Welcome To OHDSI in 2022!

OHDSI Community Call
Jan. 11, 2022 • 11 am ET

Future OHDSI Community Calls

Date	Topic
Jan. 11	Welcome To OHDSI in 2022
Jan. 18	Gathertown: Workgroup Meetings and Information Sessions
Jan. 25	Implementation of the N3C NLP Pipeline
Feb. 1	Introduction to Phenotype February
Feb. 8	Phenotype February Report #1, Workgroup Updates
Feb. 15	Phenotype February Report #2, Workgroup Updates
Feb. 22	Phenotype February Report #3, Workgroup Updates

Future OHDSI Community Calls

Date	Topic
Jan. 11	Welcome To OHDSI in 2022
Jan. 18	Gathertown: Workgroup Meetings and Information Sessions
Jan. 25	Extracting COVID Features Using NLP
Feb. 1	Introduction to Phenotype February
Feb. 8	Phenotype February Report #1, Workgroup Updates
Feb. 15	Phenotype February Report #2, Workgroup Updates
Feb. 22	Phenotype February Report #3, Workgroup Updates

Three Stages of The Journey

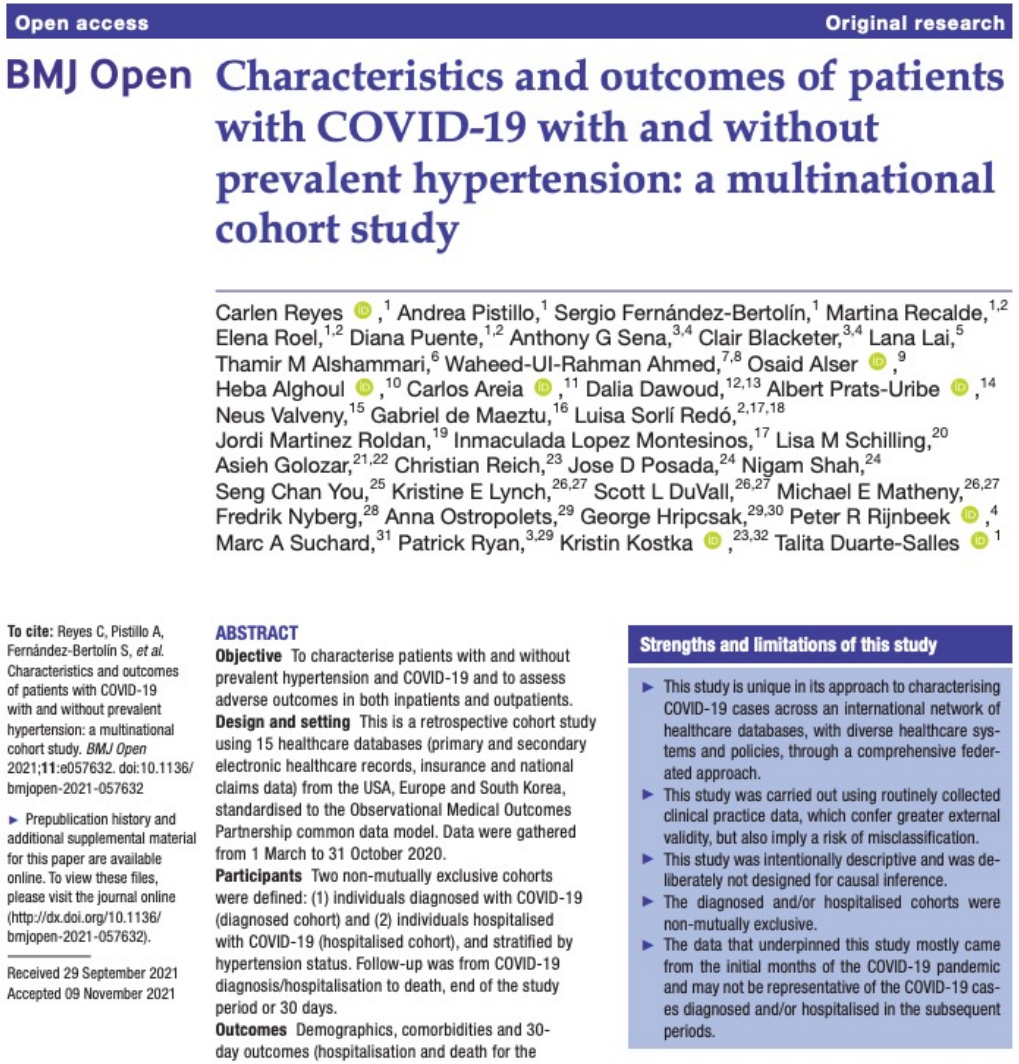
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to co-authors **Carlen Reyes, Andrea Pistillo, Sergio Fernández-Bertolín, Martina Recalde, Elena Roel, Diana Puente, Anthony Sena, Clair Blacketer, Lana Lai, Thamir Alshammari, Waheed-UI-Rahman Ahmed, Osaid Alser, Heba Alghoul, Carlos Areia, Dalia Dawoud, Albert Prats-Uribe, Neus Valveny, Gabriel de Maeztu, Luisa Sorlí Redó, Jordi Martinez Roldan, Inmaculada Lopez Montesinos, Lisa M Schilling, Asieh Golozar, Christian Reich, Jose Posada, Nigam Shah, Seng Chan You, Kristine Lynch, Scott DuVall, Michael Matheny, Fredrik Nyberg, Anna Ostropolets, George Hripcsak, Peter Rijnbeek, Marc Suchard, Patrick Ryan, Kristin Kostka, and Talita Duarte-Salles** on the study "**Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study**" which was published in BMJ Open on Dec. 22.



Open access Original research

BMJ Open

Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study

Carlen Reyes,[1] Andrea Pistillo,[1] Sergio Fernández-Bertolín,[1] Martina Recalde,[1,2] Elena Roel,[1,2] Diana Puente,[1,2] Anthony G Sena,[3,4] Clair Blacketer,[3,4] Lana Lai,[5] Thamir M Alshammari,[6] Waheed-Ul-Rahman Ahmed,[7,8] Osaid Alser,[9] Heba Alghoul,[10] Carlos Areia,[11] Dalia Dawoud,[12,13] Albert Prats-Uribe,[14] Neus Valveny,[15] Gabriel de Maeztu,[16] Luisa Sorlí Redó,[2,17,18] Jordi Martinez Roldan,[19] Inmaculada Lopez Montesinos,[17] Lisa M Schilling,[20] Asieh Golozar,[21,22] Christian Reich,[23] Jose D Posada,[24] Nigam Shah,[24] Seng Chan You,[25] Kristine E Lynch,[26,27] Scott L DuVall,[26,27] Michael E Matheny,[26,27] Fredrik Nyberg,[28] Anna Ostropolets,[29] George Hripcsak,[29,30] Peter R Rijnbeek,[4] Marc A Suchard,[31] Patrick Ryan,[3,29] Kristin Kostka,[23,32] Talita Duarte-Salles[1]

To cite: Reyes C, Pistillo A, Fernández-Bertolín S, *et al*. Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study. *BMJ Open* 2021;**11**:e057632. doi:10.1136/bmjopen-2021-057632

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057632).

Received 29 September 2021
Accepted 09 November 2021

ABSTRACT

Objective To characterise patients with and without prevalent hypertension and COVID-19 and to assess adverse outcomes in both inpatients and outpatients.

Design and setting This is a retrospective cohort study using 15 healthcare databases (primary and secondary electronic healthcare records, insurance and national claims data) from the USA, Europe and South Korea, standardised to the Observational Medical Outcomes Partnership common data model. Data were gathered from 1 March to 31 October 2020.

Participants Two non-mutually exclusive cohorts were defined: (1) individuals diagnosed with COVID-19 (diagnosed cohort) and (2) individuals hospitalised with COVID-19 (hospitalised cohort), and stratified by hypertension status. Follow-up was from COVID-19 diagnosis/hospitalisation to death, end of the study period or 30 days.

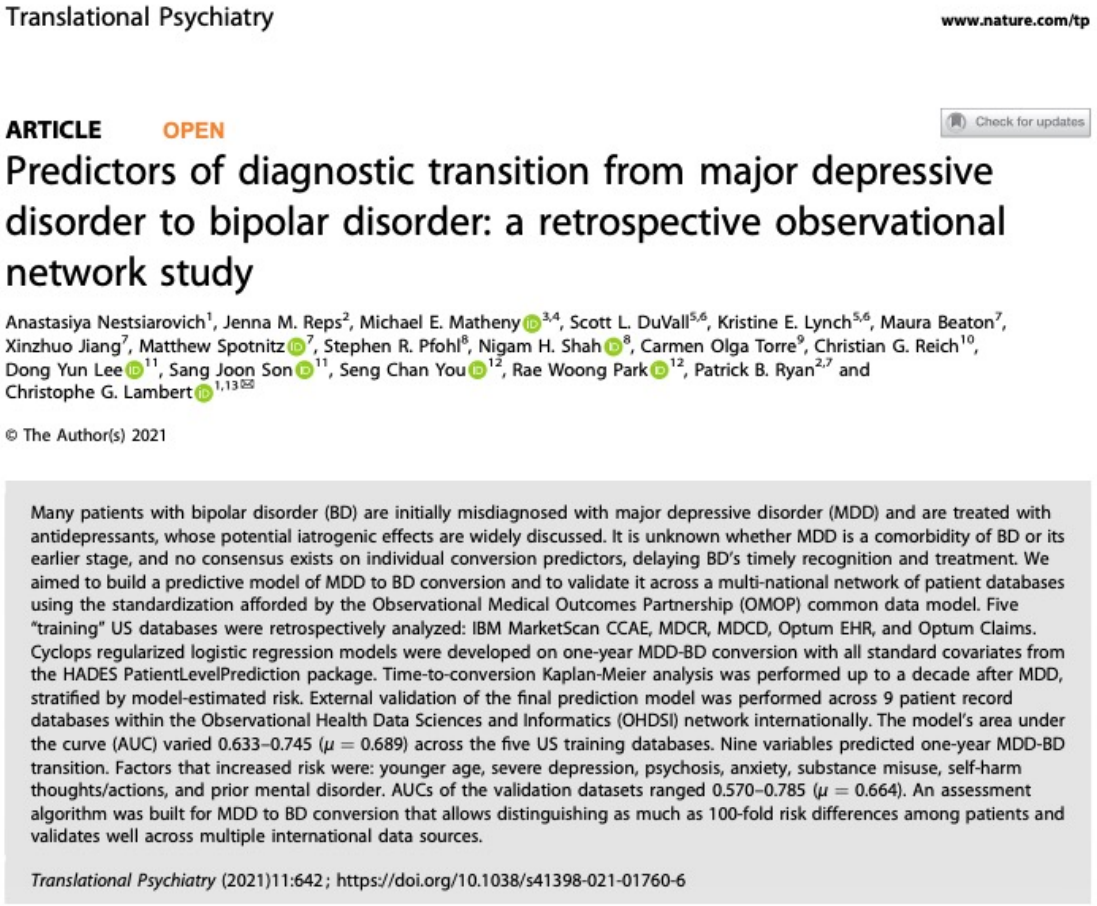
Outcomes Demographics, comorbidities and 30-day outcomes (hospitalisation and death for the

Strengths and limitations of this study

- This study is unique in its approach to characterising COVID-19 cases across an international network of healthcare databases, with diverse healthcare systems and policies, through a comprehensive federated approach.
- This study was carried out using routinely collected clinical practice data, which confer greater external validity, but also imply a risk of misclassification.
- This study was intentionally descriptive and was deliberately not designed for causal inference.
- The diagnosed and/or hospitalised cohorts were non-mutually exclusive.
- The data that underpinned this study mostly came from the initial months of the COVID-19 pandemic and may not be representative of the COVID-19 cases diagnosed and/or hospitalised in the subsequent periods.

OHDSI Shoutouts!

Congratulations to co-authors **Anastasiya Nestsiarovich, Jenna Reps, Michael Matheny, Scott DuVall, Kristine Lynch, Maura Beaton, Xinzhuo Jiang, Matthew Spotnitz, Stephen Pfohl, Nigam Shah, Carmen Olga Torre, Christian Reich, Dong Yun Lee, Sang Joon Son, Seng Chan You, Rae Woong Park, Patrick Ryan & Christophe Lambert** on the study "**Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study**" which was published in Translational Psychiatry on Dec. 20.



Translational Psychiatry www.nature.com/tp

ARTICLE OPEN

Check for updates

Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study

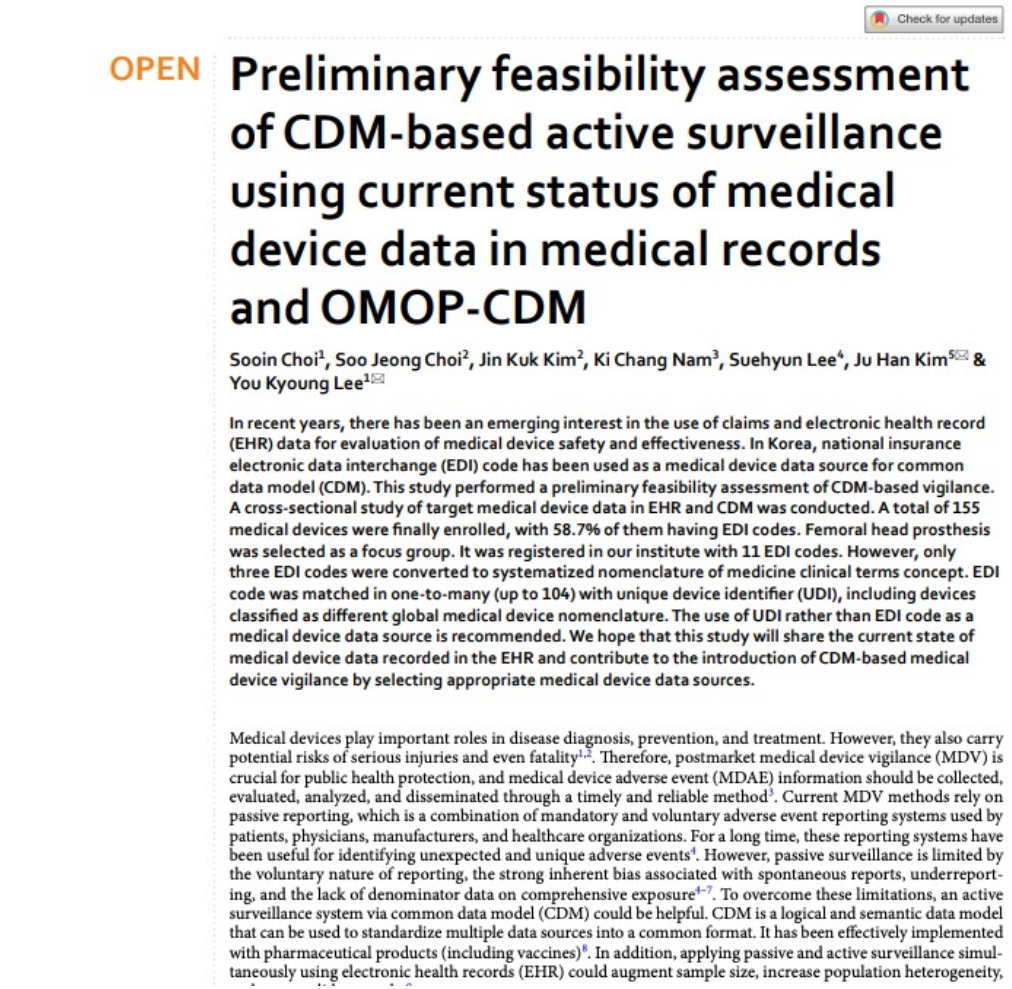
Anastasiya Nestsiarovich[1], Jenna M. Reps[2], Michael E. Matheny[3,4], Scott L. DuVall[5,6], Kristine E. Lynch[5,6], Maura Beaton[7], Xinzhuo Jiang[7], Matthew Spotnitz[7], Stephen R. Pfohl[8], Nigam H. Shah[8], Carmen Olga Torre[9], Christian G. Reich[10], Dong Yun Lee[11], Sang Joon Son[11], Seng Chan You[12], Rae Woong Park[12], Patrick B. Ryan[2,7] and Christophe G. Lambert[1,13]

Many patients with bipolar disorder (BD) are initially misdiagnosed with major depressive disorder (MDD) and are treated with antidepressants, whose potential iatrogenic effects are widely discussed. It is unknown whether MDD is a comorbidity of BD or its earlier stage, and no consensus exists on individual conversion predictors, delaying BD's timely recognition and treatment. We aimed to build a predictive model of MDD to BD conversion and to validate it across a multi-national network of patient databases using the standardization afforded by the Observational Medical Outcomes Partnership (OMOP) common data model. Five "training" US databases were retrospectively analyzed: IBM MarketScan CCAE, MDCR, MDCD, Optum EHR, and Optum Claims. Cyclops regularized logistic regression models were developed on one-year MDD-BD conversion with all standard covariates from the HADES PatientLevelPrediction package. Time-to-conversion Kaplan-Meier analysis was performed up to a decade after MDD, stratified by model-estimated risk. External validation of the final prediction model was performed across 9 patient record databases within the Observational Health Data Sciences and Informatics (OHDSI) network internationally. The model's area under the curve (AUC) varied 0.633–0.745 ($\mu = 0.689$) across the five US training databases. Nine variables predicted one-year MDD-BD transition. Factors that increased risk were: younger age, severe depression, psychosis, anxiety, substance misuse, self-harm thoughts/actions, and prior mental disorder. AUCs of the validation datasets ranged 0.570–0.785 ($\mu = 0.664$). An assessment algorithm was built for MDD to BD conversion that allows distinguishing as much as 100-fold risk differences among patients and validates well across multiple international data sources.

Translational Psychiatry (2021)11:642 ; https://doi.org/10.1038/s41398-021-01760-6

OHDSI Shoutouts!

Congratulations to co-authors **Sooin Choi, Soo Jeong Choi, Jin Kuk Kim, Ki Chang Nam, Suehyun Lee, Ju Han Kim and You Kyoung Lee** on the study **"Preliminary feasibility assessment of CDM-based active surveillance using current status of medical device data in medical records and OMOP-CDM"** which was published in Scientific Reports on Dec. 15.



scientific reports

Check for updates

OPEN

Preliminary feasibility assessment of CDM-based active surveillance using current status of medical device data in medical records and OMOP-CDM

Sooin Choi[1], Soo Jeong Choi[2], Jin Kuk Kim[2], Ki Chang Nam[3], Suehyun Lee[4], Ju Han Kim[5]✉ & You Kyoung Lee[1]✉

In recent years, there has been an emerging interest in the use of claims and electronic health record (EHR) data for evaluation of medical device safety and effectiveness. In Korea, national insurance electronic data interchange (EDI) code has been used as a medical device data source for common data model (CDM). This study performed a preliminary feasibility assessment of CDM-based vigilance. A cross-sectional study of target medical device data in EHR and CDM was conducted. A total of 155 medical devices were finally enrolled, with 58.7% of them having EDI codes. Femoral head prosthesis was selected as a focus group. It was registered in our institute with 11 EDI codes. However, only three EDI codes were converted to systematized nomenclature of medicine clinical terms concept. EDI code was matched in one-to-many (up to 104) with unique device identifier (UDI), including devices classified as different global medical device nomenclature. The use of UDI rather than EDI code as a medical device data source is recommended. We hope that this study will share the current state of medical device data recorded in the EHR and contribute to the introduction of CDM-based medical device vigilance by selecting appropriate medical device data sources.

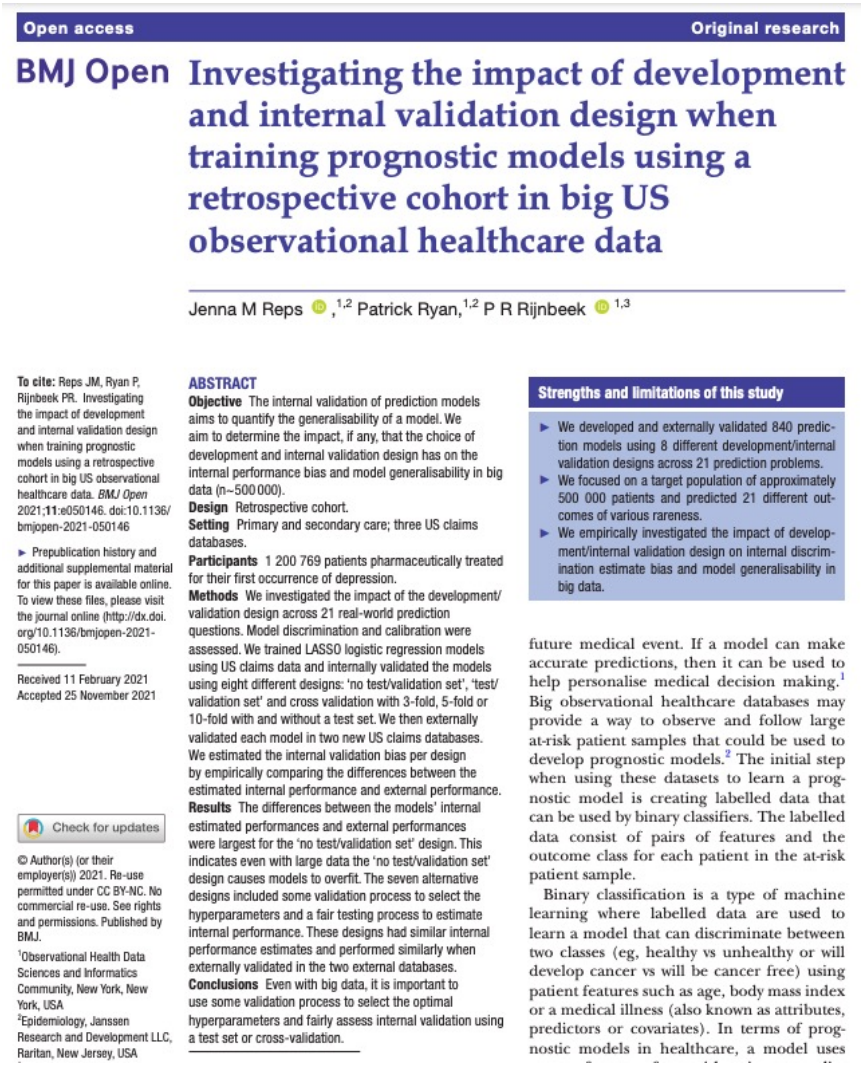
Medical devices play important roles in disease diagnosis, prevention, and treatment. However, they also carry potential risks of serious injuries and even fatality[1,2]. Therefore, postmarket medical device vigilance (MDV) is crucial for public health protection, and medical device adverse event (MDAE) information should be collected, evaluated, analyzed, and disseminated through a timely and reliable method[3]. Current MDV methods rely on passive reporting, which is a combination of mandatory and voluntary adverse event reporting systems used by patients, physicians, manufacturers, and healthcare organizations. For a long time, these reporting systems have been useful for identifying unexpected and unique adverse events[4]. However, passive surveillance is limited by the voluntary nature of reporting, the strong inherent bias associated with spontaneous reports, underreporting, and the lack of denominator data on comprehensive exposure[4–7]. To overcome these limitations, an active surveillance system via common data model (CDM) could be helpful. CDM is a logical and semantic data model that can be used to standardize multiple data sources into a common format. It has been effectively implemented with pharmaceutical products (including vaccines)[8]. In addition, applying passive and active surveillance simultaneously using electronic health records (EHR) could augment sample size, increase population heterogeneity,

OHDSI Shoutouts!

Congratulations to co-authors **Nicolas Paris, Antoine Lamer and Adrien Parrot** on the study **“Transformation and Evaluation of the MIMIC Database in the OMOP Common Data Model: Development and Usability Study”** which was published in JMIR Medical Informatics on Dec. 14.

OHDSI Shoutouts!

Congratulations to co-authors **Jenna Reps, Patrick Ryan, and Peter Rijnbeek** on the study **"Investigating the impact of development and internal validation design when training prognostic models using a retrospective cohort in big US observational healthcare data"** which was published in BMJ Open on Dec. 24.



Open access Original research

BMJ Open

Investigating the impact of development and internal validation design when training prognostic models using a retrospective cohort in big US observational healthcare data

Jenna M Reps [1,2] Patrick Ryan,[1,2] P R Rijnbeek [1,3]

To cite: Reps JM, Ryan P, Rijnbeek PR. Investigating the impact of development and internal validation design when training prognostic models using a retrospective cohort in big US observational healthcare data. *BMJ Open* 2021;**11**:e050146. doi:10.1136/bmjopen-2021-050146

► Prepublication history and additional supplemental material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-050146).

Received 11 February 2021
Accepted 25 November 2021

[1]Observational Health Data Sciences and Informatics Community, New York, New York, USA
[2]Epidemiology, Janssen Research and Development LLC, Raritan, New Jersey, USA

ABSTRACT

Objective The internal validation of prediction models aims to quantify the generalisability of a model. We aim to determine the impact, if any, that the choice of development and internal validation design has on the internal performance bias and model generalisability in big data (n~500 000).

Design Retrospective cohort.

Setting Primary and secondary care; three US claims databases.

Participants 1 200 769 patients pharmaceutically treated for their first occurrence of depression.

Methods We investigated the impact of the development/validation design across 21 real-world prediction questions. Model discrimination and calibration were assessed. We trained LASSO logistic regression models using US claims data and internally validated the models using eight different designs: 'no test/validation set', 'test/validation set' and cross validation with 3-fold, 5-fold or 10-fold with and without a test set. We then externally validated each model in two new US claims databases. We estimated the internal validation bias per design by empirically comparing the differences between the estimated internal performance and external performance.

Results The differences between the models' internal estimated performances and external performances were largest for the 'no test/validation set' design. This indicates even with large data the 'no test/validation set' design causes models to overfit. The seven alternative designs included some validation process to select the hyperparameters and a fair testing process to estimate internal performance. These designs had similar internal performance estimates and performed similarly when externally validated in the two external databases.

Conclusions Even with big data, it is important to use some validation process to select the optimal hyperparameters and fairly assess internal validation using a test set or cross-validation.

Strengths and limitations of this study

- We developed and externally validated 840 prediction models using 8 different development/internal validation designs across 21 prediction problems.
- We focused on a target population of approximately 500 000 patients and predicted 21 different outcomes of various rareness.
- We empirically investigated the impact of development/internal validation design on internal discrimination estimate bias and model generalisability in big data.

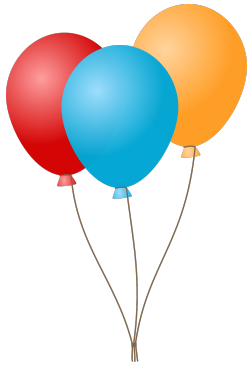
future medical event. If a model can make accurate predictions, then it can be used to help personalise medical decision making.[1] Big observational healthcare databases may provide a way to observe and follow large at-risk patient samples that could be used to develop prognostic models.[2] The initial step when using these datasets to learn a prognostic model is creating labelled data that can be used by binary classifiers. The labelled data consist of pairs of features and the outcome class for each patient in the at-risk patient sample.

Binary classification is a type of machine learning where labelled data are used to learn a model that can discriminate between two classes (eg, healthy vs unhealthy or will develop cancer vs will be cancer free) using patient features such as age, body mass index or a medical illness (also known as attributes, predictors or covariates). In terms of prognostic models in healthcare, a model uses

OHDSI Shoutouts!

Congratulations to **Mui Van Zandt**, who was recently promoted to VP & GM Real World Data & Tech at IQVIA.

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	2 pm	Health Equity
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Medical Imaging
Friday	9 am	Education
Friday	10 am	Phenotype Development and Evaluation
Friday	11 pm	China Chapter
Monday	10 am	Healthcare Systems Interest Group (formerly EHR WG)
Monday	10 am	GIS-Geographic Information System
Monday	2 pm	FHIR and OMOP – Terminologies Subgroup (ZOOM)

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

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Tuesday	12 pm	Common Data Model – Vocabulary Subgroup
Tuesday	2 pm	Health Equity
Wednesday	2 pm	Natural Language Processing
Wednesday	7 pm	Medical Imaging
Friday	9 am	Education
Friday	10 am	Phenotype Development and Evaluation
Friday	11 pm	China Chapter
Monday	10 am	Healthcare Systems Interest Group (formerly EHR WG)
Monday	10 am	GIS-Geographic Information System
Monday	2 pm	FHIR and OMOP – Terminologies Subgroup (ZOOM)

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

Healthcare Systems Interest Group

formerly the Electronic Health Record Working Group

Draft objectives:

- To provide support for transforming source EHR data to the CDM
- Support healthcare systems with building the business case for the CDM
- Provide value to healthcare systems beyond participating in OHDSI network studies
- To help each other and learn from experience

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)

Register For 2022 OHDSI Europe Symposium

www.ohdsi-europe.org/symposium-2022

Latest Edition Of “The Journey” Newsletter

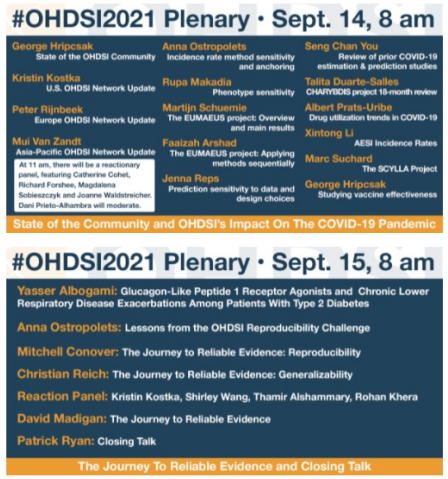
OHDSI

The Journey Newsletter (A Look Back On 2021)

Thank you to everybody who joined the journey with OHDSI in 2021 and helped our community take critical steps in our mission of empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. Join us as we reflect on 12 highlights from the last 12 months before we turn our full attention on the possibilities of 2022!

Our community is excited to get back to work in 2022. We will resume our weekly Tuesday community calls on January 11 (11 am ET), so look for a calendar invite and #JoinTheJourney!

#1: The 2021 Global Symposium



The #OHDSI2021 Global Symposium featured plenary presentations on both OHDSI’s Impact on the COVID-19 Pandemic, as well as on the Journey to Reliable Evidence. The main days included the State of the Community Presentation, the Collaborator Showcase, and a memorable Closing Ceremony that focused on OHDSI’s work through the perspective of a patient.

#2: Vaccine Surveillance Methods Research

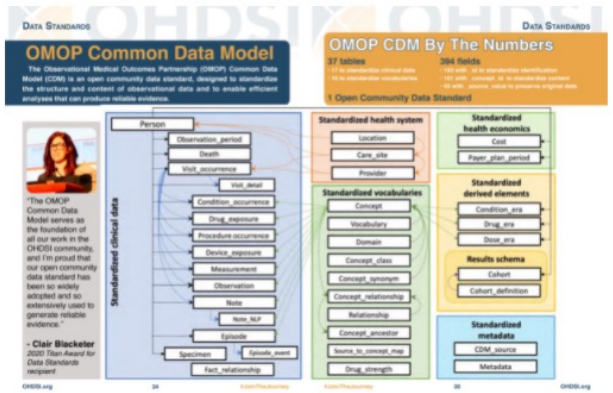
The OHDSI community collaborated on methods research on Vaccine Surveillance throughout 2021, especially within our EUMAEUS (Evaluating Use of Methods for Adverse Event Under Surveillance) Workgroup. The team provided a presentation on this work over the summer, and had its first published study shared in Frontiers of Pharmacology recently.

WATCH: EUMAEUS Presentation

Frontiers in Pharmacology Study

Press Release On Vaccine Surveillance Study

#5: CDM v5.4 Is Released

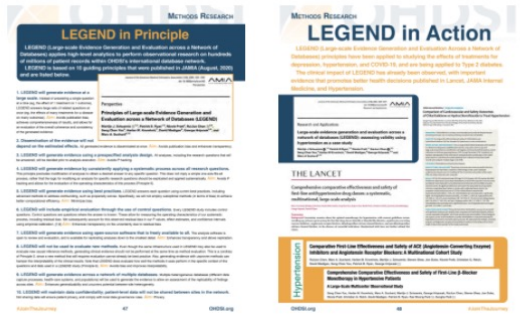


Thanks to the invaluable contributions of Clair Blacketer and the Common Data Model workgroup, the OHDSI community released v5.4 of the OMOP Common Data Model. The OMOP CDM is an open community data standard, designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence.

You can expect to hear much more about the latest version of the CDM in the new year, including workshops and other educational opportunities to learn how to best take advantage of OMOP in your research.

WATCH: CDM v5.4 Release/Brief Tutorial

#9: The LEGEND Initiative



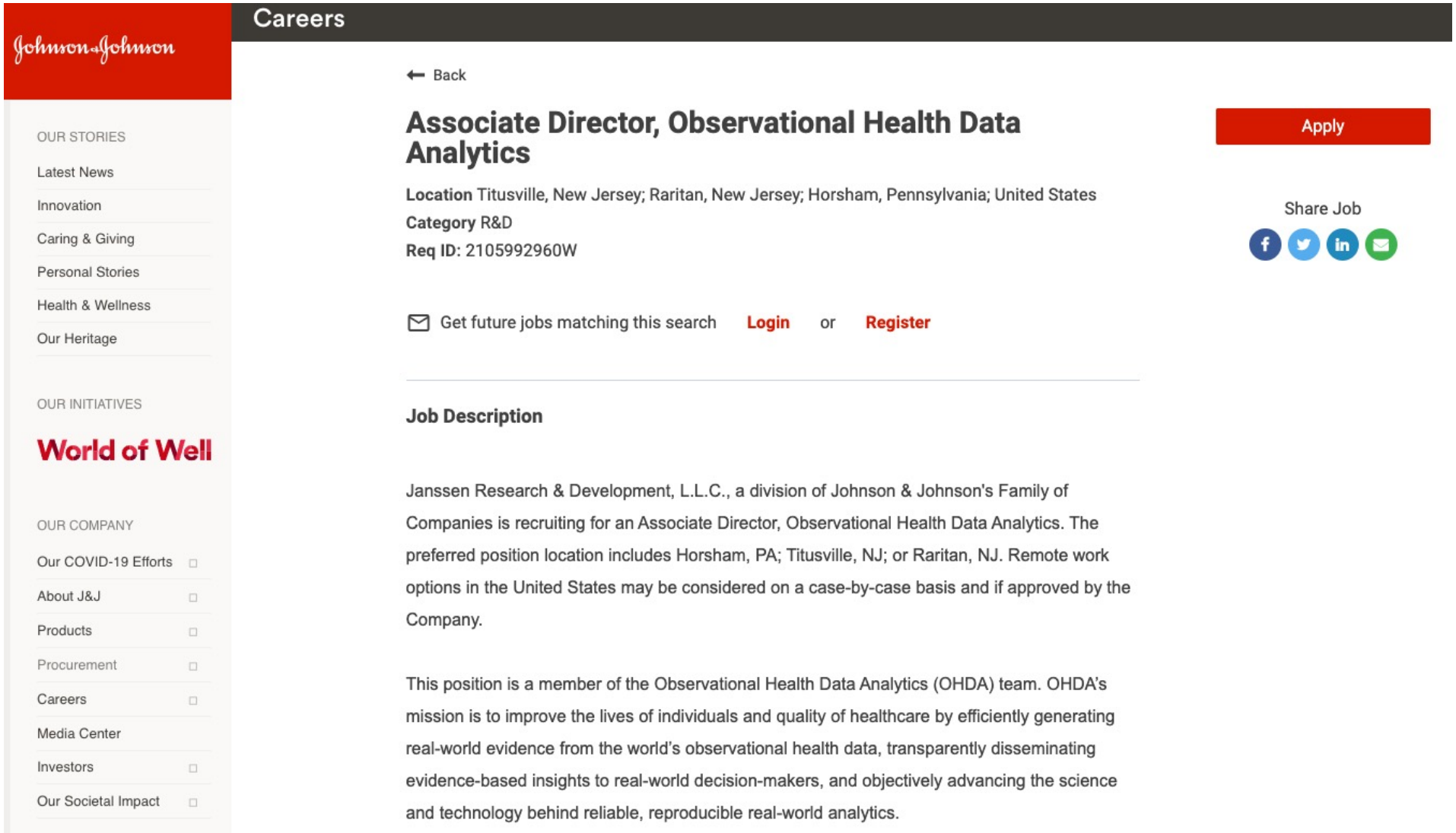
LEGEND (Large-scale Evidence Generation and Evaluation Across a Network of Databases) principles have been applied to studying the effects of treatments for depression, hypertension, and COVID-19, and are being applied to Type 2 diabetes. The clinical impact of LEGEND has already been observed, with important evidence that promotes better health decisions published in Lancet, JAMA Internal Medicine, and Hypertension.

The LEGEND team shared a presentation during an October community call that highlighted key aspects of the study, work that has been published, and announced the efforts around studying Type 2 diabetes.

The LEGEND Initiative

WATCH: The LEGEND Presentation

Openings!



Johnson & Johnson

Careers

OUR STORIES

Latest News
Innovation
Caring & Giving
Personal Stories
Health & Wellness
Our Heritage

OUR INITIATIVES

World of Well

OUR COMPANY

Our COVID-19 Efforts
About J&J
Products
Procurement
Careers
Media Center
Investors
Our Societal Impact

← Back

Associate Director, Observational Health Data Analytics

Apply

Location Titusville, New Jersey; Raritan, New Jersey; Horsham, Pennsylvania; United States
Category R&D
Req ID: 2105992960W

Share Job

Get future jobs matching this search Login or Register

Job Description

Janssen Research & Development, L.L.C., a division of Johnson & Johnson's Family of Companies is recruiting for an Associate Director, Observational Health Data Analytics. The preferred position location includes Horsham, PA; Titusville, NJ; or Raritan, NJ. Remote work options in the United States may be considered on a case-by-case basis and if approved by the Company.

This position is a member of the Observational Health Data Analytics (OHDA) team. OHDA's mission is to improve the lives of individuals and quality of healthcare by efficiently generating real-world evidence from the world's observational health data, transparently disseminating evidence-based insights to real-world decision-makers, and objectively advancing the science and technology behind reliable, reproducible real-world analytics.

Openings!

The Roux Institute
Northeastern University

Assistant Professor

Job no: **508405**
Work type: **Faculty Full-time (Tenure/Tenure Track)**
Location: **Boston Main Campus, Portland, ME Campus**
Categories: **Bouve College of Health Sciences**

Observational Health Data Sciences and Informatics Postdoctoral Fellow

Job no: 508708
Position type: Staff
Location: Portland, ME Campus
Division/Equivalent: DIV53 - Roux Institute
School/Unit: 111240 - Roux Institute-Research
Categories: Regional Campuses

About the Opportunity:

The Bouvé College of Health Sciences and The Roux Institute at Northeastern University seek candidates for **two tenure-track Assistant Professor positions** in the emerging area of health/healthcare data science. The successful candidate will have primary responsibility for working with the OHDSI Center at the Roux Institute (https://roux.northeastern.edu/ohdsi/), focusing on education, research and community support of the global Open Source OHDSI initiative (http://ohdsi.org).

Research areas of interest should encompass approaches for maximizing the value of health data for evidence generation through large-scale analytics and may include artificial intelligence (AI), machine learning (ML), computer and data sciences, digital health, life sciences, and medicine. Example: Methods that strengthen the ability to confidently draw causal inferences from comparative effectiveness research on observational healthcare data.

Other examples include real-world evidence data standardization, clinical/medical surveillance, comparative effectiveness research, personalized risk prediction and prevention, learning healthcare systems, big data, and applications of health or bio-informatics.

Aspiring candidates may be developing methods or applications that use computational modeling and large datasets to enhance our understanding of health from diagnosis, therapeutics, prevention, and health outcomes. We are also interested in efforts to understand and reduce health disparities among marginalized populations.

Our tenure and promotion process values collaborative research and teamwork. Hires will be mentored for success, with mentoring teams and group guidance. In addition, a strong and effective faculty development strategy is part of the Northeastern institutional mission. The ADVANCE Office of Faculty Development office works in conjunction with the Office of Research Development (ORD), the Office of Institutional Diversity and Inclusion (OIDI), the Center for Advancing Teaching and Learning Through Research (CATLR), and University Decision Support (UDS) to provide programs and trainings to further develop and support a thriving faculty.

At Northeastern University, we embrace a culture of respect, where each person is valued for their contribution and is treated fairly. We oppose all forms of racism. We support a culture that does not tolerate any form of discrimination and where each person may belong. We strive to have a diverse membership, one where each person is trained and mentored to promote their success.

For more information, please contact Brianne Olivieri-Mui, Assistant Professor, Department of Health Sciences: B.mui@northeastern.edu

#OHDSISocialShowcase This Week

Cohort Incidence
Robust Incidence Rates for the Common Data Model in R

PRESENTER: **Christopher Knoll**

INTRO:
- There currently isn't an open source, standardized package for calculating incidence rates.
- Atlas does contain an Incidence Rate function but is limited to a single Time-At-Risk and excludes prior outcomes.
- A new method of incidence rate and proportion is needed to allow multiple time-at-risk periods per person, allow prior outcomes and account for clean window to avoid immortal time bias.

METHODS

1. Users define target and outcome cohorts in advance.
2. Time-at-risk is defined by specifying the offsets from the cohort entry and exit.
3. Clean window (a timespan from the end of the outcome end date) is used to exclude time at risk from incidence rate calculation.
4. Subgroups can be used to isolate incidence rate calculations to specific sub-populations
5. **Incidence rate = cases / (time_at_risk – excluded time); Incidence Proportion = distinct subjects with cases / distinct subjects with at least 1 day at risk.**

RESULTS
- Source code is opensource and available at GitHub.
- A standardized object-model that can be persisted to JSON syntax.
- A standard results schema to persist results
- Test cases that verify integrity of methods

Take a picture to download materials

CohortIncidence is a new **OHDSI R package** and **standardized framework** for **characterizing Incidence Rates** at scale.

Targets	Outcomes	Time-at-Risk	Subgroup
Exposure 1 Exposure 2	Outcome 1 Outcome 2 Outcome 3	0d – 60d from start 0d – 90d from start All time post start	Age <= 30 Age 31 - 65 Age > 65

Cohort Incidence

Target	Outcome	TAR	Subgroup	Rate	Proportion
Exposure 1	Outcome 1	0d-60d	Age <= 30	103.24	24.7
Exposure 1	Outcome 1	0d-60d	Age 31 – 65	97.2	22.4
Exposure 1	Outcome 1	0d-60d	Age > 65	211.4	32.1
Exposure 1	Outcome 1	0d – 90d	Age <= 30	71.4	101.4
Exposure 1	Outcome 1	0d – 90d	Age 31 – 65	39.8	97.2
Exposure 1	Outcome 1	0d – 90d	Age > 65	101.7	132.7
Exposure 1	Outcome 1	All Time	Age <= 30	211.4	32.1
Exposure 1	Outcome 1	All Time	Age 31 – 65	71.4	101.4
Exposure 1	Outcome 1	All Time	Age > 65	39.8	97.2
Exposure 1	Outcome 1	0d-60d	Age <= 30	101.4	211.4
Exposure 1	Outcome 1	0d-60d	Age 31 – 65	97.2	71.4
...43 rows omitted...					
Exposure 2	Outcome 3	All Time	Age > 65	71.4	101.4

Example Use-case: A single person with multiple exposures and multiple outcomes:

Step 1: Derive time-at-risk episodes (from target cohort) and clean windows (from outcome cohort).

Step 2: Collapse time-at-risk and clean windows

Step 3: Exclude clean window from time-at-risk

Step 4: Calculate incidence rate (or proportion) from remaining time at risk and cases.

Christopher Knoll, Anthony Sena

janssen PHARMACEUTICAL COMPANIES OF JOHNSON & JOHNSON

OHDSI

MONDAY

CohortIncidence: A Standardized Framework for Characterizing Incidence Using OMOP Common Data Model and OHDSI tools

Authors: Christopher Knoll, Anthony Sena

#OHDSISocialShowcase This Week

Lightning Talk!

TUESDAY

Beyond Clinical: Integrating Research Assay Data into the Observational Health Data Sciences and Informatics Common Data Model (OHDSI CDM) through the Surgical Critical Care Initiative (SC2i)

Authors: Chandra Almond (presenter), William Patino, Ravi Sanka, Seth Schobel, Andrew MacKelfresh, Allan Kirk, Eric A. Elster

#OHDSISocialShowcase This Week

Comparing Data Quality Dashboard results from two ETL iterations: three new utilities.

Elena G. Lara* [(†)], Maxim Moinat [(†)]
[†] The Hyve, Arthur van Schendelstraat 650, 3511 MJ Utrecht, The Netherlands * Contact: elena@thehyve.nl / +31 (0)30 7009713

The Data Quality Dashboard (DQD) has been widely used to **evaluate the quality of an OMOP CDM** data set resulting from an ETL (extract, transform, load) process[1]. In practice, during the conversion to the OMOP CDM we perform several ETL iterations. However, interpreting the differences in quality is not always straightforward.

We developed three new utilities as part of mapping of the **UK Biobank** (UKB) data under the European Health Data Evidence Network (EHDEN) COVID19 rapid data partner call[2], and in collaboration with University College London.

1 Thresholds editing

As part of the ETL iterations, we needed to change the fail-thresholds of individual checks. We created a separate table to **list the changed thresholds in a user-friendly way**. Our utility script takes this new table to produce a customized thresholds file accepted by the default DQD scripts.

References:
[1] Blacketer C, Defalco F, Ryan P, Rijnbeek P. Increasing trust in real-world evidence through evaluation of observational data quality. medRxiv 2021.
[2] EHDEN COVID19 Rapid Data Partner Call, April 2020, retrieved May 2021. https://www.ehden.eu/open-calls/04-2020-covid19-data-partner-call/

2 Coverage per domain

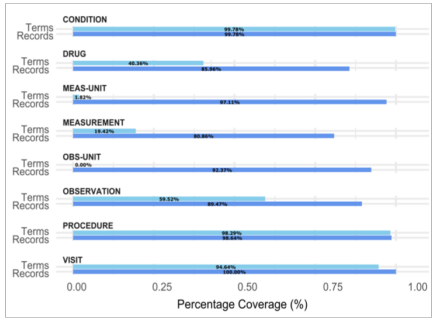


Figure 1: Barplot for the mapping coverage in an ETL. In light blue: the percentage of distinct terms mapped to a standard OMOP concept; in darker blue: the percentage of records mapped to a standard OMOP concept. ▲

An important part of the conversion quality is the concept mapping coverage. It is hard to get this overview from the DQD result tables alone. The new bar plot shows the **concept mapping coverage across all OMOP domains**. This ETL iteration achieved a high coverage throughout all domains and units (Figure 1) in terms of records mapped to standard concepts. The number of unique terms mapped was low for measurement and observation units (1.82% and 1.00%) and for measurement (19.42%).

3 Comparison of DQD results

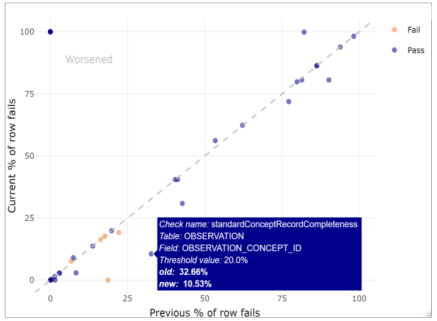


Figure 2: Each dot represents one check that has a different percentage of row fails between the iterations. This percentage of row fails is marked in the x-axis for the earliest run, and in the y-axis for the latest run. ▲

This visualization script selects the checks for which the percentage of records that satisfy said **check has changed between ETL iterations**. Here, the percentage of records satisfying the checks had modestly improved (Figure 2). As an example, there is an outlier (top left corner) that prompted us to investigate and update the ETL accordingly. On the other hand, we improved the standard record completeness in the observation table to be above 80%. Both visualizations are produced directly from DQD output.

EHDEN The Hyve OHDSI

WEDNESDAY

Comparing Data Quality Dashboard results from consecutive ETL iterations: three new utilities

Authors: Elena G. Lara, Maxim Moinat

#OHDSISocialShowcase This Week

Empirical Assessment of Alternative Methods for Identifying Seasonality in Observational Healthcare Data

PRESENTER: **Anthony Molinaro**

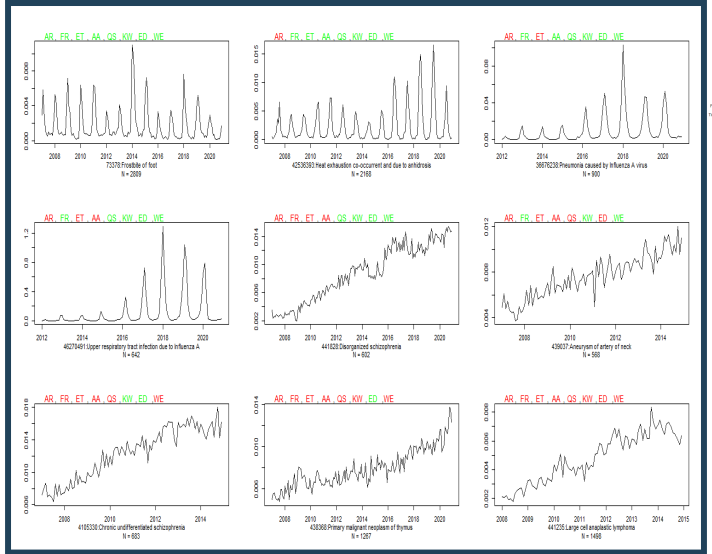
INTRO:

- Understanding the seasonality of a biological event can contribute to an improved understanding of its causes and help guide appropriate responses.
- Observational data contains timestamped diagnosis codes that can be used to create time series that serve as proxies for biological events.
- Automated methods must be employed to determine the seasonality of thousands of time series.

METHODS

- Leveraging the OHDSI packages CASTOR (https://github.com/OHDSI/Castor) and ACHILLES (https://github.com/OHDSI/Achilles), we create monthly proportion time series from condition concepts with at least four years of complete data. We compare the following eight popular methods of binary seasonality classification of time series at the .01, .05, and .1 levels of significance:
  i. Friedman's Test (FR)
  ii. Welch's Test (WE)
  iii. Kruskal-Wallis Test (KW)
  iv. Edwards' Test (ED)
  v. QS Test (QS)
  vi. ETS Test (ET)
  vii. Auto ARIMA Test (AA)
  viii. ARIMA Test (AR)

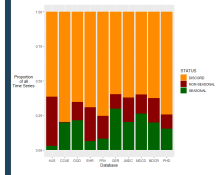
Methods for detecting seasonality in time series produce inconsistent results in observational healthcare data.



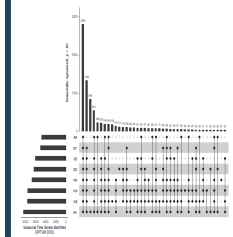
Results

- Using the 8 aforementioned statistical methods for determining seasonality, we evaluated 60,171 time series across 10 databases in the OMOP CDM format at three levels of significance. With more than 180,000 evaluations, we discovered substantial discord amongst the methods. This discord implies that these methods are not interchangeable and seasonality classification is highly dependent on the method chosen.

Concordance across all databases



UpsetR plot – concordance by method combination for OPTUM DOD, p < 0.05



Anthony Molinaro, Frank DeFalco

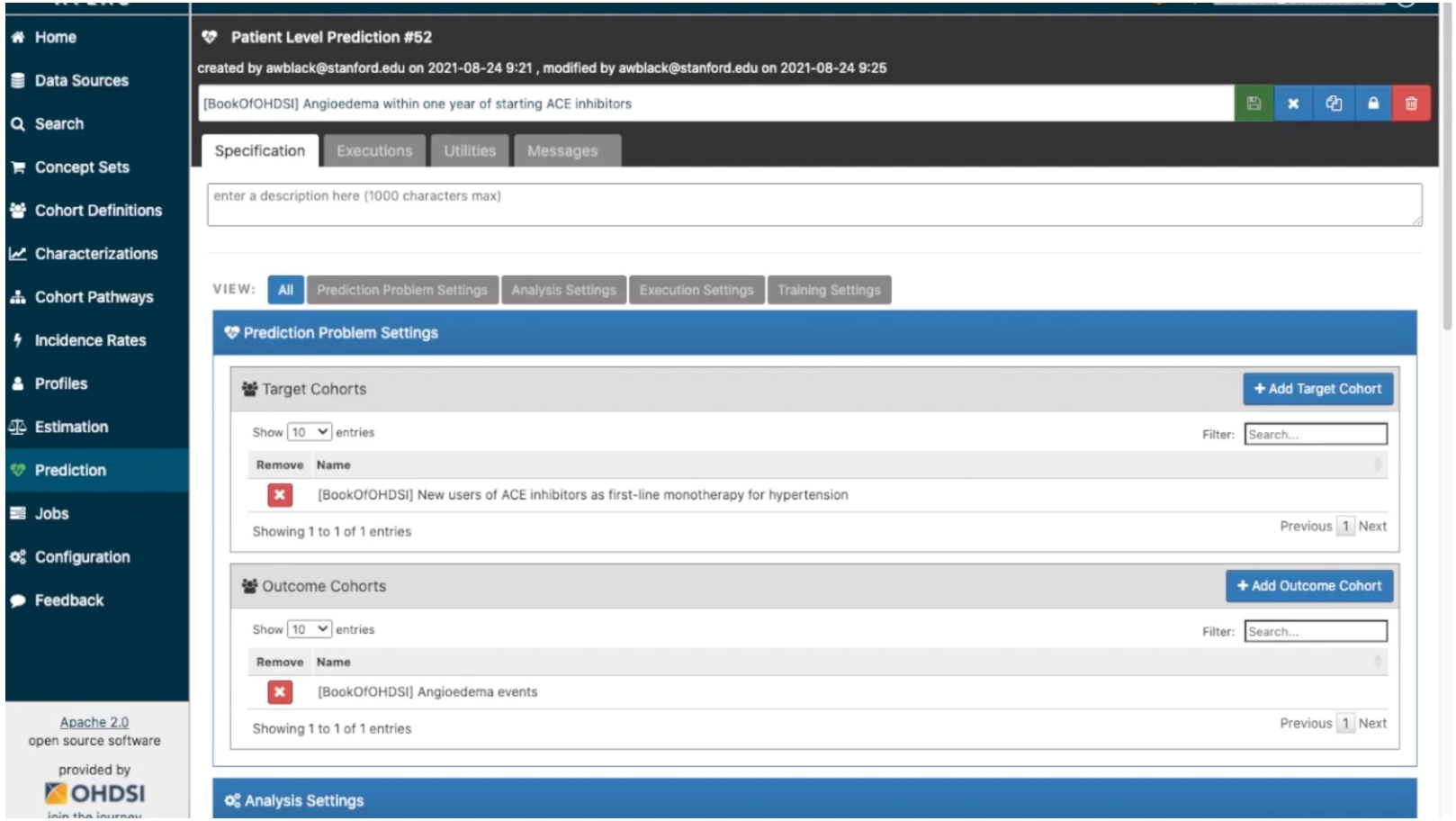
OHDSI

THURSDAY

Empirical Assessment of Alternative Methods for Identifying Seasonality in Observational Healthcare Data

Authors: Anthony Molinaro, Frank DeFalco

#OHDSISocialShowcase This Week



FRIDAY

ATLAS with a BigQuery backend running Execution Engine - a Software demo

Authors: Jose Posada, Priya Desai, Konstantin Yaroshovets, Gregory Klebanov

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?

Welcome To OHDSI Newcomers

Are there any people new to the OHDSI community call who would like to introduce themselves?

Please raise your hand,

and we will call on three people.