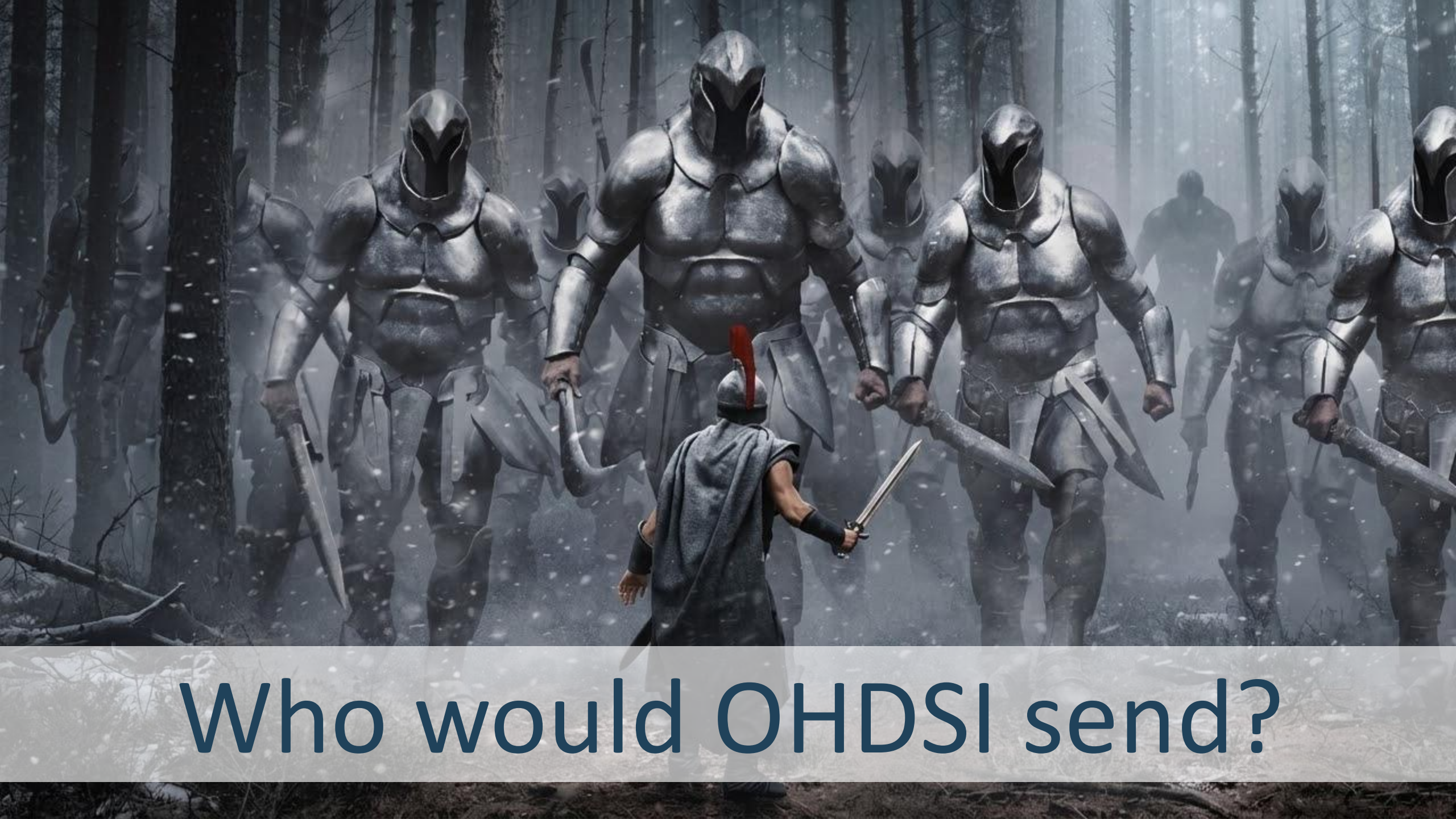


Beyond the Defaults

How the OHDSI Community is Adapting, Extending, and Reimagining Its Tools







Who would OHDSI send?


Matt Damon's Biceps in That 'Odyssey' Fight Scene? They're Hers.
Devyn Dalton



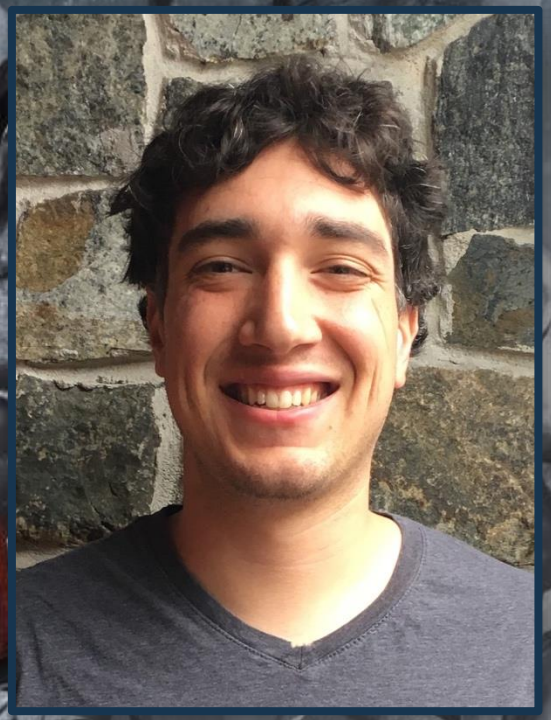


How do we get there?
We need to evolve how we collaborate

- Share what you've learned after you've done it
  - → share what you're doing while you're doing it
  - → share what you'd like to do before you do it
- Do it yourself
  - → Have courage to ask for help
- Caring to help others when asked
  - → Proactivity to seek opportunities to contribute
- Be a node in the OHDSI network
  - → Be the edge that connects the nodes
  - → Become the knowledge graph of all nodes and edges





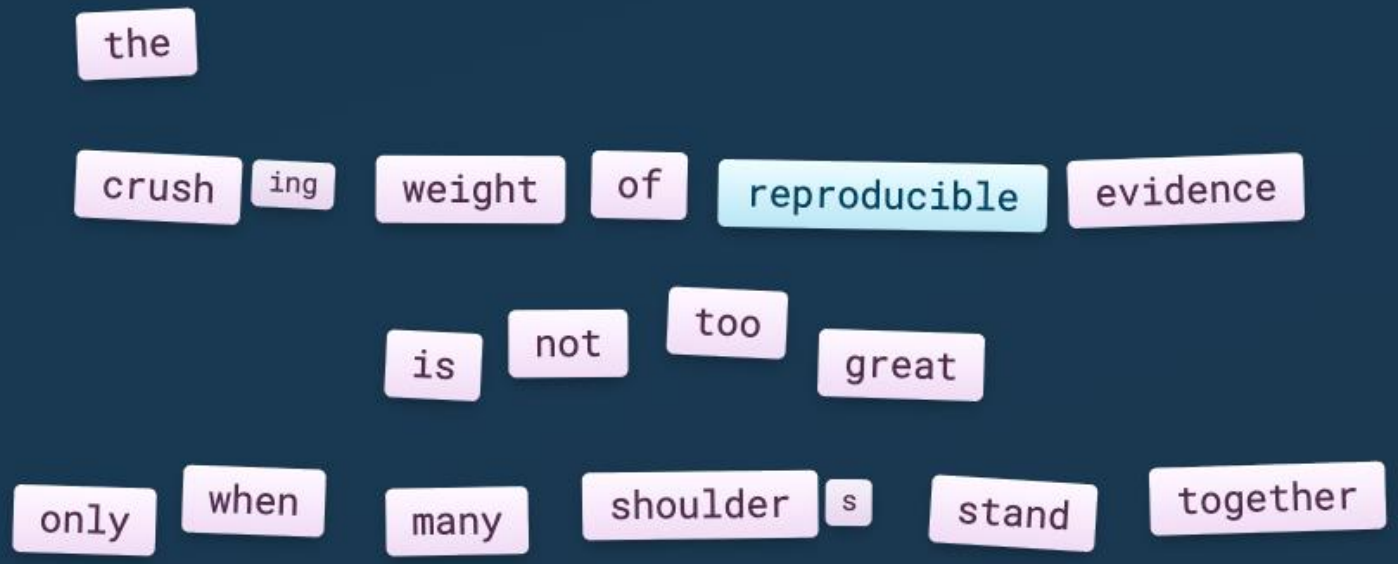


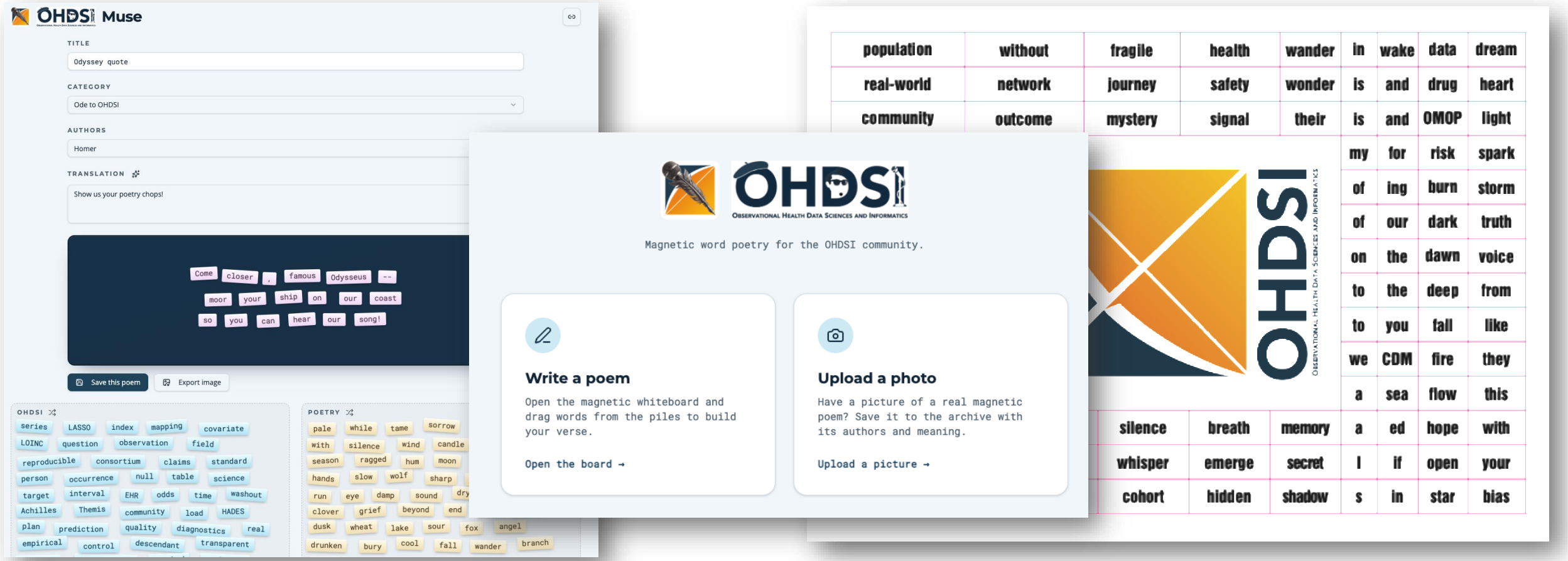

and… YOU

Who would OHDSI send?



Ode to the OHDSI Community





ohdsimuse.duvall4data.online

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



William J. O’Brien

Johnson & Johnson



Customizing Andromeda to Generate Large Scale Evidence Faster

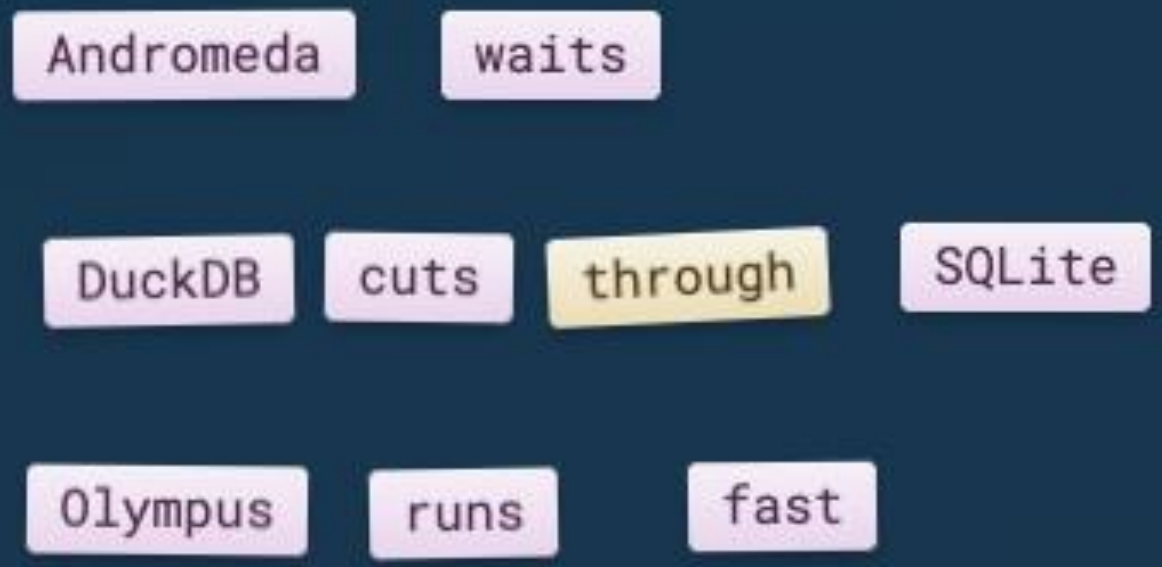
William J. O'Brien

Johnson&Johnson

Disclosures

- Current Johnson & Johnson employee
- Work was performed as VA employee

Ode to Andromeda



The challenge: sequential target trial emulation

TARGET
COVID-19 vaccination

vs

COMPARATOR
No vaccination (non-users)

→

OUTCOME
Risk of COVID-related outcomes

Repeated across sequential enrollment periods

Trial 1	Trial 2	Trial 3	Trial n
Month 1 eligible cohort	Month 2 eligible cohort	Month 3 eligible cohort	…continues monthly

Each trial requires a new propensity score model — ~14 hours each, so iteration was slow

First, we chased the wrong culprits

HYPOTHESIS 1

Was it the network?

- Worked with IT to map where storage sat relative to the R environment
- Were we pushing gigabytes across a wide area network?
- Storage layout and I/O was not the bottleneck

HYPOTHESIS 2

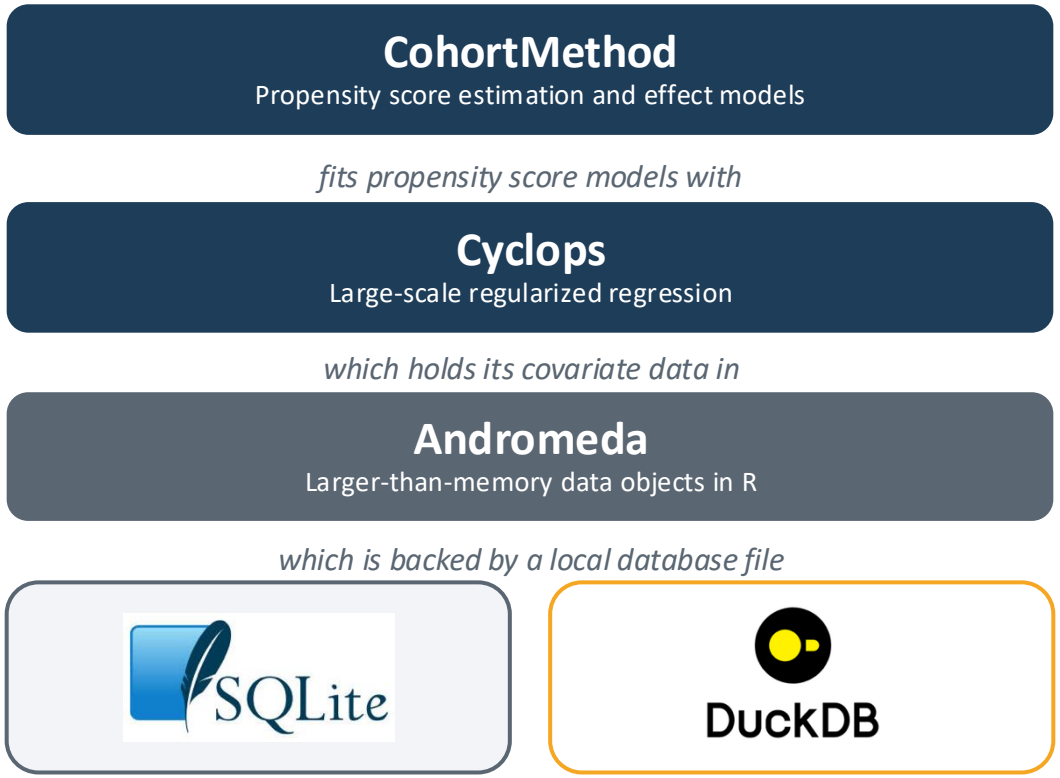
Was it SQLite tuning?

- Tuned every SQLite PRAGMA setting we could find
- Cache size, journal mode, synchronous writes, temp store
- No meaningful change in benchmarks

The bottleneck was not IT infrastructure or database configuration

Ruling these out is what pointed us toward changing the backend

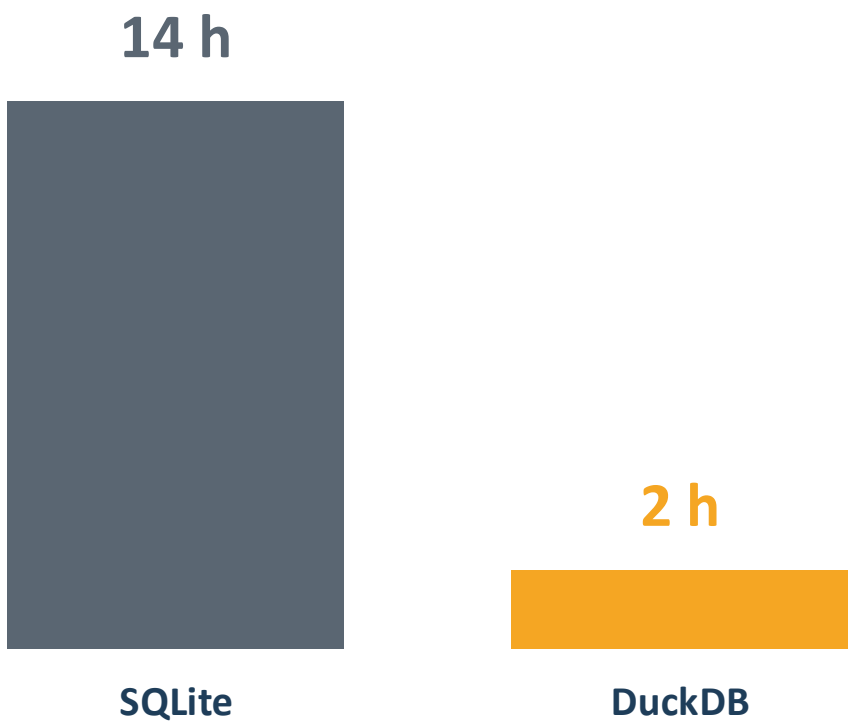
What next? A modern data science database as the back end



Same analytic workflow, only the storage engine underneath changes

- PS estimation scans covariate columns across millions of rows - DuckDB's columnar, vectorized engine is built for exactly that, while SQLite's row store must read every row in full
- DuckDB also runs queries in parallel across cores and hands data to R in bulk, so the repeated extract-and-aggregate steps behind each model shrink

Benchmark result: ~14 hours to ~2 hours per PS model



Representative VA use case

- Benchmarked in a representative VA use case
- PS model runtime fell from ~14 hours to ~2 hours
- Made the sequential target trial emulation feasible
- Broadly relevant to large-scale evidence generation

From local pain point to shared community capability

Local pain point

14-hour PS models prevented us from running our study

Community troubleshooting

We encountered some R+DuckDB inter-op bugs; the community (OHDSI *and* DuckDB developers) helped diagnose and fix them

Shared capability

The improvement was released in Andromeda v1.0

OHDSI contributions often start as local pain points, then become shared community capabilities.



Thank you

With thanks to the colleagues who made this work possible

Scott DuVall

Benjamin Viernes

Adam Black

Marc Suchard

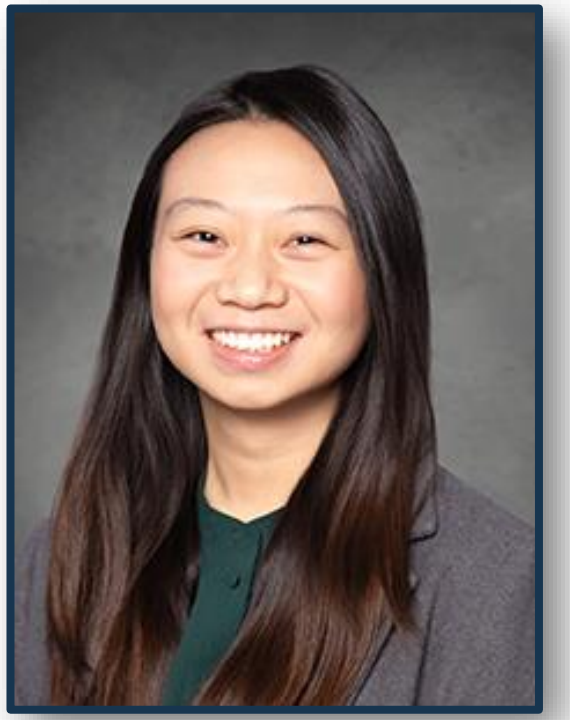
Martijn Schuemie

Anthony Sena

wobrien8@its.jnj.com

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



Hsin Yi Chen

Columbia University

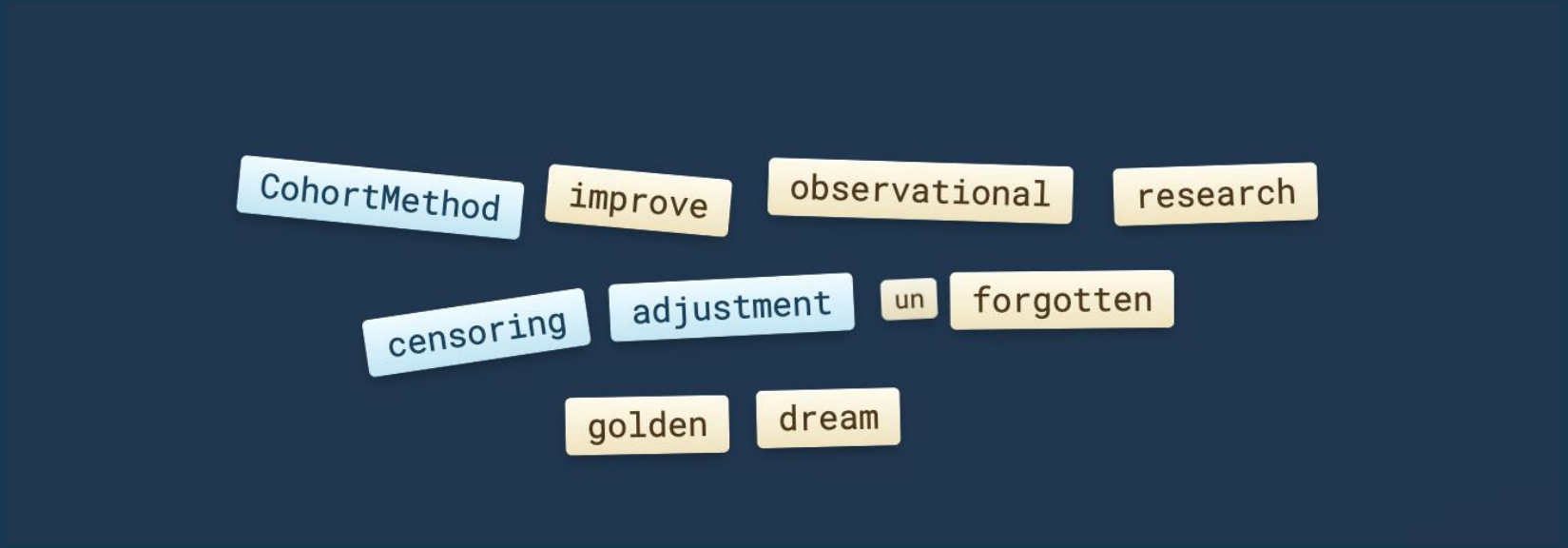


Incorporating Inverse Probability of Censoring Weighting (IPCW) into the OHDSI CohortMethod Workflow

Hsin Yi Chen, BS

Coauthors: Tara V. Anand, PhD, Linying Zhang, PhD, George Hripcsak, MD, MS

Disclaimer: I love CohortMethod!



… but sometimes there are methods we want to use that aren't implemented yet

A brief intro to CohortMethod (and Strategus)

- R package within HADES ecosystem for **new-user cohort studies** for OMOP data
- Framework behind many previous OHDSI network studies!



Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick B Ryan

JAMA Ophthalmology | Original Investigation

Semaglutide and Nonarteritic Anterior Ischemic Optic Neuropathy

Cindy X. Cai, MD, MS; Michelle Hribar, PhD; Sally Baxter, MD, MSc; Kerry Goetz, MS; Swarup S. Swaminathan, MD; Alexis Flowers, MD; Eric N. Brown, MD, PhD; Brian Toy, MD; Benjamin Xu, MD, PhD; John Chen, MD, PhD; Aiyin Chen, MD; Sophia Wang, MD, MS; Cecilia Lee, MD, MS; Theodore Leng, MD, MS; Joshua R. Ehrlich, MD, MPH; Andrew Barkmeier, MD; Karen R. Armbrust, MD, PhD; Michael V. Boland, MD, PhD; David Dorr, MD, MS; Danielle Boyce, MPH, DPA; Thamir Alshammari, PhD; Joel Swerdel, PhD, MS, MPH; Marc A. Suchard, MD, PhD; Martijn Schuemie, PhD; Fan Bu, PhD; Anthony G. Sena, BA; George Hripcsak, MD, MS; Akihiko Nishimura, PhD; Paul Nagy, PhD; Thomas Falconer, MS; Scott L. DuVall, PhD; Michael Matheny, MD; Benjamin Viernes, PhD; William O'Brien, MS; Linying Zhang, PhD; Benjamin Martin, PhD; Erik Westlund, PhD; Nestoras Mathioudakis, MD, MHS; Ruochong Fan, MA; Adam Wilcox, PhD; Albert Lai, PhD; Jacqueline C. Stocking, PhD, RN; Sahar Takkouche, MD, MBA; Lok Hin Lee, DPhil; Yangyiran Xie, BS; Izabelle Humes, PT, DPT; David B. McCoy, BA; Mohammad Adibuzzaman, PhD; Raymond G. Areaux Jr, MD; William Rojas-Carabali, MD; James Brash, PhD; David A. Lee, MD, MS; Nicole G. Weiskopf, PhD; Louise Mawn, MD; Rupesh Agrawal, MD; Hannah Morgan-Cooper, Msc; Priya Desai, Msc; Patrick B. Ryan, PhD

Comparative Effectiveness of Second-Line Antihyperglycemic Agents for Cardiovascular Outcomes

A Multinational, Federated Analysis of LEGEND-T2DM



Rohan Khera, MD, MS,[a,b,c] Arya Aminorroaya, MD, MPH,[a] Lovedeep Singh Dhingra, MBBS,[a] Phyllis M. Thangaraj, MD, PhD,[a] Aline Pedroso Camargos, PhD,[a] Fan Bu, PhD,[d] Xiyu Ding, MS,[e] Akihiko Nishimura, PhD,[e] Tara V. Anand, BS,[f] Faaizah Arshad, BS,[g] Clair Blacketer, MPH,[h] Yi Chai, PhD,[i] Shounak Chattopadhyay, PhD,[g] Michael Cook, BSc,[e] David A. Dorr, MD, MS,[j] Talita Duarte-Salles, PhD,[k,l] Scott L. DuVall, PhD,[m,n] Thomas Falconer, MS,[f] Tina E. French, RN, CPHQ,[o,p] Elizabeth E. Hanchrow, RN, MSN,[o,p] Guneet Kaur, MS,[q] Wallis C.Y. Lau, BSc, PhD,[r,s,t,u] Jing Li, MS,[v] Kelly Li, BS,[g] Yuntian Liu, MPH,[a,b] Yuan Lu, ScD,[a] Kenneth K.C. Man, BSc, MPH, PhD,[r,s,t,u] Michael E. Matheny, MD, MS, MPH,[o,p] Nestoras Mathioudakis, MD, MHS,[w] Jody-Ann McLeggon, MPH,[f] Michael F. McLemore, RN,[o,p] Evan Minty, MD, MSc,[x] Daniel R. Morales, MD,[q] Paul Nagy, PhD,[w] Anna Ostropolets, MD, PhD,[h] Andrea Pistillo, MSc,[k] Thanh-Phuc Phan, MBA,[y] Nicole Pratt, PhD,[z] Carlen Reyes, MD, PhD,[k] Lauren Richter, MD,[f] Joseph S. Ross, MD, MHS,[b,aa] Elise Ruan, MD,[f] Sarah L. Seager, BS,[bb] Katherine R. Simon, AA,[o,p] Benjamin Viernes, PhD,[m,n] Jianxiao Yang, MS,[cc] Can Yin, MS,[dd] Seng Chan You, MD, PhD,[ee,ff] Jin J. Zhou, PhD,[g,gg] Patrick B. Ryan, PhD,[f] Martijn J. Schuemie, PhD,[hh] Harlan M. Krumholz, MD, SM,[a,b,ii] George Hripcsak, MD, MS,[f] Marc A. Suchard, MD, PhD[g,m,jj,kk]

Challenge: addressing informative censoring

JACC
PUBLISHED BY ELSEVIER ON BEHALF OF THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION. THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY-NC-ND LICENSE (http://creativecommons.org/licenses/by-nc-nd/4.0/).

VOL. 87, NO. 21, 2026

Comparative Cardiovascular Effectiveness of Glucagon-Like Peptide 1 Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors in Diabetes Mellitus

Fan Bu, PhD,[a] Ruopeng Wu, BS,[b] Anna Ostropolets, MD, PhD,[c] Arya Aminorroaya, MD, MPH,[d,e] Hsin Yi Chen, BS,[f] Yi Chai, PhD,[g,h] Lovedeep Singh Dhingra, MBBS, MHS,[d,e] Thomas Falconer, MS, MS,[f] Jason C. Hsu, PhD,[i] Chungsoo Kim, PharmD, PhD,[d,j] Wallis C.Y. Lau, PhD,[g,k,l,m] Kenneth K.C. Man, PhD,[g,k,l,m] Evan Minty, MD,[n] Daniel R. Morales, PhD,[o] Akihiko Nishimura, PhD,[p] Phyllis Thangraraj, MD, PhD,[d,e] Mui Van Zandt, BS,[q] Can Yin, MS,[r] Rohan Khera, MD, MS,[d,e,j,s] George Hripcsak, MD, MS,[f,t] Marc A. Suchard, MD, PhD[t,u,v,w]

Challenge: addressing informative censoring

- If censoring reason (ex: dropping out of the study, treatment discontinuation, etc.) is related to the outcome (*informative censoring)*, that can violate key assumptions of survival analysis models
- One possible solution: Inverse Probability of Censoring (IPCW)

Challenge: addressing informative censoring

We recognize and really appreciate the importance of pre-specification. However, as noted by the statistical editor, we think that conducting IPCW-based analyses would be very important to improving the robustness of the paper. We think that this approach could be presented as a post-hoc sensitivity analyses.

- Reviewers really wanted IPCW!

Current CohortMethod Workflow

CohortGenerator and CohortMethod

1 **Generate cohorts**

Target, comparator and outcome cohorts

2 **Propensity score model**

Fit Large-scale PS model to get inverse-probability-of-treatment **(IPTW)** weights:

$$w_i^{Ps} = 1/\pi_i \text{ or } 1/(1-\pi_i)$$

where π_i is the probability of receiving treatment

3 **Outcome Model**

Fit outcome model with weights w, i.e.

$$h(t|A) = h_0(t) \cdot exp(\beta A)$$

$$HR = exp(\beta)$$

Currently implemented in Strategus!

Current CohortMethod Workflow

CohortGenerator and CohortMethod

1 **Generate cohorts**

Target, comparator and outcome cohorts

2 **Propensity score model**

Fit Large-scale PS model to get inverse-probability-of-treatment **(IPTW)** weights:

$$w_i^{Ps} = 1/\pi_i \text{ or } 1/(1-\pi_i)$$

where π_i is the probability of receiving treatment

3 **Outcome Model**

Fit outcome model with weights w, i.e.

$$h(t|A) = h_0(t) \cdot exp(\beta A)$$

$$HR = exp(\beta)$$



add fitting censoring model for censoring weights here

Proposed IPCW Implementation

Cohort Generation & PS
CohortGenerator and CohortMethod

1 Generate cohorts
Target, comparator and outcome cohorts

2 Propensity score model
Fit Large-scale PS model to get inverse-probability-of-treatment **(IPTW)** weights:

$$w_i^{Ps} = 1/\pi_i \text{ or } 1/(1-\pi_i)$$

where π_i is the probability of receiving treatment

Proposed IPCW Steps
FeatureExtraction · Cyclops

3 Censoring Cox Model
Fit large-scale penalized Cox model to get the inverse-probability of remaining uncensored **(IPCW)** weights:

$$w_i^c = 1 / G_i(T_i)$$

Where $G_i(T_i)$ is the probability of remaining uncensored at time t

4 Combine Weights
$$w_i(t) = w_i^{Ps} * w_i^c(t)$$

Outcome Modeling
CohortGenerator and CohortMethod

5 Outcome Model
Fit outcome model with weights w, i.e.

$$h(t|A) = h_0(t) \cdot exp(\beta A)$$

$$HR = exp(\beta)$$

Code: **github.com/ohdsi-studies/ExploringCensoringWeights**

We responded to the reviewer's request using OHDSI tools ☺

Supplementary Table 99: Comparing analytic results from four variant methods with IPCW (for censoring) and/or IPTW (for confounding). We summarize hazard ratio estimates and 95% confidence intervals for 3-point MACE and 4-point MACE between new-users of semaglutide and empagliflozin, across three databases, CCAE, Clinformatics and OptumEHR. Cox = Cox proportional hazards model; IPCW = inverse probability censoring weights; IPTW = inverse probability treatment weights.

	CCAE	Clinformatics	OptumEHR
3-point MACE			
Unadjusted Cox	0.87 (0.67, 1.14)	0.69 (0.56, 0.84)	0.67 (0.48, 0.93)
Cox with IPCW	0.84 (0.63, 1.11)	0.65 (0.53, 0.81)	0.66 (0.47, 0.92)
Cox with IPTW	0.98 (0.68, 1.40)	1.05 (0.82, 1.34)	1.06 (0.73, 1.55)
Cox with IPCW + IPTW	0.91 (0.62, 1.34)	0.99 (0.75, 1.29)	1.07 (0.72, 1.57)
4-point MACE			
Unadjusted Cox	0.76 (0.62, 0.94)	0.74 (0.64, 0.84)	0.60 (0.48, 0.76)
Cox with IPCW	0.75 (0.61, 0.93)	0.70 (0.61, 0.81)	0.60 (0.47, 0.76)
Cox with IPTW	0.88 (0.67, 1.15)	1.06 (0.89, 1.25)	0.90 (0.68, 1.19)
Cox with IPCW + IPTW	0.87 (0.65, 1.17)	1.01 (0.84, 1.22)	0.88 (0.66, 1.18)

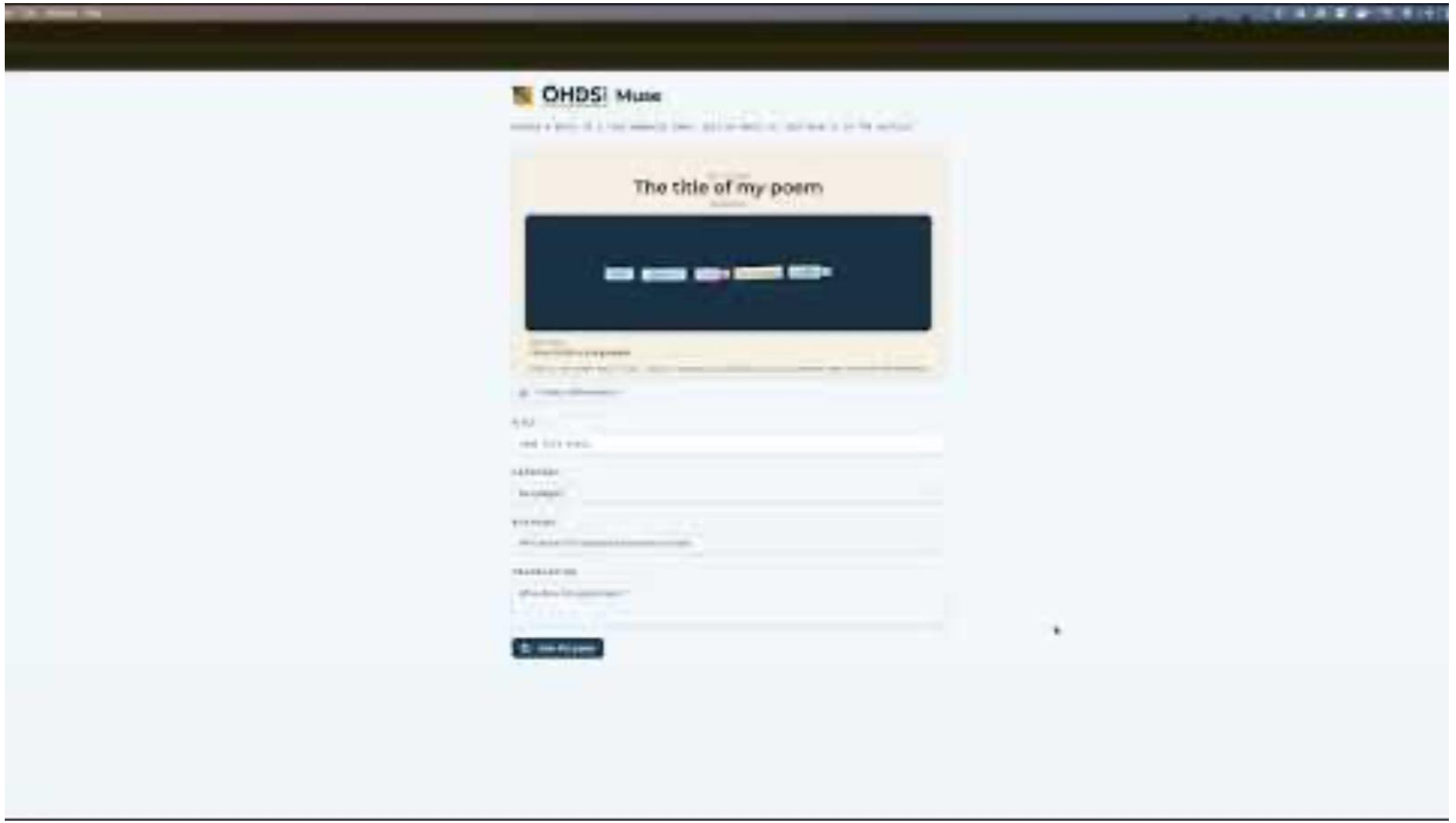
Existing OHDSI Infrastructure allows for customization

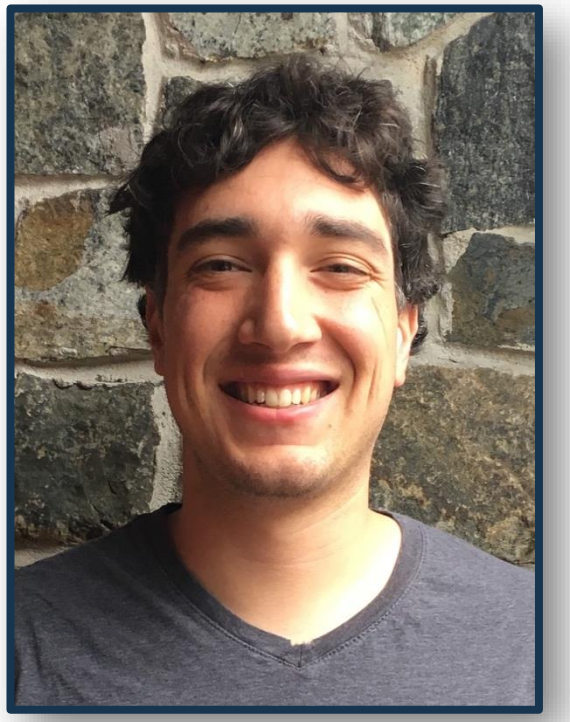
- Support packages within the HADES suite (Cyclops, FeatureExtraction, etc) makes customization of OHDSI workflows possible!

Come visit me at Poster #51!

hc3292@cumc.columbia.edu

ohdsimuse.duvall4data.online





Martin Lavallee

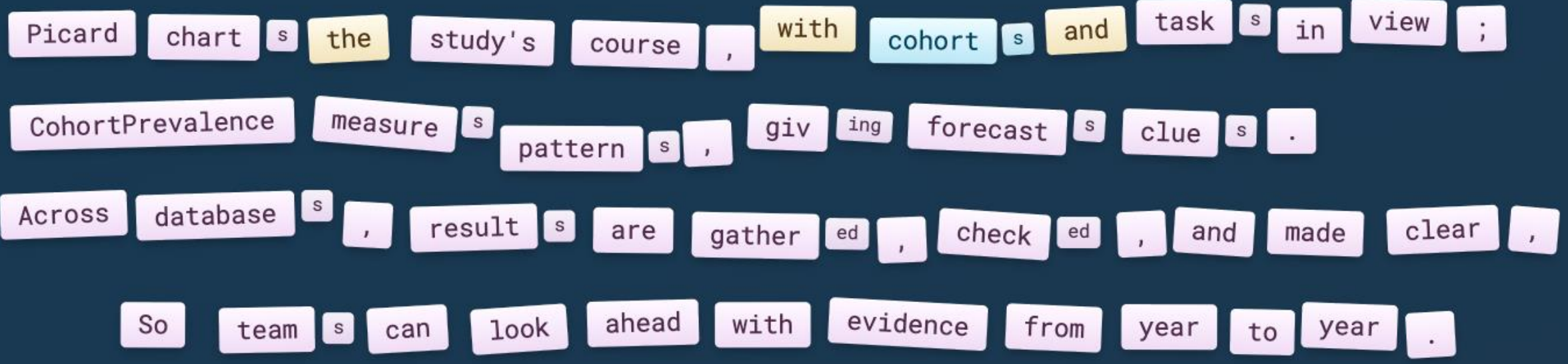
Boehringer Ingelheim



Beyond the Defaults Panel

Boehringer Ingelheim OHDSI Journey

Martin Lavallee


Picard chart s the study's course , with cohort s and task s in view ;
CohortPrevalence measure s pattern s , giv ing forecast s clue s .
Across database s , result s are gather ed , check ed , and made clear ,
So team s can look ahead with evidence from year to year .

What does the organization need?

Forecasting

- What is the annual prevalence of the indication?
- How does the population funnel from diagnosed to treated to taking a drug of interest?
- Among treated patients, which drugs are used and in what proportions?
- How can age/sex-standardized prevalence estimate the population nationally and support a rough projection into future years?

Scaling

- Can we refresh this evidence every year efficiently?
- Can we produce evidence across multiple data sources?
- How do we interpret variation of results?

Why use OMOP and OHDSI?

Quality: A standardized conversion to the OMOP CDM is an inherent data quality assessment

Reuse: a common structure allows us to write code once and use many times across data and indications

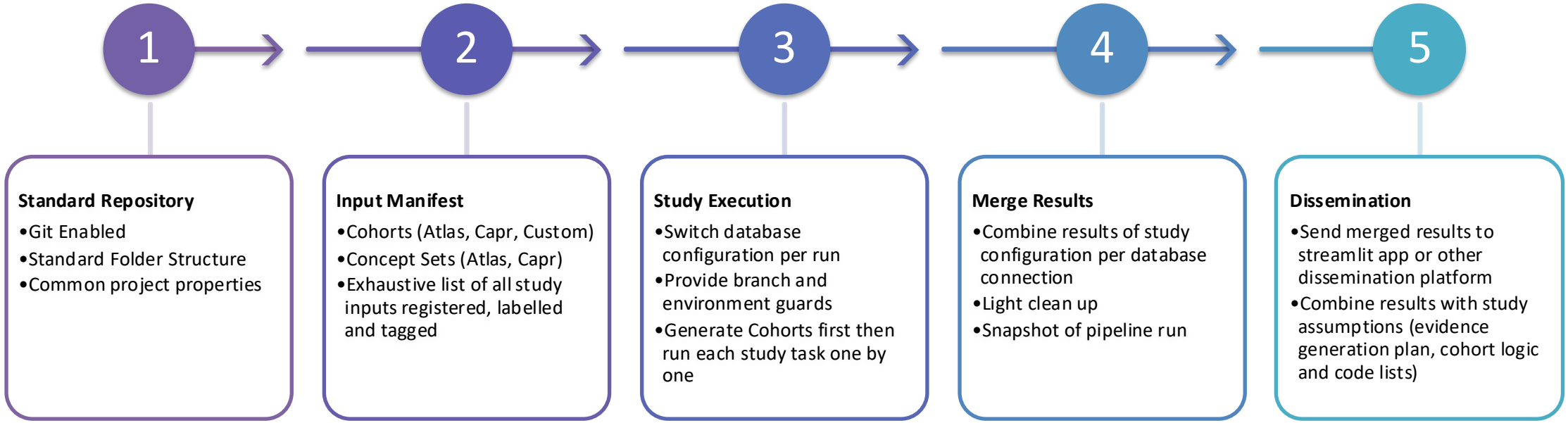
Ownership: open-source code lets us own, maintain and evolve our pipelines independently

Trust: transparent and standardized code makes it easier to scale, reproduce and explain how evidence was produced.

Gaps

Gap	Desire	Extension
Standard project scaffold and lifecycle conventions	Make repositories familiar, git-enabled and consistently managed across studies so not dependent on single analyst	Picard – standard repo structure
Manifest to organize Concept Sets and Cohorts	Ensure the totality of inputs are registered and described for cohort generation	Picard – manifest tool
Prevalence Calculator	Parameterized prevalence calculation to compare design choices using a consistent method	CohortPrevalence
Drug Utilization and Funnels	Standard method to calculate treated prevalence and understand the proportion of drug classes	CohortPrevalence
Dissemination Modality	Provide stakeholders a dashboard to review evidence	Streamlit Dashboard

Picard



Explore the *Picard* Package!

Lessons from the Journey

Adoption is possible but not always plug-and-play — It takes effort to understand and effectively operationalize OMOP and OHDSI to meet the needs of your organization.

Assess before you build — Identify what fits, what can be adapted, and what gaps remain for your organization.

Open-source what you extend — Keep analytical logic visible, inspectable, and available beyond a black box. Don’t be afraid of open-source

Shared Terminology — Simplify OMOP technical language with stakeholders and focus on common terms and methods to produce evidence.

Our OHDSI Design Contract

1. **DatabaseConnector**: the R package allows for flexible connectivity to databases using JDBC.
2. **SqlRender**: The R package that renders and translates sql to match the dialect and database parameters for any dbms enabling reproducible sql.
3. **CIRCE**: the underlying cohort builder in OHDSI that ensures that cohort definitions are always consistent and reproducible via standard sql.

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



Hanieh Razzaghi

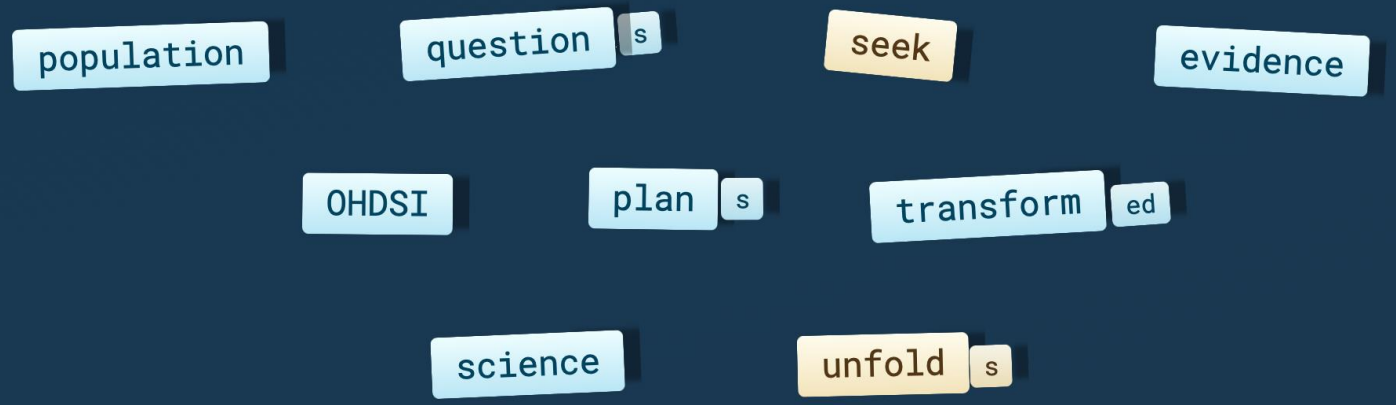
Children's Hospital of Philadelphia
University of Pennsylvania



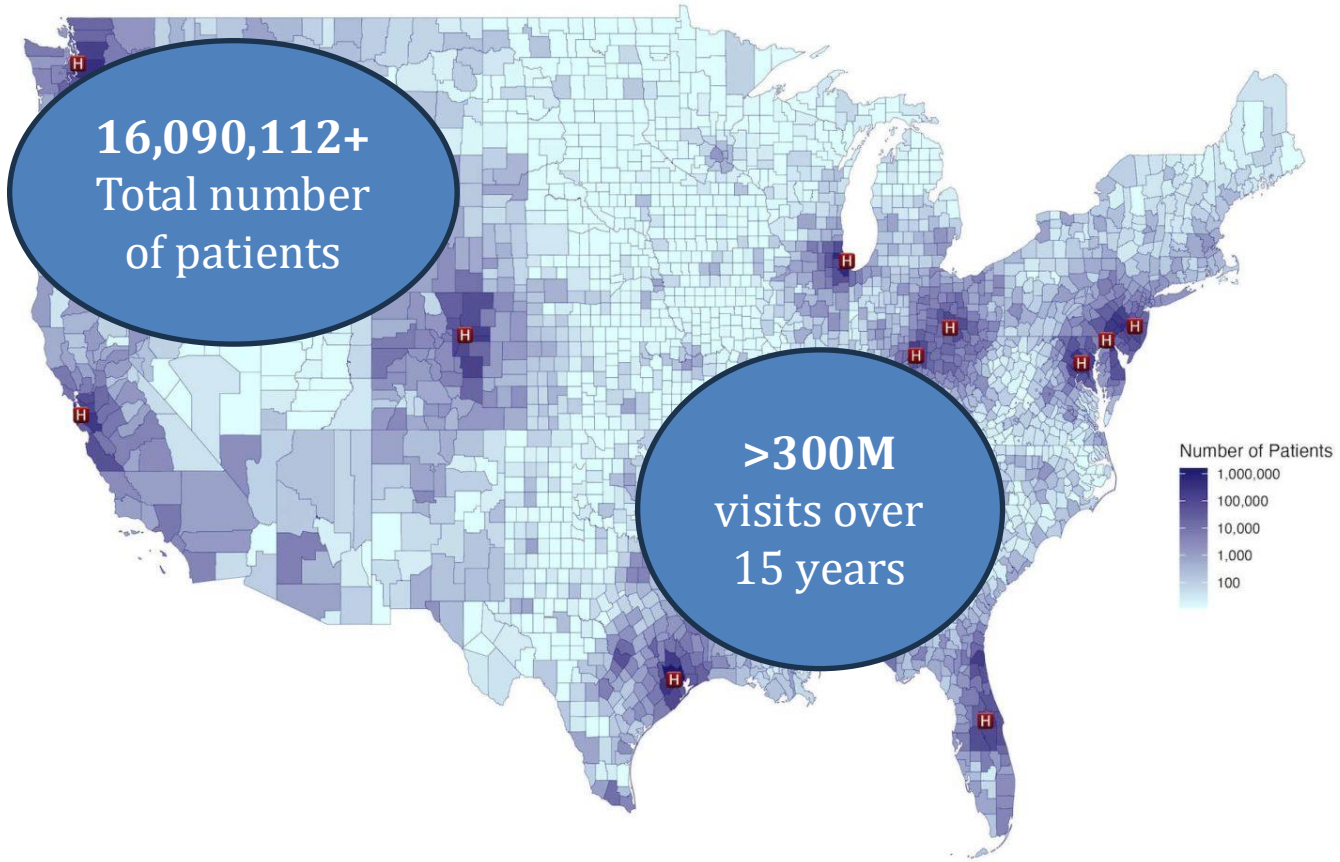
Adapting, Extending, and Reimagining the OHDSI Ecosystem for Pediatric Research: The PEDSnet Experience

Hanieh Razzaghi, PhD, MPH

Ode to OHDSI

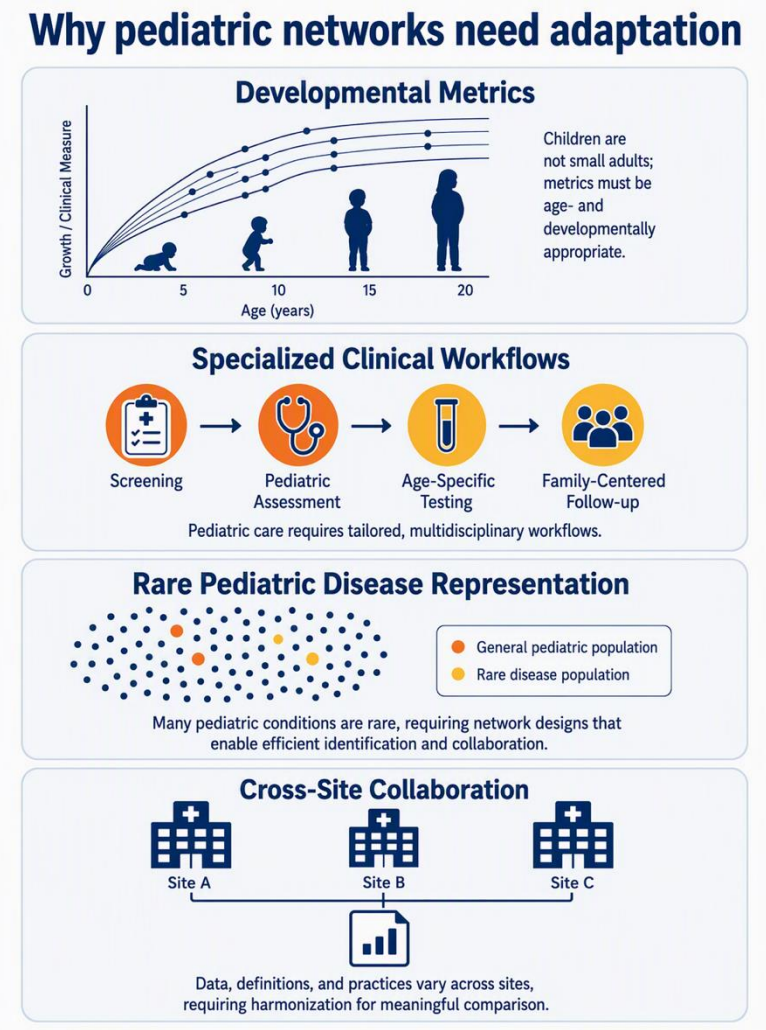

population
question
s
seek
evidence
OHDSI
plan
s
transform
ed
science
unfold
s

PEDSnet: A Multi-Institutional Learning Network



- Funded by PCORI since 2014
- Standardized to OMOP and PCORnet
- Data refreshed every 3 months
- 11 institutions and expanding

The Challenge: Adapting OHDSI for a Pediatric Collaborative Network



- Participation in pediatric network introduces requirements that are not fully addressable by default implementation
- These include examples such as developmental metrics, specialized clinical workflows, rare disease populations, and the need for rigorous cross-site comparability
- To serve a diverse portfolio of research use cases, PEDSnet has systematically adapted the OMOP Common Data Model (CDM), extending existing OHDSI tools, and developed novel data infrastructure

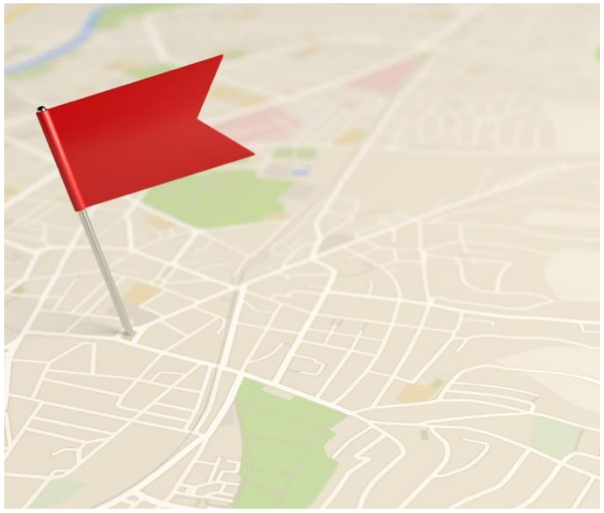
CDM Augmentation: Table Examples



Measurement Organism



Specialty



Location FIPS

PEDSnet Immunization Table

Vaccine and timing
immunization_concept_id
immunization_source_concept_id
immunization_source_value
immunization_type_concept_id
immunization_date
immunization_datetime
imm_recorded_date
imm_recorded_datetime

Product details
imm_manufacturer
imm_lot_num
imm_exp_date
imm_exp_datetime

Record links
immunization_id
person_id
provider_id
visit_occurrence_id
procedure_occurrence_id

Administration
imm_route_concept_id
imm_route_source_value
immunization_dose
imm_dose_unit_concept_id
imm_dose_unit_source_value
imm_body_site_concept_id
imm_body_site_source_value



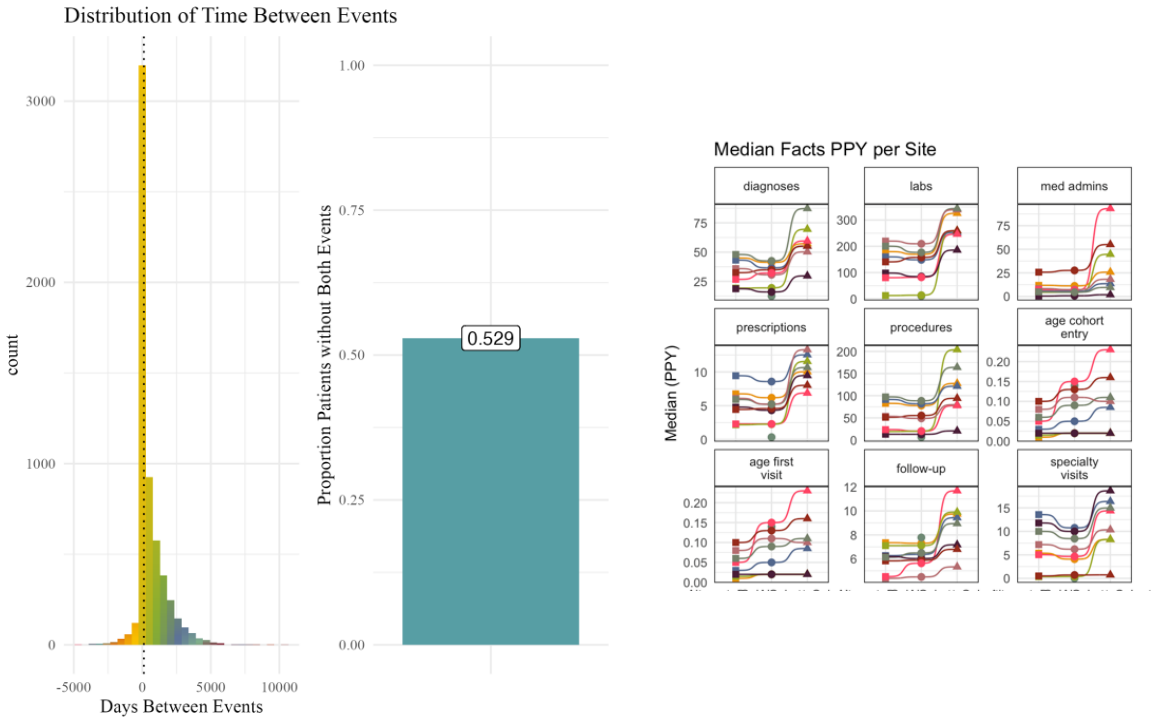
Source: PEDSnet CDM v6.2 / OMOP v5.4 ETL Conventions, §1.17, pp. 58–61. Field names grouped for presentation.

CDM Modifications

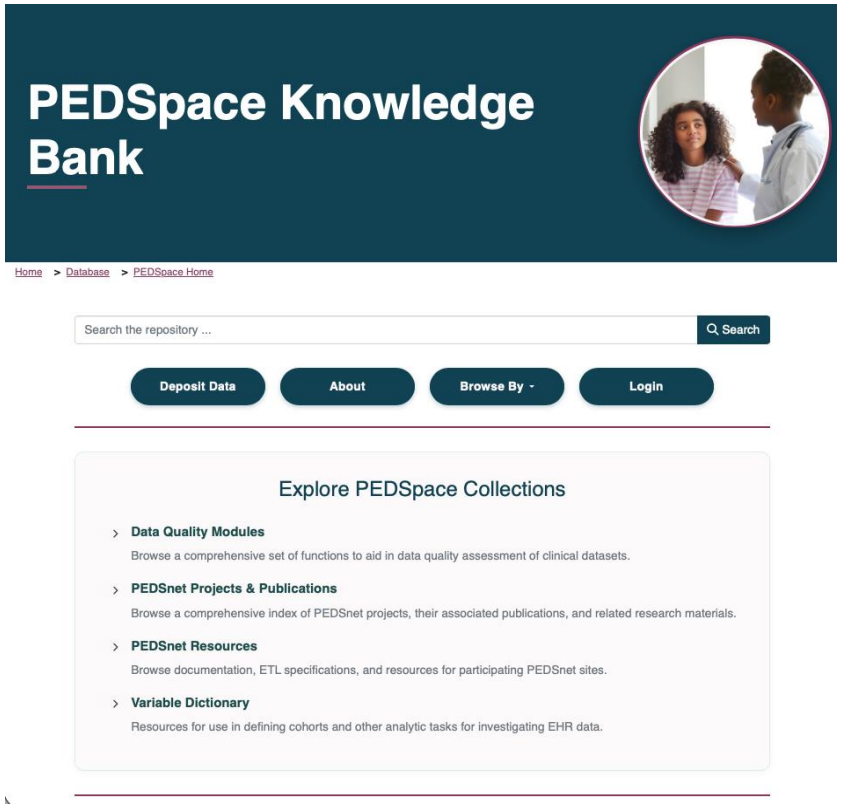
Birth Data	Derivations	ETL Conventions	Network Standardization
Gestational age	Height, weight, BMI z-scores	Unmapped concepts permitted	Standardized source values
Birth weight	Clinical complexity scores	Limited domain routing	Standard fields in Observation

Network-Wide Tools

Study- and Research- Specific DQ



PEDSnet Knowledge Bank



Acknowledgments

Patient and clinician contributors!

PEDSnet Members

Children's Atlanta
Children's Minnesota
Children's National
Children's Philadelphia
Cincinnati Children's
Colorado Children's
Lurie Children's
Nationwide Children's
Nemours Children's
Seattle Children's
Stanford Children's
Texas Children's

PEDSnet Coordinating Center

Charles Bailey
Kaleigh Wieand
Annabel Pinkney
Kimberley Dickinson
Mohan Pargi
Sahal Master
Alex Shorrock
Qiwei Shen
Lucy Shettel
Vitaly Lorman
Christopher Forrest
… and many others

Work supported by PCORnet Infrastructure and NIH ReCOVER funding



question s ?

