Workgroup Updates

OHDSI Community Call
Oct. 5, 2021 • 11 am ET

Upcoming OHDSI Community Calls

Date	Topic
Oct. 5	Workgroup Updates
Oct. 12	Meet The Titans
Oct. 19	Focus Topic: LEGEND Type 2 Diabetes Study
Oct. 26	Trick or Treat
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. 5	Workgroup Updates
Oct. 12	Meet The Titans
Oct. 19	Focus Topic: LEGEND Type 2 Diabetes Study
Oct. 26	Trick or Treat
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

Oct. 12 Community Call: Meet The Titans

Maxim Moinat

Yong Chen

Adam Black

Asieh Golozar

Erica Voss

Mui Van Zandt

Faaizah Arshad

Ross Williams

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

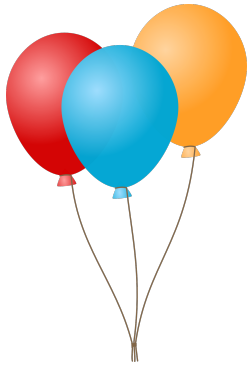
Congratulations to **Emily Pfaff, Andrew Girvin, Davera Gabriel, Kristin Kostka, Michele Morris, Matvey Palchuk, Harold Lehmann, Benjamin Amor, Mark Bissell, Katie Bradwell, Sigfried Gold, Stephanie Hong, Johanna Loomba, Amin Manna, Julie McMurry, Emily Niehaus, Nabeel Quresh, Anita Walden, Xiaohan Tanner Zhang, Richard Zhu, Richard Moffitt, Melissa Haendel, Christopher Chute, and the N3C Consortium** on the publication of "**Synergies between Centralized and Federated Approaches to Data Quality: A Report from the National COVID Cohort Collaborative**" in JAMIA.

OHDSI Shoutouts!

Good luck to a pair of new DPhil students at the University of Oxford, **Kristin Kostka and Jamie Weaver.**

Take good care of them, **Dani Prieto-Alhambra!**

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?

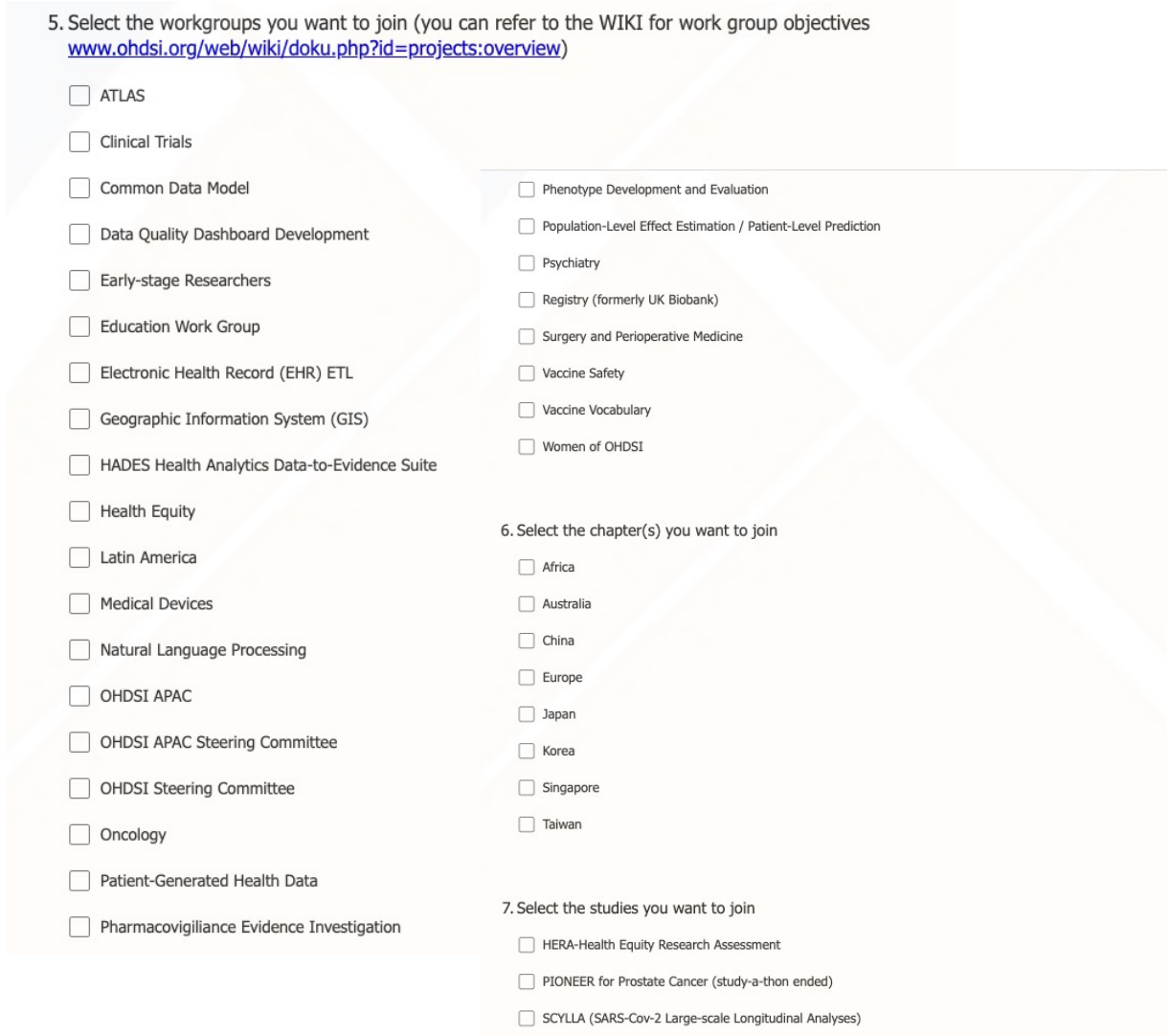
Where Are We Going?

Upcoming Workgroup Calls

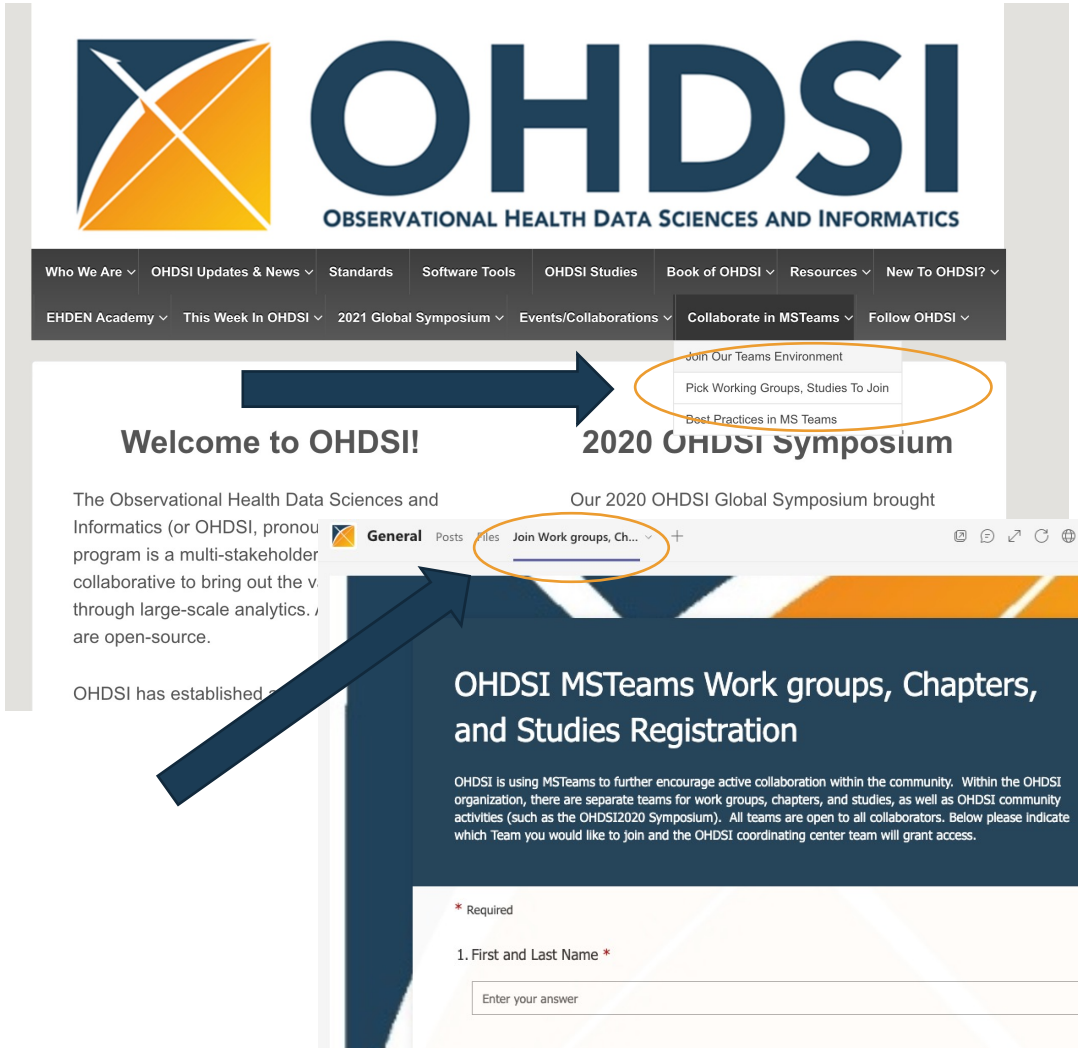
Date	Time (ET)	Meeting
Tuesday	1 pm	Common Data Model
Tuesday	2 pm	Health Equity
Tuesday	3 pm	OMOP CDM Oncology – Outreach/Research Subgroup
Wednesday	2 am (4 pm KST)	Patient-Level Prediction/Population Level Estimation (Eastern Hemi)
Wednesday	9 am	Vaccine Vocabulary
Wednesday	10 am	OMOP CDM Oncology – Development Subgroup
Thursday	12 pm	Patient-Level Prediction/Population Level Estimation (Western Hemi)
Thursday	1 pm	OMOP CDM Oncology – CDM/Vocabulary Subgroup
Friday	9 am	Education WG
Friday	1 pm	Phenotype Development & Evaluation
Monday	8 am	Early-Stage Researchers (Europe/East Coast/West Coast)
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



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)	Contents	Speaker(s)
Morning	OHDSI State of the Community	George Hripcsak/Patrick Ryan
	OHDSI APAC State of the Community	Mui Van Zandt
	EHDEN	Peter Rijnbeek
	FHIR and OHDSI Collaboration	Christian Reich
	APAC Chapter vision for 2022	APAC chapter leaders
Break		
Afternoon	Networking Session	All

www.ohdsi.org/apac

Vote For Collaborator Showcase Honors

2021 OHDSI Symposium Best Community Contribution Awards

OHDSI's mission is to improve health, by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care. In recognition of our collaborators' dedicated efforts towards achieving this mission, the annual collaborator showcase was created as an opportunity to highlight collaborators' hard work.

To help us recognize our collaborators' achievements, please select your favorite contribution (poster, lightning talk or software demonstration) to the collaborator showcase under each of the following four topics:
1) Observational Data Standards & Management
2) Methodological Research
3) Clinical Applications
4) Open-source Analytics Development

...

1. CATEGORY 1: OBSERVATIONAL DATA STANDARDS & MANAGEMENT (please choose the best contribution- posters listed by their location followed by the lightning talks)

- ◯ DS01 Conversion of UK Biobank into the OMOP CDM: New Data for Inferences Between Episodic Care (Amelia J Averitt, Alexandra Orlova, Alexander Davydov, Oleg Zhuk, Michael N Cantor, Gregory Klebanov)
- ◯ DS02 Medication dosage and exposure duration in OMOP CDM: mapping challenges (Tatiana Banokina, Dmitry Dymshyts, Alexandra Orlova, Alexander Kraynov, Alexander Davydov)
- ◯ DS03 Mapping UK Biobank to the OMOP CDM: challenges and solutions using the delphyne ETL framework (Sofia Bazakou, Maxim Moinat, Alessia Peviani, Anne van Winzum, Stefan Payralbe, Vaclav Papez, Spiros Denaxas)
- ◯ DS04 The VISIT_DETAIL: A Vehicle for Standard Visits (Clair Blacketer)
- ◯ DS05 Establishing a large COVID-19 cohort through mapping the Information System for Research in Primary Care (SIDIAP) in Catalonia to the OMOP Common Data Model (Sergio FernÃindez-BertolÃ¬n, Erica A Voss, Clair Blacketer, Maria AragÃ³n, Martina Recalde, Elena Roel, Carlen Reyes, Sebastiaan van Sandijk, Lars Halvorsen, Peter R Rijnbeek, Talita Duarte-Salles)
- ◯ MR27 Wikipedia Drug Safety Advisory Committee: Distilling A Drug Adverse Effect Reference Set Using Wisdom of The Crowd (Yonatan Bilu, Chen Yanover)
- ◯ Lightning Talk: MR LT1 Evaluating the performance of Austin's standardized difference heuristic in observational cohort studies with varying sample size (Mitchell Conover, Azza Shoaibi, Joshua Ide, Martijn Schuemie)
- ◯ Lightning Talk: MR LT2 Leveraging APHRODITE to identify bias in statistical phenotyping algorithms Juan M. Banda, Nigam H. Shah, Vyjeyanthi S Periyakoil)
- ◯ Lightning Talk: MR LT3 Assessing the impact of race on glomerular filtration rate prediction (Linying Zhang, Lauren R. Richter, George Hripcsak)
- ◯ Lightning Talk: MR LT4 A Prediction Model Library (Ross D. Williams, Jenna M. Reps, Peter R. Rijnbeek)

3. CATEGORY 3: CLINICAL APPLICATIONS (please choose the best contribution- posters listed by location followed by the lightning talks)

- ◯ CA01 Impact of the COVID-19 pandemic on pediatric utilization patterns in claims data (Alan Andryc, Rachel Weinstein, Steven Sacavage, Marsha Tharakan, Rupa Makadia)
- ◯ CA02 Short-term mortality in patients undergoing colorectal cancer surgery: A prediction study (Karoline Bendix Bräuner, Mikail Gögenur, Viviane Annabelle Lin, Andreas Weinberger Rosen, Johan Clausen, Eldar Allakhverdiiev, Rasmus Peuliche Vogelsang, Ismail Gögenur)
- ◯ CA03 Predicting early readmission after colorectal cancer surgery using only preoperative variables (Johan Clausen, Andreas Weinberger Rosen, Karoline Bendix Bräuner, Mikail Gögenur, Viviane Annabelle Lin, Eldar Allakhverdiiev, Julie Sparholt Walbech, Ismail Gögenur)
- ◯ CA04 Diagnostic Accuracy of Code-Based Algorithms to Identify Urinary Tract Infection in U.S. Administrative Claims Databases (Stephen P Fortin, Jeroen Geurtsen, Michal Sarnecki, Joachim Doua, Jamie Colasurdo, Joel Swerdel)
- ◯ CA05 Predicting risk of recurrence after surgery for colorectal cancer (Mikail Gögenur, Viviane Lin, Adamantia Tsouchnika, Eldar Allakhverdiiev, Andreas Weinberger Rosen, Karoline Bendix Bräuner, Julie Sparholt Walbech, Ismail Gögenur)

Next CBER Best Seminar Series

BEST

Webinar Registration

Topic CBER BEST Initiative Seminar Series - Exploring Vaccine Safety Datalink COVID vaccine rapid cycle analysis (RCA) methods

Description Background: The CBER BEST Initiative Seminar Series is designed to share and discuss recent research of relevance to ongoing and future surveillance activities of CBER regulated products, namely biologics. The series focuses on safety and effectiveness of biologics including vaccines, blood components, blood-derived products, tissues and advanced therapies. The seminars will provide information on characteristics of biologics, required infrastructure, study designs, and analytic methods utilized for pharmacovigilance and pharmacoepidemiologic studies of biologics. They will also cover information regarding potential data sources, informatics challenges and requirements, utilization of real-world data and evidence, and risk-benefit analysis for biologic products. The length of each session may vary, and the presenters will be invited from outside FDA. Please see the details below for our upcoming seminar. Anyone can register and join for free. Stay tuned for more details and additional webinars during the year.

Topic: Exploring Vaccine Safety Datalink COVID vaccine rapid cycle analysis (RCA) methods

Description: We will review statistical methods used in observational studies of the safety and effectiveness of COVID-19
vaccines. Topics will include:
- How to compare recent vaccinees with concurrent comparators (unvaccinated or less recently vaccinated) and
- with comparators who are not concurrent (historical rates or self-controls) to make inferences about outcome rates
- that would be expected among vaccinees had they not been vaccinated
- Methods for estimating risk ratios
- How to examine change in vaccine effectiveness (waning) or vaccine safety over time-since-vaccination
- Sequential tests

Presenter: Nicola P. Klein, MD, PhD

Time Oct 20, 2021 11:00 AM in Eastern Time (US and Canada)

Next APAC Community Call: Thursday

PHOEBE

Anna Ostropolets

PhD Student, Columbia University Dept. of Biomedical Informatics

Cohort Diagnostics

Gowtham Rao

Senior Director, Johnson & Johnson

ATC Hierarchy

Christian Reich

VP Real World Analytics Solutions, IQVIA

www.ohdsi.org/apac

October Newsletter Is Out

Community Updates

Where Have We Been

• Thank you to everybody within our community who made the 2021 Global Symposium such a memorable event. We had four full days of activities, including our first tutorial on building conceptsets, a reproducibility workshop, and then two main symposium days. The first day focused on **OHDSI's Impact on the COVID-19 Pandemic**, while the second day focused on **The Journey to Reliable Evidence**. All material from those two days is available within this newsletter.

• **Clair Blacketer** announced the release of OMOP CDM v5.4 in this recent forum post, and spoke about it briefly on the Sept. 28 community call. Clair will provide a live demo during a future OHDSI community call about v5.4, so please join the calls or follow our social channels for updates. Thank you to the entire CDM workgroup for leading this effort.

Where Are We Now

• The collaboration between OHDSI and HL7 took an important step forward this week during a two-day workshop that was jointly led by the two groups. Concepts that were explored included how to build the community and engage participants, reviewing near-term challenges regarding mapping and other issues, and working to establish a collaboration framework for moving forward, including setup of specific subgroups to advance individual use cases.

• The #OHDSISocialShowcase began this week and will continue for the next several months, as we highlight all the important and impressive research presented during our OHDSI2021 Collaborator Showcase. Each weekday, one presentation will be highlighted on both our Twitter and LinkedIn feeds. Please share with your networks to spread the word about our global efforts!

Where Are We Going

• The 2021 APAC Symposium will be held virtually Nov. 18, and some details

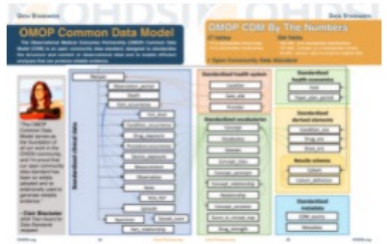
OHDSI

The Journey Newsletter (October 2021)

The #OHDSI2021 Global Symposium was another memorable event for the community. All presentations from the two main days are now available in this newsletter. We also have updates about the newest OMOP Common Data Model version, a look ahead at our October meeting schedule, and plenty more. **#JoinTheJourney**

(If images/videos don't appear, please click "View this email in your browser" link above.)

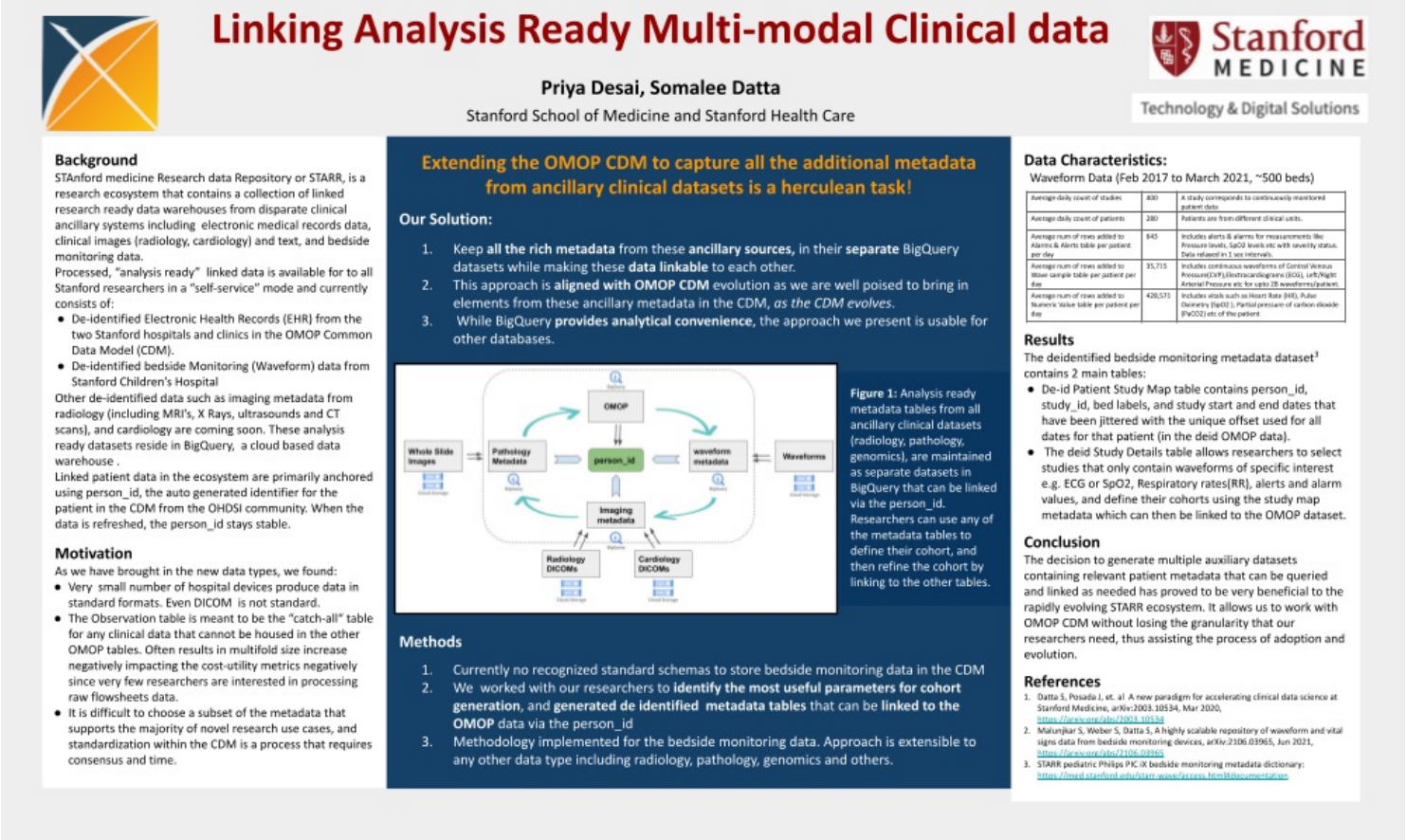
Miss Any Of #OHDSI2021? Catch Up Here!



As part of the OHDSI2021 Symposium, we unveiled a new book entitled "Our Journey: Where The OHDSI Community Has Been, And Where We Are Going." Many people ordered it as their 'Symposium Surprise' and followed along with it during the State of the Community Presentation. If you didn't order this 92-page book that focuses on all aspects of our community (research, data network, people, events and more), you can download a copy of it below.

Download "Our Journey: Where The OHDSI Community Has Been, And Where We Are Going"

#OHDSISocialShowcase This Week



Linking Analysis Ready Multi-modal Clinical data

Priya Desai, Somalee Datta

Stanford School of Medicine and Stanford Health Care

Stanford MEDICINE

Technology & Digital Solutions

Background

STAnford medicine Research data Repository or STARR, is a research ecosystem that contains a collection of linked research ready data warehouses from disparate clinical ancillary systems including electronic medical records data, clinical images (radiology, cardiology) and text, and bedside monitoring data.

Processed, "analysis ready" linked data is available for to all Stanford researchers in a "self-service" mode and currently consists of:

- De-identified Electronic Health Records (EHR) from the two Stanford hospitals and clinics in the OMOP Common Data Model (CDM).
- De-identified bedside Monitoring (Waveform) data from Stanford Children's Hospital

Other de-identified data such as imaging metadata from radiology (including MRI's, X Rays, ultrasounds and CT scans), and cardiology are coming soon. These analysis ready datasets reside in BigQuery, a cloud based data warehouse .

Linked patient data in the ecosystem are primarily anchored using person_id, the auto generated identifier for the patient in the CDM from the OHDSI community. When the data is refreshed, the person_id stays stable.

Motivation

As we have brought in the new data types, we found:

- Very small number of hospital devices produce data in standard formats. Even DICOM is not standard.
- The Observation table is meant to be the "catch-all" table for any clinical data that cannot be housed in the other OMOP tables. Often results in multifold size increase negatively impacting the cost-utility metrics negatively since very few researchers are interested in processing raw flowsheets data.
- It is difficult to choose a subset of the metadata that supports the majority of novel research use cases, and standardization within the CDM is a process that requires consensus and time.

Extending the OMOP CDM to capture all the additional metadata from ancillary clinical datasets is a herculean task!

Our Solution:

1. Keep **all the rich metadata** from these **ancillary sources,** in their **separate** BigQuery datasets while making these **data linkable** to each other.
2. This approach is **aligned with OMOP CDM** evolution as we are well poised to bring in elements from these ancillary metadata in the CDM, *as the CDM evolves*.
3. While BigQuery **provides analytical convenience**, the approach we present is usable for other databases.

Figure 1: Analysis ready metadata tables from all ancillary clinical datasets (radiology, pathology, genomics), are maintained as separate datasets in BigQuery that can be linked via the person_id. Researchers can use any of the metadata tables to define their cohort, and then refine the cohort by linking to the other tables.

Methods

1. Currently no recognized standard schemas to store bedside monitoring data in the CDM
2. We worked with our researchers to **identify the most useful parameters for cohort generation**, and **generated de identified metadata tables** that can be **linked to the OMOP** data via the person_id
3. Methodology implemented for the bedside monitoring data. Approach is extensible to any other data type including radiology, pathology, genomics and others.

Data Characteristics:

Waveform Data (Feb 2017 to March 2021, ~500 beds)

Average daily count of studies	800	A study corresponds to continuously monitored patient data
Average daily count of patients	280	Patients are from different clinical units.
Average num of rows added to Alarms & Alerts table per patient per day	845	Includes alerts & alarms for measurements like Pressure levels, SpO2 levels etc with severity status. Data released in 1 sec intervals.
Average num of rows added to Wave sample table per patient per day	35,715	Includes continuous waveforms of Central Venous Pressure(CVP),Electrocardiograms (ECG), Left/Right Arterial Pressure etc for upto 28 waveforms/patient.
Average num of rows added to Numeric Value table per patient per day	428,571	Includes vitals such as Heart Rate (HR), Pulse Oximetry (SpO2), Partial pressure of carbon dioxide (PaCO2) etc of the patient

Results

The deidentified bedside monitoring metadata dataset[3] contains 2 main tables:

- De-id Patient Study Map table contains person_id, study_id, bed labels, and study start and end dates that have been jittered with the unique offset used for all dates for that patient (in the deid OMOP data).
- The deid Study Details table allows researchers to select studies that only contain waveforms of specific interest e.g. ECG or SpO2, Respiratory rates(RR), alerts and alarm values, and define their cohorts using the study map metadata which can then be linked to the OMOP dataset.

Conclusion

The decision to generate multiple auxiliary datasets containing relevant patient metadata that can be queried and linked as needed has proved to be very beneficial to the rapidly evolving STARR ecosystem. It allows us to work with OMOP CDM without losing the granularity that our researchers need, thus assisting the process of adoption and evolution.

References

1. Datta S, Posada J, et. al A new paradigm for accelerating clinical data science at Stanford Medicine, arXiv:2003.10534, Mar 2020, https://arxiv.org/abs/2003.10534
2. Malunjkar S, Weber S, Datta S, A highly scalable repository of waveform and vital signs data from bedside monitoring devices, arXiv:2106.03965, Jun 2021, https://arxiv.org/abs/2106.03965
3. STARR pediatric Philips PIC iX bedside monitoring metadata dictionary: https://med.stanford.edu/starr-www/access.html#documentation

MONDAY

Linking Analysis Ready Multi-modal Clinical data

Authors: Priya Desai, Somalee Datta

#OHDSISocialShowcase This Week

Lightning Talk!

Data Quality Dashboard Used to Improve the Quality of the EHDEN Network

PRESENTERS: Erica Voss, Clair Blacketer

INTRO:

- The European Health Data & Evidence Network (**EHDEN**) is developing an observational research ecosystem based on federated data to enable better health outcomes (figure 1).
- Through transformation to the OMOP Common Data Model (CDM), studies can be performed that rely on the standard data structure.
- 25 data partners (DPs) were awarded a grant to convert to the OMOP CDM.
- We used the opportunity to understand how the Data Quality Dashboard (DQD) can be used to improve the quality of a network of databases.

METHODS

1. Each completed DP was required to run the DQD at least twice as they moved their data through the Data Conversion Pipeline (figure 1).
2. The DQD results from the first and last run were compared.
3. The difference in data quality passes between first and last run were quantified.

RESULTS:

- All 9 completed DPs showed improvement in the number of checks that passed between the first and last run, for the checks in common between the two (figure 3)
- The DQD highlighted issues related to the conversion to the OMOP CDM and issues with the source data that would not have been identified otherwise.
- DPs were able to participate in the inaugural Evidence-A-Thon and focus on the science because of their efforts to fix the issues highlighted by the DQD.

The **Data Quality Dashboard** improves **data networks** by **exposing issues** in both the **source** and **standardized** data.

Take a picture to download the full paper

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

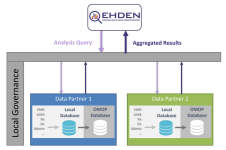


Figure 1: *The EHDEN Network*

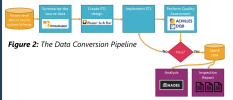


Figure 2: *The Data Conversion Pipeline*

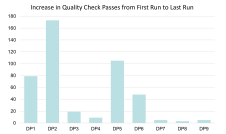


Figure 3: *Increase in data quality check passes from the first run to the last run of the DQD.*

Erica Voss, **Clair Blacketer**, Maxim Moinat, Frank DeFalco, Peter Rijnbeek

janssen The Hyve

Erasmus MC

OHDSI

TUESDAY

Data Quality Dashboard Used to Improve Data Quality in the EHDEN Network

Authors: Clair Blacketer, Erica Voss, Frank DeFalco, Maxim Moinat, Peter Rijnbeek

#OHDSISocialShowcase This Week

GPU Parallelization of Massive Sample-size Survival Analysis

Jianxiao Yang[1], Marc A. Suchard[1,2,3]
1. Department of Computational Medicine, David Geffen School of Medicine at UCLA 2. Department of Biostatistics, UCLA Fielding School of Public Health.
3. Department of Human Genetics, David Geffen School of Medicine at UCLA

Introduction

Large-scale Observational Data

- Observational databases have millions of individuals [1] with thousands of patient characteristics and up to 10 years of data per life [2].
- Resource for comparative effectiveness and safety study.
- Survival analysis is a main statistical method in comparative effectiveness and safety study.
- **Problem**: computational burden.

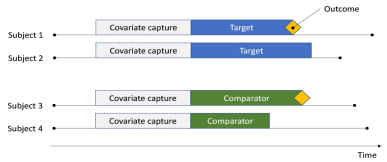


Figure 1: Comparative effectiveness study with cohort design.

GPU Parallelization

- Graphics processing units (GPUs) contain thousands of processor cores that can apply the same numerical operations simultaneously to elements of large data arrays under a "Single Instruction, Multiple Threads" (SIMT) programming paradigm.
- GPUs are relatively inexpensive, easy-to-use hardware that offers impressive potential for speeding up computations.

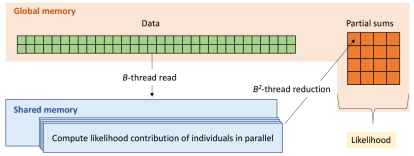


Figure 2: Massive parallelization strategy for computing the likelihood.

Methods

Cox Proportional Hazards Model

- The hazard model formula depends on a baseline survival function and a set of explanatory variables.
- Parameter estimation of the Cox proportional hazards model follows from the log-partial likelihood.

$$h(t|\boldsymbol{\beta}) = h_0(t|\boldsymbol{\beta}) \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}\right)$$

$$l_{partial}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left\{ \boldsymbol{\beta}^T \mathbf{x}_i - \log\left[\sum_{k \in R(Y_i)} \exp\left(\boldsymbol{\beta}^T \mathbf{x}_k\right) \right] \right\}$$

Fine-Gray Sub-distribution Proportional Hazards Model

- The Fine-Gray model generalizes the Cox proportional hazards model to competing risks time-to-event data that consists of more than one type of events.
- Competing risks arise when individuals can experience more than one type of event and the occurrence of one type of event will prevent the occurrence of the others.

Massive Parallelization for Parameter Estimation with Prefix Sums and Reduction

- We identify prefix sums [3] and reductions in log-partial likelihood of Cox model and Fine-Gray model due to the cumulative structure of the risk set.
- We avoid unnecessary memory transactions by fusing prefix sums and reductions operations in likelihood calculations into a single kernel.
- We minimize data movements by exploiting the sparsity of the design matrix.

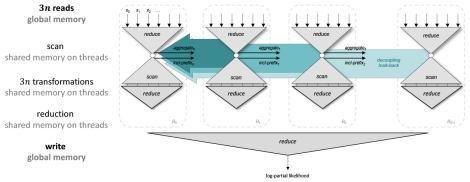


Figure 3: Fused kernel for maximum likelihood estimation.

Results

Simulation experiments

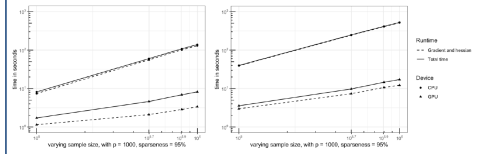


Figure 4: GPU vs CPU runtimes for Cox model (left) and Fine-Gray model (right).

Antihypertensive drug classes study

- Dataset: thiazide or thiazide-like diuretics (THZ) and angiotensin-converting enzyme inhibitors (ACEi) cohort study, reproduction of [4] in CCAE dataset.
- Cohort size: 1,065,745 hypertension patients and 7891 covariates.
- Outcome of interest: major cardiovascular events.
- GPU parallelization reduce the time of parameter estimation from 16 hours on multi-core CPU to less than one hour.

Conclusions

- We implement the massive parallelization of the Cox proportional hazards model and Fine-Gray model by using NVIDIA's CUB library for parallel computing of prefix sums and exploiting the sparsity of data.
- By saving data movement and clever manipulation of likelihood structure, our parallelization significantly reduces the runtime of large-scale survival analysis.

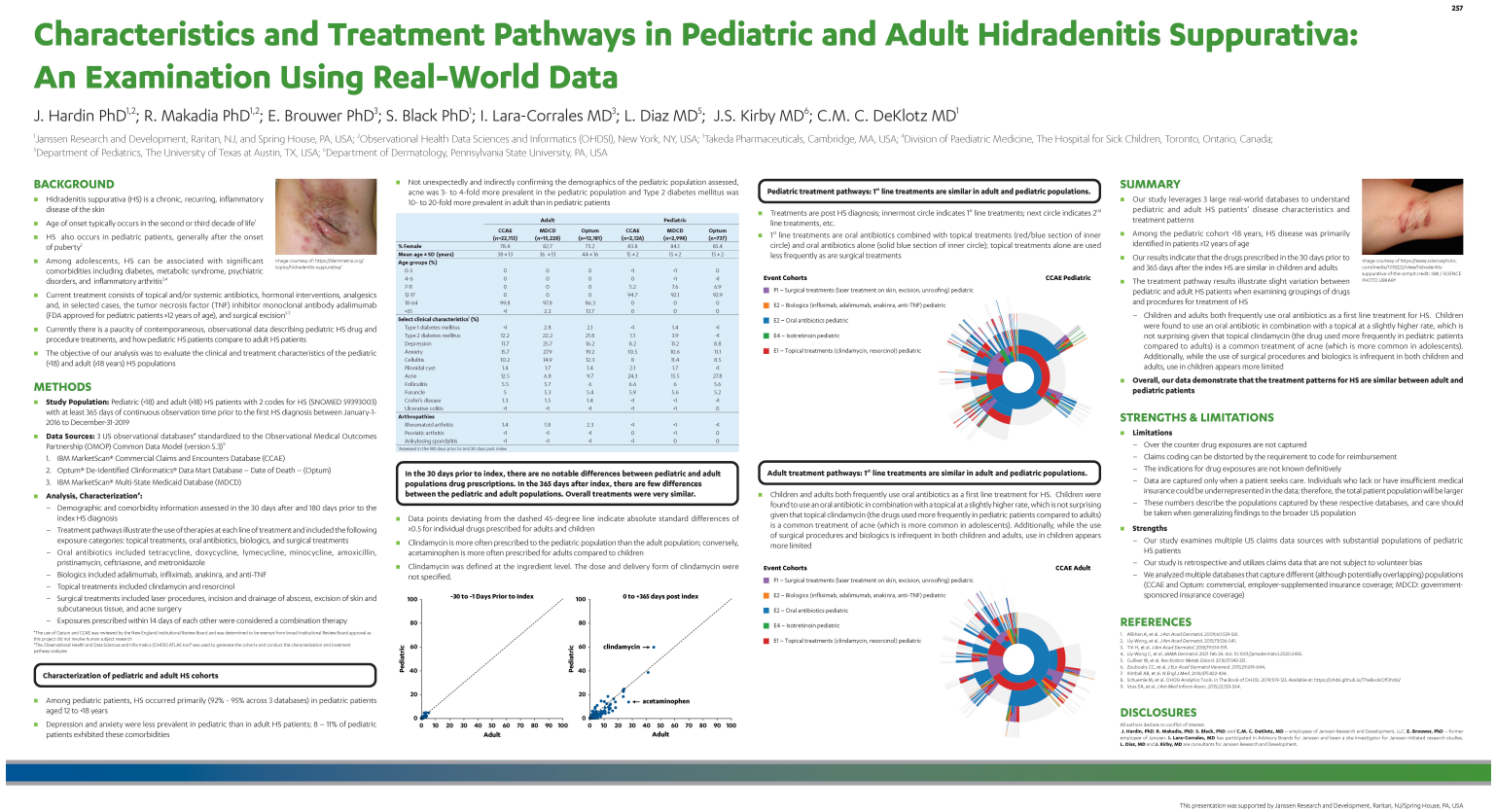
References

[1] G. Hripcsak, P. B. Ryan, J. D. Duke, N. H. Shah, R. W. Park, V. Huser, M. A. Suchard, M. J. Schuemie, F. J. DeFalco, A. Perotte, et al., "Characterizing treatment pathways at scale using the ohdsi network," Proceedings of the National Academy of Sciences, vol. 113, no. 27, pp. 7329-7336, 2016.
[2] M. A. Suchard, S. E. Simpson, I. Zorych, P. Ryan, and D. Madigan, "Massive parallelization of serial inference algorithms for a complex generalized linear model," ACM Transactions on Modeling and Computer Simulation (TOMACS), vol. 23, no. 1, p. 10, 2013.
[3] D. Merrill and M. Garland, "Single-pass parallel prefix scan with decoupled look-back," NVIDIA, Tech. Rep. NVR-2016-002, 2016.
[4] M. A. Suchard, M. J. Schuemie, H. M. Krumholz, S. C. You, R. Chen, N. Pratt, C. G. Reich, J. Duke, D. Madigan, G. Hripcsak, et al., "Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis," The Lancet, vol. 394, no. 10211, pp. 1816-1826, 2019.

WEDNESDAY GPU parallelization of massive sample-size survival analysis

Authors: Jianxiao Yang, Marc Suchard

#OHDSISocialShowcase This Week



257

Characteristics and Treatment Pathways in Pediatric and Adult Hidradenitis Suppurativa: An Examination Using Real-World Data

J. Hardin PhD[1,2]; R. Makadia PhD[1,2]; E. Brouwer PhD[3]; S. Black PhD[1]; I. Lara-Corrales MD[3]; L. Diaz MD[5]; J.S. Kirby MD[6]; C.M. C. DeKlotz MD[1]

[1]Janssen Research and Development, Raritan, NJ, and Spring House, PA, USA; [2]Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA; [3]Takeda Pharmaceuticals, Cambridge, MA, USA; [4]Division of Paediatric Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; [5]Department of Pediatrics, The University of Texas at Austin, TX, USA; [6]Department of Dermatology, Pennsylvania State University, PA, USA

BACKGROUND

- Hidradenitis suppurativa (HS) is a chronic, recurring, inflammatory disease of the skin
- Age of onset typically occurs in the second or third decade of life[1]
- HS also occurs in pediatric patients, generally after the onset of puberty[2]
- Among adolescents, HS can be associated with significant comorbidities including diabetes, metabolic syndrome, psychiatric disorders, and inflammatory arthritis[3,4]
- Current treatment consists of topical and/or systemic antibiotics, hormonal interventions, analgesics and, in selected cases, the tumor necrosis factor (TNF) inhibitor monoclonal antibody adalimumab (FDA approved for pediatric patients ≥12 years of age), and surgical excision[2,5]
- Currently there is a paucity of contemporaneous, observational data describing pediatric HS drug and procedure treatments, and how pediatric HS patients compare to adult HS patients
- The objective of our analysis was to evaluate the clinical and treatment characteristics of the pediatric (<18) and adult (≥18 years) HS populations

METHODS

- **Study Population:** Pediatric (<18) and adult (≥18) HS patients with 2 codes for HS (SNOMED 59393003) with at least 365 days of continuous observation time prior to the first HS diagnosis between January-1-2016 to December-31-2019
- **Data Sources:** 3 US observational databases* standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (version 5.3)[6]
  1. IBM MarketScan® Commercial Claims and Encounters Database (CCAE)
  2. Optum® De-Identified Clinformatics® Data Mart Database – Date of Death – (Optum)
  3. IBM MarketScan® Multi-State Medicaid Database (MDCD)
- **Analysis, Characterization*:**
  - Demographic and comorbidity information assessed in the 30 days after and 180 days prior to the index HS diagnosis
  - Treatment pathways illustrate the use of therapies at each line of treatment and included the following exposure categories: topical treatments, oral antibiotics, biologics, and surgical treatments
  - Oral antibiotics included tetracycline, doxycycline, lymecycline, minocycline, amoxicillin, pristinamycin, ceftriaxone, and metronidazole
  - Biologics included adalimumab, infliximab, anakinra, and anti-TNF
  - Topical treatments included clindamycin and resorcinol
  - Surgical treatments included laser procedures, incision and drainage of abscess, excision of skin and subcutaneous tissue, and acne surgery
  - Exposures prescribed within 14 days of each other were considered a combination therapy

*The use of Optum and CCAE was reviewed by the New England Institutional Review Board and was determined to be exempt from broad Institutional Review Board approval as this project did not involve human subject research
The Observational Health and Data Sciences and Informatics (OHDSI) ATLAS tool[7] was used to generate the cohorts and conduct the characterization and treatment pathway analyses.

Characterization of pediatric and adult HS cohorts

- Among pediatric patients, HS occurred primarily (92% - 95% across 3 databases) in pediatric patients aged 12 to <18 years
- Depression and anxiety were less prevalent in pediatric than in adult HS patients; 8 – 11% of pediatric patients exhibited these comorbidities
- Not unexpectedly and indirectly confirming the demographics of the pediatric population assessed, acne was 3- to 4-fold more prevalent in the pediatric population and Type 2 diabetes mellitus was 10- to 20-fold more prevalent in adult than in pediatric patients

	Adult			Pediatric		
	CCAE (n=22,113)	MDCD (n=15,228)	Optum (n=12,187)	CCAE (n=2,324)	MDCD (n=2,998)	Optum (n=737)
% Female	78.4	82.7	73.2	83.8	84.1	85.4
Mean age ± SD (years)	39 ± 13	36 ± 13	44 ± 16	15 ± 2	15 ± 2	15 ± 2
Age groups (%)						
0-3	0	0	0	<1	<1	0
4-6	0	0	0	0	<1	<1
7-11	0	0	0	5.2	7.6	6.9
12-17	0	0	0	94.7	92.1	92.9
18-64	99.8	97.8	86.3	0	0	0
>65	<1	2.2	13.7	0	0	0
Select clinical characteristics* (%)						
Type I diabetes mellitus	<1	2.8	2.1	<1	1.4	<1
Type 2 diabetes mellitus	12.2	22.2	25.8	1.1	3.9	<1
Depression	11.7	25.7	16.2	8.2	11.2	8.8
Anxiety	15.7	27.9	19.2	10.5	10.6	11.1
Cellulitis	10.2	14.9	12.3	8	11.4	8.5
Pilonidal cyst	1.4	1.7	1.4	2.1	1.7	<1
Acne	12.5	6.8	9.7	24.3	15.5	27.8
Folliculitis	5.5	5.7	5	6.6	6	5.6
Furuncle	5	5.3	5.4	5.9	5.6	5.2
Crohn's disease	1.3	1.5	1.4	<1	<1	<1
Ulcerative colitis	<1	<1	<1	<1	<1	0
Arthropathies						
Rheumatoid arthritis	1.4	1.8	2.3	<1	<1	<1
Psoriatic arthritis	<1	<1	<1	0	<1	0
Ankylosing spondylitis	<1	<1	<1	<1	0	0

*Assessed in the 180 days prior to and 30 days post index

In the 30 days prior to index, there are no notable differences between pediatric and adult populations drug prescriptions. In the 365 days after index, there are few differences between the pediatric and adult populations. Overall treatments were very similar.

- Data points deviating from the dashed 45-degree line indicate absolute standard differences of ±0.5 for individual drugs prescribed for adults and children
- Clindamycin is more often prescribed to the pediatric population than the adult population; conversely, acetaminophen is more often prescribed for adults compared to children
- Clindamycin was defined at the ingredient level. The dose and delivery form of clindamycin were not specified.

Pediatric treatment pathways: 1st line treatments are similar in adult and pediatric populations.

- Treatments are post HS diagnosis; innermost circle indicates 1st line treatments; next circle indicates 2nd line treatments, etc.
- 1st line treatments are oral antibiotics combined with topical treatments (red/blue section of inner circle) and oral antibiotics alone (solid blue section of inner circle); topical treatments alone are used less frequently as are surgical treatments

Adult treatment pathways: 1st line treatments are similar in adult and pediatric populations.

- Children and adults both frequently use oral antibiotics as a first line treatment for HS. Children were found to use an oral antibiotic in combination with a topical at a slightly higher rate, which is not surprising given that topical clindamycin (the drug used more frequently in pediatric patients compared to adults) is a common treatment of acne (which is more common in adolescents). Additionally, while the use of surgical procedures and biologics is infrequent in both children and adults, use in children appears more limited

SUMMARY

- Our study leverages 3 large real-world databases to understand pediatric and adult HS patients' disease characteristics and treatment patterns
- Among the pediatric cohort <18 years, HS disease was primarily identified in patients ≥12 years of age
- Our results indicate that the drugs prescribed in the 30 days prior to and 365 days after the index HS are similar in children and adults
- The treatment pathway results illustrate slight variation between pediatric and adult HS patients when examining groupings of drugs and procedures for treatment of HS
  - Children and adults both frequently use oral antibiotics as a first line treatment for HS. Children were found to use an oral antibiotic in combination with a topical at a slightly higher rate, which is not surprising given that topical clindamycin (the drug used more frequently in pediatric patients compared to adults) is a common treatment of acne (which is more common in adolescents). Additionally, while the use of surgical procedures and biologics is infrequent in both children and adults, use in children appears more limited
- **Overall, our data demonstrate that the treatment patterns for HS are similar between adult and pediatric patients**

STRENGTHS & LIMITATIONS

- **Limitations**
  - Over the counter drug exposures are not captured
  - Claims coding can be distorted by the requirement to code for reimbursement
  - The indications for drug exposures are not known definitively
  - Data are captured only when a patient seeks care. Individuals who lack or have insufficient medical insurance could be underrepresented in the data; therefore, the total patient population will be larger
  - These numbers describe the populations captured by these respective databases, and care should be taken when generalizing findings to the broader US population
- **Strengths**
  - Our study examines multiple US claims data sources with substantial populations of pediatric HS patients
  - Our study is retrospective and utilizes claims data that are not subject to volunteer bias
  - We analyzed multiple databases that capture different (although potentially overlapping) populations (CCAE and Optum: commercial, employer-supplemented insurance coverage; MDCD: government-sponsored insurance coverage)

REFERENCES

1. Alikhan A, et al. J Am Acad Dermatol. 2009;61(3):539-551.
2. Liy-Wong, et al. J Am Acad Dermatol. 2015;73:556-541.
3. Tzh H, et al. J Am Acad Dermatol. 2018;79:916-919.
4. Liy-Wong C, et al. JAMA Dermatol. 2021 Feb 24. doi: 10.1001/jamadermatol.2020.5416.
5. Gulliver W, et al. Rev Endocr Metab Disord. 2016;17:343-351.
6. Zouboulis CC, et al. J Eur Acad Dermatol Venereol. 2015;29:619-644.
7. Kimball AB, et al. N Engl J Med. 2016;375:422-434.
8. Schuemie M, et al. OHDSI Analytics Tools. In The Book of OHDSI. 2019;109-123. Available at: https://ohdsi.github.io/TheBookOfOhdsi/
9. Voss EA, et al. J Am Med Inform Assoc. 2015;23:553-564.

DISCLOSURES

All authors declare no conflict of interest.
J. Hardin, PhD; R. Makadia, PhD; S. Black, PhD; and **C.M. C. DeKlotz, MD** – employees of Janssen Research and Development, LLC. **E. Brouwer, PhD** – former employee of Janssen. **I. Lara-Corrales, MD** has participated in Advisory Boards for Janssen and been a site Investigator for Janssen initiated research studies. **L. Diaz, MD** and **J.S. Kirby, MD** are consultants for Janssen Research and Development.

This presentation was supported by Janssen Research and Development, Raritan, NJ/Spring House, PA, USA

THURSDAY

Examination of Characteristics and Treatments in Pediatric and Adult Hidradenitis Suppurativa

Authors: Jill Hardin, Rupa Makadia, Emily Brouwer, Shawn Black, Irene Lara-Corrales, Lucia Z Diaz, Joslyn Sciacca Kirby, Cynthia Marie Carver DeKlotz

#OHDSISocialShowcase This Week

Real-World Evaluation of Systematic Bias and Balance of Overall Patient Characteristics of Propensity Score Matching Versus Cardinality Matching

Stephen P Fortin[1], Martijn J Schuemie[1]
[1]Janssen R&D, LLC, Raritan, NJ, USA

Background

- Propensity score matching (PSM) is subject to limitations, especially in studies of small sample size:
  1. Susceptible to substantial bias due to limited overlap in covariate distributions
  2. Potential model overparameterization due to limited degrees of freedom
- Cardinality matching (CM) uses integer programming to find the largest matched sample meeting a set of prespecified balance criteria.
- CM overcomes the limitations of PSM by matching directly on the marginal distribution of covariates
- Prior research has shown large-scale CM achieves superior patient retention and comparable systematic bias as compared to large-scale PSM; however, large-scale methods may not be applicable in the setting of small sample sizes.

Study Objectives: To compare the performance of PSM and CM in the context of a study of new users of new users of angiotensin-converting enzyme inhibitor (ACEI) and β-blocker monotherapy at small sample sizes

Methods

Study Design: Comparative new user cohort study
Data Source: Data were from the IBM® MarketScan® Commercial Claims and Encounters database
Study Population: New users of ACEI and β-blocker monotherapy between 10-01-2014 to 01-01-2017 with a history of hypertension (index = first drug exposure)
Covariates
- **Matching covariates** – covariates included in the PS model and CM – included patient demographics (i.e., age, sex, race, ethnicity, year) and clinical characteristics (i.e., comorbidities comprising the Charlson Comorbidity Index; and the Hospital-Frailty Risk Score)
- **Observed covariates** included patient demographics, and all conditions, drug exposures and other health-service-use-behaviors observed 30 and 365 days prior to index

Statistical Analysis
- PSM was conducted through greedy matching (1:1 match, caliper=0.15)
- CM performed through 1:1 matching with the following prespecified balance criteria: max SMD=0.00, max SMD=0.01, max SMD=0.05 and max SMD=0.10

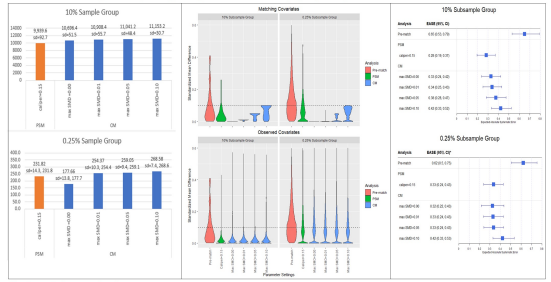
Subsample Groups
- Developed 10% and 0.25% sample groups consisting of 5 and 200 subsample draws, respectively
  - Subsample draws randomly sampled from study population without replacement

Evaluation of CM and PSM
- Post-match sample size
  - Average sample size across all subsample draws within each sample group
- Post-match matching covariate and observed covariate balance
  - Evaluated using SMDs where an absolute SMD ≤0.10 was considered balanced
  - Assessed post-match SMDs of all matching covariates and observed covariates across all subsample draws within each sample group
- Post-match residual bias
  - Performed a total of 105 negative control outcome experiments for each sample group
  - Due to the low frequency of negative control outcomes, negative control outcome experiments were conducted across a pooled sample consisting of matched patients identified across all subsample draws for each sample group
  - Assessed using the expected absolute systematic error (EASE) of the empirical null distribution of negative control outcome experiments

Results

Pre-match:
- A total of 186,233 (β-blocker: 56,871; ACEI: 129,362) patients met the study criteria
- 18,576 (β-blocker: 5,675; ACEI: 12,901) and 465 (β-blocker: 142; ACEI: 323) patients were included in each subsample draw of the 10% and 0.25% sample groups, respectively
- Average 35,458 and 8,566 observed covariates in the 10% and 0.25% sample groups, respectively

Figures 1-3. Average post-match sample size (left panel); post-match matching covariate and observed covariate balance (middle panel); and post-match EASE (right panel)



Post-match sample size:
- As shown in **Figure 1**, CM was associated with increased average post-match sample size except for analyses in the 0.25% sample group with a tightest balance criterion of (max SMD=0.00)

Post-match covariate balance:
- CM achieved balance on all matching covariates; PSM failed to achieve balance in both sample groups
- In the 10% sample group, PSM achieved improved observed covariate balance as compared to CM
- In the 0.25% sample group, as compared to PSM, observed covariate balance was improved with CM at tighter balance criteria and similar at looser balance criteria

Post-match residual confounding:
- As compared to CM, PSM was associated with improved EASE in the 10% sample group and similar EASE in the 0.25% sample group
- CM achieved improved EASE with tighter balance criteria

Conclusions

CM found the largest matched sample meeting a set of prespecified balance criteria. At smaller sample sizes, PSM and CM achieved comparable balance in overall patient characteristics and reductions in systematic bias albeit CM had improved performance at more stringent prespecified balance criteria (i.e., SMD <0.05). Improved indirect covariate balance and reductions in EASE were observed with PSM at larger sample sizes as compared to CM. We recommend CM as an alternative to PSM in studies of small sample size.

CONTACT INFORMATION
sfortin1@its.jnj.com

FRIDAY

Real-World Evaluation of Systematic Bias and Balance of Overall Patient Characteristics of Propensity Score Matching Versus Cardinality Matching

Authors: Stephen P Fortin, Martijn J Schuemie

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. 5 Community Call: Workgroup Updates

Clinical Trials

Mike Hamidi

Health Equity

Jake Gillberg

Phenotype Development & Evaluation

Gowtham Rao

Vaccine Vocabulary

Adam Black