care settings) and phenotype definitions.

METHODS:

1. We developed models using LASSO logistic regression for nine prediction tasks: predicting nine COVID-19 vaccine outcomes of interest (O) identified by the U.S. Food and Drug Administration for the general population (T) in the next 1 year (TAR).
2. Measure model stability:
   - Q1. *How many variables are selected across models?*
   - Q2. *Are the same or different variables included across models?*
   - Q3. *Is the direction of the effect of variables the same across models?*

RESULTS:

- Q1: A higher number of outcome cases generally leads to more variables being selected using (Fig 3).
- Q2: Overall model stability was poor, slightly better for top (i.e. most important) variables (Fig 4). The impact of different target/outcome phenotype definitions was limited, but the top 10 variables differed across databases (Fig 1).
- Q3: The sign of the coefficient can vary greatly even for the top variables (Fig 2), less selected variables seem more likely to switch sign.

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.

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

Figure 1. Overlap in the top 10 variables as defined by counting the number of common variables between each pair of models for same/different database (D), target population definition (T), outcome definition (O) across models

Figure 2. Graph visualizing the number of times variables are selected and the percentage of times these variables had a positive or negative sign for each prediction task.

TAKE AWAYS:

- There is substantial variation in the selected variables across models.
- Different databases lead to different 'risk factors'.
- Interpreting the effect of 'risk factors' is problematic as the sign can differ across models.
- We recommend investigating model robustness across settings or using other techniques for 'risk factor' detection (e.g. univariate analysis).

T	O	Databases
General population	Acute myocardial infarction, Anaphylaxis, Appendicitis, Disseminated intravascular coagulation, Encephalomyelitis, Guillain-Barré syndrome, Hemorrhagic stroke, Non-hemorrhagic stroke, Pulmonary embolism	CCAE, Optum EHR, Optum DoD, MDCD, IQVIA Germany, JMDC, MDCR

Figure 3. Number of outcomes vs the number of selected variables per prediction task.

Figure 4. Model stability (stability estimator Noguieira et al. 2018) vs the average number of outcome cases across prediction tasks.

Erasmus MC Cancer Institute · Health Data Science · EHDEN European Health Data & Evidence Network · OHDSI

MONDAY

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

Lead: Aniek Markus

#OHDSISocialShowcase This Week

OMOP Genomic mapping capacities in conversion of comprehensive genomic profiling results

Maria Rogozhkina[1], Vlad Korsik[1], Varvara Savitskaya[1], Alexander Davydov[1]
[1] - Odysseus Data Services

Introduction

Omics data (genomics, proteomics, metabolomics and etc) tends to be the most important data currently because of its possibility to influence decisions in a variety of medical fields.
The primary aim of this study was to evaluate the efficacy of representation of genetics data by OMOP Genomic vocabulary. We utilized a real world data depersonalized database (so that it can be applied to any enriched genomic data format such as FMI (FoundationOne), VCF (Variant Call Format), GFF (General feature format) and others) to test the conversion capacity of the vocabulary.
The secondary aim was to define the best conversion strategy for real world database test results (it is one of the most commonly used comprehensive genomic profiling systems worldwide).

Methods

We processed 3 source genomic data repositories from a real world databases including comprehensive genomic profiling systems.

- Source_data_type_1 = copy number = CN (i.e. amplification or deletion),
- Source_data_type_2 = short variant = SV (a single nucleotide variations or small insertion/deletion),
- Source_data_type_3 = multiple myeloma genetic data = MMGD.

The conversion was performed on CDM 5.4 version with OMOP Vocabulary version: v5.0 09-APR-22.

Among The CN table we analyzed 6 033 de-identified records with name of gene, synonym name, type of mutation (amplification or deletion) and copy number:

source_table_name	gene_name	gene_syn	amplificationordeletion	copynumber
CN	STK11	STK11	deletion	0
CN	CDK8	CDK8	amplification	8
CN	CUL4A	CUL4A	amplification	8

For mapping automation we applied the full-match approach with subsequent manual curation. Every distinct source name of the gene was a full name counterpart in postcoordination approach. In precoordination approach distinct source name of the gene and type of mutation were a full name counterpart. The list of targets was aggregated by concept_id list of preferred names and synonyms.

Single Nucleotide Variation table includes 233 793 records with information about the gene, substitution in DNA, RNA and protein with position, number of chromosome, type of alteration, sequencing coverage and many other (22 columns). Here are shown only the most important ones:

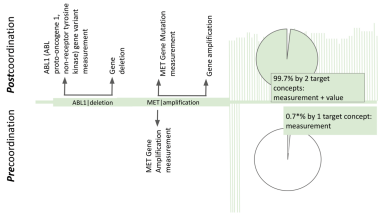
source_table_name	gene	RNA	protein	chromosome	codingtype	sequencingcoverage
SV_extended	SMO	2081delC	P694fs*82	chr7	frameshift	1622
SV_extended	APC	2078A>C	K693T	chr5	missense	4394
SV_extended	TP53	473G>T	R158L	chr17	missense	6742

Both protein and RNA columns are filled well so source_concept_name was compiled as gene, RNA, protein. At first we made a conversion on the RNA column, if nothing was found, we did matching on the protein column, and if nothing was found again, we tried to do uphill mapping on the Genetic Variation class.

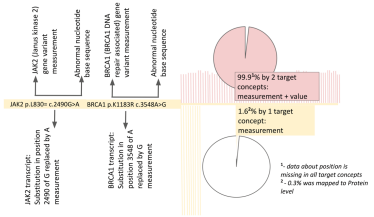
MMGD is a little table containing only 33 rows with information about specimen, method, gene and type of abnormality with all listed fields used to define source_code.

source_table_name	specimen	method	gene
MMGD	Bone Marrow Aspirate	FISH	T(14;16)
MMGD	Bone Marrow Biopsy	FISH	AMP1Q21

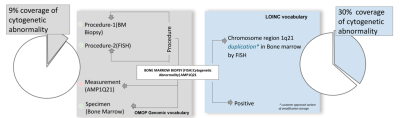
Results



For copy number table the coverage of source data was performed by postcoordination as: meaningful coverage of both genetic variation and related copy number type may be achieved in 99,7% of cases with *postcoordiantion* with semantics equivocal distribution between event and it's value.
In contrast, *precoordination* resulted in coverage of 0,7% of data, with absolute (one-to-one) semantics match coverage. I.e. *postcoordination* seems to be optimal strategy to define the changes in number of genes.



Lack of appropriate target concepts results in need to uphill mapping in most cases of nucleotide variation tables.
Despite almost perfect (99,9%) SV table's records mapping to OMOP Genomic concepts the semantic coverage remains improper.
Full semantic match is attributed to 1,3% of codes were targeted to RNA Variant and 0,3% of codes were mapped to Protein Variant. Leftover concepts (97,7%) were targeted to Genetic Variation I.e. uphill mapping is the major storing strategy.



Explicit, 1-to-1 representation request is a MMGD scenario. Impossibility of OMOP Genomic to provide the way to reflect the cytogenetic abnormality , the need to store specimen and method attributes as separate facts resulted in targeting to non-genomic terminology.

Table name	Modeling Approach	Mapping Rate	Coverage rate		
			Gene coverage	Alteration type coverage	Alteration coordinate coverage
SNV	Pre	~2%	→100%	→100%	→100%
	Post	~100%	→100%	0%	0%
CN	Pre	~1%	→100%	~1%	NA
	Post	~100%	→100%	→100%	NA

Rationale for appropriate modeling approach selection for specific alteration types

Conclusion

OMOP Genomic vocabulary covers a large number of needs well, but some improvements are also needed. It is required to:

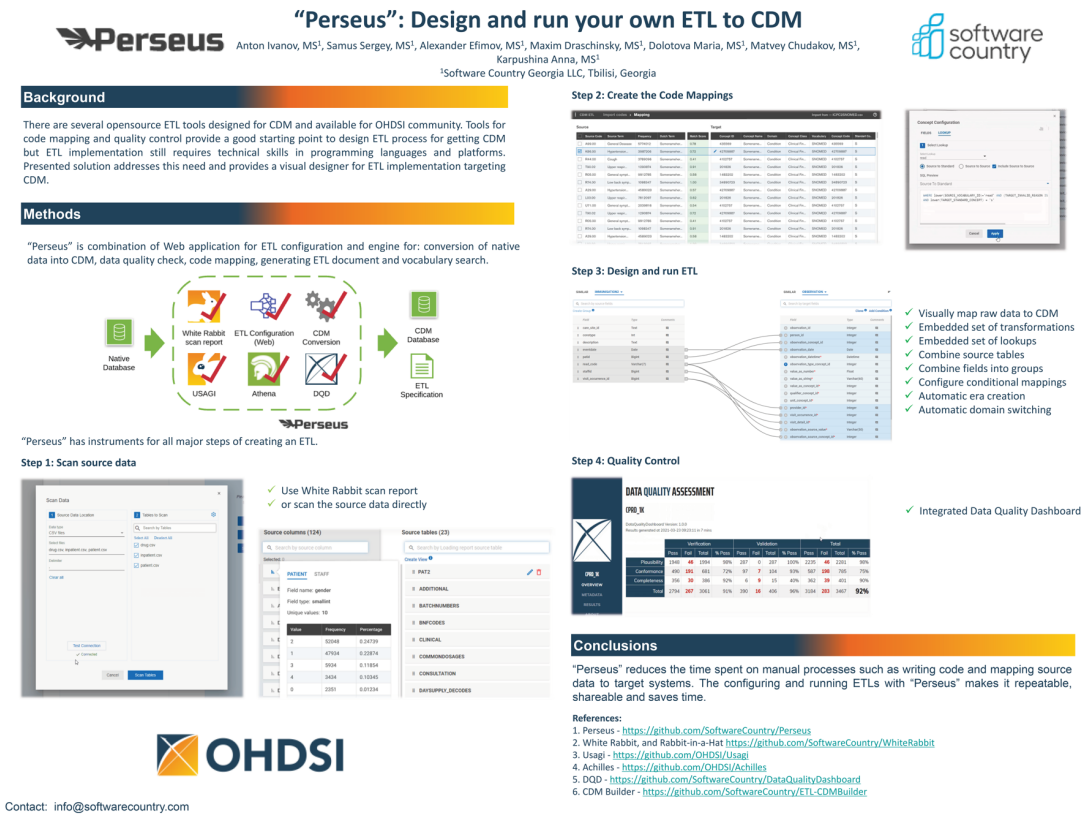
- perform deduplication and expand the list of synonyms for a better search using Clingen database
- increase the number of concepts to cover a larger number of cases: transfer information from DoCM and Cancer Hotspots, take Clinically relevant variation from ICGC to the OMOP Genomic vocabulary.
- make the genomic LOINC/SNOMED etc concepts non-standard, and then map them to the standard OMOP Genomic concepts.
- prevent "combinatorial explosion" by allowing postcoordiantion at least for Copy Number changes
- ratify the logic for storage of method and specimen

TUESDAY

OMOP Genomic mapping capacities in conversion of comprehensive genomic profiling results

Lead: Maria Rogozhkina

#OHDSISocialShowcase This Week



"Perseus": Design and run your own ETL to CDM

Perseus

Anton Ivanov, MS[1], Samus Sergey, MS[1], Alexander Efimov, MS[1], Maxim Draschinsky, MS[1], Dolotova Maria, MS[1], Matvey Chudakov, MS[1], Karpushina Anna, MS[1]
[1]Software Country Georgia LLC, Tbilisi, Georgia

software country

Background

There are several opensource ETL tools designed for CDM and available for OHDSI community. Tools for code mapping and quality control provide a good starting point to design ETL process for getting CDM but ETL implementation still requires technical skills in programming languages and platforms. Presented solution addresses this need and provides a visual designer for ETL implementation targeting CDM.

Methods

"Perseus" is combination of Web application for ETL configuration and engine for: conversion of native data into CDM, data quality check, code mapping, generating ETL document and vocabulary search.

Native Database
White Rabbit scan report
ETL Configuration (Web)
CDM Conversion
USAGI
Athena
DQD
CDM Database
ETL Specification
Perseus

"Perseus" has instruments for all major steps of creating an ETL.

Step 1: Scan source data

- ✓ Use White Rabbit scan report
- ✓ or scan the source data directly

Step 2: Create the Code Mappings

Step 3: Design and run ETL

- ✓ Visually map raw data to CDM
- ✓ Embedded set of transformations
- ✓ Embedded set of lookups
- ✓ Combine source tables
- ✓ Combine fields into groups
- ✓ Configure conditional mappings
- ✓ Automatic era creation
- ✓ Automatic domain switching

Step 4: Quality Control

DATA QUALITY ASSESSMENT

- ✓ Integrated Data Quality Dashboard

Conclusions

"Perseus" reduces the time spent on manual processes such as writing code and mapping source data to target systems. The configuring and running ETLs with "Perseus" makes it repeatable, shareable and saves time.

References:
1. Perseus - https://github.com/SoftwareCountry/Perseus
2. White Rabbit, and Rabbit-in-a-Hat https://github.com/SoftwareCountry/WhiteRabbit
3. Usagi - https://github.com/OHDSI/Usagi
4. Achilles - https://github.com/OHDSI/Achilles
5. DQD - https://github.com/SoftwareCountry/DataQualityDashboard
6. CDM Builder - https://github.com/SoftwareCountry/ETL-CDMBuilder

OHDSI

Contact: info@softwarecountry.com

WEDNESDAY

Perseus Design and run your own ETL to CDM

Lead: Anton Ivanov

#OHDSISocialShowcase This Week

Using geospatial approaches and machine learning for asthma and COPD outcomes: a systematic review

Enriching OMOP CDM

PRESENTER: **Daniel Jeannetot**
d.jeannetot@erasmusmc.nl

INTRO:
Asthma & COPD are major contributor to morbidity and mortality worldwide. OMOP CDM databases provide a unique opportunity to **enrich Electronic health records with geospatial data and machine learning approaches to improve patient-level prediction**. This systematic review shows that this is still an untapped approach which large potential for exploration

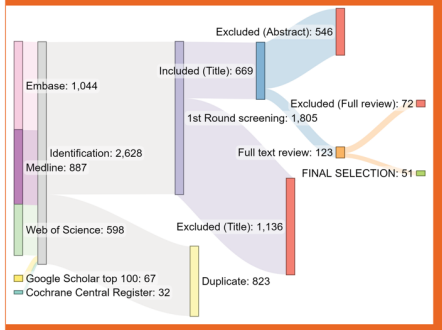
METHODS

1. Systematic review following PRISMA guideline
2. 4 databases queried
3. 3 reviewers involved in full text review
4. 12 specific characteristics for data extraction including type of models (ML/non-ML), spatial scale, spatial approach.

RESULTS

- 1805 papers screened.
- 123 Papers fully reviewed
- 51 Articles making use of geospatial approach and modelling to measure and predict asthma/related outcomes.

Asthma/COPD research has a lot of potential to benefit from machine learning algorithms and geospatial approaches, especially if combined with observational data.



Most papers approached spatial information in two steps: use of geospatial models to estimates an environmental factor exposure (e.g. specific air pollutants values at specific location) to then integrate values in non-spatial regressions models (.e.g linear, multivariate, etc.), removing specific geospatial and geographical processes information.

The type of scale used varied greatly, with most papers using a local administrative level (e.g. counties, neighbourhoods), thus local, but hard to compare or generalize. Only a few used grid-based spatial data, and even then the resolutions ranged from 5m grid to 1km grid and beyond, leading to widely disparate estimates and areas.

OHDSI provides a coherent and and readily available infrastructure to help Asthma/COPD research leverage observational data, machine learning, and geospatial approaches for very large-scale analyses

List of included studies available in annex

Inclusion criteria

- Has modelling/prediction methods
- Has geospatial/geostatistical approaches
- Explanatory variables include geographical/environmental (air pollution, green/blue space, etc.)
- Main outcome is COPD and/or Asthma related
- Population should be 18 years old or above.

Search term categories

1. Asthma and/or COPD **AND**
2. Prediction models (OR. Modelling, Machine Learning etc.) **AND**
3. Spatial (OR geostatistical; geo*, etc.) **AND**
4. ADULT (NOT children, etc.)

Key points

- Population varied greatly in age groups and sample size (min= 105, max= +50000)
- Scale greatly varied but generally local
- <10 papers used Machine learning algorithms
- Most geospatial approaches are 2 steps
- < 10 papers used specific geostatistical tools
- Inconsistent quality and application of geospatial tools

Daniel Jeannetot, Johnmary Arinze, Victor Pera, Peter Rijnbeek, Katia Verhamme

Erasmus MC

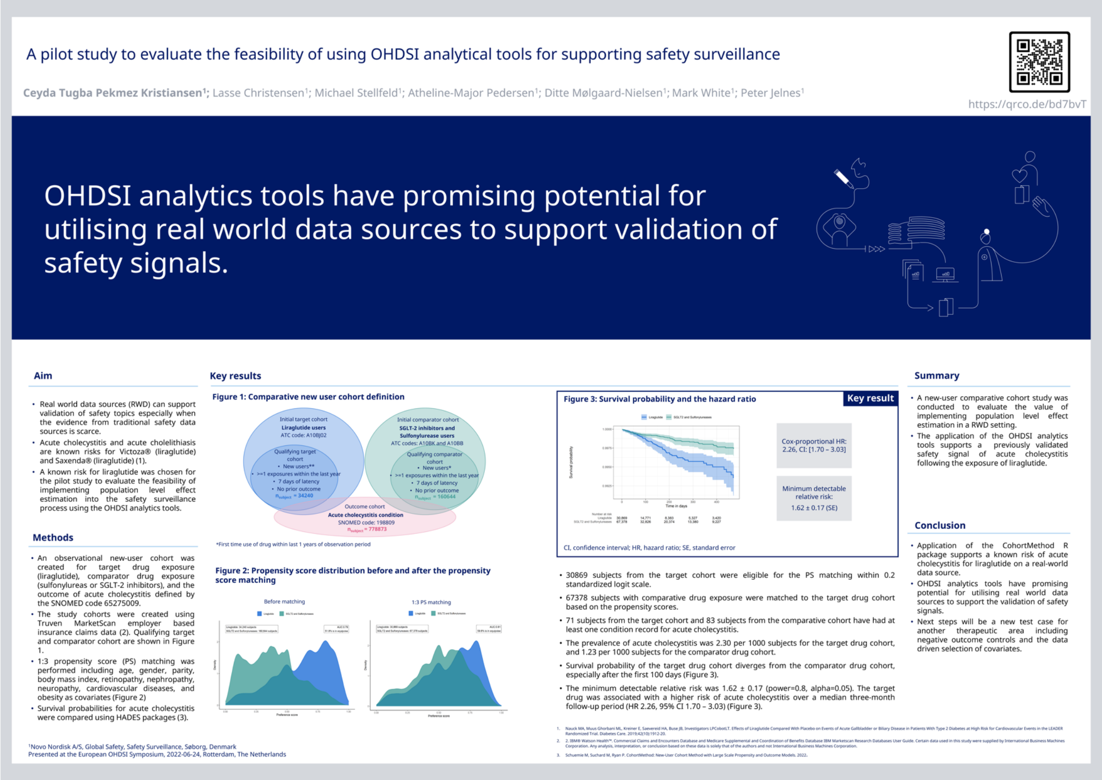
OHDSI

THURSDAY

Using geospatial approaches and machine learning for asthma and COPD outcomes: a systematic review

Lead: Daniel Jeannetot

#OHDSISocialShowcase This Week



FRIDAY

A pilot study to evaluate the feasibility of using Observational Health Data Sciences and Informatics analytics tools for supporting the validation of safety signals

Lead: Ceyda Pekmez

#OHDSISocialShowcase This Week

TROY: Trials Replication through Observational study by Yonsei

PRESENTER: **Jaehyeong Cho**

INTRO

When it is difficult to generalize the results of RCTs, a retrospective study with well-controlled confounding factors through replication of the RCT design using observational data can be a complementary alternative.

When it is difficult to generalize the results of RCTs, a retrospective study with well-controlled confounding factors through replication of the RCT design using observational data can be a complementary alternative.

METHOD

1. The two common data model (CDM) databases used in this study are Yonsei University Healthcare System (YUHS) CDM database and Ajou University School of Medicine (AUSOM) CDM database.
2. The population of interest are as follows: 1) a group of patients that were as close as possible to the recruitment target of each pivotal RCT to be replicated (group 1), and 2) a group of all patients used by having an indication for the drug (group 2).
3. The 15 randomized clinical trials to be replicated in the TROY project are shown in Table 1.

Table 1. Replication list of pivotal randomized clinical trials considered in this study.

Study	Target drug (class)	Comparator drug (class)	Note
LEADER	Liraglutide (GLP-1)	DPP-4	Placebo-controlled RCT
DECLARE-TIMI 58	Dapagliflozin (SGLT-2)	DPP-4	Placebo-controlled RCT
EMPA-REG OUTCOME	Empagliflozin (SGLT-2)	DPP-4	Placebo-controlled RCT
CANVAS	Canagliflozin (SGLT-2)	DPP-4	Placebo-controlled RCT
CARMELINA	Linagliptin (DPP-4)	SU	Placebo-controlled RCT
TECOS	Sitagliptin (DPP-4)	SU	Placebo-controlled RCT
SAVOR-TIMI 53	Saxagliptin (DPP-4)	SU	Placebo-controlled RCT
CAROLINA	Linagliptin (DPP-4)	Glimepiride (SU)	
TRITON-TIMI 38	Prasugrel + Aspirin	Clopidogrel + Aspirin	
PLATO	Ticagrelor + Aspirin	Clopidogrel + Aspirin	
ROCKET AF	Rivaroxaban	Warfarin	
ARISTOTLE	Apixaban	Warfarin	
ENGAGE AF-TIMI 48	Edoxaban	Warfarin	
ORAL	Tofacitinib	TNF inhibitor	
STAR-RA	Tofacitinib	TNF inhibitor	

4. We compare outcomes between drugs (and classes) for each pivotal RCT emulated using the packages included in HADES (formally known as the OHDSI Methods Library).

This is ongoing research

We initiated the 'Trials Replication through Observational study by Yonsei (TROY)' project to generate large population-level evidence for 15 pivotal RCTs in the real world:

Type 2 diabetes mellitus, atrial arrhythmia, acute coronary syndrome, and rheumatoid arthritis

Scan QR code link to **GitHub repository**

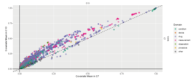
RESULTS

This is ongoing research. CohortDiagnostics was performed by replicating pivotal RCTs for three anticoagulants (ARTISTOTLE, ROCKET AF, and ENGAGE AF-TIMI 48), and the number of identified patients is presented in Table 2.

Table 2. Number of patients identified in the YUHS and AUSOM database.

Study	Target drug	Comparator drug	YUHS CDM (group 2)	AUSOM CDM (group 2)
Exposure				
ARISTOTLE	Apixaban	Warfarin	1,807(6,735) / 2,434(8,122)	135(487) / 195(961)
ROCKET AF	Rivaroxaban	Warfarin	463(2,763) / 1,864(6,355)	112(630) / 37(554)
ENGAGE AF-TIMI 48	Edoxaban	Warfarin	988(4973) / 1,796(8,450)	252(1,186) / 170(995)
Outcome				
3P MACE	LEADER, EMPA-REG OUTCOME, CANVAS, CARMELINA, SAVOR-TIMI 53, CAROLINA, TRITON-TIMI 38, PLATO, ORAL		55,792	47,333
4P MACE	TECOS		73,573	75,205
HHF + CV death	DECLARE-TIMI 58		29,945	20,664
Stroke + systemic embolism	ROCKET AF, ARISTOTLE, ENGAGE AF-TIMI 48		45,996	34,382
Cancer	STAR-RA, ORAL		70,308	41,586
MI + Stroke	STAR-RA		51,146	42,277

YUHS CDM



AUSOM CDM

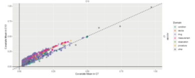


Figure 1 shows the comparison of patient characteristics between group 1 and group 2 for Apixaban.

Jaehyeong Cho[1,2] (boyinai03@gmail.com), Chungsoo Kim[3] (ted9219@ajou.ac.kr), Kyulee Jeon[2,3] (kyulee.jeon@gmail.com), Ju-Young Shin[4] (shin.jy@skku.edu), Rae Woong Park[3,5] (veritas@ajou.ac.kr), Kyung Won Kim[2,6] (kwkim@yuhs.ac), Seng Chan You[1,2] (seng.chan.you@ohdsi.org)

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
[2]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea
[3]Department of Biomedical Sciences, Ajou University School of Medicine, Suwon, Republic of Korea
[4]School of Pharmacy, Sungkyunkwan University, Suwon, Republic of Korea
[5]Department of Biomedical Informatics, Ajou University Graduate School of Medicine, Suwon, Republic of Korea
[6]Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea

YONSEI UNIVERSITY

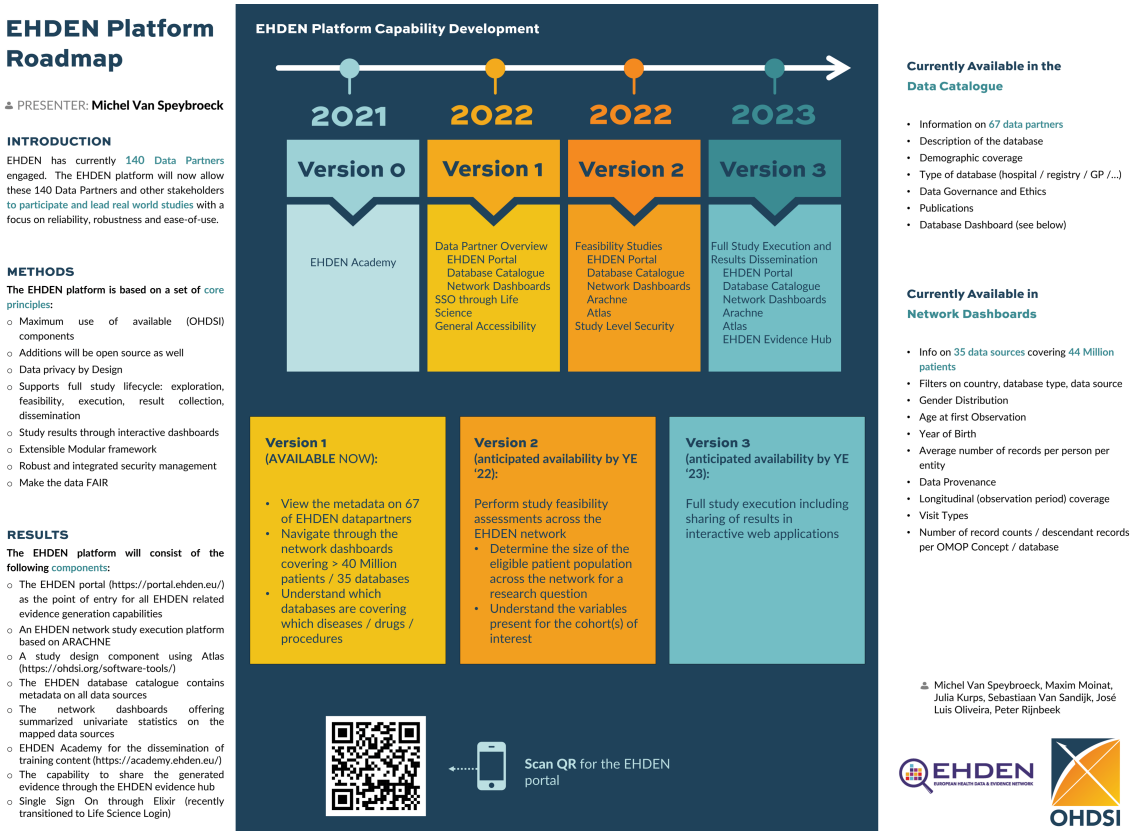
OHDSI

MONDAY

TROY: Trials Replication through Observational study by Yonsei

Lead: Jaehyeong Cho

#OHDSISocialShowcase This Week



EHDEN Platform Roadmap

PRESENTER: **Michel Van Speybroeck**

INTRODUCTION

EHDEN has currently 140 Data Partners engaged. The EHDEN platform will now allow these 140 Data Partners and other stakeholders to participate and lead real world studies with a focus on reliability, robustness and ease-of-use.

METHODS

The EHDEN platform is based on a set of core principles:

- Maximum use of available (OHDSI) components
- Additions will be open source as well
- Data privacy by Design
- Supports full study lifecycle: exploration, feasibility, execution, result collection, dissemination
- Study results through interactive dashboards
- Extensible Modular framework
- Robust and integrated security management
- Make the data FAIR

RESULTS

The EHDEN platform will consist of the following components:

- The EHDEN portal (https://portal.ehden.eu/) as the point of entry for all EHDEN related evidence generation capabilities
- An EHDEN network study execution platform based on ARACHNE
- A study design component using Atlas (https://ohdsi.org/software-tools/)
- The EHDEN database catalogue contains metadata on all data sources
- The network dashboards offering summarized univariate statistics on the mapped data sources
- EHDEN Academy for the dissemination of training content (https://academy.ehden.eu/)
- The capability to share the generated evidence through the EHDEN evidence hub
- Single Sign On through Elixir (recently transitioned to Life Science Login)

Currently Available in the Data Catalogue

- Information on 67 data partners
- Description of the database
- Demographic coverage
- Type of database (hospital / registry / GP /...)
- Data Governance and Ethics
- Publications
- Database Dashboard (see below)

Currently Available in Network Dashboards

- Info on 35 data sources covering 44 Million patients
- Filters on country, database type, data source
- Gender Distribution
- Age at first Observation
- Year of Birth
- Average number of records per person per entity
- Data Provenance
- Longitudinal (observation period) coverage
- Visit Types
- Number of record counts / descendant records per OMOP Concept / database

Michel Van Speybroeck, Maxim Moinat, Julia Kurps, Sebastiaan Van Sandijk, José Luis Oliveira, Peter Rijnbeek

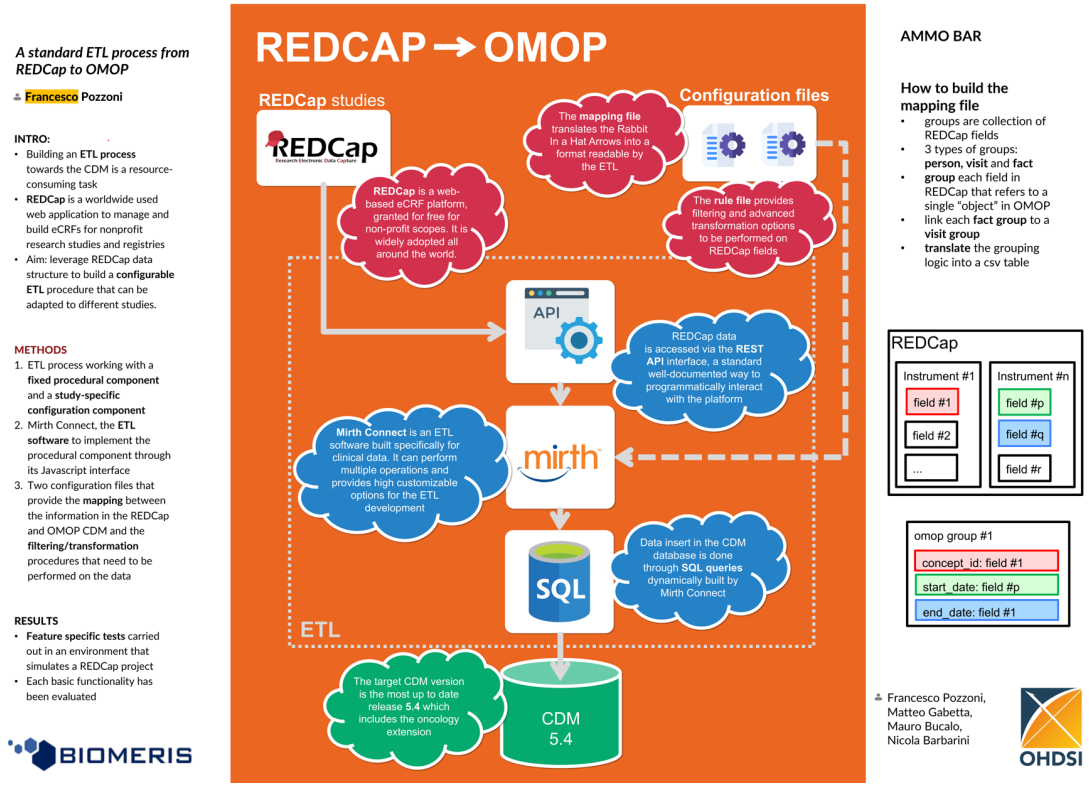
EHDEN OHDSI

TUESDAY

The EHDEN Platform Roadmap

Lead: Michel Van Speybroeck

#OHDSISocialShowcase This Week

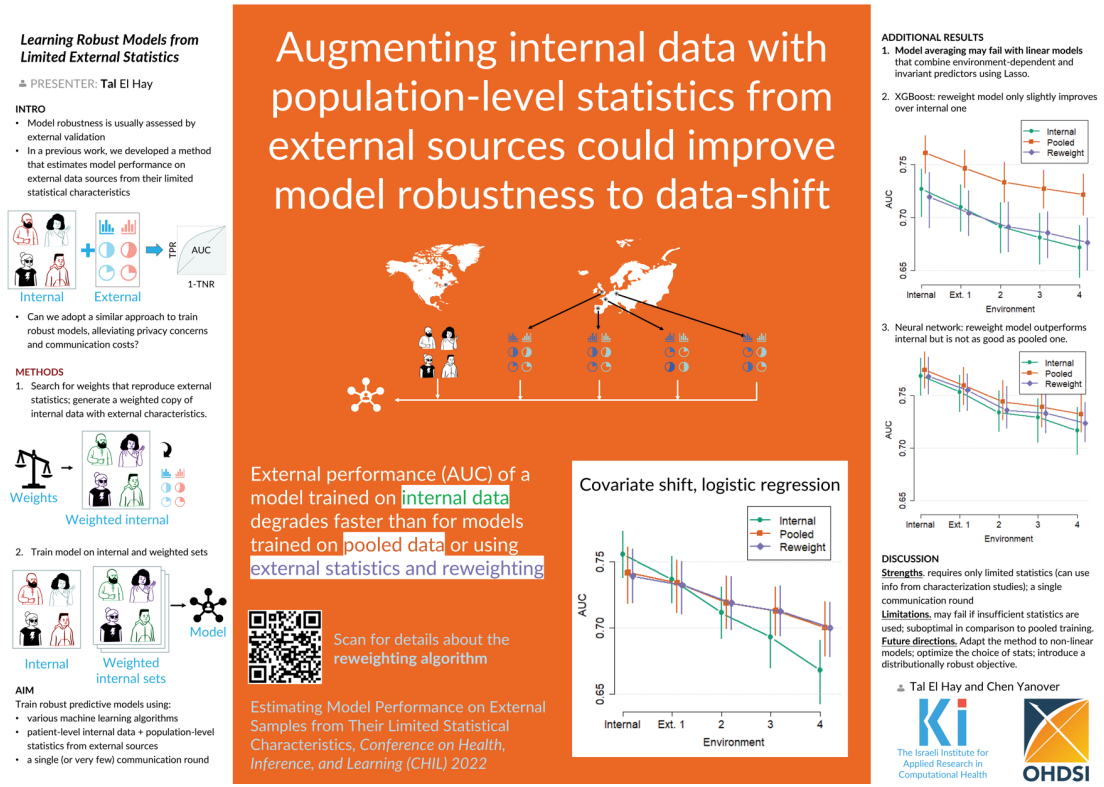


WEDNESDAY

A standard ETL process from REDCap to OMOP

Lead: Francesco Pozzoni

#OHDSISocialShowcase This Week

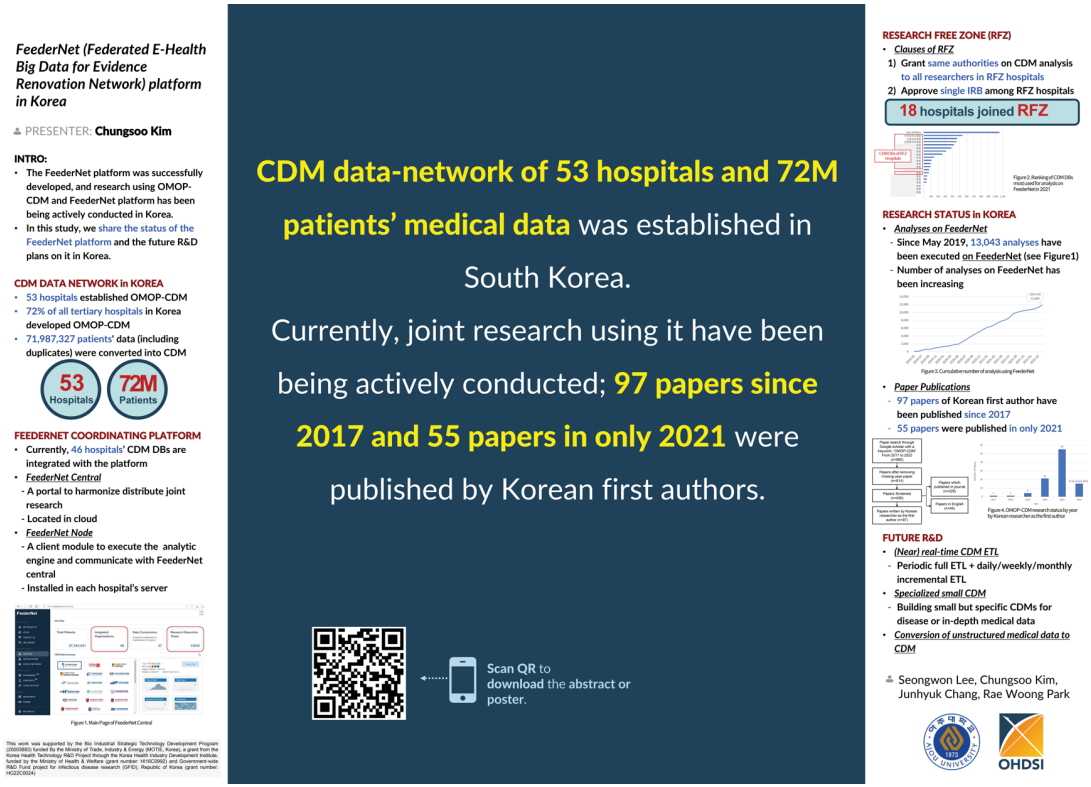


THURSDAY

Learning robust models from limited external statistics

Lead: Tal El Hay

#OHDSISocialShowcase This Week



FRIDAY

FeederNet (Federated E-Health Big Data for Evidence Renovation Network) platform in Korea

Lead: Chungsoo Kim

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

June 28: The European Symposium

Presenter:

Nigel Hughes • Director, Observational Health Data Analytics at Janssen Research & Development