

Lighting Talks #2

Presenters

- 1) ORBIT: One-shot Reconstruction-Based Inference via Tensorization **Qiaoshi Yang (Univ. of Pennsylvania)**
- 2) Comparative performance of the concurrent comparator design with existing vaccine safety surveillance approaches on real-world observational health data **Shounak Chattopadhyay (Univ. of Virginia)**
- 3) Implementing Population Health Risk Analytics on a Standardized EHR Data Layer: Evaluation and Framework Development **Haeun Lee (Johns Hopkins Univ.)**
- 4) UNO: Neural Pruning Enhances AI Reliability of Observational Data-based Findings **Yue Wu (Univ. of Pennsylvania)**
- 5) A Multi-Database, Retrospective, Long-term, Observational Safety Study of Pediatric Patients with Inflammatory Bowel Disease **Kevin Haynes (Johnson and Johnson)**

Penn Medicine

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Department of Biostatistics, Epidemiology and Informatics

ORBIT: One-shot Reconstruction-Based Inference via Tensorization

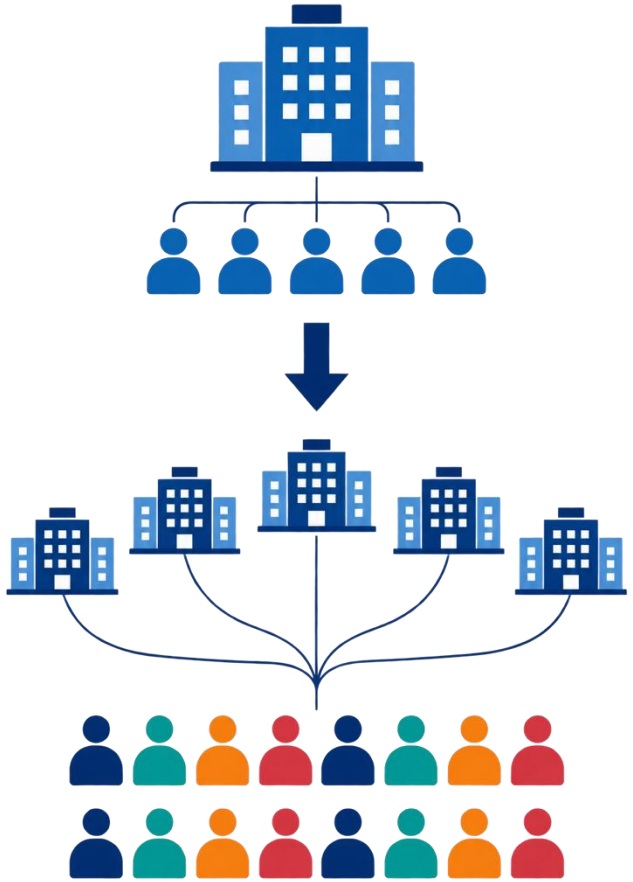
Qiaoshi Yang, Ph.D. candidate at the University of Pennsylvania
Advisor: Dr. Yong Chen

2026 OHDSI Symposium

CHASE
Center for Health AI & Synthesis of Evidence

PennCIL
A Computing · Inference · Learning lab at University of Pennsylvania

Motivation: Multi-institutional EHRs enable broader and more reliable evidence



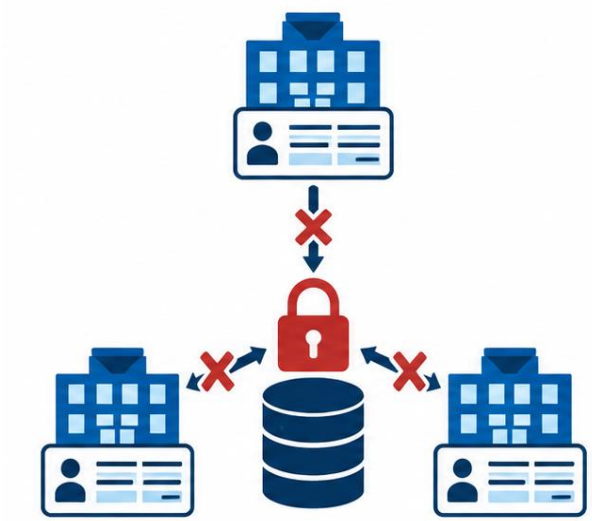
Single institution: Limited sample size and population coverage

Multi-institutional network
Larger and more diverse patient populations

- Broader population coverage: Multi-institutional EHR networks capture patients from different demographic, geographic, and healthcare settings.
- Greater statistical power: Larger combined sample sizes improve the ability to study uncommon outcomes, heterogeneous effects, and clinically important subgroups.
- More generalizable evidence: Findings supported across multiple institutions are potentially more reproducible and transportable than results from a single site.

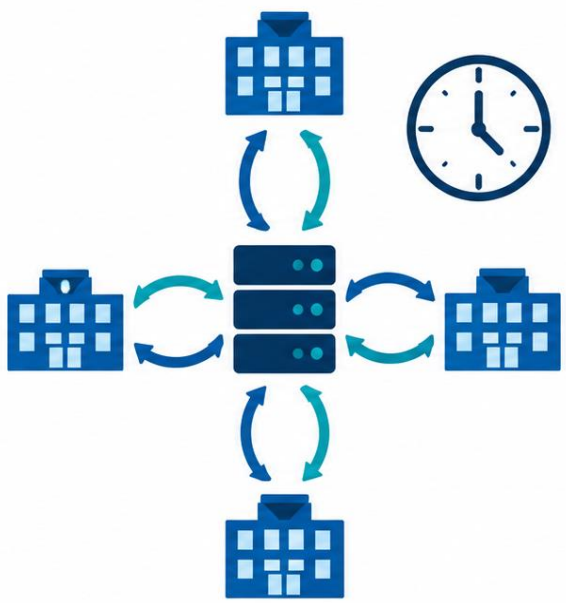
Realizing this potential requires learning from distributed institutions without directly pooling patient-level records.

Multi-institutional inference faces both data-sharing constraints and cross-site heterogeneity



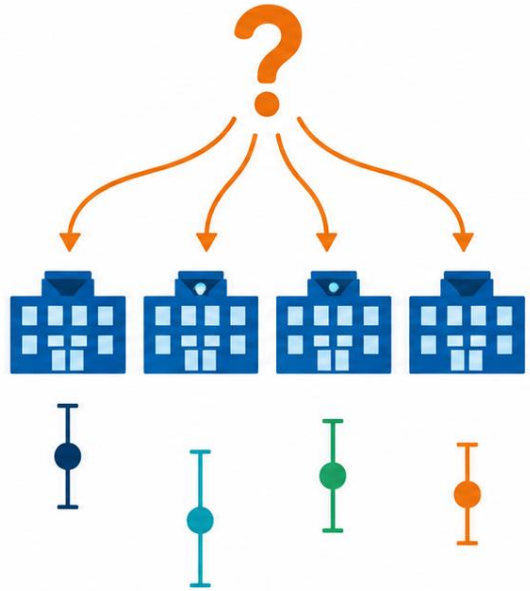
Data-sharing constraints:
Patient-level EHR data cannot be transferred across institutions.

Privacy regulations · data-use agreements · institutional governance



Communication burden:
Repeated exchanges between local sites and the coordinating center can be costly and impractical.

Computation · Coordination · Governance approval



Unknown cross-site heterogeneity:
Some clinical effects are shared across sites, whereas others are site-specific.

The heterogeneity structure is usually unknown in advance.

We need a one-shot federated method that approximates pooled-data analysis while learning which effects are shared across sites.

Existing federated inference methods address only part of these challenges

- Communication-efficient distributed inference enables multi-site analysis without sharing individual-level records. Allows to enlarge the study sample size to incorporate diverse population
- Existing methods typically communicate local gradients, Hessians, or other summary statistics to construct a surrogate likelihood.



Jordan, Lee and Yang (2019), JASA:
Constructs a global surrogate objective using local derivative information.
Limit:
Primarily homogeneous models



Duan et al. (2020), JAMIA:
Enables one-shot distributed logistic regression using aggregated derivatives.
Limit: Not exactly pooled in finite samples



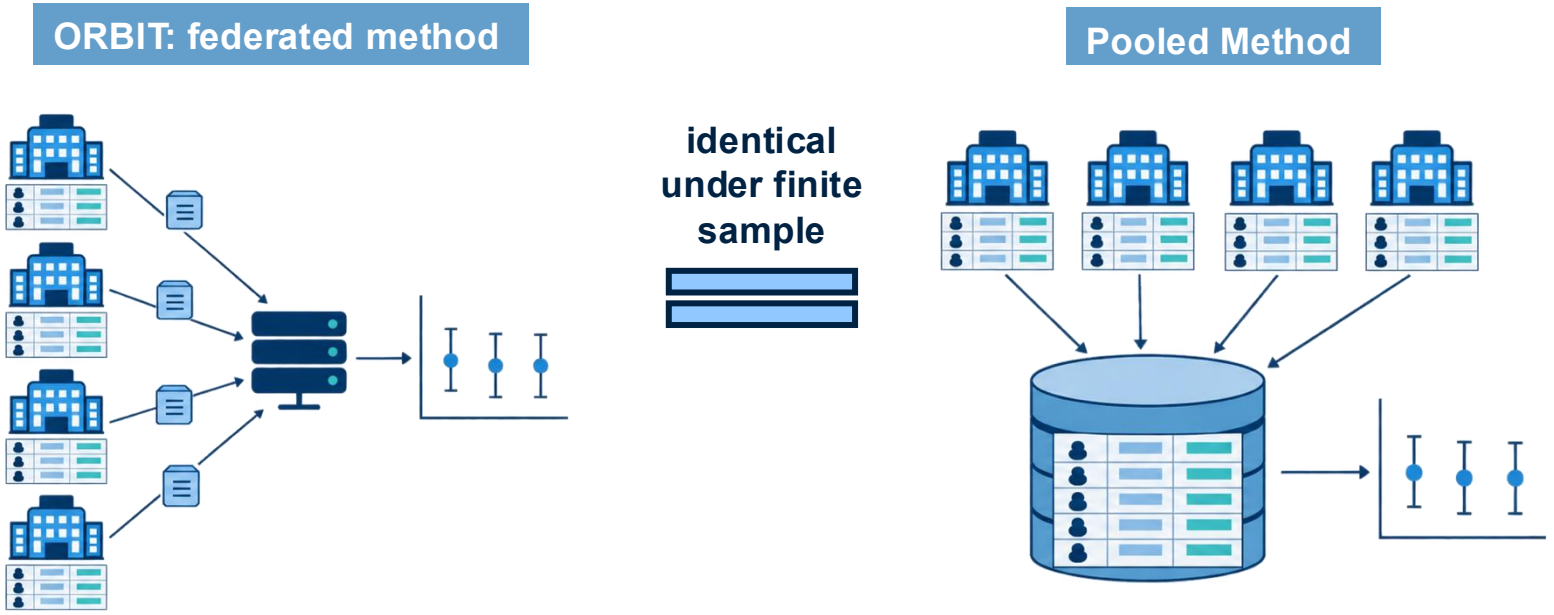
Duan, Ning and Chen (2022), Biometrika:
Allows shared parameters of interest and site-specific nuisance parameters.
Limit:
Requires the heterogeneity structure to be prespecified

Local rather than functional communication · Asymptotic rather than finite-sample equivalence (to pooled) · Prespecified heterogeneity structure

A unified framework combining one-shot communication, pooled-equivalent recovery, and data-driven heterogeneity learning remains needed.

What should an ideal federated inference method achieve?

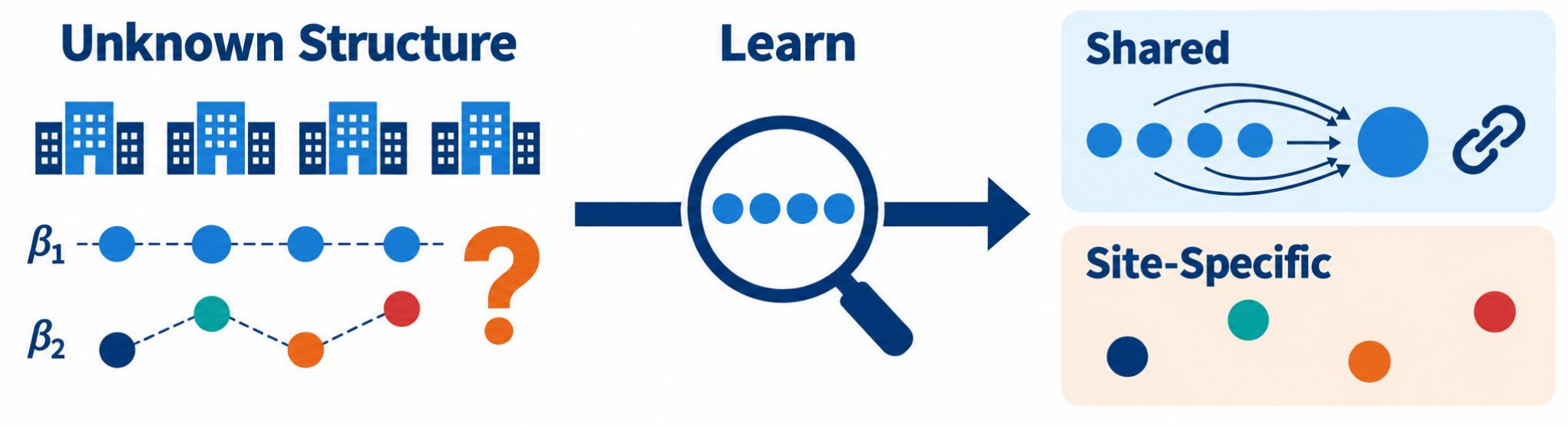
- **One-shot:** each site communicates with the central server only once, no iterative exchange of model updates or derivatives.
- **Lossless:** The federated estimator recovers the pooled-data estimator up to arbitrarily small numerical approximation error under finite samples.



However, these two properties alone are not sufficient when the heterogeneity structure is unknown.

What should an ideal federated inference method achieve?

- **One-shot:** each site communicates with the central server only once, no iterative exchange of model updates or derivatives.
- **Lossless:** The federated estimator recovers the pooled-data estimator up to arbitrarily small numerical approximation error under finite samples.
- **Data-adaptive heterogeneity:** Automatically learns which parameters are shared and which are site-specific. (The heterogeneity structure does not need to be known in advance.)

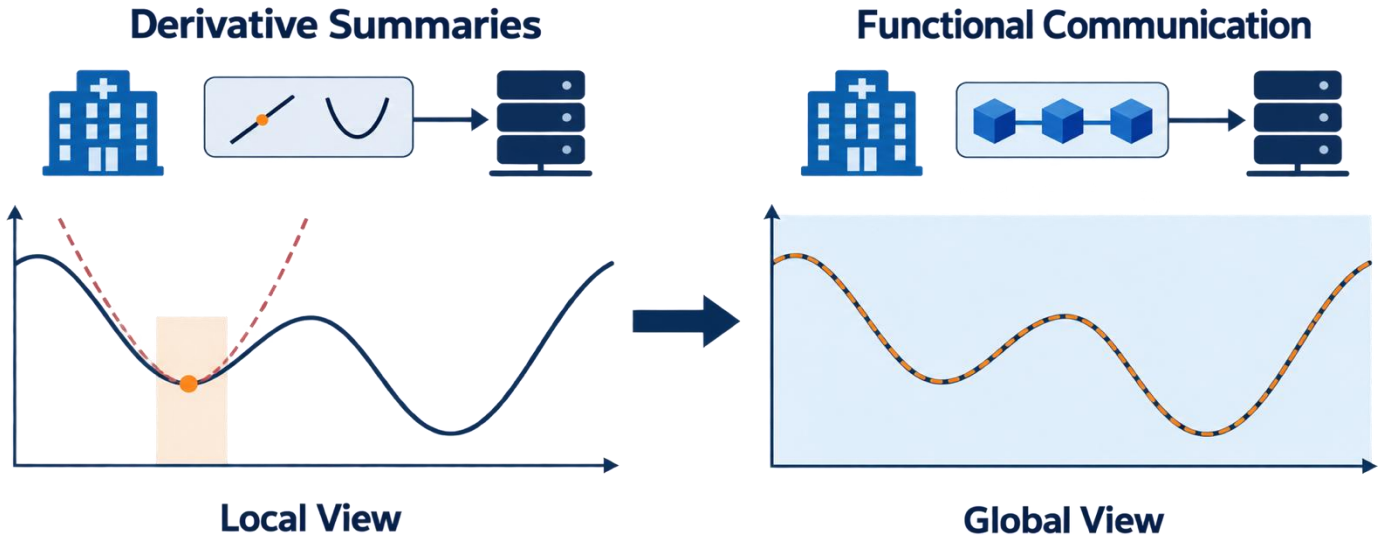


Existing methods may achieve some of these properties, but not all three simultaneously.

ORBIT communicates the objective itself—not only local derivatives

ORBIT: One-shot Reconstruction-Based Inference via Tensorization

One-shot communication · Pooled-equivalent recovery · Data-adaptive heterogeneity learning

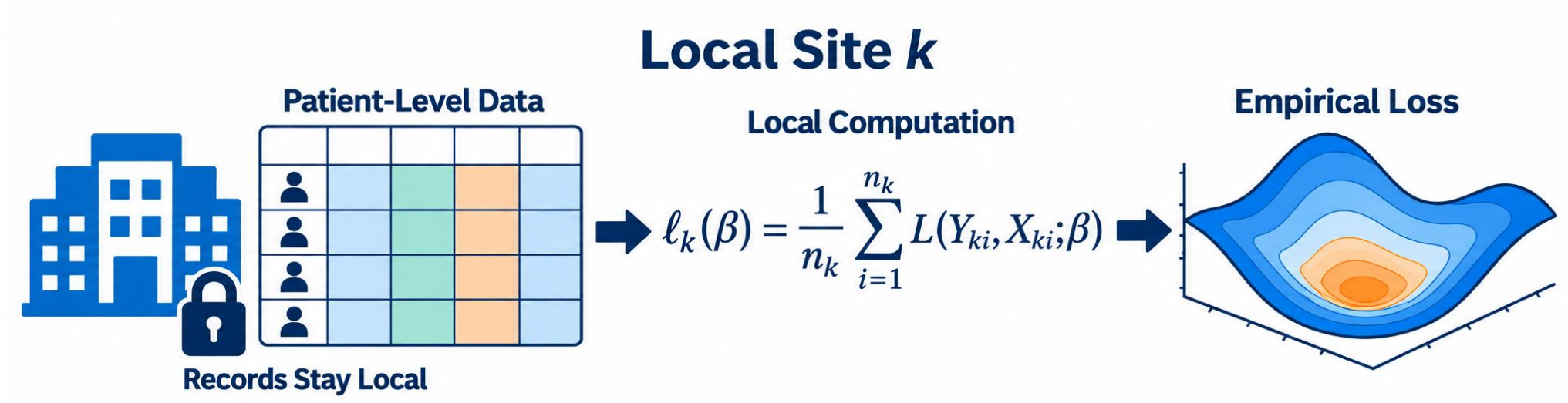


Each site communicates a compressed representation of its empirical loss function in a single round.

Overview of the ORBIT framework: local site

- **Step 1:** Each site compute its local empirical loss

Patient-Level Data → Empirical-loss calculation in local site → Empirical Loss Surface

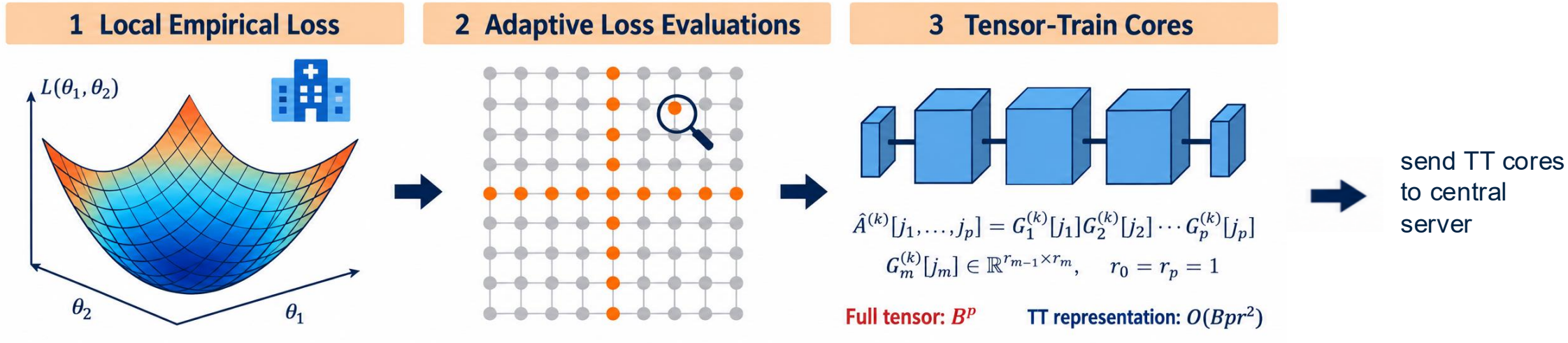


Storage Problem: The next question is how to represent a multivariate loss function without evaluating it on a very large grid (tensor).

Overview of the ORBIT framework: local site

TT-cross constructs a compressed representation of the entire loss surface

- **Step 1:** Each site compute its local empirical loss
- **Step 2:** Each site store its local empirical loss using TT-cross algorithm



Why TT-cross (a quick way to obtain TT cores of a big tensor)?

1) Storage saving: without computing loss function on many points (tensor)

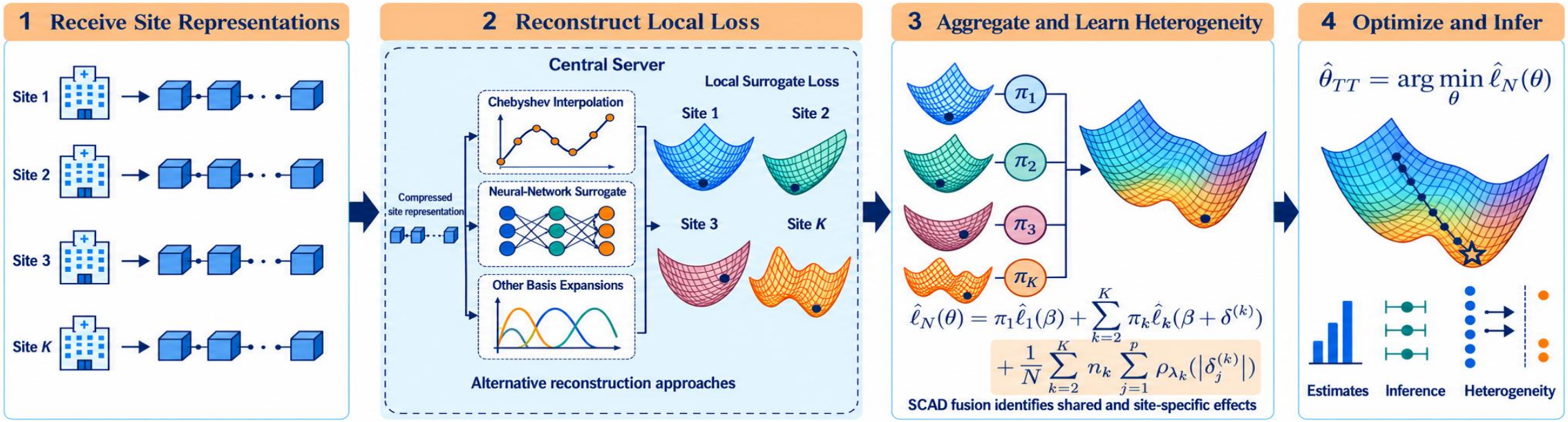
2) Communication efficiency: only need to send TT cores (some small tensor) to the central server

3) Theoretical guarantee: the error between the loss function value from TT-cross and the true loss function value can be arbitrary small

Overview of the ORBIT framework: Central Server

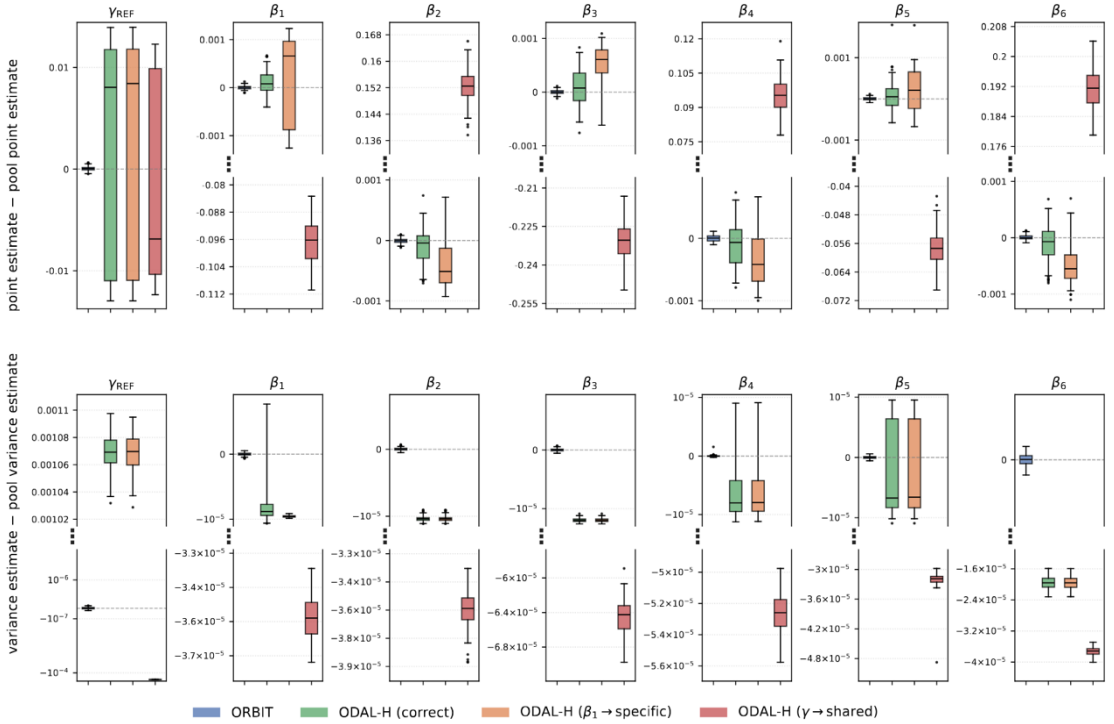
Fitting, synthesizing and optimizing

- **Step 1:** Each site compute its local empirical loss
- **Step 2:** Each site store its local empirical loss using TT-cross algorithm
- **Step 3:** Central server reconstructs all sites' loss surfaces, synthesis them as pooled loss surface and then minimize this pooled loss



Simulation studies: ORBIT closely reproduces pooled point and variance estimates in finite samples

10 sites · 5,000 observations per site · Logistic regression model · Site-specific intercepts · Six shared effects · 100 replications

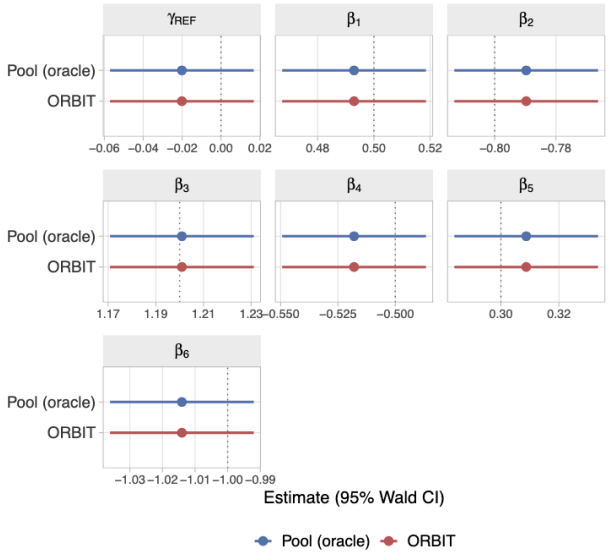


Model in site k (logistic regression):

$$\Pr(Y_{ki} = 1 \mid X_{ki}) = \sigma\left(\gamma_k^* + X_{ki}^\top \beta^*\right)$$

Results:

- (Left) Compared with ODAL-H and its variations, ORBIT can give better point and variance estimations which are identical to pooled estimations under finite sample
- (Right) ORBIT can give identical statistical inference (confidence interval) as pooled



Real-world application: ORBIT identifies clinically interpretable heterogeneity across Penn Medicine hospitals

Q: Does the association between telemedicine use and 30-day follow-up visits remain stable across hospitals?

A: ORBIT finds Telemedicine may substitute for subsequent in-person visits in some hospitals, while serving as part of an ongoing care pathway in others.

Data: 163,308 outpatient encounters from five Penn Medicine hospitals.
CCH: 12,972 · HUP: 86,167 · PMC: 19,670 · PAH: 26,134 · PPMC: 18,365

Outcome: At least one follow-up visit within 30 days of the index encounter

Covariates: Telemedicine · Age · Sex · Race/ethnicity · Travel distance · Mental-health comorbidity

Learned structure:

Shared: Age, sex, and race/ethnicity

Site-specific: Telemedicine, travel distance, and mental-health comorbidity

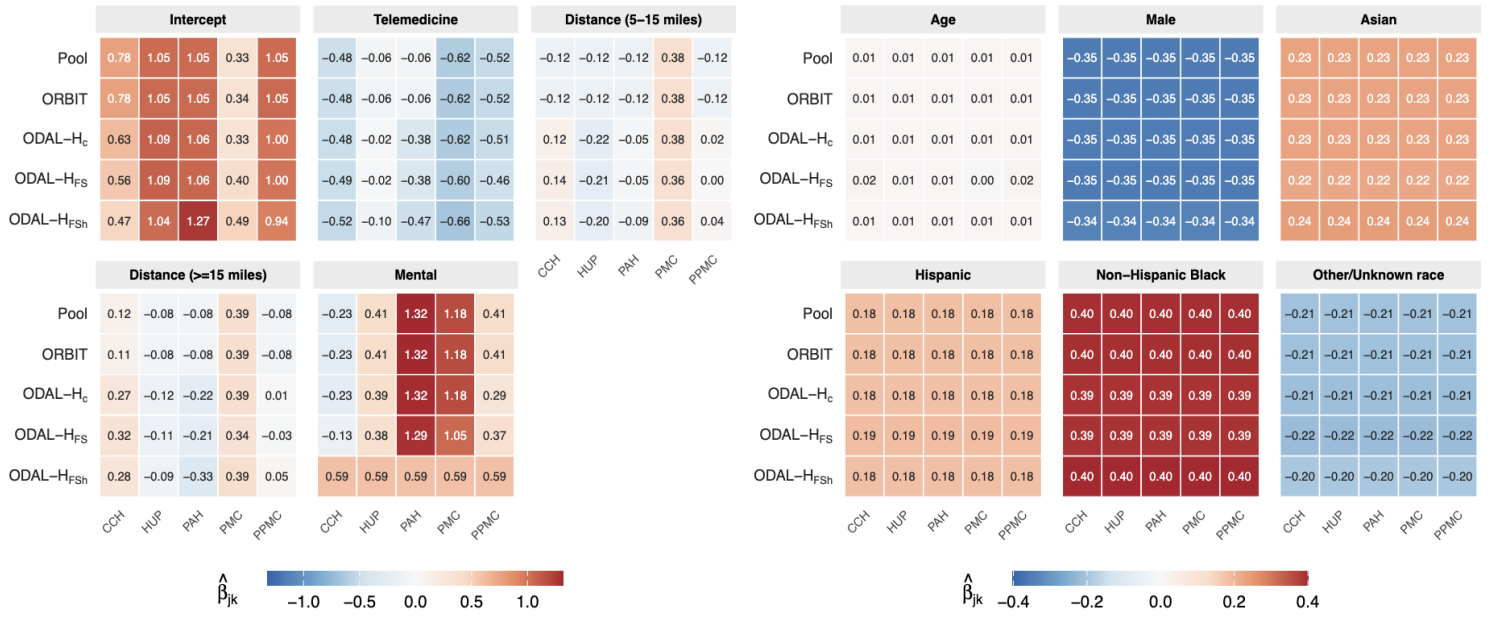


Real-world application

Finding: Demographic associations are stable, while care-utilization associations vary across hospitals

Site-specific care-delivery effects

Shared demographic effects



C: correct; FS: forced site-specific; FSh: forced shared

Findings 1: Stable demographic associations

Age, sex, and race/ethnicity show consistent coefficient patterns across hospitals.

Findings 2: Heterogeneous telemedicine association

Telemedicine is associated with fewer 30-day follow-up visits, but the magnitude varies substantially across hospitals.

Findings 3: Local care environments matter

Associations involving travel distance and mental-health comorbidity depend on hospital-specific geography, workflows, and care coordination.

Conclusion: Without prior knowledge of the heterogeneity structure, ORBIT nearly reproduces both the structure and coefficient estimates obtained from pooled analysis.

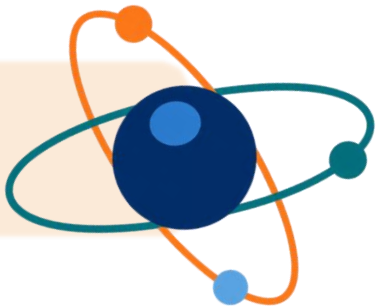
Summary

“ORBIT enables pooled-quality inference from distributed and heterogeneous data”

- **Functional communication:** TT-cross compresses each local empirical loss into compact TT cores.
- **One-shot and pooled-equivalent:** one upload closely reproduces pooled point estimates, variance estimates, and Wald inference.
- **Data-adaptive heterogeneity:** SCAD fusion learns shared and site-specific effects without a prespecified partition.
- **Real-world value:** ORBIT reveals clinically interpretable cross-hospital heterogeneity while nearly matching pooled analysis.

ORBIT: One-shot Reconstruction-Based Inference via Tensorization

One-shot communication · Pooled-equivalent recovery · Data-adaptive heterogeneity learning



Acknowledgments

- Jianqing Fan, Princeton University
- Yong Chen, University of Pennsylvania
- Qiaoshi Yang, University of Pennsylvania
- Jie Hu, University of Pennsylvania
- Yudong Wang, University of Macau

Correspondence to:

Yong Chen, ychen123@pennmedicine.upenn.edu

Qiaoshi yang, yangqs@sas.upenn.edu



Real-world performance of the concurrent comparator design

Shounak Chattopadhyay
University of Virginia

Background

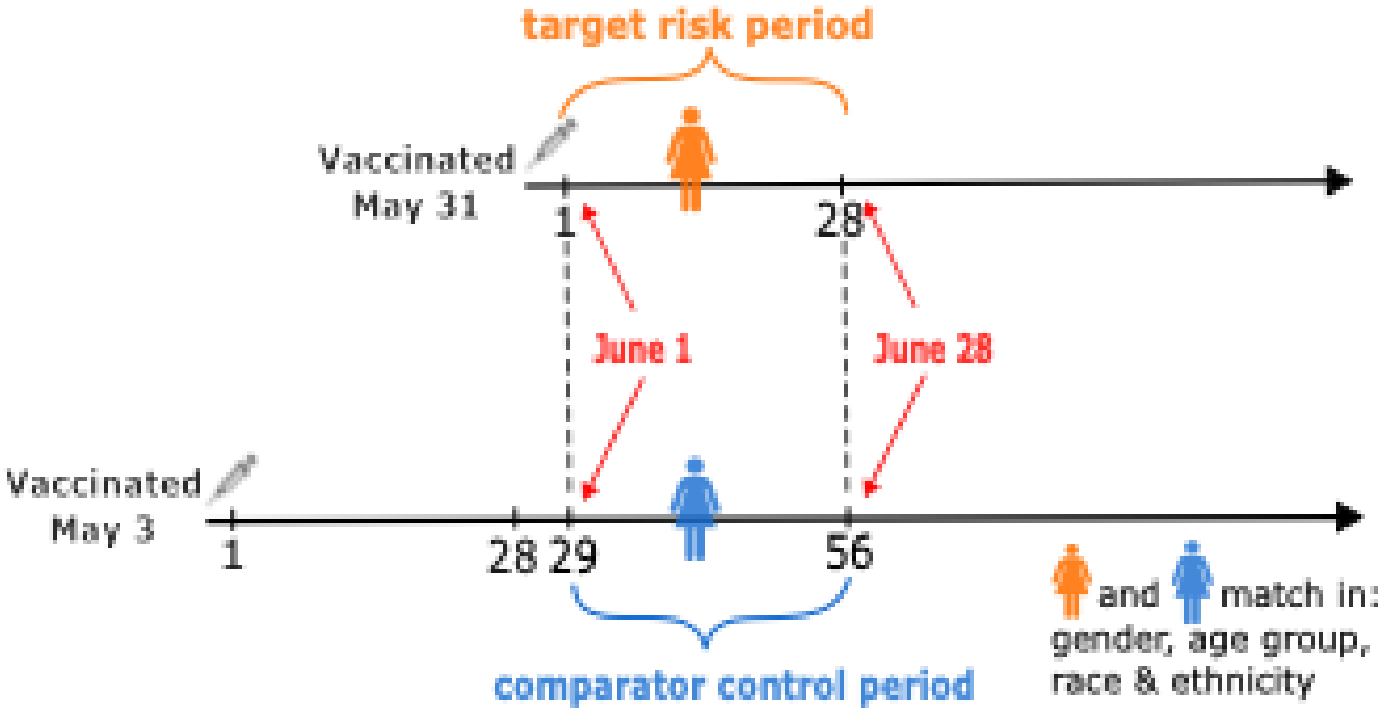
- Earlier work* carried out large-scale comparative evaluation of existing state-of-the-art methods for vaccine safety surveillance.
- These methods may produce biased estimates based on differences in patient preference, regarding (1) whether they receive the vaccine and/or (2) when they receive the vaccine.
- Klein et al. (2021)** proposed a new method called concurrent comparator, which aims to control for this bias.
- **Question**: How does the concurrent comparator design perform based on real-world observational health data?

*Schuemie MJ, Arshad F, Pratt N, et al. Vaccine safety surveillance using routinely collected healthcare Data—An empirical evaluation of epidemiological designs. Frontiers in Pharmacology. 2022;13.

**Klein NP, Lewis N, Goddard K, et al. Surveillance for adverse events after COVID-19 mRNA vaccination. JAMA: The Journal of the American Medical Association. 2021;326:1390–9.

Concurrent comparator

- Predefines a risk interval post-vaccination (1-28 days) and matches each vaccinated patient for whom an outcome is observed during their **target risk period**, with a vaccinated patient in their **comparator control period** on the same calendar day.
- This matching is done based on gender, age group, race, and ethnicity.



- Estimation of risk ratio is carried out using a conditional Poisson regression.

Real-world evidence with 118M estimates

Exposures of interest:

- H1N1pdm (`09-`10)
- Seasonal influenza (Fluvirin, `17-`18)
- Seasonal influenza (Fluzone, `17-`18)
- Seasonal influenza (all, `17-`18)
- Zoster (2018, 2 doses)
- HPV (2018, 2 doses)
- Covid-19 (BNT126b2, `20-`21)
- Covid-19 (mRNA-1273, `20-`21)

Data sources:

- CCAE
- MDCR
- MDCD
- Optum EHR
- Optum DOD

Negative control outcomes (93):

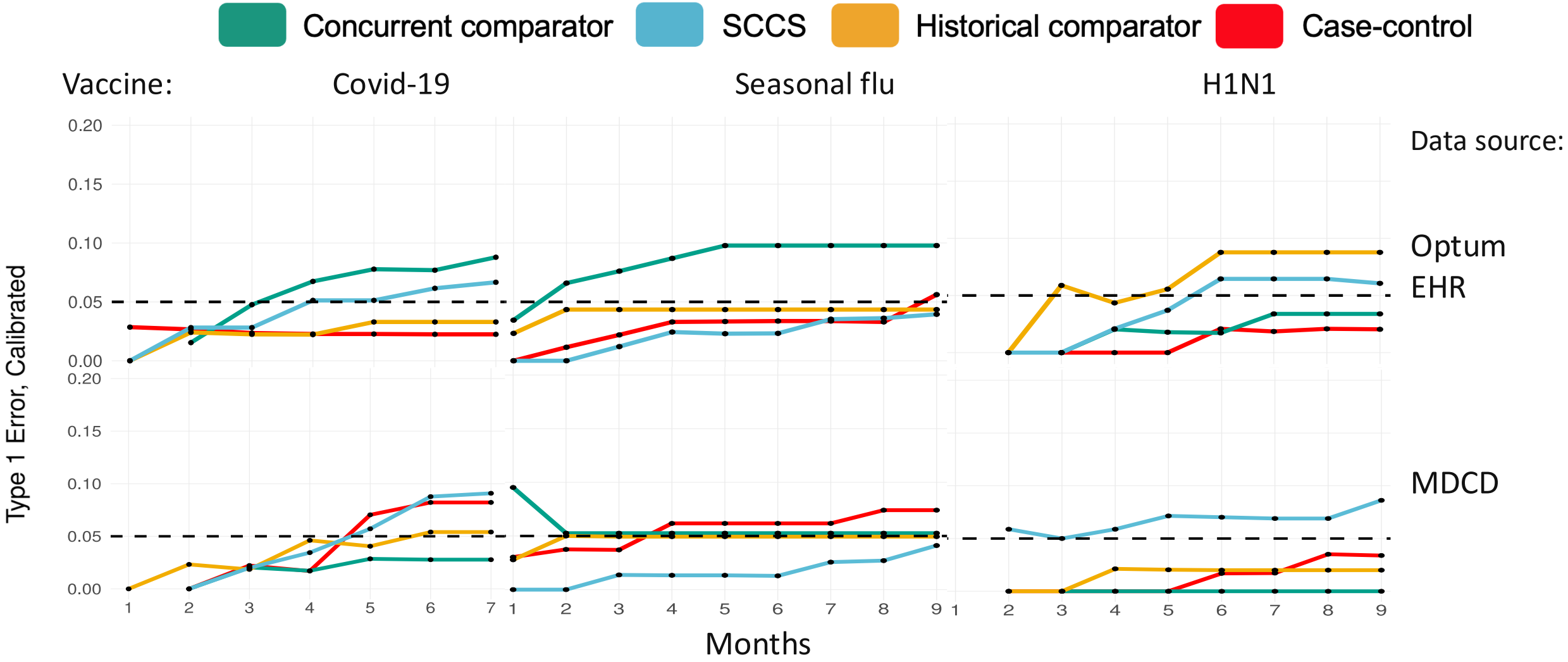
- Not related to any of these vaccines
- Similar prevalence and %-inpatient diagnoses (severity) to adverse events
- Clinical expert review

Positive control outcomes:

- Imputed from negative controls
- Known effect sizes (1.5, 2, 4 x)

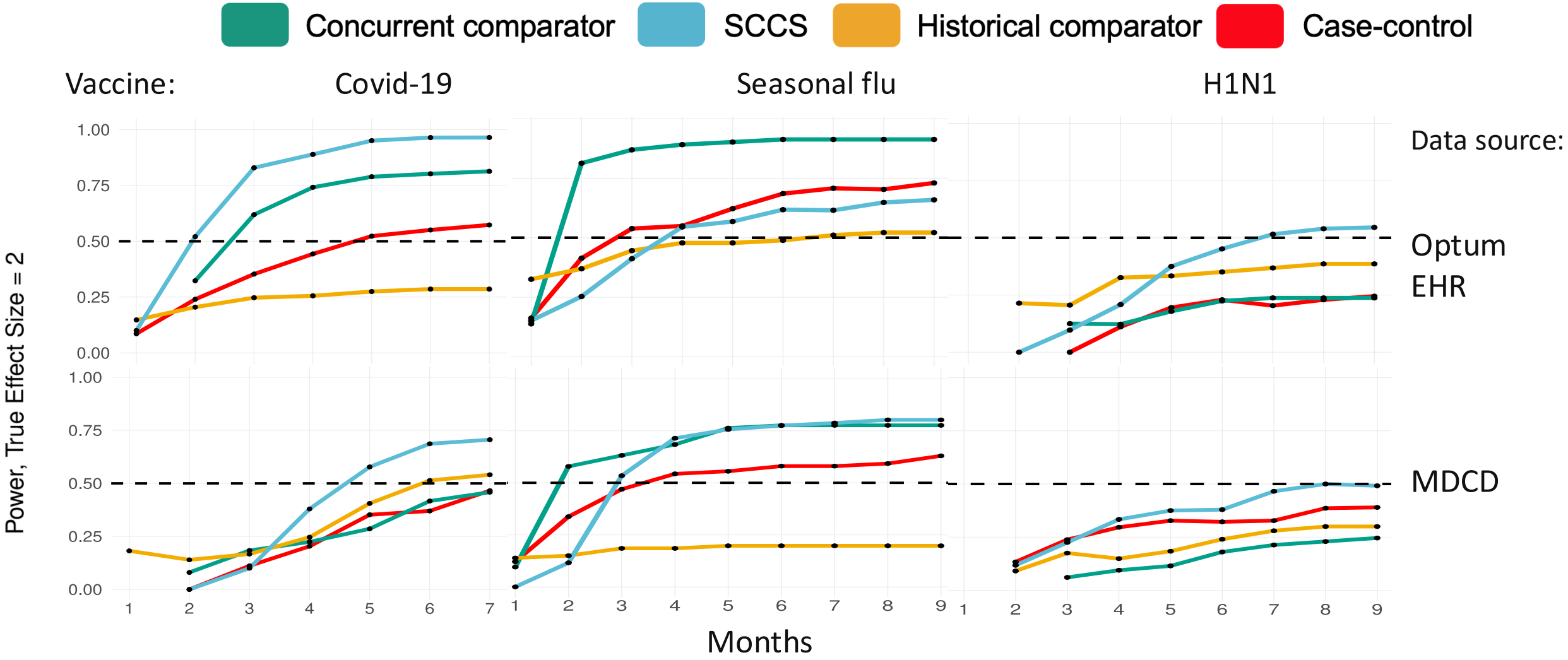
Study protocol link: https://ohdsi-studies.github.io/Eumaeus/Protocol.html

Type 1 Error (calibrated)



- To restore type 1 errors to near-nominal values, we carry out empirical calibration.

Power of detection (true effect size = 2)



Proportion of non-finite estimates

- We also evaluate the methods based on the proportion of non-finite effect size estimates.
- Concurrent comparator has a higher proportion than the other methods.

Method	Optum EHR	MDCD
Case-control	16.95%	4.84%
Concurrent comparator	**36.56%**	**34.41%**
Historical comparator	26.88%	33.33%
SCCS	1.49%	0.00%

Exposure: H1N1pdm vaccine

Conclusion

- After empirical calibration to restore nominal type 1 error, SCCS performs the best overall, with concurrent comparator close to SCCS in terms of power of detection.
- The relative performance of the concurrent comparator decreases for smaller data sources.
- Compared with other approaches, concurrent comparator produces non-finite estimates more frequently.

Preprint

Authors: Shounak Chattopadhyay, Fan Bu, Martijn J. Schuemie, Jody-Ann McLeggon, Erik Westlund, Elise L. Boventer, Tara V. Anand, George Hripcsak, Patrick B. Ryan, Marc A. Suchard

Implementing Population Health Analytics on a Standardized EHR Data Layer: Evaluation and Framework Development

Haeun Lee[1], Tom Richards[2], Benjamin Martin [1], Paul Nagy[1], Hadi Kharrazi[1,2], Jonathan Weiner[2], Chintan Pandya[2]

[1] Biomedical Informatics and Data Science, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

[2] Center for Population Health Information Technology, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

2026 OHDSI Global Symposium – Lightning Talk

October 21, 2026

Haeun Lee, PhD Candidate

Population Health Analytics in the Era of Interoperability

Value-based care and population health programs rely on accurate population health analytics — including risk stratification, to identify high-risk patients, allocate resources, support care management, and improve quality.

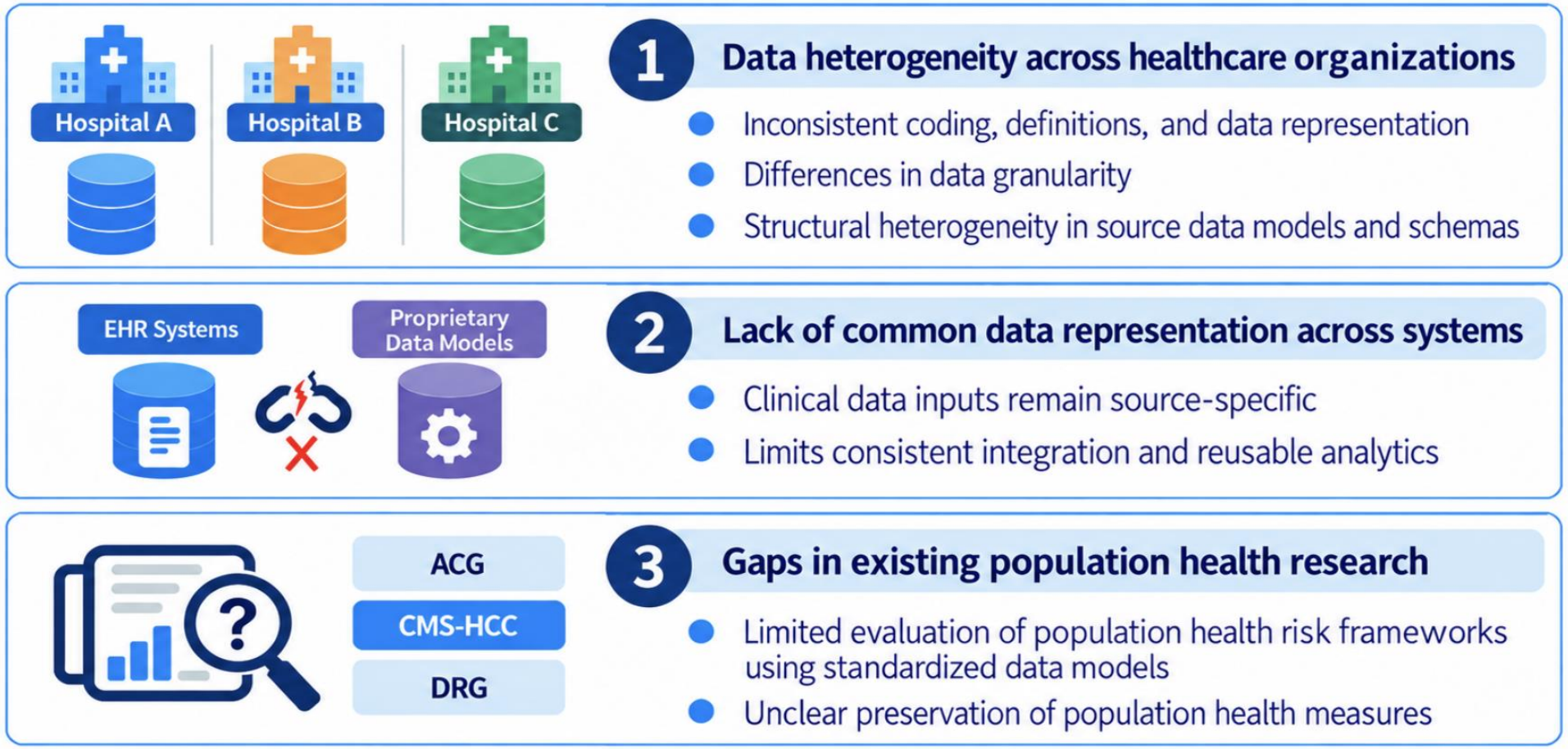
EHR data are increasingly used for population health analytics — but heterogeneity across systems challenges consistent and interoperable measurement.

OMOP CDM provides a standardized EHR data layer — that could support scalable, cross-system population health analytics.

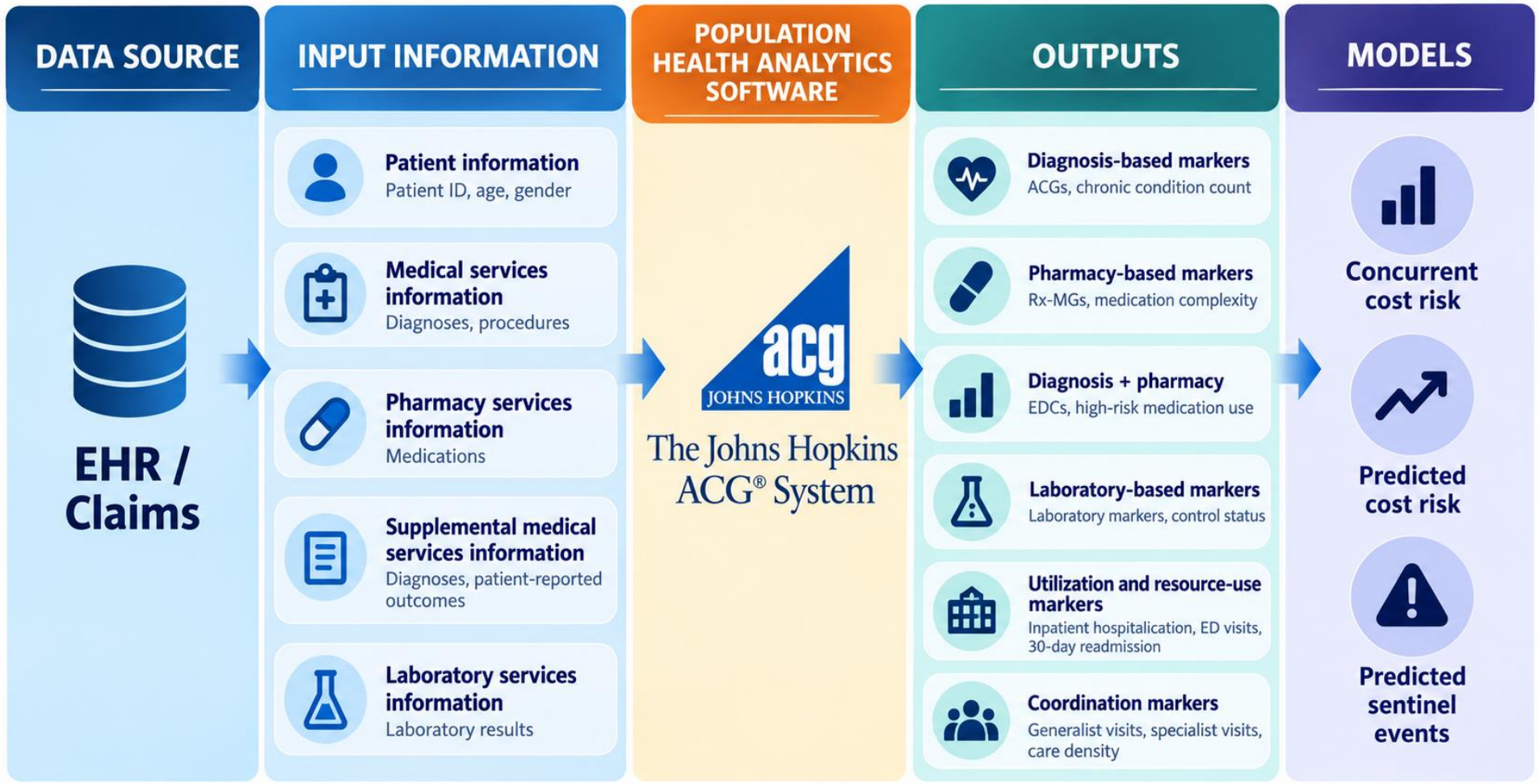
Preserving clinically meaningful population health measures across EHR-to-OMOP transformation remains critical.



Data Heterogeneity Challenges for Population Health Analytics and Integration



What are ACG (Adjusted Clinical Groups) System?

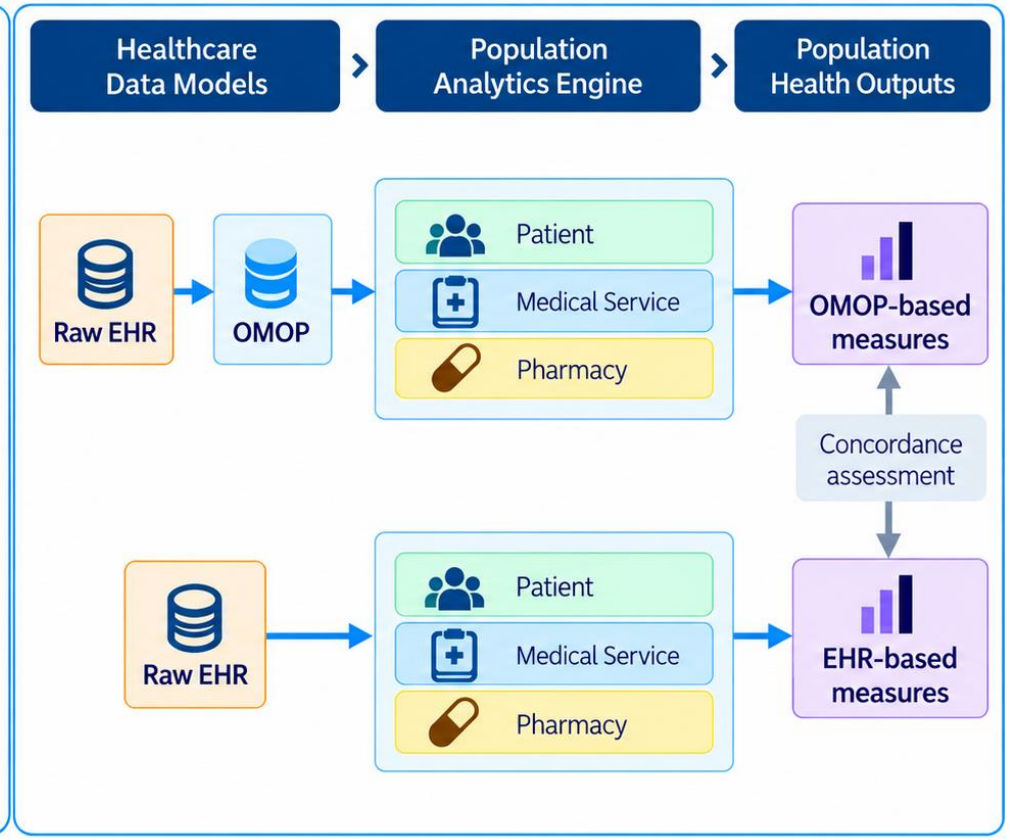


Study Objectives

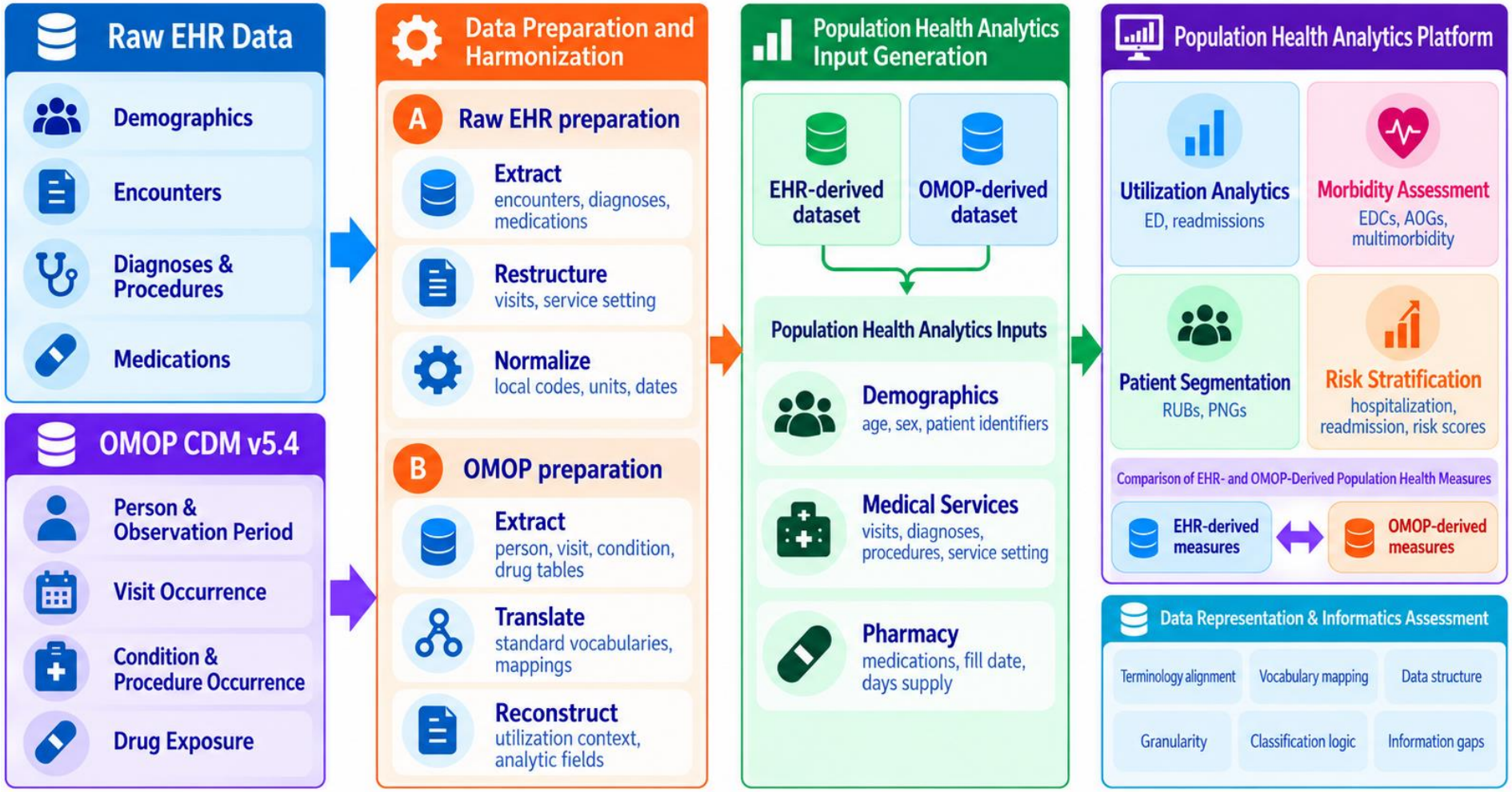
Knowledge gap

Limited empirical evidence exists on whether OMOP standardization preserves population health measures relative to source EHR data.

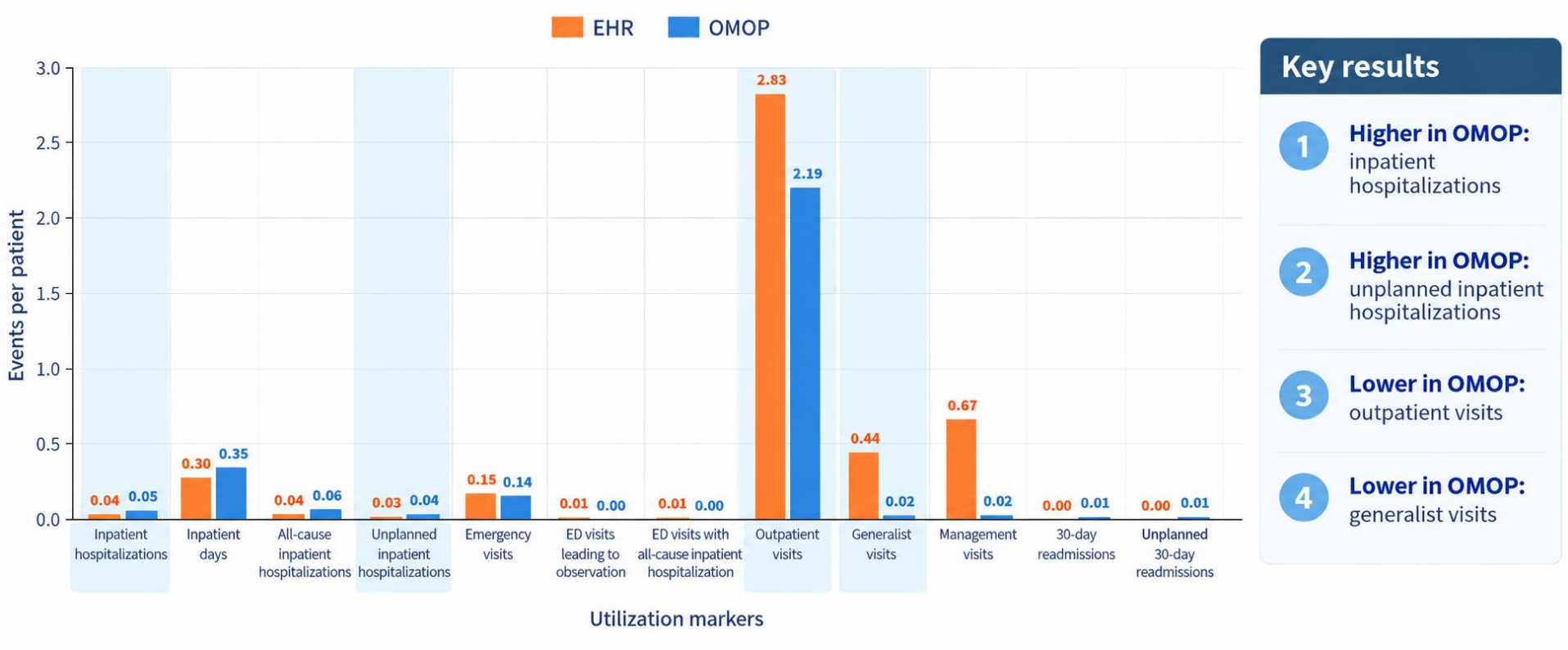
1. Derive comparable population health input variables from source EHR and OMOP data
2. Generate population health measures using a common analytics platform
3. Evaluate concordance of OMOP- and EHR-derived population health measures and risk outputs



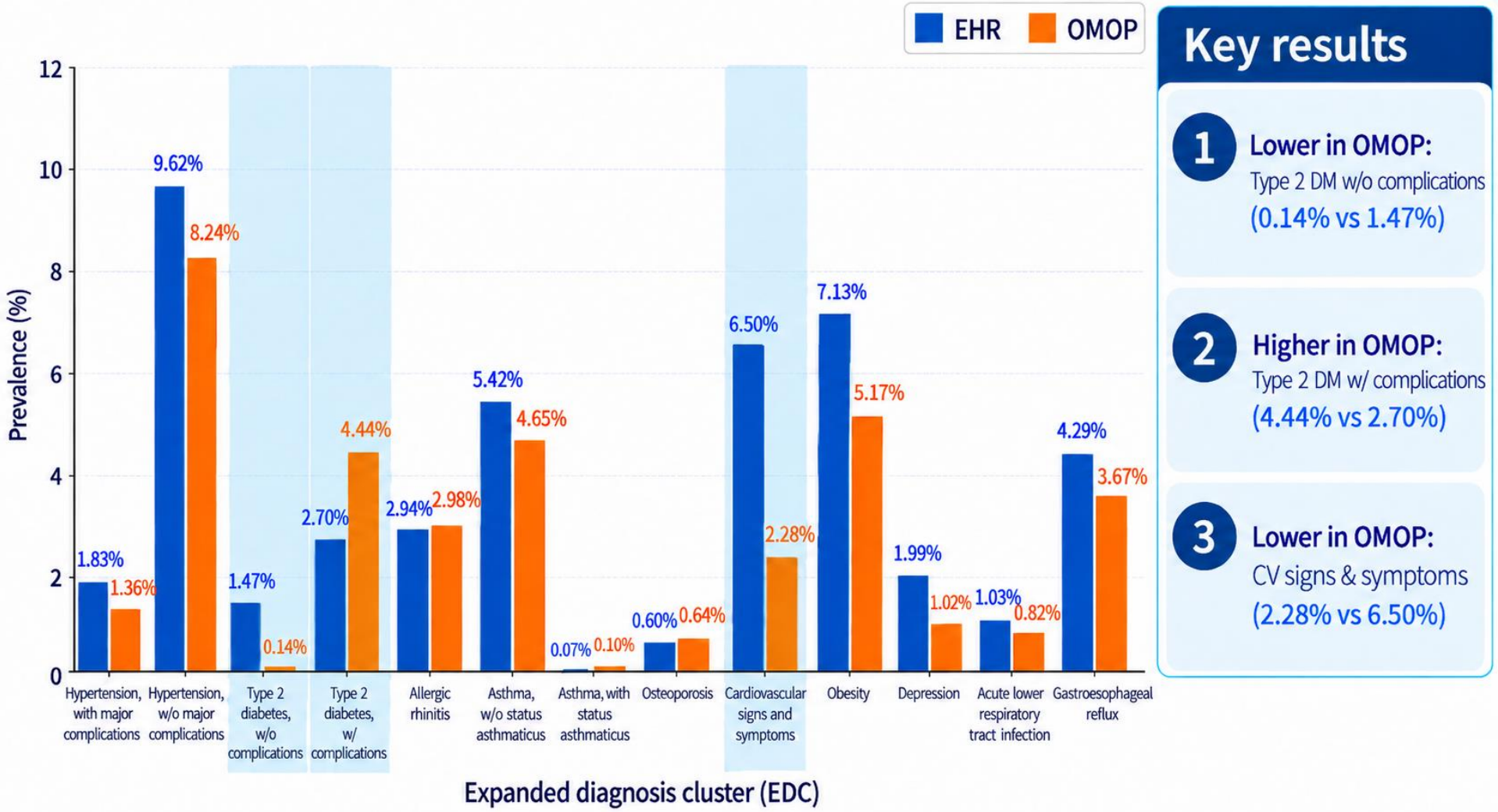
OMOP CDM-Based Data Transformation Framework



Comparison of Healthcare Utilization between EHR vs. OMOP



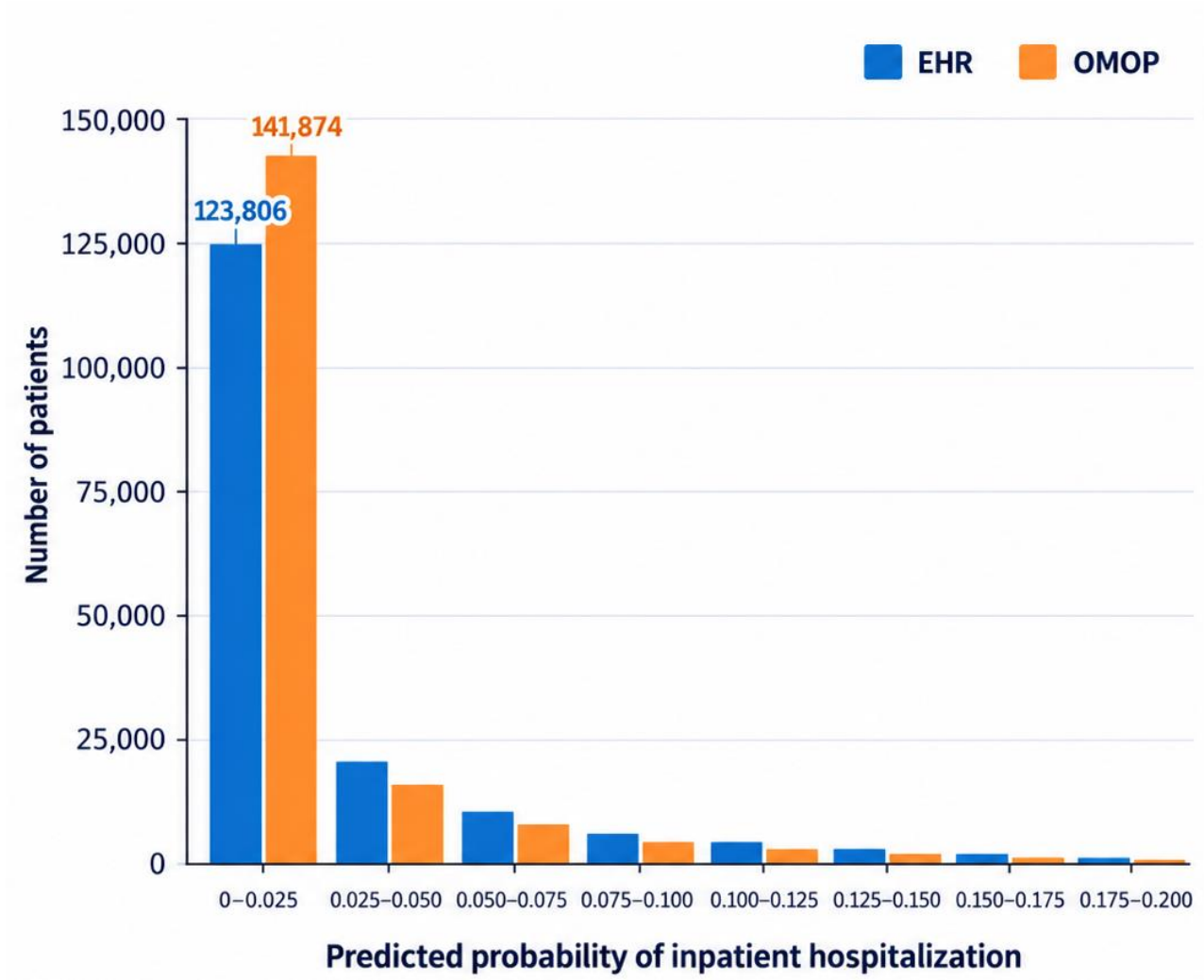
Comparison of Chronic Conditions Prevalence Between EHR vs. OMOP



Key results

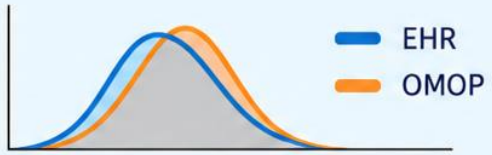
1. **Lower in OMOP:** Type 2 DM w/o complications (0.14% vs 1.47%)
2. **Higher in OMOP:** Type 2 DM w/ complications (4.44% vs 2.70%)
3. **Lower in OMOP:** CV signs & symptoms (2.28% vs 6.50%)

Distribution of Predicted Inpatient Hospitalization Probability



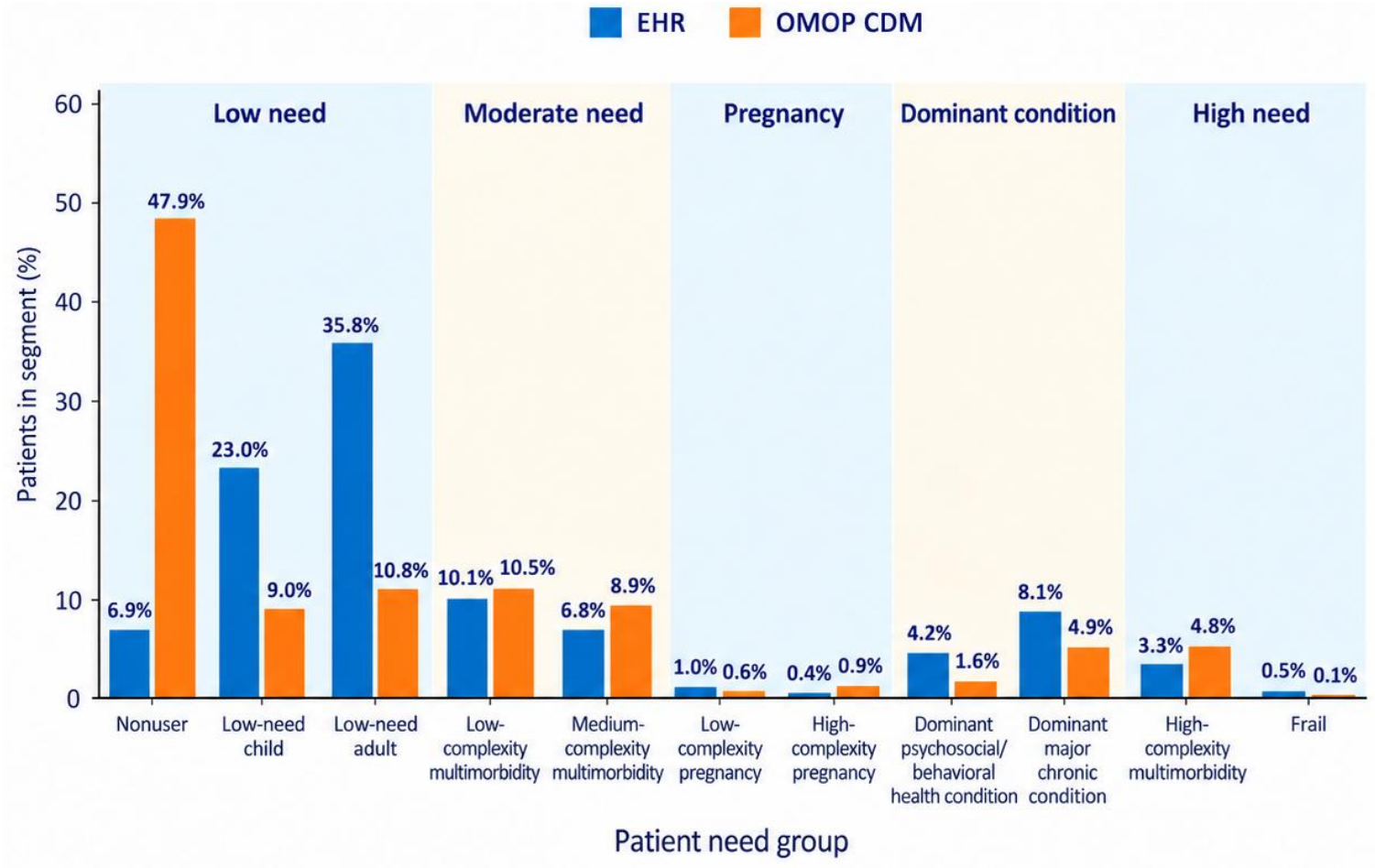
Key results

1. **More patients in the lowest-risk bin in OMOP**
   86.4% vs 75.4%
2. **Overall risk distribution broadly similar**



3. **Fewer patients in higher-risk strata in OMOP**

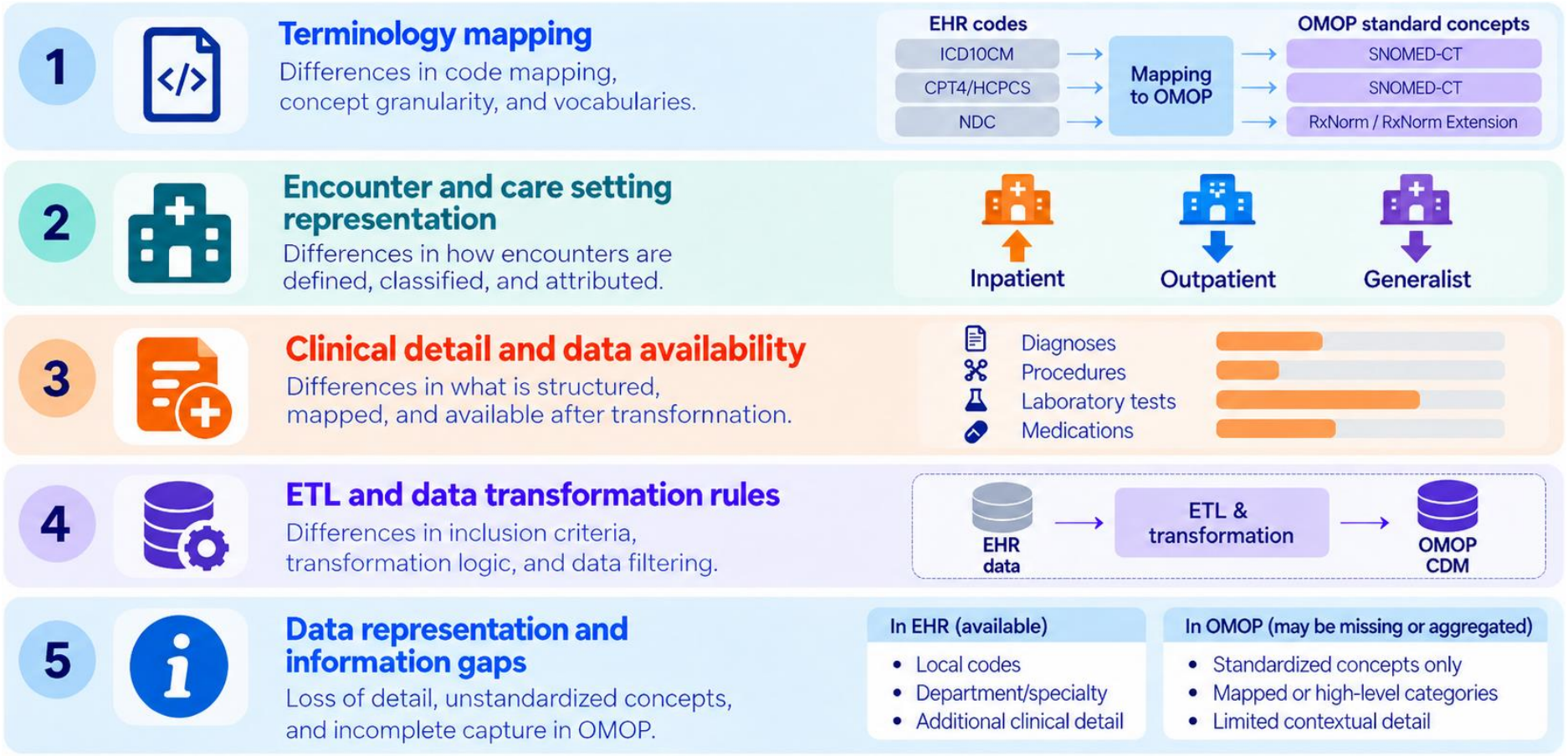
Comparison of Patient Need Group Distribution: EHR vs. OMOP



Key results

1. More nonusers in OMOP; larger low-need child and adult segments in EHR
2. Higher-need groups broadly similar; slightly higher multimorbidity in OMOP

Where the differences come from



Conclusion

1.  **OMOP CDM** supports population health analytics at scale
2.  **Most OMOP-derived** population health measures were broadly aligned with EHR-derived results
3.  Selected population health outputs, particularly **risk stratification**, showed meaningful differences
4.  Differences in **data transformation and representation** may contribute to the observed difference
5.  **Validation, transparent transformation processes, and data governance** are needed for operational use of OMOP-derived population health measures.

Future Directions

1



Identify and quantify sources of discordance

Link EHR–OMOP input differences to changes in population health and risk outputs

2



Refine OMOP transformation rules

Address terminology, encounter context, granularity, structure, and information gaps

3



Evaluate transferability across healthcare organizations

Strengthen the existing OMOP-based framework through standardized quality checks and transparent transformation processes

Project Team and Acknowledgements

- Chintan Pandya, MD, PhD, MPH (Project PI)
- Tom Richards, MS
- Benjamin Martin, PhD
- Paul Nagy, PhD
- Hadi Kharrazi, MD, PhD
- Jonathan Weiner, DrPH
- The project was partially supported from royalties for non-academic use ACG® software methodology

Penn Medicine

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Department of Biostatistics, Epidemiology and Informatics

UNO: Neural Pruning Enhances AI Reliability of Observational Data-based Findings

Yue Wu
University of Pennsylvania

2026 OHDSI Symposium

CHASE
Center for Health AI & Synthesis of Evidence



ennCIL
A Computing · Inference · Learning lab at University of Pennsylvania

Motivation: Reliable AI for Observational Studies

1. **Observational data enable large-scale drug research.**

   They support evaluation of treatment effects across diverse populations.

2. **Systematic error can persist despite careful study design.**

   Even with target trial emulation, incompletely captured clinical factors can leave residual bias.

3. **Neural networks may learn spurious correlations from biased data.**

   These learned patterns can undermine the reliability of downstream findings.

> **How can we mitigate systematic error in AI models?**

Negative control outcomes (NCO): signals of potential systematic error

1 NCO definition

No causal effect of treatment.

Treatment ⊘→ **NCO**

May share unmeasured confounders with the primary outcome.

2 Estimate a signal

Compare predicted NCO outcomes.

Treated − **Control**

Same observed covariates

Differences can signal potential systematic error.

3 Guide refinement

Combine signals from multiple NCOs.

High-NCO-signal subset

Low-NCO-signal subset

Signal groups, not bias labels.

UNO uses NCO-derived signals to guide model refinement.

UNO: NCO-guided pruning and fine-tuning

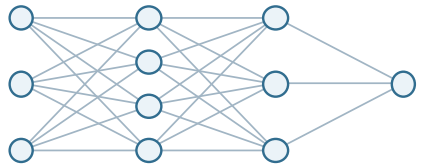
Here: a propensity-score neural network predicts treatment from baseline covariates.

NCO-derived subsets (from the previous slide)

High-NCO-signal

Low-NCO-signal

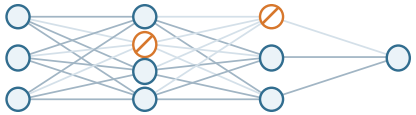
1 Base model



Predict treatment using covariates

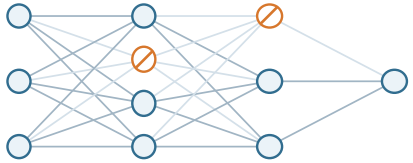
2 Score & prune

Mean activation² ratio (high / low)



Set top-scoring neurons to 0

3 Fine-tune



Low-signal training
Fixed pruning mask

4 Full cohort



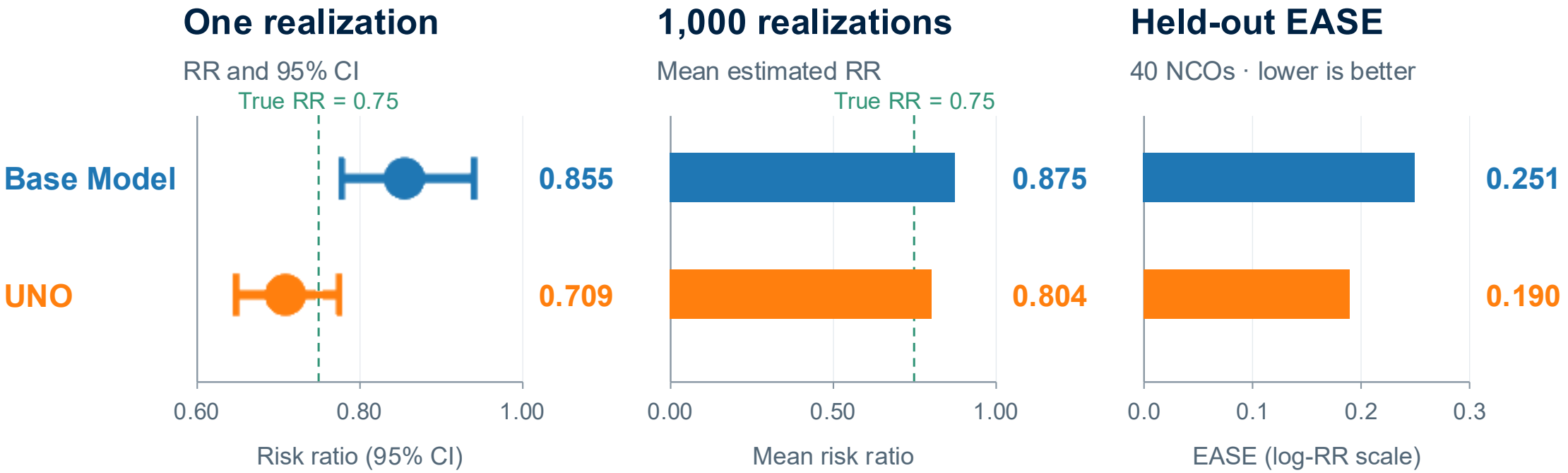
Propensity scores for every patient

Treatment-effect estimation uses the full cohort—both subsets retained.

An illustrative simulation example

10,000 individuals | True causal risk ratio (RR) = 0.75 | 40 held-out NCOs

NCOs and target outcome share unmeasured variables through different nonlinear functions.



Mean RR: 1,000 paired target-outcome realizations, with the cohort and fitted models held fixed.

In this example, UNO had a mean RR closer to 0.75 and lower held-out EASE than the Base Model.

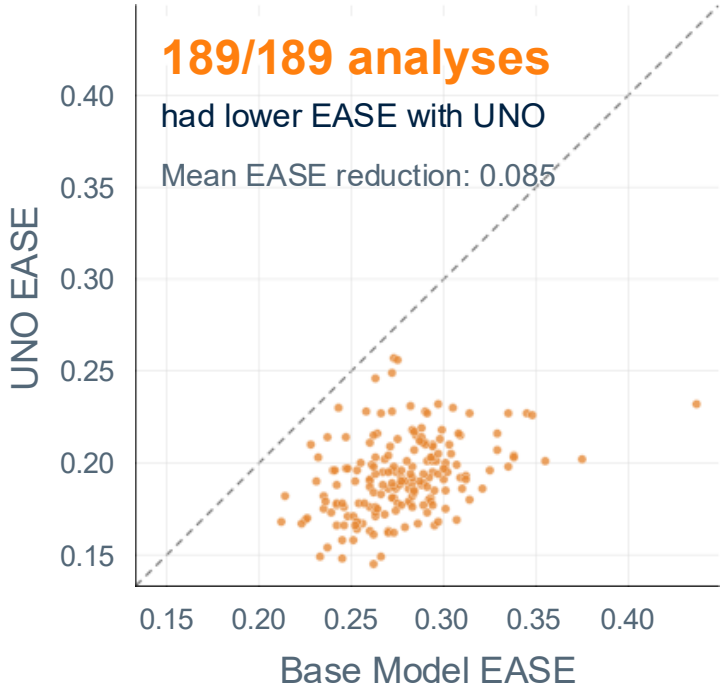
Result 1: GLP-1RAs vs DPP-4is

Pharmacologic outcome profiling | TriNetX, 2019–2024

303,580 treatment initiators with T2D | 189 outcome-specific analyses

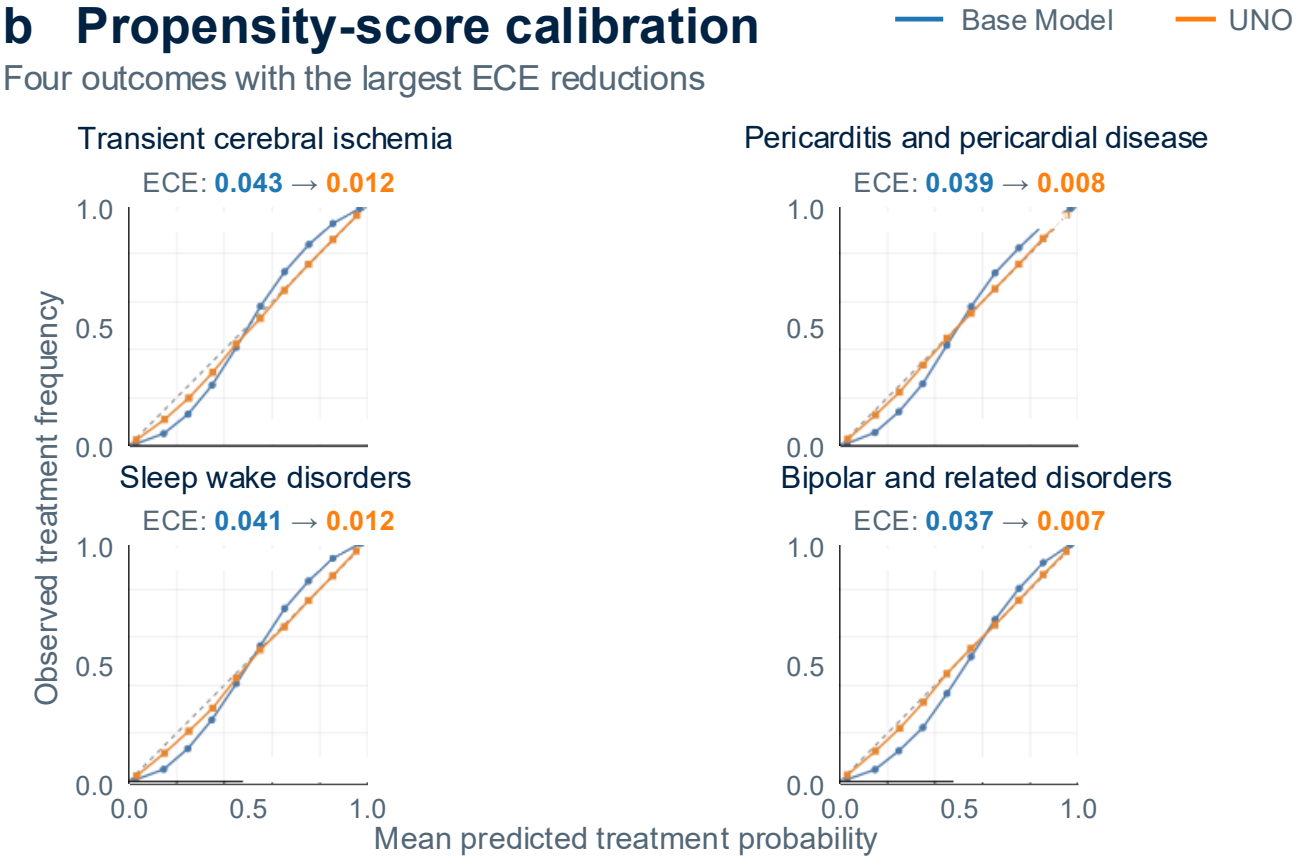
a Held-out EASE

50% of eligible NCOs reserved for evaluation



b Propensity-score calibration

Four outcomes with the largest ECE reductions



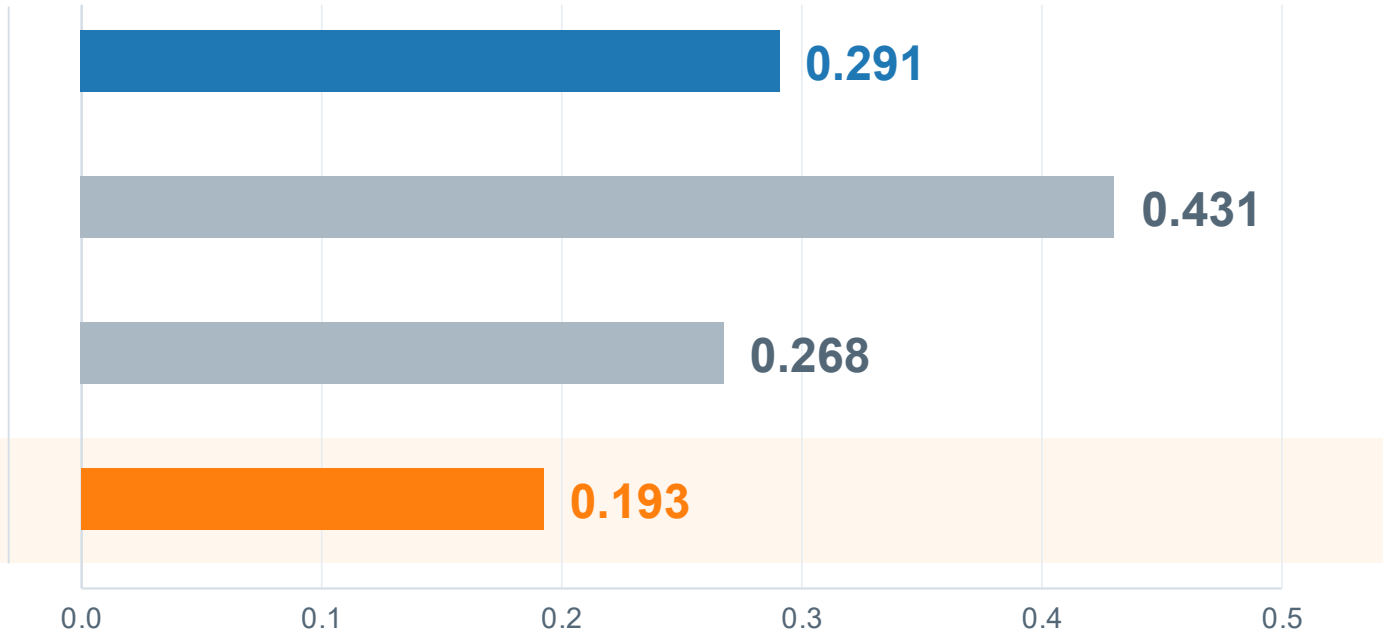
EASE: expected absolute systematic error. ECE: expected calibration error.

Penn Medicine

Ablation: Contributions of pruning and fine-tuning

GLP-1RAs vs DPP-4is | 10 outcome-specific analyses

Variant	Pruning	Fine-tuning
Base Model	—	—
Pruning only	✓	—
Fine-tuning only	—	✓
Full UNO	✓	✓



Pruning: UNO-selected mask. Fine-tuning: low-NCO-signal subset. Means are equally weighted across analyses.

Full UNO had lower mean EASE than either step alone in these analyses.

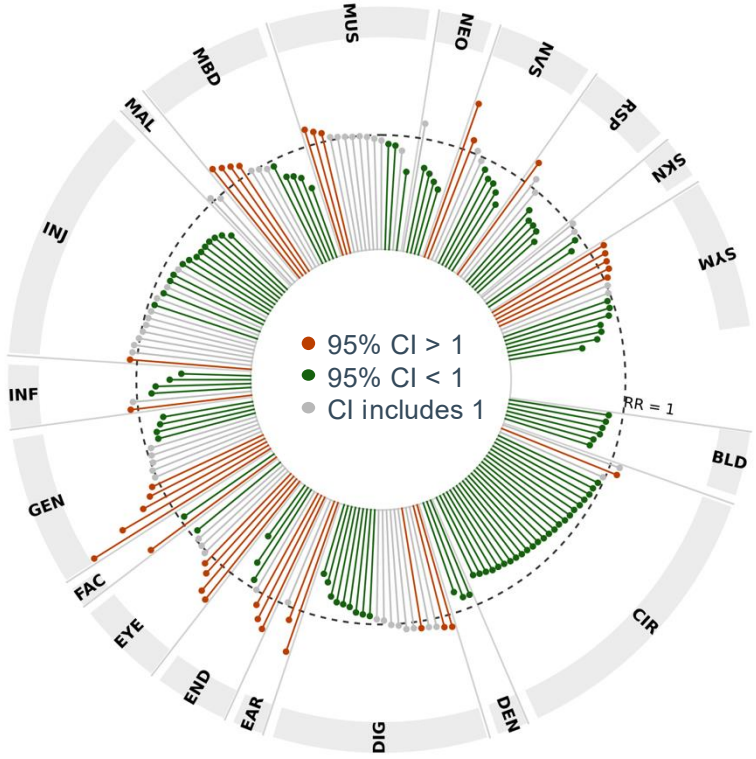
Same base-model checkpoint, evaluation cohort, and held-out NCO set within each analysis.

Penn Medicine

Result 1: Outcome-wide profiling results

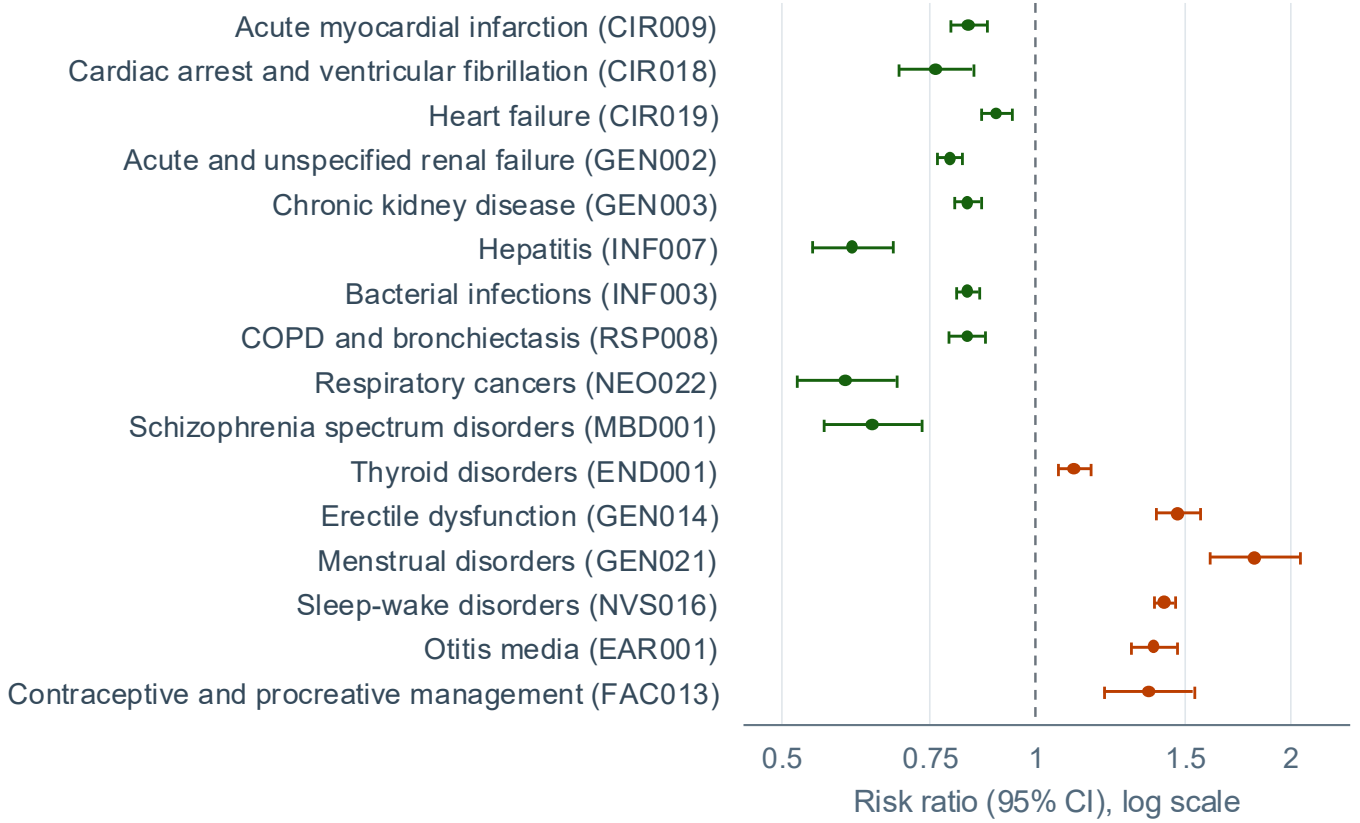
GLP-1RAs vs DPP-4is | 189 CCSR-defined clinical outcomes

a 189 outcomes by body system



Dashed circle: RR = 1

b 16 outcomes discussed



One treatment comparison → an outcome-wide association profile.

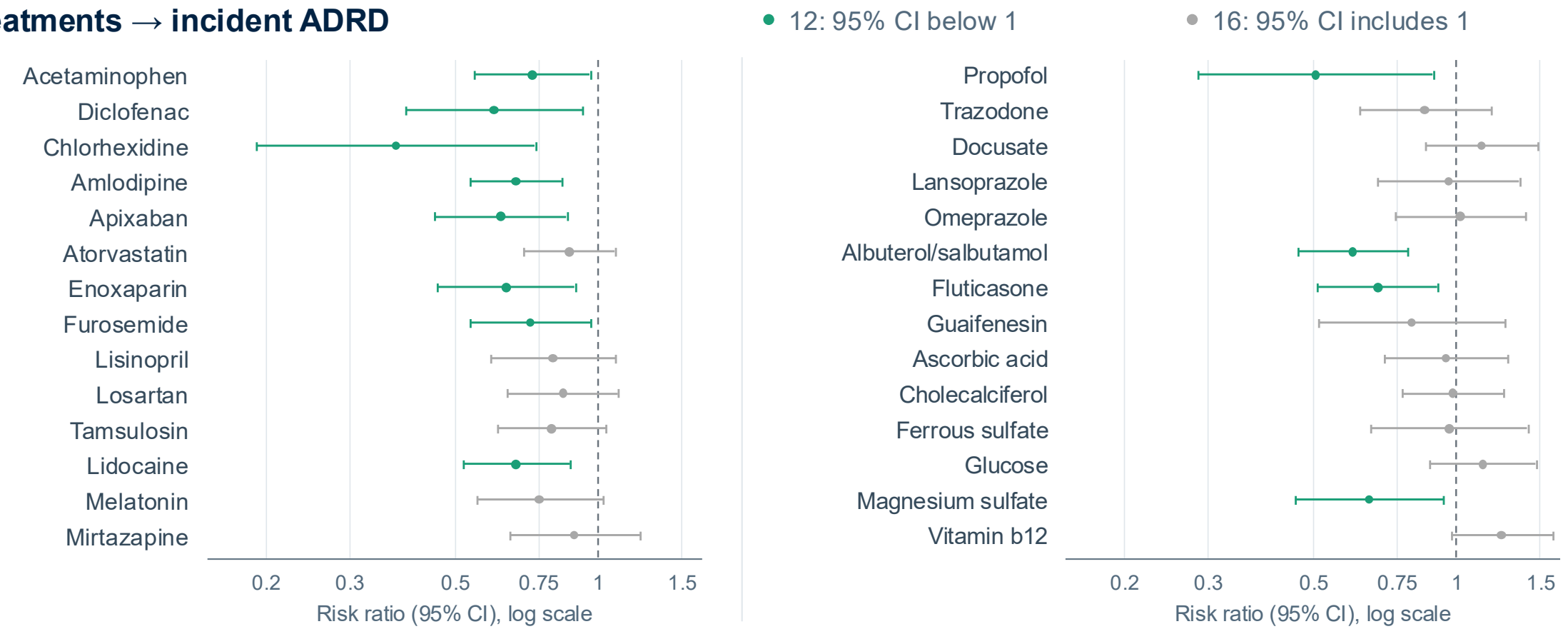
all 189 outcomes in panel a and all 16 main-text outcomes in panel b. CCSR: Clinical Classifications Software Refined.

Penn Medicine

Result 2: ADRD drug repurposing

Penn Medicine, 2012–2020 | 16,184 patients with MCI | 28 candidate medications

28 treatments → incident ADRD



Prioritize candidates for further study.

MCI: mild cognitive impairment. ADRD: Alzheimer's disease and related dementias.

Penn Medicine

Summary

1 NCO-guided model refinement

UNO uses negative control outcomes to guide neural pruning and fine-tuning.

2 Improved reliability

Lower held-out EASE and better propensity-score calibration than the Base Model.
Performance depends on relevant NCOs. Residual confounding may remain.

3 Compatible with existing analysis pipelines and large-scale studies

Applicable across tasks and neural network architectures.

Acknowledgments

*† Co-first authors * Senior authors*

- **Yiwen Lu†** University of Pennsylvania
- Huiyuan Wang University of Pennsylvania
- Dazheng Zhang University of Pennsylvania
- Huilin Tang University of Pennsylvania
- Bingyu Zhang University of Pennsylvania
- Lu Li University of Pennsylvania
- Wenjie Hu University of Pennsylvania
- Jiajie Chen University of Pennsylvania
- Tingyin Wang University of Pennsylvania
- Han Wu The Harker School
- Ting Zhou University of Pennsylvania
- Yuqing Lei University of Pennsylvania
- Xinyao Jian University of Pennsylvania
- Tsai Hor Chan University of Pennsylvania
- Siqi Chen University of Pennsylvania
- Yong Lin Eli Lilly and Company
- Demissie Alemayehu Pfizer Inc.
- **James Zou*** Stanford University
- **David Asch*** University of Pennsylvania
- **Yong Chen*** University of Pennsylvania

github.com/Penncil/UNO

Thank you!

A Multi-Database, Retrospective, Long-term, Observational Safety Study of Pediatric Patients with Inflammatory Bowel Disease

Kevin Haynes
Tim Hoops
Pankti Bhatt
Robert Suruki
Patrick Ryan
Jared Houghtaling

Why this study was needed

Background

Infliximab is widely used for pediatric Crohn's disease and ulcerative colitis

- CD pediatric approval: May 19, 2006
- UC pediatric approval: Sept 23, 2011

Long-term pediatric safety data remain limited, analysis complements post-marketing registry commitment

Safety outcomes of interest included hospitalization, serious infection, tuberculosis, malignancy, and autoimmune disease

Clinical Question

What is the comparative safety of infliximab among pediatric patients with IBD?

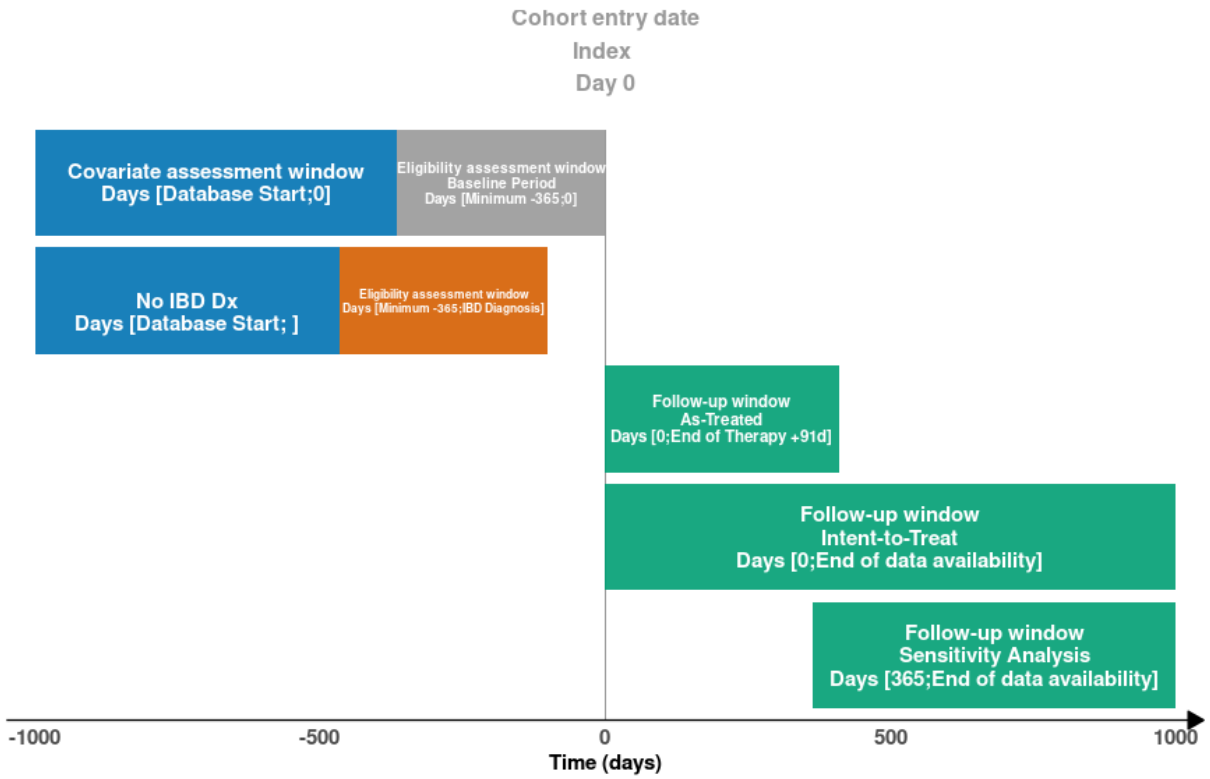
Comparators

Non-TNF therapies

Adalimumab

What did a standardized, multi-database RWE framework reveal about pediatric infliximab safety?

Study design: new-user, active-comparator safety evaluation



- Retrospective, observational cohort study across five U.S. electronic healthcare databases
- Databases mapped to OMOP CDM v5.4
- Patients aged <17 years with IBD and ≥1 IBD complication
- Two primary comparisons: infliximab vs non-TNF therapies and infliximab vs adalimumab
- As-treated risk windows for hospitalization, serious infection, TB, and autoimmune disease; intent-to-treat for malignancy

Design principle:
Emulate a credible safety comparison before looking at results

Standardized RWE analytics

What makes this study different: validity diagnostics before interpretation

1 **OMOP CDM**

Common structure across databases

2 **OHDSI HADES tools**

Reproducible distributed analytics

3 **Prespecified diagnostics**

Power, bias, equipoise, balance

4 **Bayesian meta-analysis**

Synthesize only reliable estimates

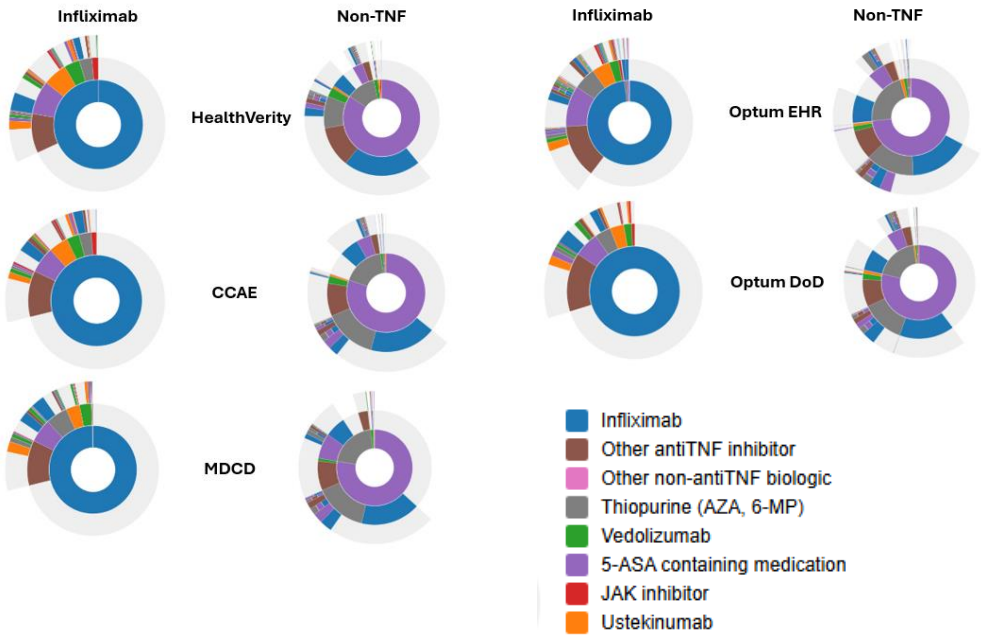
Diagnostic thresholds

MDRR <10 | EASE <0.25 | Preference-score equipoise | Max SMD <0.1

Only analyses passing diagnostics were unblinded and synthesized.

Population: Pediatric IBD cohorts across five databases

>9,000	3.1 yrs	~12.5	5
new infliximab pediatric patients	mean follow-up	mean age at cohort entry	U.S. OMOP databases



Real-world treatment dynamics

Patients largely adhered to initial treatments

Therapeutic switching was observed in routine practice

Non-TNF initiators most commonly switched to infliximab or other anti-TNF agents

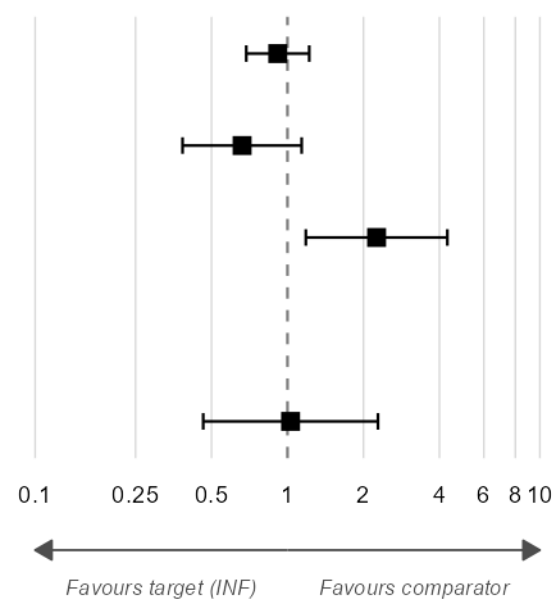
Switching is informative context

Infliximab vs non-TNF initiators

Primary results: no increased risk for hospitalization, serious infection, or malignancy

Meta-analysis Summary — nonTNF

Outcome	Number of databases	INF	nonTNF	INF Events	nonTNF Events	INF IR per 1,000 PY	nonTNF IR per 1,000 PY	HR (95% CI)
On-Treatment Meta-analysis								
Hospitalization	4	2,990	6,183	496	857	98.3	130.7	0.91 (0.69, 1.22)
Serious Infection	4	7,317	10,540	198	171	14.0	14.4	0.66 (0.38, 1.14)
Autoimmune	4	7,838	10,869	446	106	30.1	8.7	2.25 (1.18, 4.30)
Intent-to-Treat Meta-analysis								
Malignancy	4	7,879	10,932	<52	<40	--	--	1.03 (0.46, 2.28)



No increase

No increase

Elevated

No increase

TB was too rare for inferential analysis.

Focused interpretation

Interpreting the autoimmune signal: primarily psoriasis and consistent with class effect

Infliximab vs non-TNF

HR 2.26
95% CI 1.18 - 4.31

Autoimmune disease increase was driven predominantly by new-onset psoriasis

Psoriasis accounted for >70% of autoimmune events among infliximab users

Other autoimmune outcomes were uncommon

Infliximab vs adalimumab

HR 0.96
95% CI 0.53 - 1.72

Interpretation

No differential risk between anti-TNF agents

Take-home message

Large standardized multi-database RWE study of pediatric IBD: Infliximab was not associated with increased risk of hospitalization, serious infection, or malignancy

Scalable path to generate transparent, reproducible, regulatory-relevant pediatric safety evidence



ORIGINAL ARTICLE | Open Access

Active Safety Surveillance Using Real-World Evidence (ASSURE) Implementation: Transparent and Reproducible Real-World Evidence Standardized Framework to Support Safety Signal Evaluation[1]

Thank you

Backup slide

Infliximab versus adalimumab results

Outcome	HR (95% CI)	Interpretation
Any hospitalization	0.95 (0.54, 1.66)	No differential risk
Serious infection	0.77 (0.37, 1.62)	No differential risk
New malignancy	1.18 (0.45, 3.08)	No differential risk
New autoimmune disease	0.96 (0.53, 1.72)	No differential risk

Backup slide

Objective study validity diagnostics

