

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

Haeun Lee¹, Tom Richards², Benjamin Martin¹, Paul Nagy¹, Hadi Kharrazi^{1,2}, Jonathan Weiner², Chintan Pandya²

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.





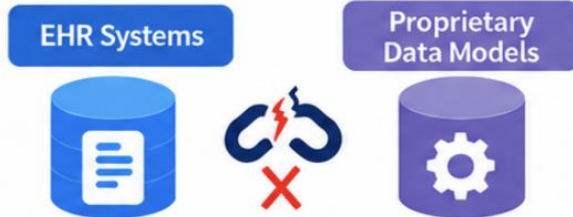
Data Heterogeneity Challenges for Population Health Analytics and Integration



1

Data heterogeneity across healthcare organizations

- Inconsistent coding, definitions, and data representation
- Differences in data granularity
- Structural heterogeneity in source data models and schemas



2

Lack of common data representation across systems

- Clinical data inputs remain source-specific
- Limits consistent integration and reusable analytics



ACG

CMS-HCC

DRG

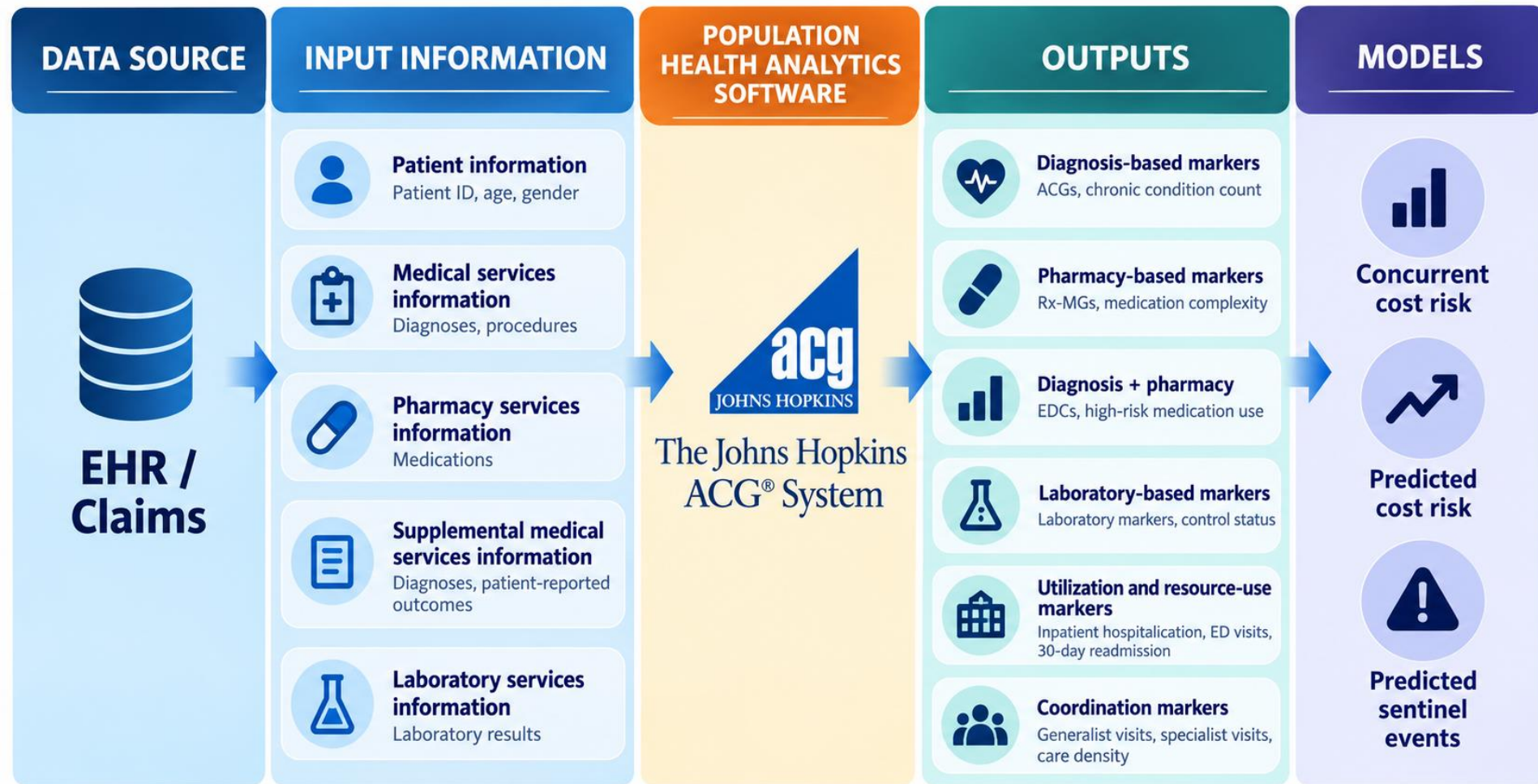
3

Gaps in existing population health research

- Limited evaluation of population health risk frameworks using standardized data models
- Unclear preservation of population health measures



What are ACG (Adjusted Clinical Groups)





Comprehensive Population Health Analytics

Johns Hopkins ACG System quantifies the **health of a population**, identifies **targeted patient needs**, and drives action to **improve outcomes**.



Utilization Markers

- Avoidable ED Visits
- Unnecessary inpatient & readmissions
- Specialist overuse
- Care Coordination Needs



Pharmacy Markers

- Novel Medication Complexity Score (MCS)
- Untreated conditions
- Pharmacy-based needs markers
- Opioid markers



Predictive Models

- 10+ risk models allowing for proactive interventions:
 - High cost
 - Admissions and Readmissions
 - Persistent High Utilizer



Patient Health Needs

- Actionable Care Modifiers
- Frailty Markers
- Multimorbidity & Complexity
- Social Needs
- Psychiatric Needs

Key Applications and Use Cases



Population health monitoring



Clinical screening and care management



Provider performance assessment



Capitation and rate setting



Study Objectives



Knowledge gap

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

1

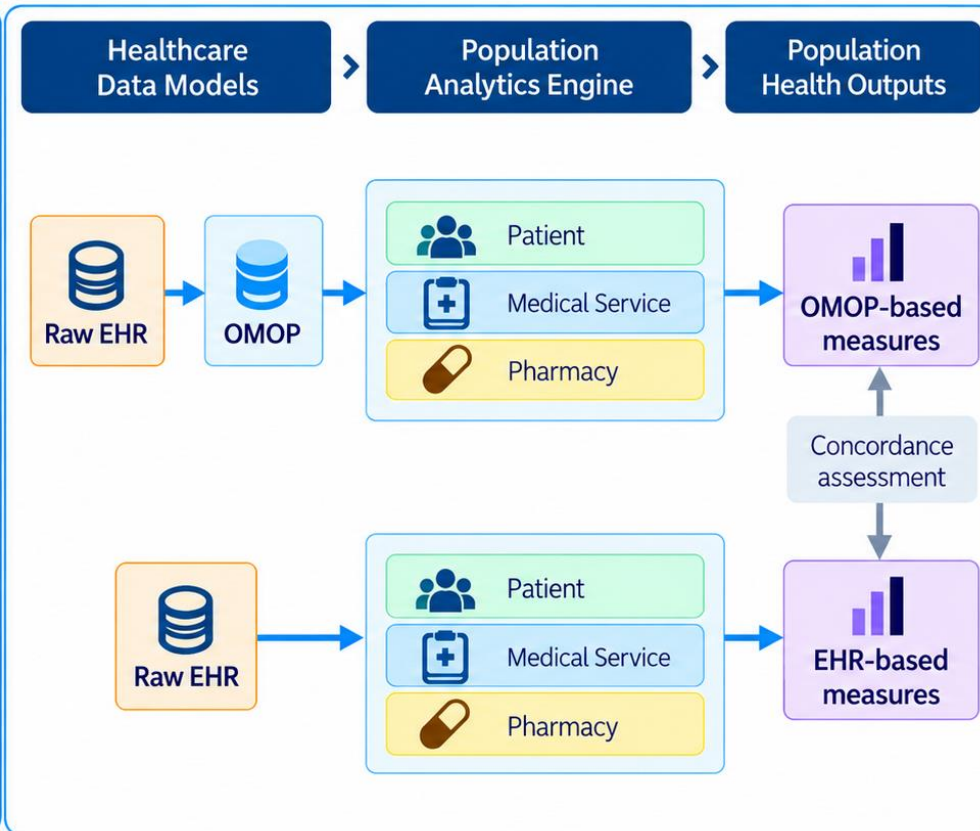
Derive comparable population health input variables from source EHR and OMOP data

2

Generate population health measures using a common analytics platform

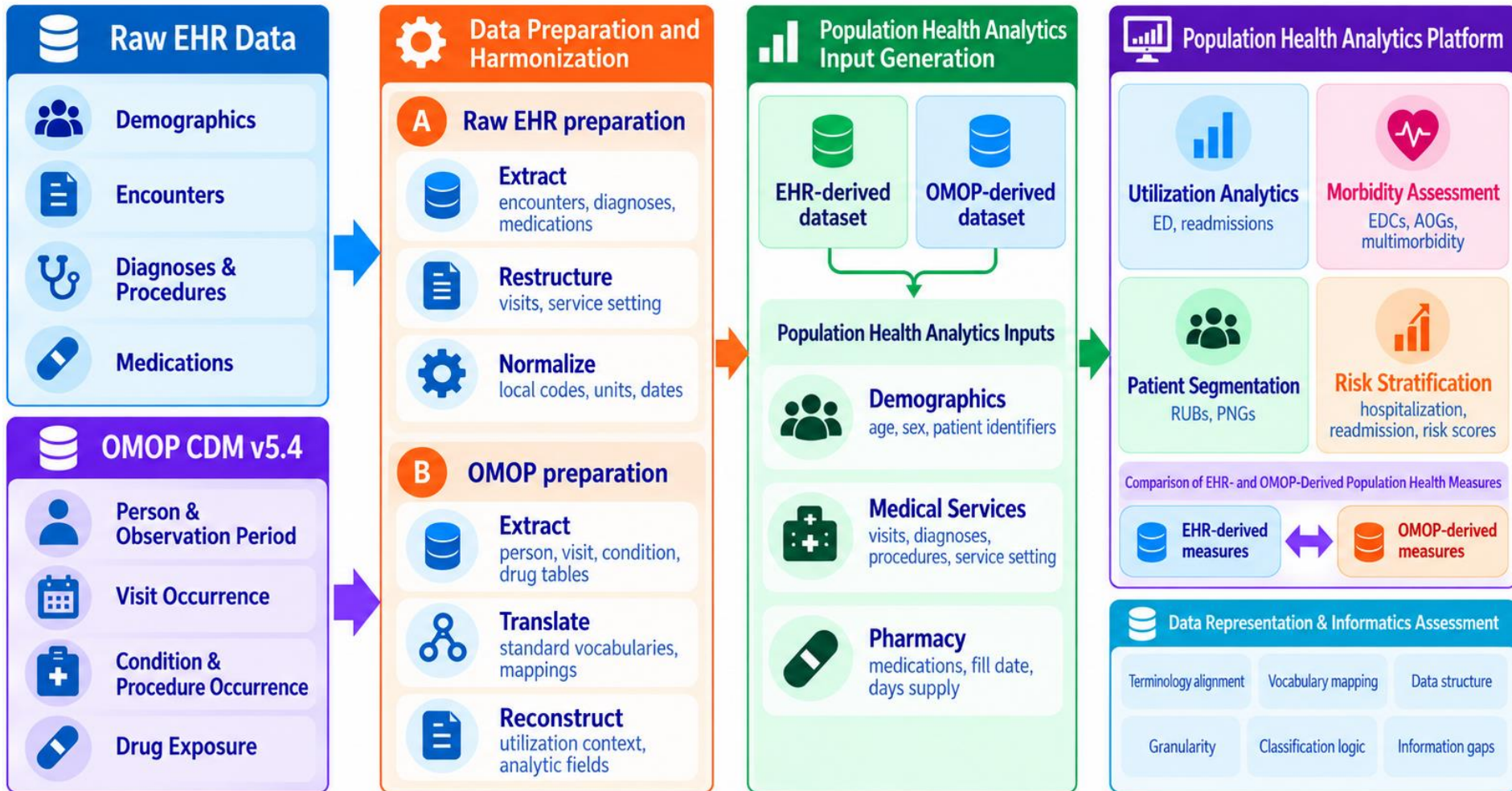
3

Evaluate concordance of OMOP- and EHR-derived population health measures and risk outputs



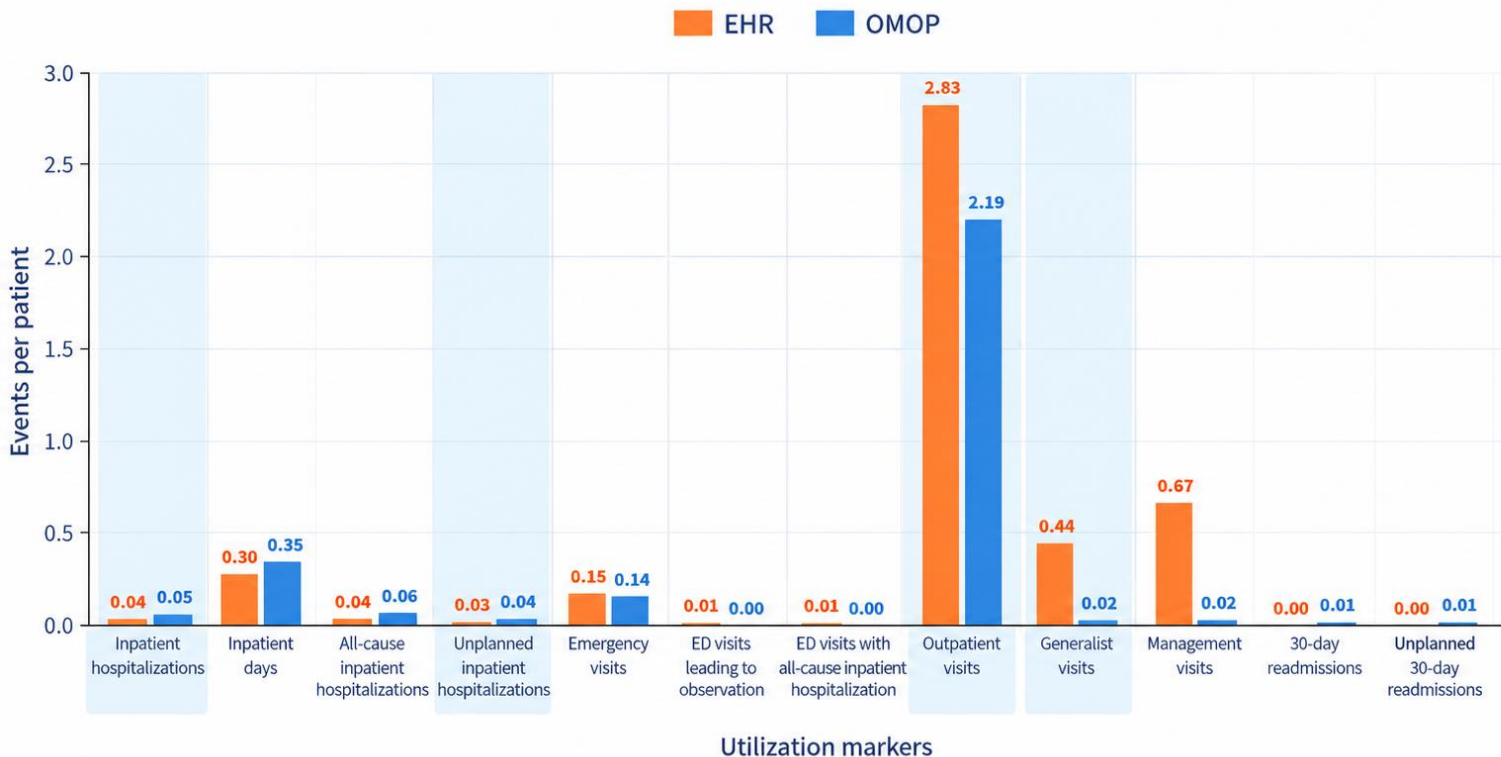


OMOP CDM-Based Data Transformation Framework





Comparison of Healthcare Utilization between EHR vs. OMOP

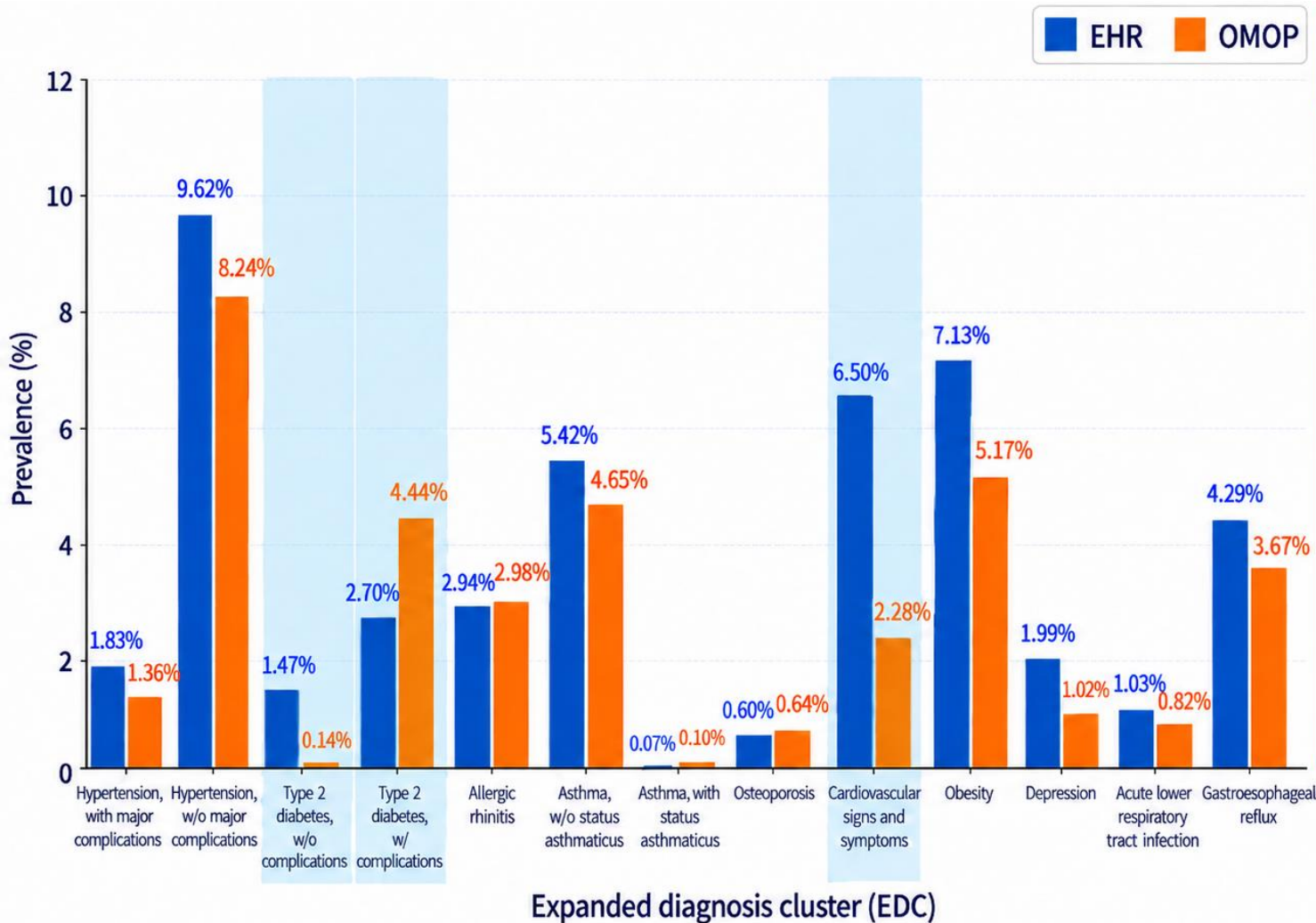


Key results

- 1 Higher in OMOP:** inpatient hospitalizations
- 2 Higher in OMOP:** unplanned inpatient hospitalizations
- 3 Lower in OMOP:** outpatient visits
- 4 Lower in OMOP:** generalist visits



Comparison of Chronic Conditions Prevalence Between EHR vs. OMOP

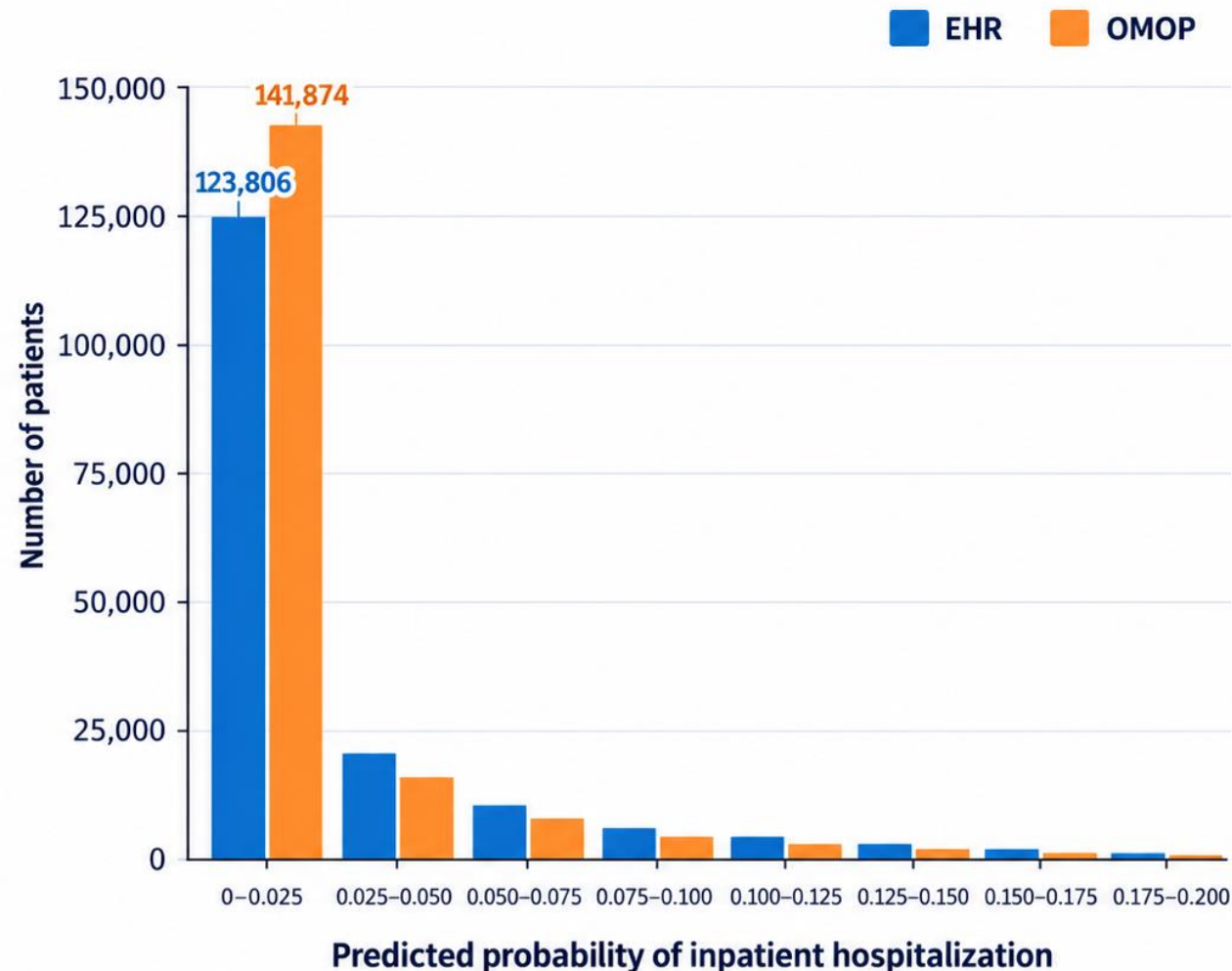


Key results

- 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%)



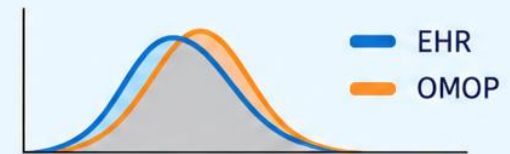
Distribution of Predicted Inpatient Hospitalization Probability



Key results

1 More patients in the lowest-risk bin in OMOP
86.4% vs 75.4%

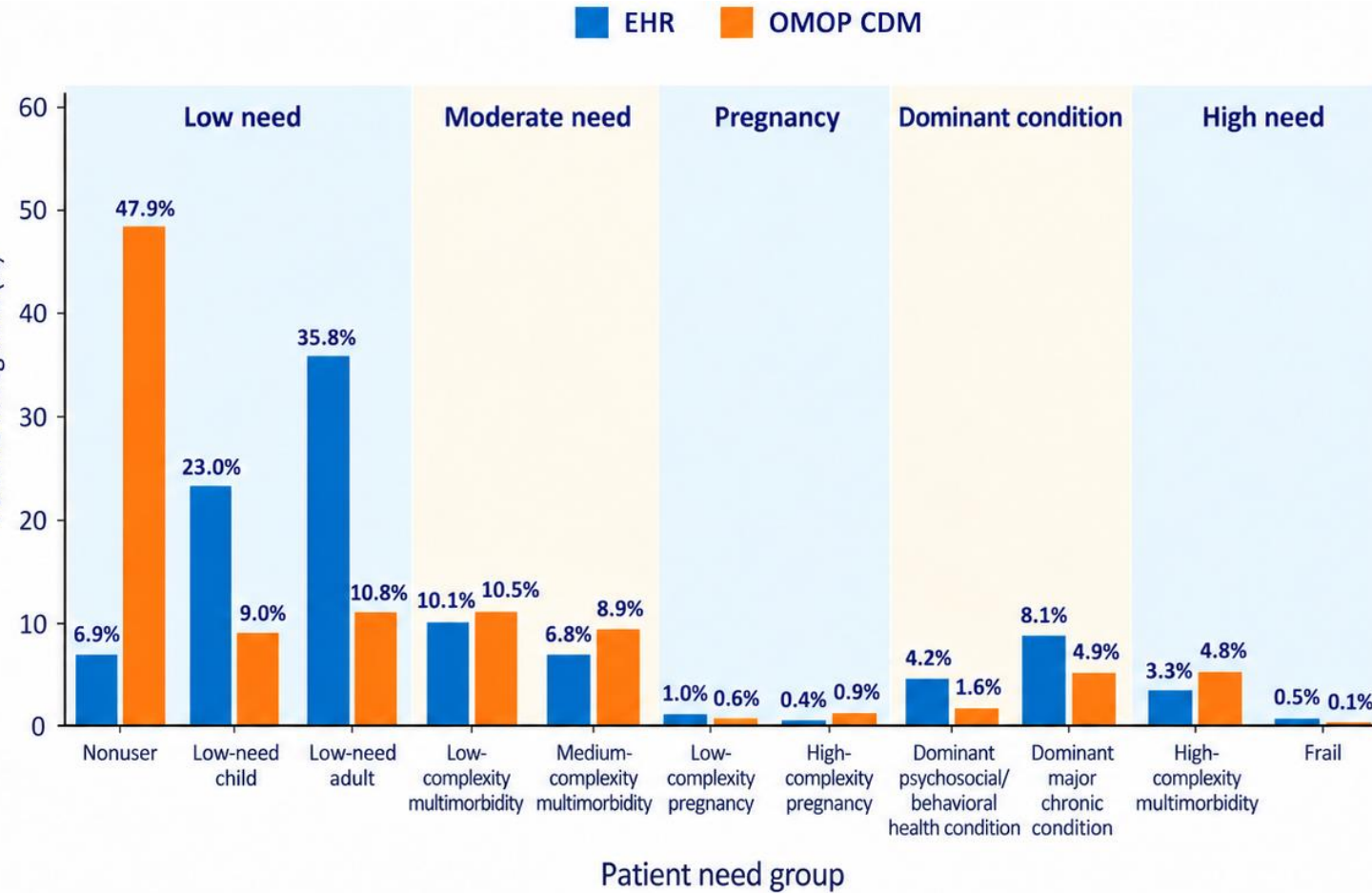
2 Overall risk distribution broadly similar



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

Classification of Data Mapping Differences

OMOP CDM	Mapping Difference	EHR	Key results
 PERSON.person_id	 None (Direct Mapping) Identical field structure and data type Direct mapping: no transformation	 patient_id	1  Most OMOP fields mappable
 gender_concept_id	 Semantic Difference Same concept, different values, units, or conventions Translate values or normalize	 sex (M/F/U)	
 RxNorm drug_concept_id	 Terminology Difference Same clinical concept, different code systems Vocabulary mapping / crosswalk	 NDC (rx_cd)	
 ZIP3	 Granularity Difference Same information, different granularity Aggregate or enrich to align	 ZIP5	
 visit_concept_id	 Categorization Difference Different classification schemes or business rules Rule-based regrouping	 CMS service place	
 visit_occurrence_id	 Structural Difference Different table structure or relational formats Reshape data (joins / aggregation)	 outpatient_visit_id	
 Not represented	 Information Gap Available in source EHR but absent from OMOP CDM Cannot be derived from OMOP	 prescriptions dispensed medications	
			2  Granularity loss and information gaps matter
			3  Terminology and categorization require transformation



Where the differences come from

1



Terminology mapping

Differences in code mapping, concept granularity, and vocabularies.



2



Encounter and care setting representation

Differences in how encounters are defined, classified, and attributed.



3



Clinical detail and data availability

Differences in what is structured, mapped, and available after transformation.



4



ETL and data transformation rules

Differences in inclusion criteria, transformation logic, and data filtering.



5



Data representation and information gaps

Loss of detail, unstandardized concepts, and incomplete capture in OMOP.

In EHR (available)

- Local codes
- Department/specialty
- Additional clinical detail

In OMOP (may be missing or aggregated)

- Standardized concepts only
- Mapped or high-level categories
- Limited contextual detail



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
-