Collaborator Showcase Software Demos

OHDSI Community Call
Nov. 21, 2023 • 11 am ET

Upcoming Community Calls

Date	Topic
Nov. 21	Showcase Software Demos
Nov. 28	OHDSI Coordinating Center
Dec. 5	Recent Publications
Dec. 12	Happy Birthday OHDSI! Where Have We Come In 10 Years, and in 12 Months?
Dec. 19	Holiday-Themed Goodbye to 2023!

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

OHDSI Shoutouts!

Charity for Ukrainian kids “Saint Nicolas Reindeers”

General

9 MONTHS LATER

Dymshyts Dmytry Dymshyts 1 11h

Dear OHDSI community, I encourage you to participate in the “Saint Nicolas Reindeers" charity this year as well.
The idea is that you can send the gifts for kids displaced due to the war in Ukraine.
Please see the last year reports in posts above.
Although it’s not easy for non-Ukrainian speakers to navigate Ukrainian online stores and organize the delivery, you can send the money to me or Anna,

- My Paypal: dimshitc@gmail.com
- If you don’t have Paypal, you can use Venmo: **@Anna-Ostropolets**
  and then we chose kids and send the gifts to the organization.

The beauty of this initiative is that the “Saint Nicolas Reindeers”, the group of Ukrainian volunteers not just deliver presents but organize the whole New Year / Christmas celebration with games and parties for kids.

Of course you can have preferences of what presents and which kids to choose. Then we can work on this together.
Here 1 **you can see the list of letters to the Saint Nicolas** with wishes we will fulfill.

Do not hesitate to contact me, if you have any questions: dymshyts@ohdsi.org or in teams.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

Upcoming Workgroup Calls

Date	Time (ET)	Meeting
Wednesday	12 pm	Latin America
Wednesday	4 pm	Vulcan/OHDSI Meeting
Friday	9 am	Phenotype Development & Evaluation
Friday	9 am	GIS – Geographic Information System Development
Friday	9 am	Vaccine Vocabulary
Monday	10 am	Africa Chapter
Monday	4 pm	Eyecare & Vision Research
Tuesday	9 am	OMOP CDM Oncology Genomic Subgroup

Global Symposium Homepage

2023 OHDSI Symposium

Oct. 20-22 · East Brunswick, New Jersey

The 2023 OHDSI Global Symposium welcomed more than 440 of our global collaborators together for three days of sharing research, forging new connections and pushing forward together the OHDSI mission of improving health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

This page will be home to all materials from the global symposium. Check back in the coming days for all video presentations from the event! #JoinTheJourney #OHDSI2023




State of the Community

Various leaders within OHDSI shared a presentation on the state of the community, with specific focuses on data standards, vocabulary enhancements and open-source development. **Speakers included**:

George Hripcsak, Columbia University
Clair Blacketer, Johnson & Johnson
Alexander Davydov, Odysseus Data Services
Katy Sadowski, Boehringer Ingelheim
Peter Rijnbeek, Erasmus MC
Mengling 'Mornin' Feng, National University of Singapore

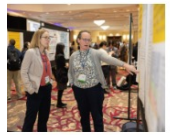
State of the Community Slides

Collaborator Showcase Posters & Software Demos

ceived a record number of submissions for the 2023 Collaborator case, following detailed review by community volunteers in the Scientific w Committee, there were 137 posters and 24 software demos that were nted during the collaborator showcase.

e visit the link below to visit the posters, brief reports and other ementary materials for each showcase submission. Each submission will filed in the #OHDSISocialShowcase, so please make sure you follow l on Twitter/X, LinkedIn and Instagram.

2023 Collaborator Showcase Posters & Demos




Tutorial: Introduction to OHDSI

urney from data to evidence can be challenging alone but is greatly nabled through community collaboration. In this half-day tutorial, we will introduce newcomers to OHDSI. Specifically, about the tools, practices, and open-science approach to evidence generation that the OHDSI community has developed and evolved over the past decade.

Faculty will highlight the ways community individuals can participate as well as receive value from the community's outputs. The course will include topics such as open community data standards – including the OMOP Common Data Model and OHDSI Standardized Vocabularies, opensource analytic tools

2023 Global Collaborator Showcase

Observational Data Standards & Management

2 – FinOMOP – a population-based data network (Javier Gracia-Tabuenca, Perttu Koskenvesa, Pia Tajanen, Sampo Kukkurainen, Gustav Klingstedt, Anna Hammais, Persephone Doupi, Oscar Brück, Leena Hakkarainen, Annu Kaila, Marco Hautalahti, Toni Mikkola, Marianna Niemi, Pasi Rikala, Simo Ryhänen, Anna Virtanen, Arto Mannermaa, Arto Vuori, Joanne Demmler, Eric Fey, Terhi Kilpi, Arho Virkki, Tarja Laitinen, Kimmo Porkka)
3 – From OMOP to CDISC SDTM: Successes, Challenges, and Future Opportunities of using EHR Data for Drug Repurposing in COVID-19 (Wesley Anderson, Ruth Kurtycz, Tahsin Farid, Shermarke Hassan, Kalynn Kennon, Pam Dasher, Danielle Boyce, Will Roddy, Smith F. Heavner)
4 – Augmenting the National COVID Cohort Collaborative (N3C) Dataset with Medicare and Medicaid (CMS) Data, Secure and Deidentified Clinical Dataset (Stephanie Hong, Thomas Richards, Benjamin Amor, Tim Schwab, Philip Sparks, Maya Choudhury, Saad Ljazouli, Peter Leese, Amin Manna, Christophe Roeder, Tanner Zhang, Lisa Eskenazi, Bryan Laraway, James Cavallon, Eric Kim, Shijia Zhang, Emir Amaro Syailendra, Shawn O'Neil, Davera Gabriel, Sigfried Gold, Tricia Francis, Andrew Girvin, Emily Pfaff, Anita Walden, Harold Lehmann, Melissa Haendel, Ken Gersing, Christopher G Chute)
5 – Integrating clinical and laboratory research data using the OMOP CDM (Edward A. Frankenberger, Chun Yang, Vamsidhar Reddy Meda Venkata, Alyssa Goodson)
6 – Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension (Woo Yeon Park, Kyulee Jeon, Teri Sippel Schmidt, Haridimos Kondylakis, Seng Chan You, Paul Nagy)
7 – Conversion of a Myositis Precision Medicine Center into a Common Data Model: A Case Study (Zachary Wang, Will Kelly, Paul Nagy, Christopher A Mecoli)
8 – Implementing a common data model in ophthalmology: Comparison of general eye examination mapping to standard OMOP concepts across two major EHR systems (Justin C. Quon, William Halfpenny, Cindy X. Cai, Sally L. Baxter, Brian C. Toy)
9 – Enhancing Data Quality Management: Introducing Capture and Cleanse Modes to the Data Quality Dashboard (Frank DeFalco, Clair Blacketer)
10 – "OMOP Anywhere": Daily Updates from EHR Data Leveraging Epic's Native Tools (Mujeeb A Basit, Mereeja Varghese, Aamirah Vadsariya, Bhavini Nayee, Margaret Langley, Ashley Huynh, Jennifer Cai, Donglu Xie, Cindy Kao, Eric Nguyen, Todd Boutte, Shiby Antony, Tammye Garrett, Christoph U Lehmann, Duwayne L Willett)
11 – A Toxin Vocabulary for the OMOP CDM (Maksym Trofymenko, Polina Talapova, Tetiana Nesmiian, Andrew Williams, Denys Kaduk, Max Ved, Inna Ageeva)
12 – Challenges and opportunities in adopting OMOP-CDM in Brazilian healthcare: a report from Hospital Israelita Albert Einstein (Maria Abrahao, Uri Adrian Prync Flato, Mateus de Lima Freitas, Diogo Patrão, Amanda Gomes Rabelo, Cesar Augusto Madid Truyts, Gabriela Chiuffa Tunes, Etienne Duin, Gabriel Mesquita de Souza, Soraya Yukari Aashiro, Adriano José Pereira, Edson Amaro)
13 – Transforming the Optum® Enriched Oncology module to OMOP CDM (Dmitry Dymshyts, Clair Blacketer)
14 – Mapping Multi-layered Oncology Data in OMOP (John Methot, Sherry Lee)
15 – Development of psychiatric common data model (P-CDM) leveraging psychiatric scales (Dong Yun Lee, Chungsoo Kim, Rae Woong Park)
16 – Brazilian administrative data for real-world research: a deterministic linkage procedure and OMOP CDM harmonization (Jessica Mayumi Maruyama, Julio Cesar Barbour Oliveira)
17 – Integration of Clinical and Genomic Data Mapped to the OMOP Common Data Model in a Federated Data Network in Belgium (Tatjana Jatsenko, Murat Akand, Joris Robert Vermeesch, Dries Rombaut, Michel Van Speybroeck, Martine Lewi , Valerie Vandeweerd)

ohdsi.org/OHDSI2023

#OHDSISocialShowcase This Week

MONDAY

Conversion of the Canadian Observational Study on Epilepsy (CANOE) REDCap Registry to the OMOP Common Data Model

(Danielle Boyce, Colin Bruce Josephson, Ray Jiang, Samuel Wiebe)

Conversion of the Canadian Observational Study on Epilepsy (CANOE) REDCap Registry to the OMOP Common Data Model

PRESENTER: **Danielle Boyce**

INTRO:

- Epilepsy is a **deadly** medical condition affecting millions of people worldwide.
- There are many **unanswered questions** about causes and treatments.
- Many organizations create "home grown" (non-EHR based) patient registries which were not designed with OMOP in mind.
- Widespread use and promotion of the OMOP CDM could result in collaborative research and contribute to our understanding of epilepsy and development of treatments.

GOALS:

- Gain better understanding of OMOP CDM
- Use lessons learned to improve future data capture and lead international efforts
- Participate in OHDSI software testing and network studies

METHODS

- Longitudinal REDCap Registry with 6,547 people living with epilepsy
- Clinician-entered data
- Met with international team to discuss common data elements of interest to larger network study
  - Selected minimum viable product (MVP) subset of data
- Performed ETL (see Figure 1)

Biggest challenges:
Understanding REDCap data structure, evolution

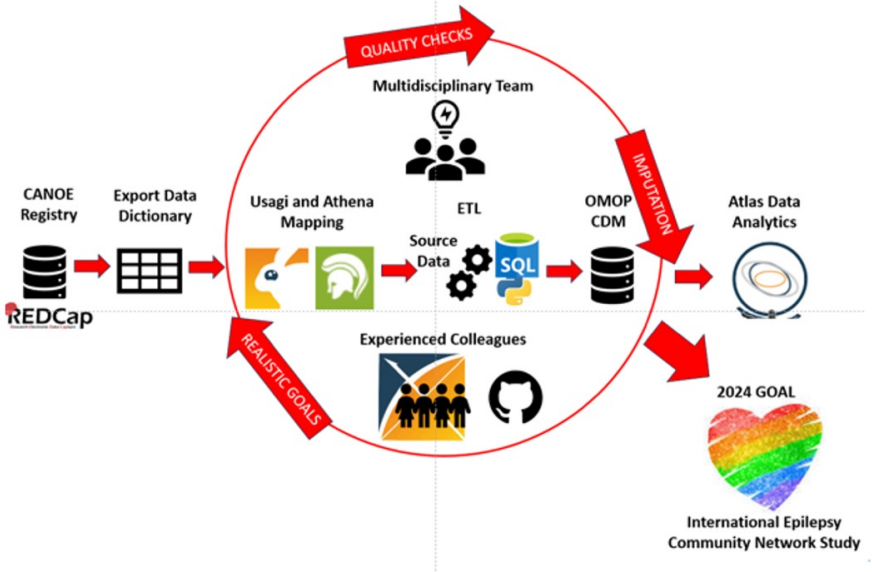
Identifying validated instruments vs. "homemade" surveys

Conditions and drug history prior to baseline = no start dates, e.g.
Do you take X medication?
"1, Yes (current) | 2, Yes (past)"

Training technical team on OHDSI tools

It is possible to ETL a "home grown" patient registry into the OMOP CDM with a **multidisciplinary team**. Consider the **use case** and select "minimum viable product" (**MVP**). **Date imputation may be required** but must be **well documented**.

Figure 1
CANOE Registry ETL Process



RESULTS

MVP: high priority concepts:
Age, race, tests, medications, epilepsy surgery, comorbidities, validated instruments, e.g. General Anxiety Disorder – 7 (GAD-7)

Lower priority concepts:
Driving status, occupation, language, site-specific (non-validated) questionnaires not used at other sites, provider, location

Established wonderful connection with OHDSI Registry Workgroup

Started discussions with international epilepsy community colleagues about planning an OHDSI network study.

AUTHORS
Danielle Boyce[1,2], Colin Bruce Josephson[3], Ray Jiang[3], Samuel Wiebe[3]

[1] Tufts University School of Medicine, [2]Johns Hopkins University, [2]University of Calgary

Take a picture to download the full paper

Reference
Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, Wetherill G, Brand M, Didden EM. (2021). Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. BMC Medical Research Methodology, 21(1):238. https://doi.org/10.1186/s12874-021-01434-3

#OHDSISocialShowcase This Week

TUESDAY

Streamlining Cytogenetic Data Processing with ISCN Parsing and OMOP

(Ben Smith, Trent Peterson, Jessica Manzyuk)

Streamlining Cytogenetic Data Processing with ISCN Parsing and OMOP

PRESENTER: **Ben Smith**

INTRODUCTION

Capturing cytogenetic data for research is often a time-consuming effort beset with errors given that source information in most health systems is found on scanned lab reports requiring manual data entry. Human chromosome data is commonly summarized in strings using the standardized International System for Human Cytogenomic Nomenclature (ISCN), but it is difficult to capture and process these strings in a format suitable for collaborative research.

Cytogenetic research is needed to improve diagnosis and treatment of numerous cancers. ISCN strings describe genomic rearrangements identified by techniques such as karyotyping, Fluorescence In Situ Hybridization (FISH), and microarrays.

Our team explored ways to streamline the process of capturing cytogenetic data and integrating it with other clinical information for cancer research. The result involved development of an ISCN string parser and validator that stores the values and metadata in OMOP.

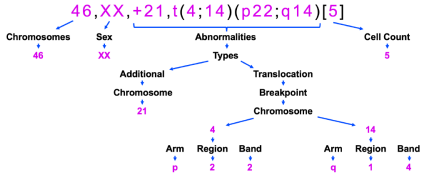
METHODS

We found several existing ISCN string parsers, but each was limited in functionality needed for common use cases. We developed a parser and validator using the ANTLR (ANother Tool for Language Recognition) programming language and decision tree logic. The parsed results are initially output to JSON for subsequent storage in OMOP.

Within OMOP, we decided to store the captured ISCN strings in the NOTE table and the parsed results in the NOTE_NLP table. Some ISCN strings can be much longer than the 50-character limit found in other OMOP fields that were considered. Additionally, parsed string results align well with the purpose of the NOTE_NLP table.

We developed an accompanying API service that can be integrated with various systems for large-scale processing of ISCN strings. We also developed a simple web client that can be publicly accessed at cyto.principia.health.

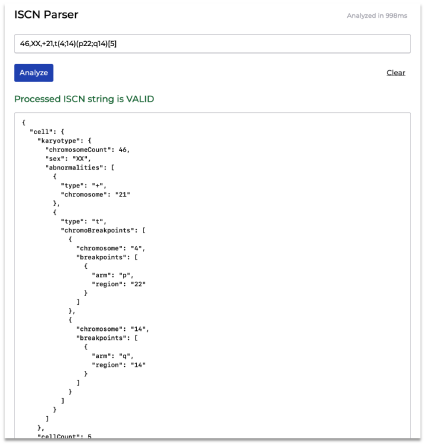
Figure 1 Example ISCN String Parsing Analysis



Streamlining **cytogenetic** data integration for research through **ISCN string parsing** and storage in OMOP

Principia Health Sciences develops collaborative research environments to engage patients, providers, researchers, and other stakeholders in our quest to discover, develop, and deliver curative therapies

Figure 2 Example Parsed String Validation and JSON output



RESULTS

Our comparison of source ISCN strings and processed results confirmed that the parser successfully recognized all target abnormalities within complex strings and stored them in appropriate OMOP CDM tables—both the string (in NOTE) and the parsed abnormality values (in NOTE_NLP). This data could be analyzed alongside corresponding clinical information for given synthetic patient records.

CONCLUSION

This automated approach to cytogenetic data processing represents a significant opportunity to scale cytogenetic research by reducing manual effort and enable inclusion of cytogenetics data in a wider range of studies. Ideally, researchers will be able to use ATLAS with this parsed data, but currently ATLAS does not work properly with NOTE and NOTE_NLP tables. The parser's development roadmap includes the addition of optical character recognition (OCR) for extracting ISCN strings directly from scanned lab reports, and extending support beyond traditional karyotyping to include FISH.

Authors: Ben Smith[1], Trent Peterson[1], Jessica Manzyuk[1],
[1]Principia Health Sciences, Inc

PRINCIPIA HEALTH SCIENCES

OHDSI

#OHDSISocialShowcase This Week

WEDNESDAY

An initial investigation into more complex stacking methods to improve transportability of prediction models developed across multiple databases

(Cynthia Yang, Egill Fridgeirsson, Jan Kors, Jenna Reps, Peter Rijnbeek, Ross Williams, Jenna Wong)

An initial investigation into more complex stacking methods to improve the transportability of prediction models developed across multiple databases

PRESENTER: Jenna Wong

INTRO AND OBJECTIVES

- "Stacking" is an ensemble learning method where the predicted probabilities from a set of base learners are used as features to train a meta-learner. Logistic regression is most often used for the meta-learner.
- Stacking was previously used to combine base learners developed in different databases to improve model transportability.
- We investigated if two incremental enhancements could improve the transportability of stacking ensembles: 1) using random forest for the meta-learner, and 2) additionally including age and sex as features in the random forest meta-learner to allow for tailored combinations of base learner predictions by these basic patient features (interactions).

METHODS

1. We used the OHDSI Patient-Level Prediction framework and sampled up to 500K patients from each of 5 databases: CCAE, MDCD, MDCR, Optum Claims, and Optum EHR.
2. We investigated 21 binary outcomes within a target population of people suffering from depression.
3. We iteratively combined lasso logistic regression base learners from four of the databases and stacked them using $x = \{2{,}000, 20{,}000\}$ observations from the fifth database. The remaining data in the fifth database were used to evaluate performance of the meta-learners and custom models; we evaluated discrimination using the area under the receiver operator characteristic curve (AUC).

The more complex stacking methods investigated in this study did not improve transportability of prediction models developed across multiple databases.

Take a picture to download the full paper

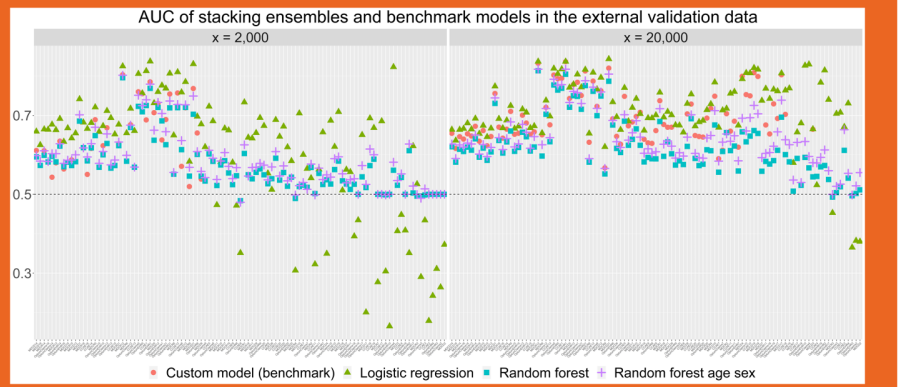


Figure 1. Results when using $x = 2{,}000$ and $x = 20{,}000$ observations from the external validation database for training, across all outcomes and external validation database combinations in order of decreasing observed outcome risk (from left to right).

RESULTS

- The logistic regression meta-learners generally performed best (but performed very poorly in a few cases).
- The random forest meta-learners generally performed worse than the logistic regression meta-learners (but unlike the logistic regression meta-learners, were not prone to very poor performance in some cases).
- Adding age and sex to the random forest meta-learner did not have an appreciable impact.
- A custom model often could not be developed when x=2,000. When x=20,000, the custom models often performed better than the random forest meta-learners.

CONCLUSIONS

- Using random forest instead of logistic regression to develop a meta-learner does not appear to improve model transportability.
- When the amount of available training data in a new database is small, developing a stacking ensemble to combine base learners from other databases may be more feasible than developing a new custom model.
- Future work will explore other more parametric approaches for developing complex meta-learners targeting the types of associations and interactions that were of interest in this study.

Cynthia Yang, MSc,
Egill A. Fridgeirsson, MSc,
Jenna M. Reps, PhD,
Jan A. Kors, PhD,
Peter R. Rijnbeek, PhD,
Ross D. Williams, PhD,
Jenna Wong, PhD

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.

#OHDSISocialShowcase This Week

THURSDAY

Developing a Personalized Clinical Decision Support System for Statin Therapy for Primary Prevention using OMOP-CDM and Deep Learning Techniques

(Su Min Kim, Ju-Hyeon Kim, Yunjin Yum, Eunbeen Jo, Jose Moon, Jong-Ho Kim, Yong Hyun Kim, Eung Ju Kim, Hyung Joon Joo)

Developing a Personalized Clinical Decision Support System for Statin Therapy for Primary Prevention using OMOP-CDM and Deep Learning Techniques

PRESENTER: Hyung Joon Joo

INTRO:
Dyslipidemia is considered a major risk factor for cardiovascular disease (CVD). Statins are the cornerstone of dyslipidemia treatment. However, their clinical use may be hindered by side effects like myalgia and an increased risk of diabetes[1]. Current guidelines suggest specific target low-density lipoprotein cholesterol (LDL-C) levels for statin therapy based on patient characteristics but may not account for individual variability in statin response or tolerability[2]. This highlights the need for personalized approaches in statin therapy. This study aimed to develop a deep learning algorithm utilizing OMOP-CDM data to recommend the optimal statin therapy to achieve target LDL-C levels based on individual cardiovascular risk in primary prevention (Figure 1).

METHODS
We extracted data from patients initiating statin therapy for the first time between 2003 and 2022 from the electronic health record databases of three tertiary university hospitals. Exclusions were patients with overt atherosclerotic cardiovascular disease, those without direct LDL-C measurements, and those unlikely to adhere to statin therapy. A total of 22,447 patient datasets including demographic, clinical information, statin prescription data, diagnostic information, and laboratory data were used. The extracted data were mapped to the OMOP-CDM. This includes mapping local codes to standard OMOP-CDM concepts, and storing them in the appropriate tables (Person, Observation, Drug_exposure, Drug_strength, Condition_occurrence, and Measurement tables). a total of 18,073 patients from two hospitals and 4,374 patients from another hospital were allocated to the development dataset and the external validation dataset, respectively. Multiple predictive models were developed using the development dataset employing multiple linear regression, KNN, SVM, Random Forest, XGB, and Multi-Layer Perceptron (MLP) with 5-fold cross-validation.

Figure 1. Scheme for the development and clinical implementation

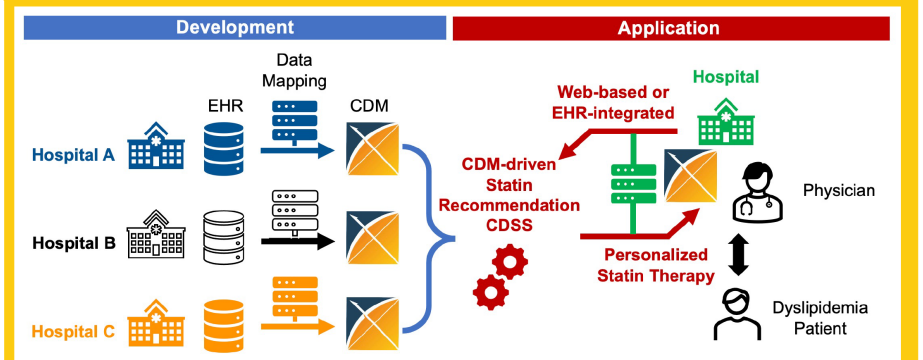
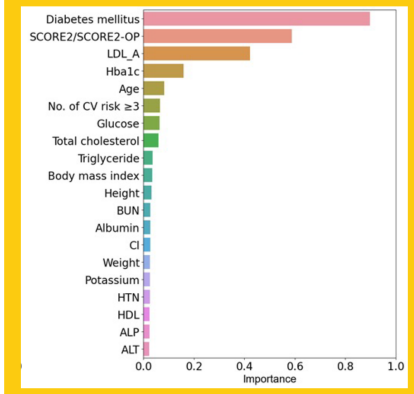


Figure 2. Feature importance



RESULTS:
The study findings revealed that a significant proportion of initial statin prescriptions did not achieve the desired low-density lipoprotein cholesterol (LDL-C) target level, highlighting the limitations of empirical prescription practices. Feature importance analysis identified diabetes mellitus, SCORE2/SCORE2-OP, baseline LDL-C level, Hba1c level, and age as the critical features for achieving target LDL-C levels with statin therapy (Figure 2).

Summary

- **Who cares ?**
  Doctors and patients who don't like to change medications or dosages frequently. When prescribing statins for the first time in patients with dyslipidemia, there are not many cases where the LDL-C target is achieved at once (Target LDL-C achievement rate only **30% in real-word** (DA VINCI study), and **20% in this dataset**).
- **How good?**
  By using this model, the likelihood of success on the first attempt rises to **80%**.

XGBoost demonstrated the highest predictive performance. In the 5-fold cross-validation, the XGB model exhibited an Area Under the Receiver Operating Characteristic (AUROC) of 0.84, accuracy of 0.86, precision of 0.80, recall of 0.52, and F1-score of 0.63 (Table 1).

Table 1. Model performance

Dataset	Model	Accuracy	Precision	Recall	F1-score	AUROC
5-fold cross validation	MLR	0.82 ± 0.01	0.65 ± 0.02	0.47 ± 0.02	0.54 ± 0.02	0.79 ± 0.00
	KNN	0.79 ± 0.01	0.54 ± 0.03	0.40 ± 0.03	0.46 ± 0.03	0.73 ± 0.02
	SVM	0.83 ± 0.01	0.70 ± 0.01	0.46 ± 0.02	0.55 ± 0.01	0.79 ± 0.01
	Random Forest	0.85 ± 0.01	0.80 ± 0.02	0.49 ± 0.03	0.60 ± 0.03	0.82 ± 0.01
	XGBoost	0.86 ± 0.01	0.80 ± 0.02	0.52 ± 0.03	0.63 ± 0.02	0.84 ± 0.01
	MLP	0.85 ± 0.01	0.75 ± 0.01	0.48 ± 0.03	0.59 ± 0.02	0.82 ± 0.01
External validation	MLR	0.79	0.67	0.39	0.49	0.78
	KNN	0.78	0.60	0.44	0.51	0.73
	SVM	0.80	0.69	0.39	0.50	0.74
	Random Forest	0.82	0.79	0.42	0.55	0.82
	XGBoost	0.82	0.71	0.48	0.57	0.81
	MLP	0.84	0.80	0.49	0.60	0.83

Conclusion:
The results highlight the potential of leveraging standardized OMOP-CDM data and deep learning techniques for personalized statin therapy in dyslipidemia. The scalable CDSS platform developed in this study can be readily implemented, considering the increasing adoption of the OMOP-CDM schema and vocabulary in healthcare institutions. Future work will focus on prospective validation of the algorithm's effectiveness and enhancing scalability by upgrading the platform based on HL7/FHIR.

We use OMOP-CDM data from 3 tertiary hospitals in Korea ^^

Take a picture to **download** the **full paper**

Su Min Kim, Ju-Hyeon Kim, Yunjin Yum, Eunbeen Jo, Jose Moon, Jong-Ho Kim, Yong Hyun Kim, Eung Ju Kim, Hyung Joon Joo

Contact:
Hyung Joon Joo
drjoohj@gmail

We are deeply waiting for the collaborators regarding this project as well as the other project using OMOP-CDM.

#OHDSISocialShowcase This Week

FRIDAY

Prediction of Hospital Length of Stay for Planned Admissions Using OMOP CDM

(Haeun Lee, Seok Kim, Hui-Woun Moon, Se Young Jung, Ho-Young Lee, Sooyoung Yoo)

Prediction of Hospital Length of Stay for Planned Admissions Using OMOP CDM

Haeun Lee [1,2], Seok Kim [2], Hui-Woun Moon [2], Ho-Young Lee [2,3], Se Young Jung [2,4], Sooyoung Yoo [2]

[1] Johns Hopkins University School of Medicine, Baltimore, US
[2] Healthcare ICT Research Center, Seoul National University Bundang Hospital, Seongnam, South Korea
[3] Department of Nuclear Medicine, Seoul National University Bundang Hospital, Seongnam South Korea
[4] Department of Family Medicine, Seoul National University Bundang Hospital, Seongnam, South Korea

Background

Accurate prediction of hospital length of stay (LOS) is essential for efficient resource management and healthcare planning, yet these models are often compromised by limited covariates and the heterogeneity in healthcare data, hospital systems, and patient populations. This study aims to develop and validate machine learning-based LOS prediction models for planned admissions using the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Methods

Data source:
Electronic health record (EHR) data from Seoul National University Bundang Hospital (SNUBH) in South Korea, converted to the OMOP CDM (version 5.3), were used in the analysis. The database contains the EHR data of 1,903,603 million patients (40,723,280 admissions) accumulated from 2003 to 2020.

Target cohort:
Two cohorts are analyzed: planned admissions and planned admissions with surgical operations. The study included 137,123 planned hospital admissions from January 2016 to December 2020, including 80,180 surgery admissions.

Feature selection:
The covariates included the (1) demographic information, (2) condition occurrence, (3) medication, (4) observation, (5) measurement, (6) procedure, (7) severity index and (8) visit occurrence. Lasso regularization was applied in Logistic regression.

Outcome: The primary outcome was hospitalization with a length of stay of seven days or more.

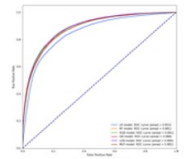
Algorithms: Six algorithms including Logistic regression (LR), random forest (RF), extreme gradient boosting (XGB), light gradient boosting (LGB), gradient boosting (GB) and multilayer perceptron (MLP) algorithms for classification were used to develop the prediction models.

Evaluation: The performance of the models was evaluated based on the area under the receiver operating characteristic curve (AUC), area under Precision-Recall curve (AUPRC), sensitivity (Recall), specificity, positive predictive value (precision), and accuracy. The SHapley Additive exPlanations (SHAP) analysis was further performed to analyze features importance, while calibration plots were used to assess the reliability of the prediction models.

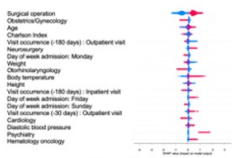
Results

Table 1. Performance metrics of LoS prediction models in planned admissions and surgical patients.

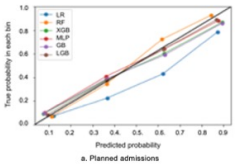
Cohorts	Models	AUROC	AUPRC	Sensitivity	Specificity	PPV	NPV
Planned admissions	LR	0.853	0.744	**0.771**	0.785	0.642	**0.873**
	RF	0.881	0.802	0.604	**0.922**	**0.794**	0.824
	XGB	**0.891**	**0.819**	0.686	0.896	0.768	0.852
	GB	0.888	0.811	0.681	0.894	0.763	0.849
	LGB	0.889	0.816	0.661	0.906	0.778	0.843
	MLP	0.882	0.804	0.635	0.910	0.778	0.833
Surgical patients	LR	0.935	0.836	0.88	0.853	0.865	0.930
	RF	0.931	0.841	0.864	0.839	0.866	0.920
	XGB	0.947	0.854	**0.889**	**0.866**	0.879	0.942
	GB	0.943	0.852	0.877	0.853	0.876	0.937
	LGB	**0.948**	**0.856**	0.884	0.861	**0.880**	**0.943**
	MLP	0.944	0.849	0.885	0.861	0.875	0.938



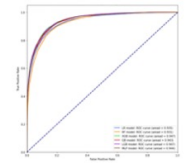
a. Planned admissions



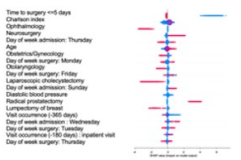
a. Planned admissions



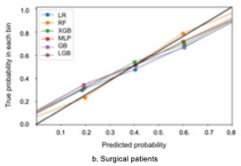
a. Planned admissions



b. Surgical patients



b. Surgical patients



b. Surgical patients

Figure 1. Receiver operating characteristics curves for LOS prediction

Figure 2. SHAP summary plot for LOS prediction model

Figure 3. Calibration of LOS prediction models in the test set

Conclusions

We demonstrated the use of the OMOP CDM to predict LOS for both planned and surgery admissions. Predictors such as surgical operations, admitting specialty, age have been identified as potential contributors, while time to surgery, severity score, admitting specialty are predictive in surgery admissions. These models could provide insights and strategies for optimizing resource utilization across various healthcare facilities that implement OMOP CDM, potentially leading to reduced surgical mortality rates and improved overall operational efficiency. Additional research necessary to validate the models' performance across different institutions is currently underway, using external validations within the OHDSI network community.

Contact: hlee292@jh.edu

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI22C0471).

Opening: Limerick Digital Cancer Research Centre

University of Limerick local time: 21-November-2023 14:39

Navigation Section

- Search
  Click here to search for vacancies. Please email erecruitment@ul.ie if you experience difficulties.
- Login
  Click to Login. UL Staff, you must be logged in to see Internal Vacancies
- Register
  Click to Register
- Vacancy Detail
  Click to view the current job specification
- Terms & Conditions
  Agree to the terms and conditions
- Human Resources Website
  "THE UNIVERSITY OF LIMERICK IS AN EQUAL OPPORTUNITIES EMPLOYER AND COMMITTED TO SELECTION ON MERIT"

corehr

Job Spec

Advertisement/Information for Applicants

Please click on Information for Applicants/Job Description link below for full job

Post Doctoral Researcher (Level 1 or 2) in Cancer Digital Health Real World Evidence (2 Positions)

With over 18,000 students and 2,000 members of staff, the University of Limerick (UL) is an energetic, research led and enterprising institution with a proud record in innovation and excellence in education, research and scholarship. The dynamic, entrepreneurial and pioneering values which drive UL's mission and strategy ensure that we capitalise on local, national and international engagement and connectivity. We are renowned for providing an outstanding student experience and conducting leading-edge research. Our commitment is to make a difference by shaping the future through educating and empowering our students.
With the River Shannon as a unifying focal point, UL is situated on a superb riverside campus of over 130 hectares. Outstanding recreational, cultural and sporting facilities further enhance the campus's exceptional learning and research environment.

Applications are invited for the following position:

Faculty of Education & Health Sciences

School of Medicine

Post Doctoral Researcher (Level 1 or 2) in Cancer Digital Health Real World Evidence (2 Positions) Specific Purpose Contract

Salary Scales: PD1 €42,033 - €48,427 p.a. pro rata

PD2 €49,790 – €54,153 p.a. pro rata

Informal enquires regarding the post may be directed to:
Professor Aedin Culhane
School of Medicine
University of Limerick
Email: aedin.culhane@ul.ie

"This is a professional training and development role and the training and development relevant to this position will be completed within the period of the contract. Postdoctoral Researchers appointed will be expected to complete the Researcher Career Development Programme."
The closing date for receipt of applications is Friday, 15th December 2023.
Applications must be completed online before 12 noon, Irish Standard Time on the closing date.

The University of Limerick supports blended working

Job Opening: Stanford University

Stanford | Office of Postdoctoral Affairs
All Postdocs. All the Time.

Home About Prospective Postdocs Current Postdocs Faculty Mentors Postdoc Admins

Prospective Postdocs

- How To Apply
- Open Postdoctoral Positions
- Finding a Faculty Mentor
- Cost of Living
- Housing
- Fellowships at Stanford
- Fellowships outside Stanford

Open Postdoctoral position, faculty mentor Brian Bateman

Our research team is looking for a postdoctoral scholar in perinatal pharmacoepidemiology. The scholar will work closely with Drs. Brian Bateman and Stephanie Leonard on NIH-funded research projects on the comparative safety and effectiveness of medications in pregnancy and related research topics. Our projects employ advanced analytical methods in large databases, which include claims data and electronic health record data in conventional structures and in common data models. Current topical focus areas include mental health, behavioral health and cardiovascular health of people who are pregnant or postpartum.

Our research group prioritizes a collaborative and inclusive team environment. The principal investigators are experienced mentors who are highly committed to supporting the postdoctoral scholar in advancing their career as a future independent investigator. The

Important Info

Faculty Sponsor (Last, First Name):
Bateman, Brian

Other Mentor(s) if Applicable:
Stephanie Leonard

Stanford Departments and Centers:
Anesthes, Periop & Pain Med

Postdoc Appointment Term:
Initial appointment is 1 year with renewal after the first year for an additional 1-2 years by mutual agreement

Appointment Start Date: Flexible start date

Group or Departmental Website:

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?

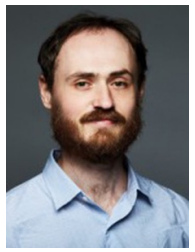
Where Are We Going?

Nov. 21: Collaborator Showcase Demos

Sarah Gasman

Senior Data Analyst • Boston Medical Center

Leveraging the OMOP Common Data Model to Support Distributed Health Equity Research



Boudewijn Aasman

Senior Data Science Engineer • Montefiore Medical Center

Integrating ATLAS Cohorts with DICOM Images and ECG Waveforms to Enrich Real-World Evidence Research

Laurence Lawrence-Archer

Data Scientist • Odysseus Data Services

Introducing KOIOS: removing impediments in genomic variant identification and mapping

Andrey Soares

Assistant Professor • University of Colorado Anschutz Medical Campus

OMOP-to-BULK FHIR: A tool to convert population level clinical data into standardized FHIR batch data