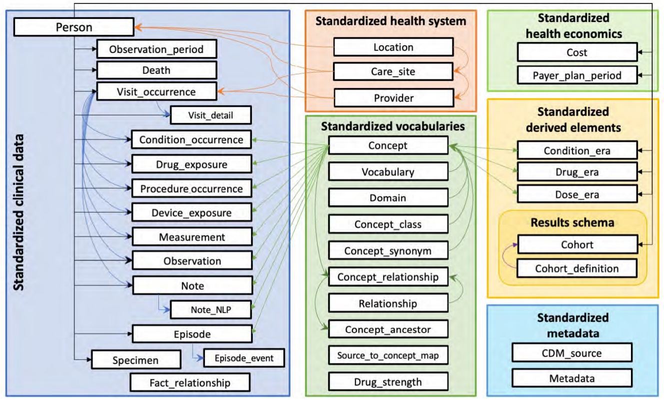


Trick or Treat

How to use OHDSI tools to quickly generate insights from your OMOP CDM

OHDSI Community Call
Oct. 26, 2021 • 11 am ET

Upcoming OHDSI Community Calls

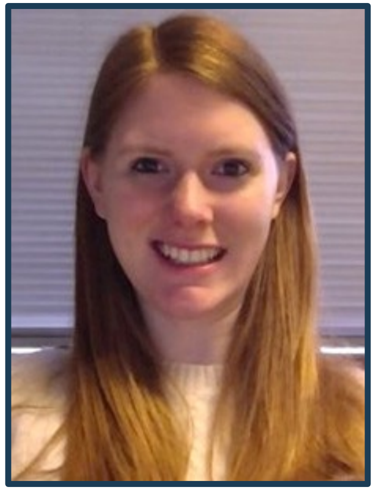
Date	Topic
Oct. 26	Trick or Treat: How to use OHDSI tools to quickly generate insights from your OMOP CDM
Nov. 2	Collaboration Opportunities: Methods Res., Data Standards, Open-Source, Clinical App.
Nov. 9	Demos: Tools for Adoption of OHDSI Data Standards
Nov. 16	Open Network Studies
Nov. 23	History of OHDSI
Nov. 30	Collaborator Showcase Presentations

Upcoming OHDSI Community Calls

Date	Topic
Oct. 26	Trick or Treat: How to use OHDSI tools to quickly generate insights from your OMOP CDM
Nov. 2	Collaboration Opportunities: Methods Res., Data Standards, Open-Source, Clinical App.
Nov. 9	Demos: Tools for Adoption of OHDSI Data Standards
Nov. 16	Open Network Studies
Nov. 23	History of OHDSI
Nov. 30	Collaborator Showcase Presentations

Nov. 2: Future Collaboration Opportunity Breakouts

Methods Research


Jenna Reps


Martijn Schuemie

Data Standards


Clair Blacketer


Maxim Moinat

Open-Source Development


Adam Black


Anthony Sena

Clinical Applications


Talita Duarte-Salles


Asieh Golozar

Three Stages of The Journey

Where Have We Been?

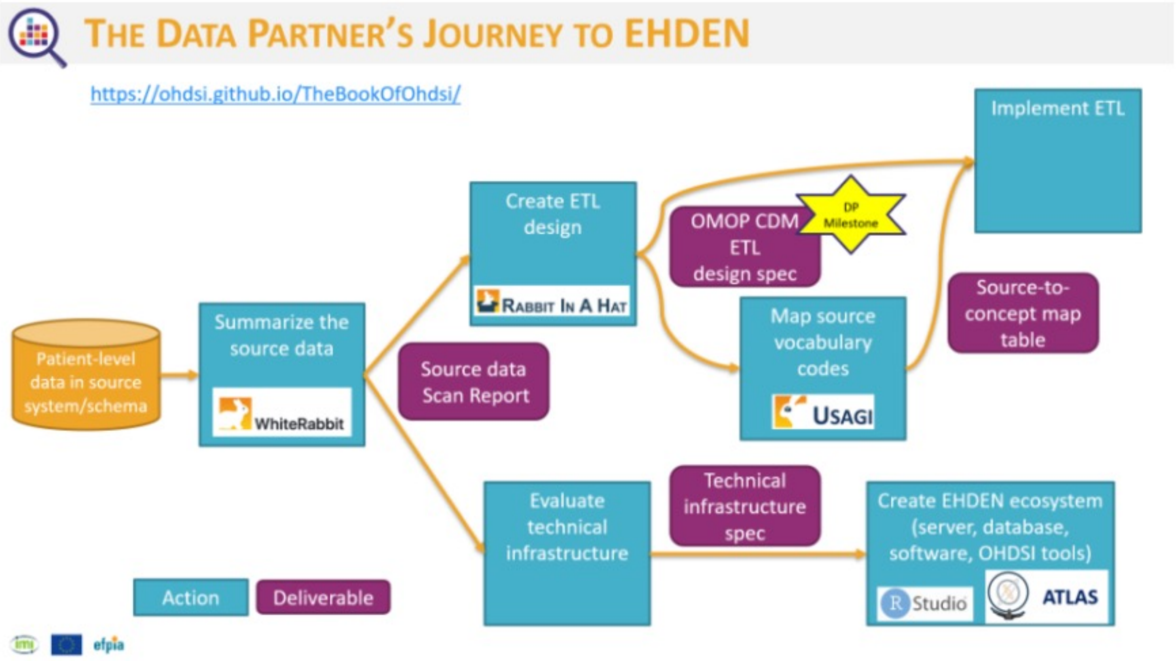
Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to **the EHDEN Consortium** on welcoming 21 new SMEs to support mapping to the OMOP Common Data Model, and perform services in the ecosystem of the EHDEN federated data network.

EHDEN now has a total of 47 SMEs across 19 European nations to assist in real world evidence generation within the community.



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
Wednesday	10 am	FHIR and OMOP - Digital Quality Measures Subgroup (ZOOM)
Thursday	8 am	Psychiatry
Thursday	1 pm	OMOP CDM Oncology – CDM/Vocabulary Subgroup
Friday	10 am	Electronic Health Record
Friday	10:30 am	Clinical Trials
Monday	10 am	GIS-Geographic Information System
Tuesday	9 am	OMOP CDM Oncology – Genomic Subgroup

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

Get Access To Different Teams/WGs/Chapters

5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- ☐ ATLAS
- ☐ Clinical Trials
- ☐ Common Data Model
- ☐ Data Quality Dashboard Development
- ☐ Early-stage Researchers
- ☐ Education Work Group
- ☐ Electronic Health Record (EHR) ETL
- ☐ Geographic Information System (GIS)
- ☐ HADES Health Analytics Data-to-Evidence Suite
- ☐ Health Equity
- ☐ Latin America
- ☐ Medical Devices
- ☐ Natural Language Processing
- ☐ OHDSI APAC
- ☐ OHDSI APAC Steering Committee
- ☐ OHDSI Steering Committee
- ☐ Oncology
- ☐ Patient-Generated Health Data
- ☐ Pharmacovigiliance Evidence Investigation
- ☐ Phenotype Development and Evaluation
- ☐ Population-Level Effect Estimation / Patient-Level Prediction
- ☐ Psychiatry
- ☐ Registry (formerly UK Biobank)
- ☐ Surgery and Perioperative Medicine
- ☐ Vaccine Safety
- ☐ Vaccine Vocabulary
- ☐ Women of OHDSI

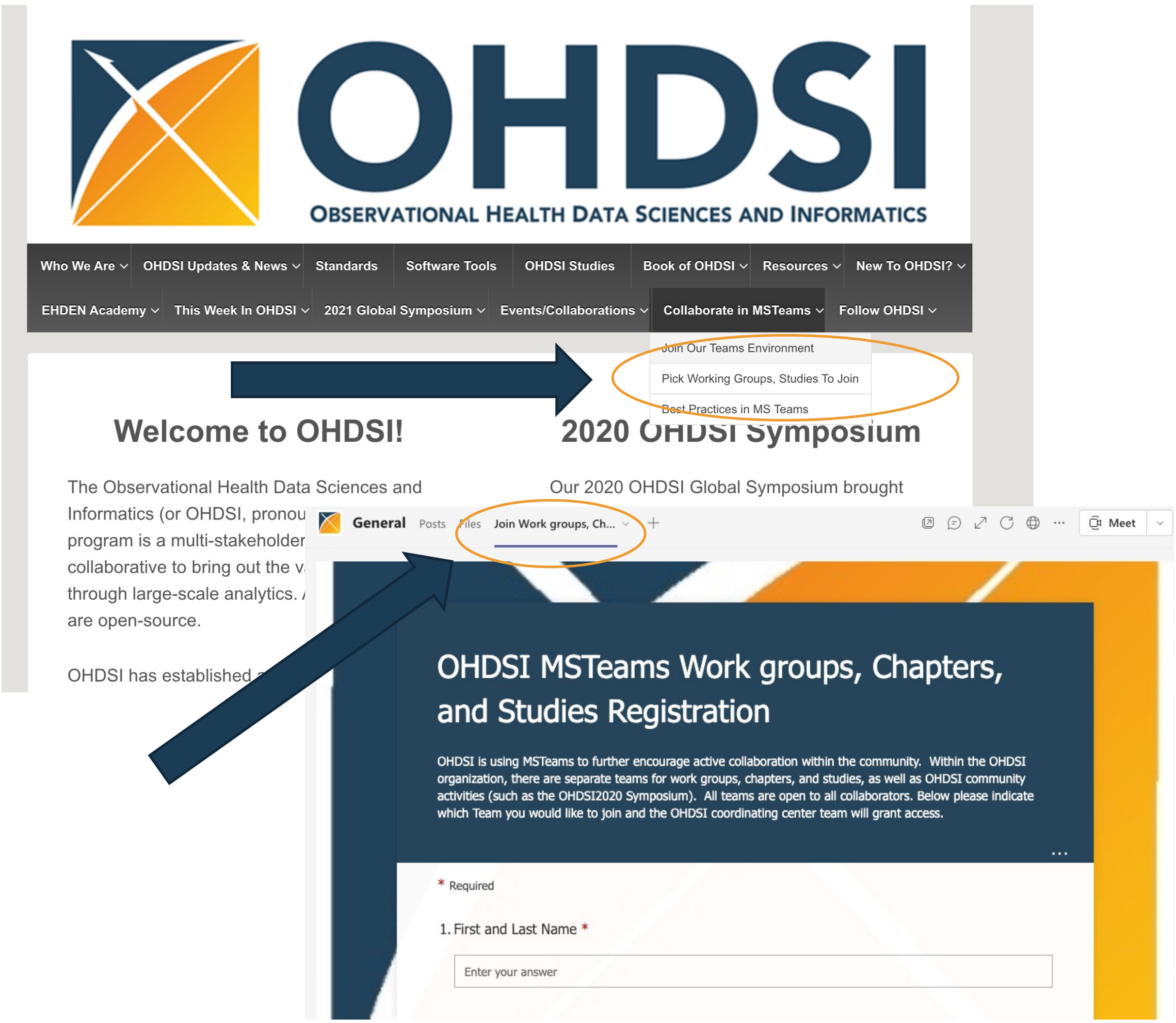
6. Select the chapter(s) you want to join

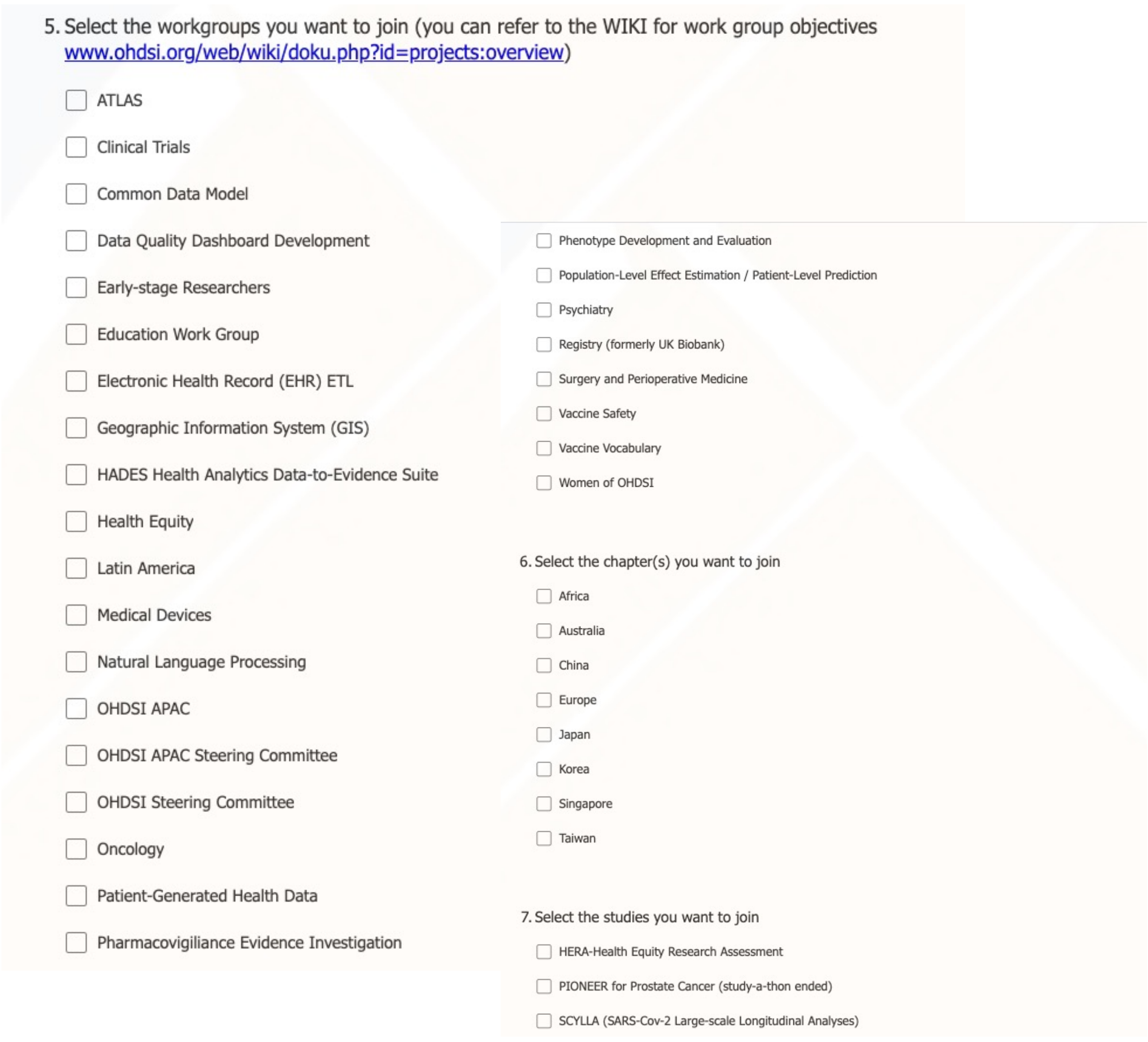
- ☐ Africa
- ☐ Australia
- ☐ China
- ☐ Europe
- ☐ Japan
- ☐ Korea
- ☐ Singapore
- ☐ Taiwan

7. Select the studies you want to join

- ☐ HERA-Health Equity Research Assessment
- ☐ PIONEER for Prostate Cancer (study-a-thon ended)
- ☐ SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses)

Get Access To Different Teams/WGs/Chapters





5. Select the workgroups you want to join (you can refer to the WIKI for work group objectives www.ohdsi.org/web/wiki/doku.php?id=projects:overview)

- ☐ ATLAS
- ☐ Clinical Trials
- ☐ Common Data Model
- ☐ Data Quality Dashboard Development
- ☐ Early-stage Researchers
- ☐ Education Work Group
- ☐ Electronic Health Record (EHR) ETL
- ☐ Geographic Information System (GIS)
- ☐ HADES Health Analytics Data-to-Evidence Suite
- ☐ Health Equity
- ☐ Latin America
- ☐ Medical Devices
- ☐ Natural Language Processing
- ☐ OHDSI APAC
- ☐ OHDSI APAC Steering Committee
- ☐ OHDSI Steering Committee
- ☐ Oncology
- ☐ Patient-Generated Health Data
- ☐ Pharmacovigiliance Evidence Investigation
- ☐ Phenotype Development and Evaluation
- ☐ Population-Level Effect Estimation / Patient-Level Prediction
- ☐ Psychiatry
- ☐ Registry (formerly UK Biobank)
- ☐ Surgery and Perioperative Medicine
- ☐ Vaccine Safety
- ☐ Vaccine Vocabulary
- ☐ Women of OHDSI

6. Select the chapter(s) you want to join

- ☐ Africa
- ☐ Australia
- ☐ China
- ☐ Europe
- ☐ Japan
- ☐ Korea
- ☐ Singapore
- ☐ Taiwan

7. Select the studies you want to join

- ☐ HERA-Health Equity Research Assessment
- ☐ PIONEER for Prostate Cancer (study-a-thon ended)
- ☐ SCYLLA (SARS-Cov-2 Large-scale Longitudinal Analyses)

2021 APAC Symposium • Nov. 18

Nov. 18 (APAC Time Zone)	Time (Korea time)	Contents	Speaker(s)
Morning	9:00 – 9:25 am	OHDSI State of the Community	George Hripcsak/Patrick Ryan
	9:25 – 9:50 am	OHDSI APAC State of the Community	Mui Van Zandt
	9:50 – 10:00 am	Energy Break	
	10:00 – 10:25 am	EHDEN	Peter Rijnbeek
	10:25 – 10:50 am	FHIR and OHDSI Collaboration	Christian Reich
	10:50 – 11:00 am	Energy Break	
	11:00 - 12:30 pm	APAC Chapter Visions for 2022	Chapter Leads
Lunch Break	12:30 – 13:00 pm		
Afternoon (in GatherTown)	13:00 – 14:00 pm	Workgroup Sessions (Medical Image, FHIR, CDM Tables)	
	14:00 – 15:00 pm	Collaboration Showcase	
	15:00 – 16:00 pm	APAC Study Sessions	

www.ohdsi.org/apac

#OHDSISocialShowcase This Week

Association Rule and Frequent Pattern Mining using the OMOP- CDM

PRESENTER: **Solomon Ioannou**

INTRODUCTION

To better understand the cooccurrence of data elements and their sequence, association rules analysis and frequent pattern analysis are powerful tools.

An **Association Rule** analysis answers the question "Given a cohort of patients, what are the most associated concepts that occur together?"

A **Frequent Pattern** analysis answers the question "What are the most common sequences of concepts observed in a cohort of patients"?

Potentially, they are also promising tools to improve other data mining tasks such as patient-level prediction.

We introduce here an open-source analytics framework, an R package, for performing Association Rule and Frequent Pattern mining using data in the OMOP-CDM.

METHODS

1. The AssociationRuleMining R package makes use of the open source SPMF Java library by Phillippe Fournier-Viger that implements a large collection of association rule and frequent pattern mining algorithms.
2. Using standard HADES packages the user can connect to a database, create the cohort/s of interest and extract relevant covariates.
3. Functionalities within the package allow efficient preparation of the input datasets and analysis using the algorithm of choice.

Workflow Description

1 **Create a cohort using one of OHDSI's tool of choice**

2 **Extract covariates using the FeatureExtraction package.**
- For Association Rule Mining, extracting the first occurrence of an event (diagnosis, drug subscription, etc) will suffice to perform the analysis.
- For Frequent Pattern Mining, the order of events matters, therefore extracting temporal covariates is essential.

3 **Choose an algorithm for the relevant analysis and set its parameters.**
- A required parameter to extract highly occurring itemsets or frequent patterns is minimum support, which acts as the threshold for the minimum number of patients that should have the concept set in their medical history, e.g., {obesity, diabetes}
- Algorithms that extract either association rules or frequent patterns require also to specify minimum confidence, which is the threshold for determining how often the left side of the rule occurs together with the right side, e.g., {obesity, diabetes} -> {heart failure}

4 **Prepare input datasets and run the analysis.**
- The package provides specific functionalities to prepare the input datasets to the necessary format and execute the algorithm.
- Based on the size of the cohort, an iterative procedure to select the optimal value for minimum support and minimum confidence may be applied.

5 **Viewing and exploring the results through interactive plots.**

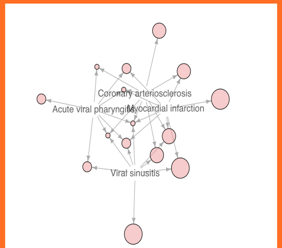


Figure 1: Highly associated concept sets.

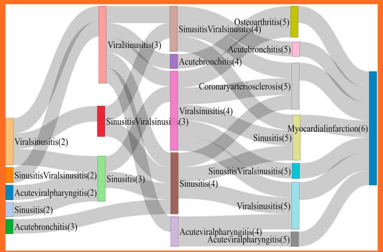


Figure 2: Frequent patterns indicating the chronological ordering of events.

RESULTS

1. Depending on the size of the cohort to be analyzed, the number of concepts included, and the values of predefined parameters of minimum support and minimum confidence, a huge number of rules or patterns can be revealed.
2. Currently results are presented in lists for further processing and use, such as, covariates in prediction problems.
3. Interactive visualizations are also implemented to explore the results graphically.

How can this tool be used?

1. We are exploring the possibilities of using these methods for characterization purposes.
2. Another research direction is the added predictive value, especially of frequent patterns, in clinical prediction problems.

Clinical relevance

1. Characterising frequent patterns and associations in health data can help to identify different types of patients that may need different types of treatment.
2. Frequent pattern analyses could help to generate new hypothesis for the pathogenesis of diseases.

The European Health Data & Evidence Network 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.

Solomon Ioannou,
Egill Fridgeirsson,
Jan Kors,
Peter Rijnbeek

Erasmus MC

EHDEN

OHDSI

MONDAY

Association Rule and Frequent Pattern Mining using the OMOP CDM

Authors: Solomon Ioannou, Egill Fridgeirsson, Jan Kors, Peter Rijnbeek

#OHDSISocialShowcase This Week

Lightning Talk!

TUESDAY

Extending the OMOP CDM to store the output of natural language processing pipelines

Authors: Monica Arrue (presenter), Sandra Pulido, Alvaro Abella, Gabriel Maeztu, Alberto Labarga

#OHDSISocialShowcase This Week

Title: CQL Scripting From Atlas Cohort Definitions

Michael Riley

INTRO:

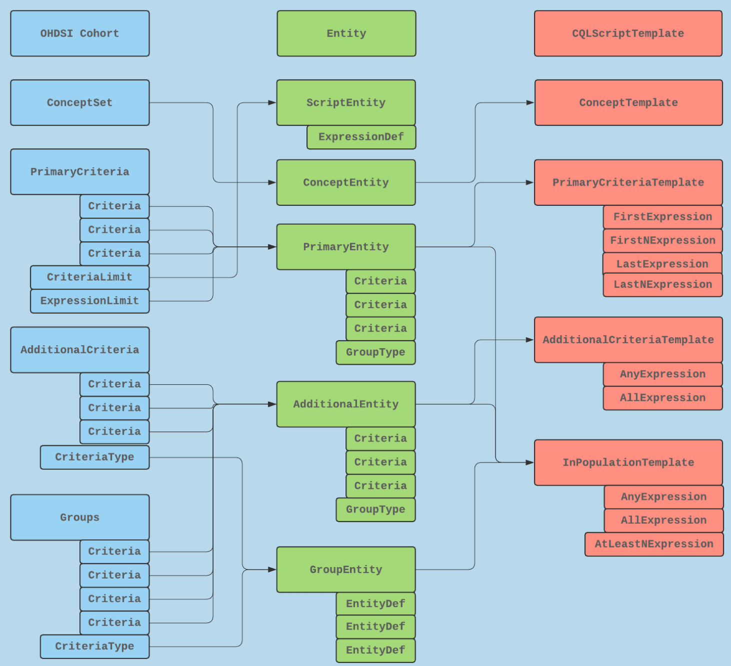
- **Who cares?** CQL is a complete language for defining logical inferences from medical datasets. Atlas Cohort Definitions are a quick structure-based definition to quickly and easily define a population set. By translating Atlas Cohort Definitions to CQL, we can use the various measure libraries available in CQL, as well as expand cohort definitions into a form that invites comparison to other measures.

METHODS

We captured the cohort definitions from atlas-ohdsi as a JSON definition. From this definition we parsed to a set of interal entities describing the concepts, primary criteria, and additional criteria used. We then use a set of cql formatting templates to convert the criteria into CQL definitions and craft a final InPopulation definition given the criteria groupings from the original cohort.

RESULTS

Translation was tested on 11 different cohort definitions with a variety of criteria sets. Preliminary results comparing the accuracy of the CQL definition vs the original cohort are pending.



- ConceptSet With SystemURI Definitions were automatically expanded into ConceptEntities using VSAC Terminology Service
- ExpressionLimit(First/Last/FirstN/LastN) Used as a global definition applied to primaryEntity
- GroupEntity Defines Grouping of AdditionalCriteria while AdditionalEntity collects temporal and value based filtering on the entity
- InPopulation Subsumes final definition from Primary, Additional, and Group Entities
- Patient CQL Context used, Population CQL Context feature in development

Michael Riley (Michael.Riley@gtri.gatech.edu)

OHDSI

Georgia Tech Research Institute

WEDNESDAY

Automated Translation of Cohort Definitions from Atlas JSON to CQL-FHIR

Authors: Michael Riley, Jon Duke

#OHDSISocialShowcase This Week

Best of Intent, Worst of Both Worlds: Why Sequentially Combining Epidemiological Methods Does Not Improve Signal Detection in Vaccine Surveillance

PRESENTER: **Faaizah** Arshad

INTRO
- There is a clinical intuition that combining sensitive and specific methods will improve vaccine safety signal detection.
- Little is known on the comparative performance of methods with real-world data.

METHODS
1. We evaluated six vaccine exposures: H1N1pdm, seasonal flu (Fluvirin), seasonal flu (Fluzone), seasonal flu (All), zoster (Shingrix), HPV (Gardasil 9) across four databases (CCAE, IBM MDCR, IBM MDCD, Optum EHR).
2. All data partners used the Observational Medical Outcomes Partnership (OMOP) common data model (CDM).
3. We generated a set of negative control and imputed positive control outcomes.
4. We defined a time-at-risk of 1-28 days after vaccination.
5. We used R programming to compute and compare the one-sided *p* values and type I and II errors (with and without empirical calibration) of a highly sensitive method (historical comparator), a highly specific method (self-controlled case series), and a method that sequentially combines the two.

RESULTS
- Using a highly sensitive method followed by a highly specific method did not compensate for the individual flaws of each method alone.

Applying a sensitive method followed by a specific method **does not improve signal detection** for adverse events under vaccine surveillance.

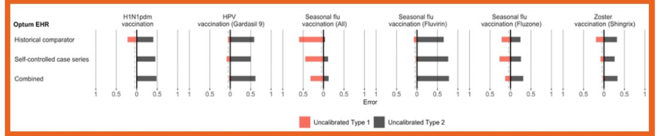


Figure 1. Type I and II errors (before empirical calibration) for all outcomes in Optum EHR. Historical comparator tends to be more sensitive, and SCCS tends to be more specific. Sequentially combining them increases specificity and decreases sensitivity.

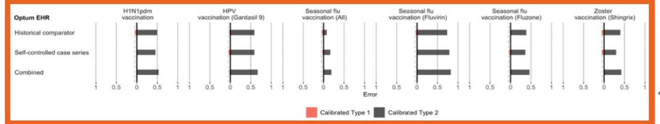


Figure 2. Type I and II errors (with empirical calibration) for all outcomes in Optum EHR. Type I errors return to nominal. Even with calibration, combining historical comparator and SCCS using the serial approach does not improve both sensitivity and specificity.

1. Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, U.S.A.
2. Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, U.S.A.
3. Observational Health Data Analytics, Janssen R&D, Titusville, NJ, U.S.A.

ADDITIONAL RESULTS

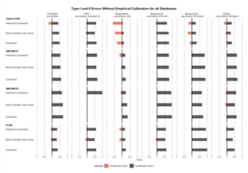


Figure 3. Type I and II errors without empirical calibration across databases. Historical comparator is not always more sensitive than SCCS, and SCCS is not always more specific than historical comparator.

Figure 4. Type I and II errors for all outcomes with empirical calibration across databases. Type I error returns to nominal.

DISCUSSION
- The use of real-world data mapped to the CDM allows for replicability and transparency.
- One limitation was the lack of COVID-19 vaccine exposures.

CONCLUSION
- Our findings oppose clinical advice to use a serial method in signal detection.

AUTHORS
Faaizah Arshad[1],
Martijn J. Schuemie[1,3], Marc A. Suchard[1,2]
on behalf of the EUMAEUS task force

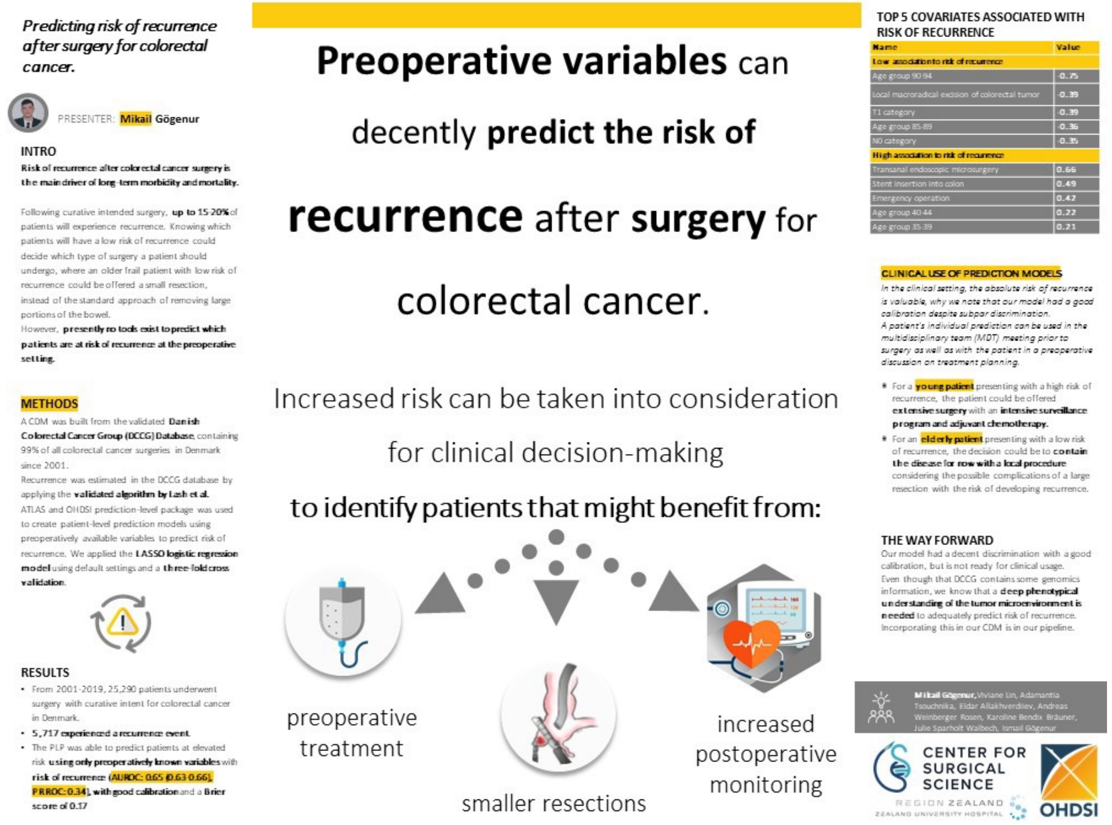
OHDSI

THURSDAY

Best of intent, worst of both worlds: why sequentially combining epidemiological methods does not improve signal detection in vaccine surveillance

Authors: Faaizah Arshad, Lana YH Lai, George Hripcsak, Daniel Prieto-Alhambra, Martijn J. Schuemie, Marc A. Suchard

#OHDSISocialShowcase This Week



Predicting risk of recurrence after surgery for colorectal cancer.

PRESENTER: Mikail Gögenur

INTRO

Risk of recurrence after colorectal cancer surgery is the main driver of long-term morbidity and mortality.

Following curative intended surgery, **up to 15-20%** of patients will experience recurrence. Knowing which patients will have a low risk of recurrence could decide which type of surgery a patient should undergo, where an older frail patient with low risk of recurrence could be offered a small resection, instead of the standard approach of removing large portions of the bowel.
However, **presently no tools exist to predict which patients are at risk of recurrence at the preoperative setting.**

METHODS

A CDM was built from the validated **Danish Colorectal Cancer Group (DCCG) Database**, containing 99% of all colorectal cancer surgeries in Denmark since 2001.
Recurrence was estimated in the DCCG database by applying the **validated algorithm by Lash et al.**
ATLAS and OHDSI prediction-level package was used to create patient-level prediction models using preoperatively available variables to predict risk of recurrence. We applied the **LASSO logistic regression model** using default settings and a **three-fold cross validation.**

RESULTS

- From 2001-2019, 25,290 patients underwent surgery with curative intent for colorectal cancer in Denmark.
- **5,717 experienced a recurrence event.**
- The PLP was able to predict patients at elevated risk **using only preoperatively known variables** with **risk of recurrence (AUROC: 0.65 (0.63-0.66), PRROC: 0.34), with good calibration** and a **Brier score of 0.17**

Preoperative variables can decently **predict the risk of recurrence** after **surgery** for colorectal cancer.

Increased risk can be taken into consideration for clinical decision-making to identify patients that might benefit from:

preoperative treatment

smaller resections

increased postoperative monitoring

TOP 5 COVARIATES ASSOCIATED WITH RISK OF RECURRENCE

Name	Value
Low association to risk of recurrence	
Age group 90-94	-0.75
Local macroradical excision of colorectal tumor	-0.39
T1 category	-0.39
Age group 85-89	-0.36
N0 category	-0.35
High association to risk of recurrence	
Transanal endoscopic microsurgery	0.66
Stent insertion into colon	0.49
Emergency operation	0.42
Age group 40-44	0.22
Age group 35-39	0.21

CLINICAL USE OF PREDICTION MODELS

In the clinical setting, the absolute risk of recurrence is valuable, why we note that our model had a good calibration despite subpar discrimination.
A patient's individual prediction can be used in the multidisciplinary team (MDT) meeting prior to surgery as well as with the patient in a preoperative discussion on treatment planning.

- For a **young patient** presenting with a high risk of recurrence, the patient could be offered **extensive surgery** with an **intensive surveillance program and adjuvant chemotherapy.**
- For an **elderly patient** presenting with a low risk of recurrence, the decision could be to **contain the disease for now with a local procedure** considering the possible complications of a large resection with the risk of developing recurrence.

THE WAY FORWARD

Our model had a decent discrimination with a good calibration, but is not ready for clinical usage.
Even though that DCCG contains some genomics information, we know that a **deep phenotypical understanding of the tumor microenvironment is needed** to adequately predict risk of recurrence.
Incorporating this in our CDM is in our pipeline.

Mikail Gögenur, Viviane Lin, Adamantia Tsouchnika, Eldar Allakhverdiev, Andreas Weinberger Rosen, Karoline Bendix Bräuner, Julie Sparholt Walbech, Ismail Gögenur

CENTER FOR SURGICAL SCIENCE
REGION ZEALAND
ZEALAND UNIVERSITY HOSPITAL
OHDSI

FRIDAY

Predicting risk of recurrence after surgery for colorectal cancer

Authors: Mikail Gögenur, Viviane Lin, Adamantia Tsouchnika, Eldar Allakhverdiiev, Andreas Weinberger Rosen, Karoline Bendix Bräuner, Julie Sparholt Walbech, Ismail Gögenur

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?

Oct. 26 Community Call: Trick or Treat

On Tuesday, Oct. 26 (11 am ET), Patrick Ryan will lead a Halloween-themed interactive demonstration of how you can use the OHDSI tools to quickly generate insights from your OMOP CDM.

We hope you'll learn a TRICK or two, and that it will be a TREAT.