

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

Haeun Lee, MS¹, Tom Richards, MS², Benjamin Martin, PhD¹, Paul Nagy, PhD¹, Hadi Kharrazi, MD, PhD^{1,2}, Jonathan Weiner, DrPH², Chintan Pandya, PhD, MPH, MBBS²,

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

Background

Patient-level clinical data generated within healthcare systems form the foundation of population health management and risk stratification, yet these analytics have traditionally been implemented within heterogeneous, site-specific electronic health record (EHR) environments^{1,2}. Population health studies require data harmonization to support consistent, comparable, and scalable analyses across institutions, highlighting the need for standardized analytical infrastructures³. Common data models (CDM) such as the Observational Medical Outcomes Partnership (OMOP) CDM integrate disparate EHR data into a unified structure for multi-institutional real-world data research³. However, morbidity groupers, risk markers, and predicted risk scores generated by population risk stratification frameworks are often produced outside the standardized data model, limiting their scalability within harmonized data environments. This study implemented population risk stratification directly within the OMOP framework, generated and evaluated risk outputs within the CDM, and incorporated population health analytics into a standardized data infrastructure.

Methods

We conducted a retrospective cohort study using raw EHR data and corresponding OMOP data derived from the same patient population at a large urban academic medical center in 2021. Core population health data elements were independently extracted from raw EHR records and the OMOP data, and each dataset was analyzed separately using the Johns Hopkins Adjusted Clinical Groups (ACG) System, a validated population health analytics platform that utilizes diagnosis, pharmacy, and utilization data to generate measures of health care utilization, morbidity burden, and risk stratification (Figure 1). Bivariate analyses were performed to compare population health metrics derived from EHR and OMOP data and to inform the design of an OMOP extension for structured incorporation of key risk measures.

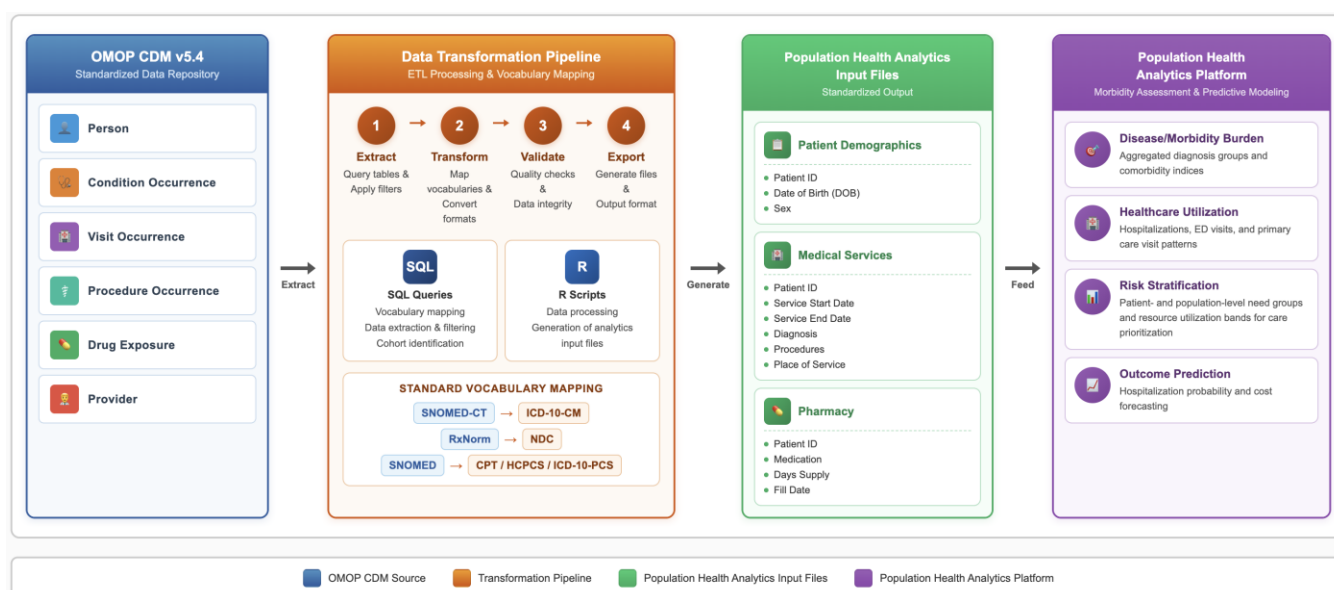


Figure 1. OMOP CDM-Based Data Transformation Framework for Population Health Analytics

Results

Compared with EHR-based analyses, OMOP-based analyses identified higher inpatient hospitalization rates (0.05 vs. 0.04) and higher 30-day readmission rates (0.006 vs. 0.004). Overall emergency department visit rates were similar; however, OMOP data captured relatively more primary care–treatable and injury-related ED visits, whereas EHR data captured more non-emergent ED visits (Table 1). OMOP analyses identified fewer outpatient visits (2.2 vs. 2.8 visits per patient) and substantially fewer generalist visits (0.02 vs. 0.44). Overall morbidity burden was captured higher in EHR-based analyses, with fewer patients classified with zero Adjusted Diagnosis Groups (44% vs. 51%) and higher prevalence of psychosocial conditions (7.8% vs. 4.2%). EHR-based analyses identified more hypertension, asthma, and cardiovascular conditions, while OMOP-based analyses identified relatively more Type 2 diabetes with complication and osteoporosis (Figure 2). OMOP analyses classified fewer patients into high-need groups and more into moderate-need groups, with a substantially higher proportion of non-users (48% vs. 7%) (Figure 3). EHR-based analyses identified more patients at moderate-to-high hospitalization risk compared to OMOP-based analyses (Figure 4).

Table 1. Study Population Characteristics, Morbidity Burden, and Healthcare Utilization in EHR and OMOP CDM		
N (%)	EHR	OMOP
Demographics		
Age Group		
Children (0-17 years)	52,141 (31.8%)	52,135 (31.7%)
Youth/young adults (18-44 years)	65,470 (39.9%)	65,451 (39.9%)
Adults (45-64 years)	32,447 (19.8%)	32,527 (19.8%)
Seniors (≥ 65 years)	14,195 (8.7%)	14,140 (8.6%)
Sex		
Male	69,002 (42.0%)	68,996 (42.0%)
Female	95,251 (58.0%)	95,257 (58.0%)
Morbidity Groupers		
Patient with 0 ADGs	72,645 (44.2%)	84,248 (51.3%)
Psychosocial conditions	33,294 (7.8%)	16,235 (4.2%)
Hypertension, with major complications	3006 (1.8%)	2226 (1.4%)
Type 2 diabetes, w/ complication	4433 (2.7%)	7292 (4.4%)
Healthcare Utilization, N (%)		
Inpatient Hospitalizations	6,427 (0.04%)	7,922 (0.05%)
Outpatient Visits	464,953 (2.8%)	360,346 (2.2%)
Patients with Major Procedures	1464 (0.0%)	2857 (0.0%)
Emergency Visit		
ED Visit: Non-emergent	23,856 (0.15%)	23,099 (0.14%)
ED Visit: Emergent, primary care treatable	11,605 (44.3%)	8,199 (35.1%)
ED Visit: Emergent, primary care treatable	5,841 (22.3%)	5,298 (22.7%)
ED Visit: Injury, severe	211 (0.8%)	763 (3.3%)
Generalist visits	72,486 (0.44%)	2,921 (0.02%)
30-Day Readmissions	702 (0.004%)	1,131 (0.006%)
Resource utilization band, N (%)		
Non-user	11,475 (7.0%)	78,707 (47.9%)
Healthy users	80,065 (48.7%)	10,888 (6.6%)
Low	21,410 (13.0%)	22,632 (13.8%)
Moderate	33,753 (20.6%)	29,014 (17.7%)
High	11,479 (7.0%)	14,826 (9.0%)
Very high	6,071 (3.7%)	8,186 (5.0%)



Figure 2. Comparison of Expanded Diagnosis Cluster (EDC) Prevalence Between EHR and OMOP

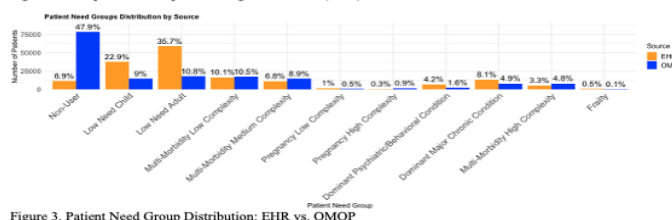


Figure 3. Patient Need Group Distribution: EHR vs. OMOP

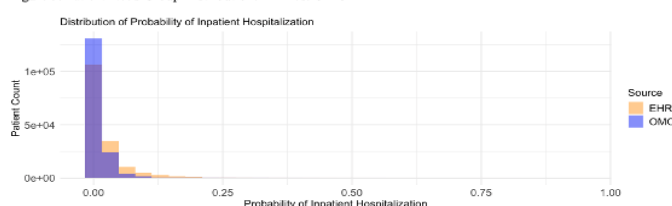


Figure 4. Comparison of Predicted Hospitalization Risk Between EHR and OMOP

Conclusions

Analyses comparing population health outputs generated from OMOP CDM and raw EHR data revealed meaningful differences in healthcare utilization measurement, morbidity assessment, and risk stratification. OMOP-based analyses identified slightly higher inpatient hospitalization and 30-day readmission rates, but substantially fewer outpatient encounters and generalist visits, suggesting differences in encounter classification during OMOP transformation. Morbidity burden was consistently higher in EHR-based analyses, with greater prevalence of psychosocial conditions, hypertension, asthma, and cardiovascular disease, while OMOP captured more type 2 diabetes with complications and osteoporosis, indicating potential variation in diagnosis capture and vocabulary mapping. These differences translated into divergent risk stratification: OMOP classified fewer patients into high-need groups and more as non-users, whereas EHR identified larger proportions at moderate-to-high hospitalization risk. These findings underscore the importance of validating population health analytics and risk adjustment models when implementing them within OMOP-standardized environments to ensure clinical fidelity and operational reliability. This study demonstrates the feasibility of implementing a population health analytic framework within the OMOP CDM, providing an interoperable framework for population health analytics to support research and operational goals within and across healthcare organizations.

References

1. Gamache R, Kharrazi H, Weiner JP. Public and population health informatics: the bridging of big data to benefit communities. *Yearb Med Inform.* 2018;27(1):199-206. doi:10.1055/s-0038-1667081.
2. Wells BJ, Downs SM, Ostasiewski B. Using electronic health records for the learning health system: creation of a diabetes research registry. *JMIR Med Inform.* 2022;10(9):e39746. doi:10.2196/39746.
3. Ochola M, Kiwuwa-Muyingo S, Bhattacharjee T, et al. Harmonizing population health data into OMOP common data model: a demonstration using COVID-19 sero-surveillance data from Nairobi Urban Health and Demographic Surveillance System. *Front Digit Health.* 2025;7:1423621. doi:10.3389/fdgh.2025.1423621.