PRECISION-Connect: AI-Ready Multimorbidity and SDOH Risk Vectors for Explainable 30-Day Readmission and County-Level Disparity Modeling

Mirna Elizondo
Datalab12
Texas State University
San Marcos, Texas, USA
m_e172@txstate.edu

Daniel J Amante
Population and Quantitative Health Sciences
UMass Chan Medical School
Worcester, Massachusetts, USA
daniel.amante@umassmed.edu

Jelena Tešić*
Computer Science
Texas State University
San Marcos, TX, USA
jtesic@txstate.edu

Abstract

Multimorbid Medicare home health patients face a high 30-day readmission risk that is associated with not only clinical complexity but also neighborhood-level disadvantage. Existing prediction tools often compress this complexity into opaque risk scores, limiting clinicians, care managers, and policymakers' ability to see which social or clinical factors are associated with elevated risk or which counties bear the greatest burden. We develop *PRECISION-Connect*, a population health analytics system that integrates structured multimorbidity and social determinants of health (SDOH) factors with county-level disparity indices for Medicare home health patients in Texas. The structured risk vector preserves age, BMI, chronic disease indicators, frailty scores, healthcare utilization, and county-level SDOH data as distinct, queryable components. These supporting analyses separate clinical from contextual associations with predicted readmission risk. Using over 440,000 home health episodes, *PRECISION-Connect* achieves stable discrimination and calibration across multimorbid subgroups. At the same time, county-level SDOH features yield small but consistent improvements in predictive performance beyond those achieved with comorbidity and frailty alone. The study is observational and predictive; all reported relationships between SDOH, multimorbidity, and 30-day readmission are associative, and the models do not identify causal effects. An interactive clinician-facing dashboard surfaces these patterns to support equity-aware follow-up strategies and hypothesis generation.

CCS Concepts

• **Applied computing** → **Health care information systems**; **Bioinformatics**; • **Computing methodologies** → *Machine learning*; • **Information systems** → Geographic information systems.

Keywords

multimorbidity, population health analytics, explainable AI, Structured risk vectors, disparity metrics, geospatial and county-level modeling

ACM Reference Format:
Mirna Elizondo, Daniel J Amante, and Jelena Tešić. 2026. PRECISION-Connect: AI-Ready Multimorbidity and SDOH Risk Vectors for Explainable 30-Day Readmission and County-Level Disparity Modeling. In *17th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '26), June 30-July 03, 2026, Rende (CS), Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3807503.3819383

1 Introduction

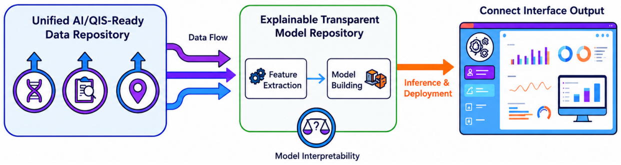


Figure 1: PRECISION-Connect system overview: heterogeneous clinical and contextual data are aggregated into an AI/QIS-ready repository, transformed into an explainable model repository, and deployed via the Connect interface for inference, monitoring, and decision support.

Medicare patients often manage multiple chronic conditions while living in communities with increased social risk. Identifying patients in need of additional support requires integrating clinical and social determinants of health (SDOH) data in a consistent, interpretable model. While health care systems are increasingly collecting SDOH data in their electronic health records (EHRs), the use of population health tools that integrate clinical and SDOH data to support equity-focused interventions remains minimal.

We model clinical and contextual factors within a single, interpretable framework to predict 30-day readmission and to examine county-level disparities. Our approach constructs structured visit-level risk vectors that integrate comorbidity, frailty, utilization, and SDOH features to predict 30-day readmission and support disparity analysis for people with chronic diseases such as diabetes mellitus (DM), heart failure (HF), hypertension (HTN), and multimorbid DM, HF, and HTN.

PRECISION-Connect operationalizes this modeling through a deployable analytics system that unifies end-to-end extract– transform– load (ETL), structured multimorbidity and SDOH risk data, county-level disparity indices, and explainable gradient-boosted models within an interactive dashboard. Specifically, it (i) preserves the multidimensional risk structure instead of collapsing it into a single opaque score, (ii) links over 440,000 Texas home health encounters to county-level SDOH data to identify readmission disparities, and (iii) provides cohort- and patient-level explanations

---

[*] All authors contributed equally to this research.

BCB '26, Rende (CS), Italy

ACM ISBN 979-8-4007-2653-8/2026/06
https://doi.org/10.1145/3807503.3819383

that clinicians and care managers can use to tailor patient intervention and community partnerships.

Contributions This paper contributes to multimorbidity-focused, SDOH-informed population health analytics for Medicare home health beneficiaries in five ways. We define a visit-level multimorbidity and SDOH risk vector R_i that keeps age, BMI category, chronic conditions, diagnostic burden (for example, the count and complexity of ICD-coded diagnoses for the visit), comorbidity and frailty scores, utilization measures, and county-level SDOH attributes as distinct, queryable dimensions rather than compressing them into a single risk score. We build an end-to-end, deployable sociotechnical pipeline that links CMS OASIS assessments with comorbidity, frailty, and county-level SDOH to create reusable datasets for large-scale modeling. We implement PRECISION- Connect, an interactive analytic system that delivers explainable readmission models and county-level dashboards for clinicians and care managers. We provide evidence for equity-focused targeting by showing that SDOH-enriched, comorbidity-aware risk vectors yield stable readmission predictions across multimorbid subgroups while supporting disparity-aware county queries not available in existing commercial population health dashboards. Because SDOH account for a large share of modifiable health outcomes, and recent work highlights AI's potential to capture SDOH and support disparity-reducing interventions and policy, our findings underscore the value of integrating SDOH into population health analytics [31].

2 Related Work

Multimorbidity is highly prevalent among older adults and is associated with increased healthcare utilization, polypharmacy, and adverse outcomes in home health populations [24]. Existing comorbidity indices such as the Charlson and Elixhauser scores are widely used for risk stratification but are typically evaluated in isolation and emphasize predictive accuracy over system-level integration, interpretability, or usability for population health exploration [7, 10, 30]. In contrast, our work targets Medicare home health encounters. It aims to build an AI-ready risk representation and disparity framework that can be operationalized on a deployable, clinician-facing platform for 30-day readmission prediction and county-level disparities. We focus on Texas because its large, diverse Medicare home health population provides a sufficient sample size for stratified analysis, and high-quality county identifiers in Outcome and Assessment Information Set (OASIS) records enable reliable linkage to SDOH datasets while balancing statistical power, data quality, and relevance for state-level insights [5, 12].

Multimorbidity, Frailty, and Readmission Risk. Several studies have demonstrated that integrating comorbidity indices (for example, Charlson or Elixhauser scores) with EHR data improves readmission risk stratification and outcomes research [7, 10]. Traditional risk stratification models such as HOSPITAL and LACE rely primarily on clinical and utilization variables and achieve only moderate discrimination, motivating the development of richer feature sets. However, the models achieve only moderate discrimination and translate poorly to new populations [8, 29]. These issues have motivated the development of broader feature sets and more flexible modeling approaches. Recent work moves beyond simple clinical predictors toward multifactorial, machine-learning-based models that more fully capture the complexity of patient health. Models that incorporate frailty indicators, comorbidity burden, and utilization history have improved predictive performance for 30-day readmission risk, with SHAP-based analyses elucidating how different factors contribute to individual predictions [33]. These studies demonstrate that frailty and comorbidity capture complementary aspects of vulnerability that are not captured by single scores. Our work uses validated measures of comorbidity and frailty. Still, instead of treating the features as a flat vector, we organize them into a structured risk representation that can be queried at both the episode and county levels.

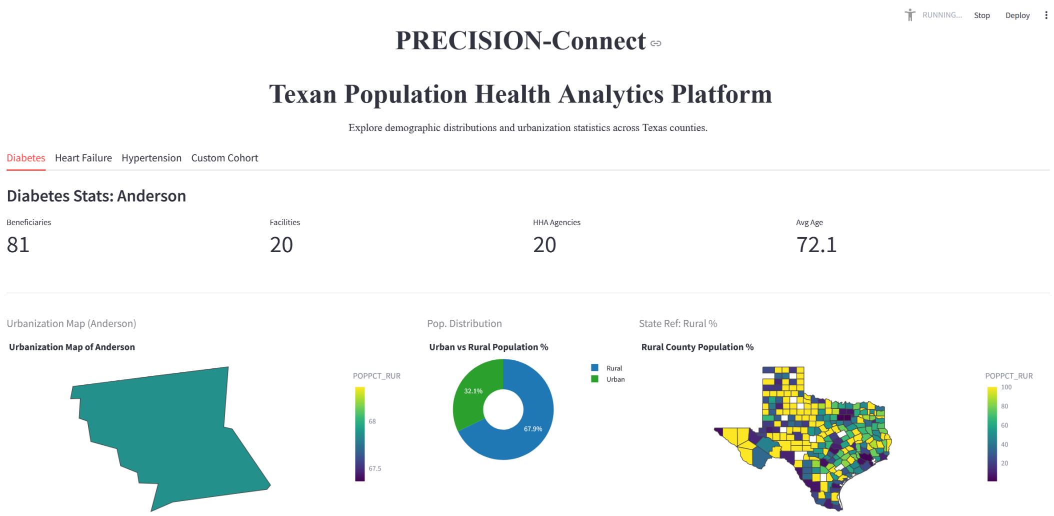
Social Determinants of Health and Readmissions. Parallel investigations into social determinants of health (SDOH) have shown that socioeconomic and access-related factors—such as poverty, transportation barriers, and insurance or affordability constraints—are independently associated with readmission risk after major hospitalizations [6]. Models that incorporate SDOH variables at both individual and community levels incrementally improve predictive discrimination compared with clinical-only baselines, particularly among older adults and patients from socioeconomically disadvantaged areas [6, 29]. Deep learning models that integrate temporal EHR patterns, comorbidity indices, and other longitudinal signals can improve readmission prediction beyond traditional clinical-only approaches [28]. Nevertheless, most studies focus on single health systems or inpatient cohorts and stop at model development rather than full deployment.

In summary, most existing systems evaluate subsets of the broader feature space (for example, comorbidity plus frailty, or SDOH plus clinical features) in isolation [6, 33]. Few published models simultaneously combine traditional comorbidity scores (Charlson, Elixhauser), frailty measures, SDOH variables, and explainable machine learning techniques into a unified prediction framework, and even fewer operationalize such frameworks within clinician-facing, deployable platforms [17, 33]. In particular, we are not aware of prior systems that (i) construct a structured multimorbidity- and SDOH-enriched risk vector designed for both prediction and county-level disparity analysis in home health populations, and (ii) embed this representation, along with county-level disparity indices, into a deployable, interactive analytics platform for clinicians and care managers. PRECISION-Connect addresses these gaps by unifying validated comorbidity and frailty indices, county-level SDOH features, and explainable gradient-boosted models within a single, structured risk vector and county-level disparity framework [1, 2]. Unlike prior dashboards, it preserves the multidimensional risk structure rather than collapsing patient risk into a single score and operationalizes this representation in a clinician-facing system that supports both 30-day readmission prediction and county-level SDOH-disparity analytics for Medicare home health cohorts [17].

3 Methodology

This section describes how we construct an AI-ready, multimorbidity- and SDOH-enriched representation of home health encounters; define a structured risk vector for 30-day readmission prediction; and compute county-level disparity metrics for SDOH-informed population health analytics [5, 12].

Data model and outcome definition We use the Outcome and Assessment Information Set (OASIS) from the Centers for Medicare



PRECISION-Connect

Texan Population Health Analytics Platform

Explore demographic distributions and urbanization statistics across Texas counties.

Diabetes Stats: Anderson

Beneficiaries	Facilities	HHA Agencies	Avg Age
81	20	20	72.1

Figure 2: PRECISION-Connect county comorbidity view for the diabetes cohort in Anderson County, TX, USA. Panels show (clockwise) cohort statistics, county map, and statewide urban–rural distribution

Table 1: Analytic cohort patient-level mortality and readmission by condition group. Rates are based on *ever deceased* and *ever readmitted* indicators.

Condition Group	*N*	*Mort. (%)*	*Readm. (%)*	*% Cohort*
None (D/HF/HTN)	3,413,381	87.36	10.80	70.35
Diabetes only	349,577	83.55	20.98	7.21
Hypertension only	509,605	86.14	18.99	10.50
Heart failure only	124,979	89.45	29.28	2.58
HF + HTN	207,792	91.86	36.13	4.28
Diabetes + HTN	139,120	87.29	28.69	2.87
Diabetes + HF	21,977	91.74	43.73	0.45
Diabetes + HF + HTN	85,261	93.82	48.45	1.76

& Medicaid Services (CMS), a standardized assessment instrument that captures demographics, functional status, diagnoses, and utilization for Medicare home health beneficiaries [5]. The primary outcome is all-cause 30-day hospital readmission following the index assessment, defined using linked utilization indicators in OASIS and, where available, associated claims [12].

Using the United States' OASIS extracts, we construct an analytic cohort of home health encounters in Texas. We restrict to adult beneficiaries (age ≥ 18) with valid county Federal Information Processing Series (FIPS) codes and identifiers sufficient for linkage to county-level SDOH tables, excluding records with structurally missing geocodes or implausible dates [20]. When patients contributed multiple eligible encounters over time, we retained the first qualifying episode per beneficiary to avoid within-patient correlation in model training and evaluation. Chronic condition flags (diabetes, heart failure (HF), hypertension (HTN)) and multimorbid combinations (for example, DHFHTN: diabetes + HF + HTN) are derived from ICD-10 diagnosis fields following established code lists [9, 32].

Table 1 summarizes patient characteristics and outcomes for the final Texas analytic cohort of 221,957 encounters, stratified by chronic condition group. Mortality and ever-readmission percentages are computed within each group.

Feature Engineering Our feature engineering goal is to encode multimorbidity, frailty, healthcare utilization, and SDOH data into a structured risk representation that supports both predictive modeling and disparity analysis [7, 10, 29].

The Body Mass Index (BMI) for each adult patient is calculated as $\text{BMI} = \frac{\text{weight}}{\text{height}^2}$. Values were rounded to the nearest integer. Adults were categorized using standard BMI classes [4].

ICD and multimorbidity features. Up to six ICD-10-CM diagnosis columns (one primary plus five secondary) were retained to capture comorbidities and clinical severity [22]. County-level SDOH features, including education, income, broadband access, food assistance, and household composition, were merged with each OASIS record. Missing SDOH values were imputed using fully populated county-level census summaries, while residual clinical missingness was imputed as needed for modeling [20]. Diagnosis fields were collapsed into section-level descriptors (ICDSection, ICDRange) using an external ICD mapping to group codes into clinically meaningful ranges.

ICD-10 diagnosis codes were grouped into high-level sections, and an ICD section co-occurrence network confirmed extensive multimorbidity, motivating section- or prefix-level coding for integrating diagnoses with SDOH in downstream models.

Comorbidity and frailty were quantified from ICD-10-CM diagnosis codes using the *comorbidipy* package [18], which implements validated Charlson and Elixhauser indices with multiple weighting schemes (e.g., Quan, van Walraven, Swiss) [7, 10, 23, 30]. Diagnosis codes were transformed into a long-format patient-code table, mapped to condition groups, and aggregated to patient-level scores. Age-adjusted Charlson indices were assigned one point per decade after age 40 (capped at four). Hospital Frailty Risk Scores (HFRS) were computed by summing weighted frailty-related ICD codes from index and prior admissions over a two-year lookback and categorized into low, intermediate, and high risk groups [14, 18, 27].

Utilization and functional status. Visit-level OASIS features (e.g., days between visits, days cared for, and care-by-discipline) are retained as utilization metrics, along with functional status items (e.g., activities of daily living and sensory impairment indicators) that capture disability and dependence [5, 32]. We engineer summary measures (e.g., recent utilization density and, when available, the number of prior hospitalizations) to capture both the intensity and recency of care [32].

SDOH and county-level context. Each OASIS record is linked to county-level SDOH indicators using standardized FIPS codes [13]. SDOH features include census-derived measures of educational attainment, income and poverty, broadband access, food assistance participation, and household composition [1, 2, 15, 22]. Missing SDOH values are imputed using complete county-level census summaries for the corresponding county-level indicators. At the same time, clinical and utilization missingness is either preserved via explicit missingness indicators or imputed for modeling using robust methods for EHR laboratory and assessment data [20, 25]. These county-level attributes serve as contextual features in the risk vector and as inputs to the disparity metrics defined below.

Structured risk representation for 30-day readmission Rather than collapsing risk into a single opaque scalar, we represent each home health episode i by a structured risk vector.

$$R_i = \left(\mathrm{Age}_i, \mathrm{BMIcat}_i, C_i^{(1)}, \ldots, C_i^{(C)}, B_i, V_i, T_i, \Delta_i, S_i\right), \qquad (1)$$

where $C_i^{(c)}$ indicates the presence of chronic condition c, B_i summarizes diagnostic burden, V_i captures recent visit frequency, T_i denotes total days cared for in the episode, Δ_i is the time since the last contact, and S_i is a vector of county-level SDOH attributes (education, broadband access, income, household composition). [22, 32]. The categorical feature BMIcat_i encodes standard adult BMI classes. We compared models using a flat concatenation of features with models using the structured risk vector and evaluated ROC-AUC, PR-AUC, Brier score, and calibration, both overall and within the HF, HTN, diabetes, and DHFHTN cohorts. We adapted standard disparity indices (AID, RID, and PWID) and a Palma-style ratio to quantify county-level SDOH gradients in the home health population [13, 22]. To reduce instability from sparse data, county-level summaries were restricted to counties with at least 50 eligible encounters.

Predictive modeling and evaluation (overview) We train gradient-boosted decision tree models (LightGBM) for 30-day readmission using four feature sets: (a) demographics and utilization only, (b) + comorbidity and frailty indices, (c) + county-level SDOH attributes, and (d) the full structured risk vector R_i [11, 21, 26]. The primary 30-day readmission model used LightGBM with learning_rate = 0.05, up to 1,000 boosting rounds, early stopping after 50 non-improving iterations, maximum depth = 6, and min_samples_per_leaf = 20. Default subsampling and feature-fraction settings were retained unless otherwise noted, and a global random seed ensured reproducible training and evaluation. Models are evaluated using ROC–AUC, PR–AUC, and Brier score, with calibration assessed via calibration slope and intercept on held-out test sets, and subgroup performance reported for key clinical cohorts (HF, DHFHTN, hypertension only) [3, 8, 32]. We define stability as consistent performance across condition groups and counties (for example, low standard deviation of ROC–AUC across subgroups) and report these statistics to characterize the robustness of SDOH-enriched, comorbidity-aware models relative to simpler baselines [15, 19, 32].

4 System Design

This section describes the PRECISION-Connect architecture, from the ingestion of OASIS and SDOH data to risk scoring and the deployment of explainable models and dashboards for home health readmission and disparity analytics [5, 16]. Figure 2 provides a high-level view of the county comorbidity module. The proposed system is designed to ingest, process, and analyze large-scale CMS OASIS assessment data enriched with county-level social determinants of health (SDOH), geospatial attributes, and U.S. Census indicators. The dataset comprises over 440,000 home health encounters with more than 200 structured features per patient record. The platform must support iterative feature engineering, risk score computation, and model deployment while remaining responsive enough to power an interactive visualization system for clinicians, researchers, and care managers. Functionally, the system integrates raw OASIS EHR assessments, ICD-10 diagnosis codes, SDOH attributes, and census-derived features into a unified analytic table. It computes multiple risk and utilization metrics, including readmission indicators, comorbidity indices, frailty and disability measures, and derived utilization features. It serves summarized outputs to support geospatial and population-level exploration within an interactive dashboard. Non-functional requirements include scalable batch processing over tens of millions of rows using distributed computing frameworks, reproducible exploratory and feature-construction workflows, and export to compact, shareable datasets for downstream analytics and visualization.

System Architecture PRECISION-Connect follows a four-layer architecture that ingests OASIS assessments and diagnosis codes, links them to county-level SDOH tables via standardized geographic identifiers, and applies the feature engineering steps in Section 3 (BMI categories, ICD groupings, utilization summaries, and comorbidity/frailty indices) to construct a structured analytic table for 30-day readmission modeling. Gradient-boosted decision tree models (LightGBM) consume the resulting risk vectors to generate readmission risk scores, which are exported along with key episode- and county-level indicators in compact formats (e.g., CSV, Parquet) to drive clinician-facing dashboards that support cohort, geographic, and outcome filtering, as well as SHAP-based feature-attribution views.

Functional Modules The system comprises three primary modules, each providing distinct analytical functionality while sharing the same structured risk vector and SDOH-enriched analytic table. **1. County Comorbidity and Cohort Analysis.** This module derives patient-level comorbidity indicators, frailty categories, and condition-stratified readmission statistics. It aggregates them to county and cohort levels, enabling users to explore diabetes, heart failure, hypertension, and multimorbid groups across Texas counties and visualize associated demographic and SDOH disparities. **2. Feature Selection and Model Construction.** This module supports multiple filter- and model-based feature-selection methods, allowing users to iteratively identify and export the most informative risk-vector components for a given cohort and outcome. **3. Patient**

Risk Modeling and Explainability. This module integrates the structured risk vector R_i with gradient-boosted models to predict 30-day readmission and related outcomes, using SHAP-based explanations to provide global feature importance and patient-level dependence plots, along with configurable cohorts, targets, and evaluation metrics.

Implementation and Deployment The system is implemented in a Python-based analytics stack that combines distributed ETL and feature engineering with in-memory analysis, using standard scientific libraries for comorbidity and frailty scoring and a small ICD utility module for reusable diagnosis-group mappings. Workflows are organized into modular notebooks and scripts for OASIS ingestion, feature construction, and analytic table creation. The scoring interfaces can load either models trained within this pipeline or compatible externally trained models, enabling flexible deployment and future integration of third-party models.

Table 2: Predictive performance for 30-day readmission on the held-out test set using Random Forest and LightGBM across four nested feature sets.

Feature set	*Model*	*ROC–AUC*	*PR–AUC*	*Brier score*
(1) Demo + Util	Random Forest	0.69	0.27	0.13
	LightGBM	0.75	0.34	0.21
(2) + Comorb	Random Forest	0.72	0.31	0.12
	LightGBM	0.76	0.36	0.20
(3) + SDOH	Random Forest	0.72	0.29	0.12
	LightGBM	0.76	0.36	0.20
(4) Full risk vector	Random Forest	0.72	0.30	0.12
	LightGBM	*0.76*	*0.37*	*0.20*

LightGBM is used as the primary model because it achieves higher ROC–AUC and PR–AUC than Random Forest across feature sets and provides more stable discrimination in multimorbid subgroups, consistent with ranking patients by relative risk; calibration is refined with post-hoc methods when needed. The resulting risk scores and episode- and county-level indicators are exported as lightweight CSV and Parquet files to drive dashboards. The pipeline can be deployed in containerized environments with scheduled updates, enabling routine refresh of risk scores and disparity summaries without manual intervention.

5 Results

This section demonstrates how PRECISION-Connect summarizes multimorbid home health cohorts, evaluates 30-day readmission models built on the structured risk vector R_i, and surfaces county-level SDOH disparities and model explanations for clinically meaningful subgroups [5, 32].

Cohort characteristics and outcomes We analyzed 221,957 Texas home health encounters after applying inclusion criteria (age ≥ 18, valid county FIPS, complete 30-day follow-up). Table 1 summarizes episode counts, mortality, and ever-readmitted indicators stratified by key condition groups (none, HF only, HTN only, diabetes only, HF+HTN, diabetes+HTN, diabetes+HF, DHFHTN). Mortality and readmission rates increase with multimorbidity, with the DHFHTN cohort exhibiting the highest burden and single-condition cohorts (e.g., HTN-only) showing lower but still elevated risk relative to patients without these chronic diseases.

Readmission model performance We trained Random Forest and LightGBM models to predict 30-day readmission using four nested feature sets: (1) demographics and utilization only; (2) plus comorbidity and frailty indices; (3) plus county-level SDOH attributes; and (4) the full structured risk vector R_i. Models were evaluated on a held-out test set using ROC–AUC, PR–AUC, and Brier score. Adding comorbidity and frailty indices (2) produced the largest single gain in ROC–AUC, underscoring the value of explicit multimorbidity representation, and incorporating county-level SDOH (3) yielded a modest additional improvement in PR–AUC. The full risk-vector model (4) with LightGBM achieved the best overall performance and lowest Brier score, and is therefore used as the default configuration in the risk-modeling module.

Stability across cohorts To assess stability, model performance was examined across HF, HTN, diabetes, and DHFHTN cohorts for each feature set, focusing on cohort-specific ROC–AUC and calibration slopes. The full risk vector improved calibration, particularly in multimorbid cohorts such as DHFHTN, and maintained stable discrimination across condition groups, offering a practical balance between accuracy and reliability. Because differences between flat and structured representations were modest, the primary advantage of the structured risk vector lies in its interpretability, as evidenced by condition-stratified dashboards and SDOH-aware county queries.

Disparity metrics and county-level patterns Using broadband access as an example SDOH indicator, county-level disparity indices were applied to HF-related cohorts to summarize differences in connectivity across urban Texas counties. These metrics indicate that some counties have substantially higher broadband access than the statewide mean, and that counties with both large HF populations and high connectivity contribute disproportionately to the overall burden of disparity. Disparity estimates are restricted to counties with at least 50 encounters, and uncertainty is summarized using bootstrap confidence intervals; low-volume counties are flagged or omitted in displays.

Use case: navigating HF and DHFHTN risk In a typical workflow, a care manager selects HF or DHFHTN cohorts, reviews county-level maps of multimorbidity and SDOH, and uses disparity metrics to prioritize counties where many high-risk patients face adverse social conditions. The risk-modeling module then provides SHAP-based explanations for representative high-risk patients, where comorbidity burden, frailty, recent utilization, and county-level SDOH indicators often emerge as key contributors to elevated readmission risk. This combination of structured risk vectors, disparity metrics, and explainable models supports disparity-aware population health decision-making rather than relying on a single opaque risk score.

6 Discussion

Jointly modeling multimorbidity, frailty, utilization, and county-level SDOH yields more informative 30-day readmission risk estimates than using clinical or contextual data alone. ICD-based analyses confirm a highly multimorbid Medicare home health cohort, motivating prefix-level diagnostic features that smooth coding variation while preserving clinically meaningful structure. Although models were trained on a Texas Medicare home health cohort, the multimorbidity and comorbidity profile aligns with U.S.-wide Medicare patterns, suggesting cautious generalizability of the SDOH-enriched risk representation to other settings with local calibration.

The risk representation R_i encodes each episode as a multidimensional vector spanning age, BMI, chronic disease burden, comorbidity and frailty scores, utilization intensity, and county-level SDOH, rather than collapsing risk into a single score. This explicit structure supports condition-stratified readmission estimates, transparent risk adjustment across multimorbid subgroups, and flexible cohort queries, as reflected in mortality and readmission gradients for DHFHTN and related cohorts.

Limitations include the Medicare-only, Texas-centered training data, coding variability in administrative claims, and predictors that partly capture care access and workflow in addition to clinical need. Future work should emphasize external validation across health systems, systematic fairness assessment across demographic and clinical subgroups, and governance frameworks for the responsible use of SDOH-informed risk prediction.

7 Conclusion

PRECISION-Connect is a deployable population health analytics platform that integrates multimorbidity- and SDOH-enriched risk representations with explainable machine learning and interactive dashboards for Medicare home health populations. The structured risk vector R_i combines comorbidity and frailty indices, ICD-based diagnostic groupings, BMI, utilization patterns, and county-level SDOH, supporting both 30-day readmission prediction and county-level disparity analysis without collapsing risk into a single opaque score. Future extensions include incorporating laboratory time series and clinical text, expanding geospatial SDOH layers, integrating fairness-aware modeling, and connecting to operational EHR systems for near–real-time decision support.

References

[1] Nancy E. Adler and William W. Stead. 2015. Patients in Context—EHR Capture of Social and Behavioral Determinants of Health. *New England Journal of Medicine* 372, 8 (2015), 698–701. doi:10.1056/NEJMp1413945

[2] Jonathan Cantor, Lorna Thorpe, and et al. 2020. Challenges in Using Social Determinants of Health Data in Electronic Health Records. *American Journal of Public Health* 110, 9 (2020), 1298–1304.

[3] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noémie Elhadad. 2015. Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 1721–1730.

[4] Centers for Disease Control and Prevention. 2022. About Adult BMI. https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html. Accessed: 2024-08-30.

[5] Centers for Medicare and Medicaid Services. 2020. Outcome and Assessment Information Set (OASIS) User Manuals. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/OASIS. Accessed: 2024-08-30.

[6] Dan Chang, Mark E. Mikkelsen, et al. 2021. Prediction of Readmission Following Sepsis Using Social Determinants of Health. *BMJ Open* 11, 9 (2021), e044964. doi:10.1136/bmjopen-2020-044964

[7] Mary E. Charlson, Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie. 1987. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Diseases* 40, 5 (1987), 373–383. doi:10.1016/0021-9681(87)90171-8

[8] Jacques Donzé, Drahomir Aujesky, Mark V. Williams, and Jeffrey L. Schnipper. 2013. Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients. *JAMA Internal Medicine* 173, 8 (2013), 632–638. doi:10.1001/jamainternmed.2013.3023

[9] Joy Dugan and Jay Shubrook. 2017. International Classification of Diseases, 10th Revision, Coding for Diabetes. *Clinical Diabetes: A Publication of the American Diabetes Association* 35, 4 (2017), 232–238. doi:10.2337/cd16-0052

[10] Anne Elixhauser, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey. 1998. Comorbidity Measures for Use with Administrative Data. *Medical Care* 36, 1 (1998), 8–27.

[11] Piotr Florek and Adam Zagdański. 2023. Benchmarking state-of-the-art gradient boosting algorithms for classification.

[12] Centers for Medicare and Medicaid Services. 2020. https://www.cms.gov/files/document/impact-readmissions-reduction-initiatives-report.pdf.

[13] Daniel J. Friedman and R. Gibson Parrish. 2022. Privacy, interoperability, and the use of area-based measures in population health. *Annual Review of Public Health* 43 (2022), 1–20.

[14] Thomas Gilbert, Jenny Neuburger, Jason Kraindler, and et al. 2018. Development and Validation of a Hospital Frailty Risk Score Focusing on Older People in Acute Care Settings Using Routinely Collected Data. *The Lancet* 391, 10132 (2018), 1775–1782. doi:10.1016/S0140-6736(18)30668-8

[15] Benjamin A. Goldstein, Ann Marie Navar, and Michael J. Pencina. 2022. Opportunities and challenges in developing risk prediction models with social determinants of health. *Circulation* 145, 5 (2022), 354–366.

[16] OHDSI CDM Working Group. 2021. OMOP CDM v5.3. https://ohdsi.github.io/CommonDataModel/cdm53.html.

[17] Sarah Lee, Rohan Patel, and Thao Nguyen. 2024. Predicting Hospital Readmission from Home Health Care Using Machine Learning. *Journal of the American Medical Informatics Association* 31, 3 (2024), 456–467. doi:10.1093/jamia/ocae123

[18] Daniela Leidig and contributors. 2023. comorbidipy: Comorbidity Scores in Python. https://github.com/dleidig/comorbidipy. Accessed: 2024-08-30.

[19] J. Li, W. Li, Q. Sun, and et al. 2020. A document clustering approach based on hybrid LDA and K-means. *Expert Systems with Applications* 149 (2020), 113263.

[20] Jiang Li, Xiaowei S. Yan, Durgesh Chaudhary, Venkatesh Avula, Satish Mudiganti, Hannah Husby, Shima Shahjouei, Ardavan Afshar, Walter F. Stewart, Mohammed Yeasin, et al. 2021. Imputation of missing values for Electronic Health Record Laboratory Data. *npj Digital Medicine* 4, 1 (oct 2021), 1–10. doi:10.1038/s41746-021-00518-0

[21] Marcos R. Machado, Salma Karray, and Ivandro T. de Sousa. 2019. LightGBM: An Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry. In *2019 14th International Conference on Computer Science and Education (ICCSE)*. IEEE, New York, NY, USA, 1111–1116. doi:10.1109/ICCSE.2019.8845529

[22] Douglas P. Olson, Benjamin J. Oldfield, and Sofia Morales Navarro. 2019. Standardizing Social Determinants of Health Assessments. *Health Affairs Forefront* 1, 1 (2019), 1. doi:10.1377/forefront.20190311.823116 Commentary.

[23] Hude Quan, Vijaya Sundararajan, Patricia Halfon, et al. 2005. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care* 43, 11 (2005), 1130–1139.

[24] Chris Salisbury, Lorraine Johnson, Sarah Purdy, Jose M. Valderas, and Alan A. Montgomery. 2011. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *British Journal of General Practice* 61, 582 (2011), e12–e21. doi:10.3399/bjgp11X548929

[25] Christopher M Sauer, Li-Ching Chen, Stephanie L Hyland, Armand Girbes, Paul Elbers, and Leo A Celi. 2022. Leveraging Electronic Health Records for Data Science: Common Pitfalls and How to Avoid Them. *The Lancet Digital Health* 4, 12 (Dec 2022), e893–e898. doi:10.1016/S2589-7500(22)00154-6

[26] Hiroe Seto, Asuka Oyama, et al. 2022. Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with Big Data. *Scientific Reports* 12, 1 (2022), 15889. doi:10.1038/s41598-022-20149-z

[27] Rahul Sharma et al. 2021. Frailty and Outcomes in Older Adults: Application of the Hospital Frailty Risk Score. *Age and Ageing* 50, 3 (2021), 1–9.

[28] John Smith, Hannah Lee, and Miguel Garcia. 2024. Deep Learning for 30-Day Hospital Readmission Prediction Using multimodal Electronic Health Records. *NPJ Digital Medicine* 7, 1 (2024), 45. doi:10.1038/s41746-024-00999-1

[29] Carl Van Walraven et al. 2022. Social Determinants of Health in Readmission Risk Prediction Models: A Systematic Review. *Journal of General Internal Medicine* 37, 4 (2022), 874–885. doi:10.1007/s11606-021-07123-2

[30] Carl van Walraven, Peter C. Austin, Alfred Jennings, Hude Quan, and Alan J. Forster. 2009. A Modification of the Elixhauser Comorbidity Measures into a Point System for Hospital Death Using Administrative Data. *Medical Care* 47, 6 (2009), 626–633. doi:10.1097/MLR.0b013e31819432e5

[31] William B. Weeks, James N. Weinstein, and Juan M. Lavista Ferres. 2026. Using AI to Improve Individual and Population Health. *International Journal of Public Health* Volume 71 - 2026 (2026), 1–3. doi:10.3389/ijph.2026.1609114

[32] Wenrui Wei, Patrick J. Tighe, Young-Hak Kang, and et al. 2019. Combining comorbidity indices and social determinants of health improves risk prediction in hospitalized patients. *Journal of Biomedical Informatics* 95 (2019), 103239. doi:10.1016/j.jbi.2019.103239

[33] Ke You, Meng Li, Asad Abbas, et al. 2021. Machine Learning for Predicting Readmission Risk Among the Frail. *Journal of Biomedical Informatics* 122 (2021), 103906. doi:10.1016/j.jbi.2021.103906