SOCIALLYDETERMINED

Netrin Health

Analyzing the Impact of Social Risks on High Healthcare Utilization through ML
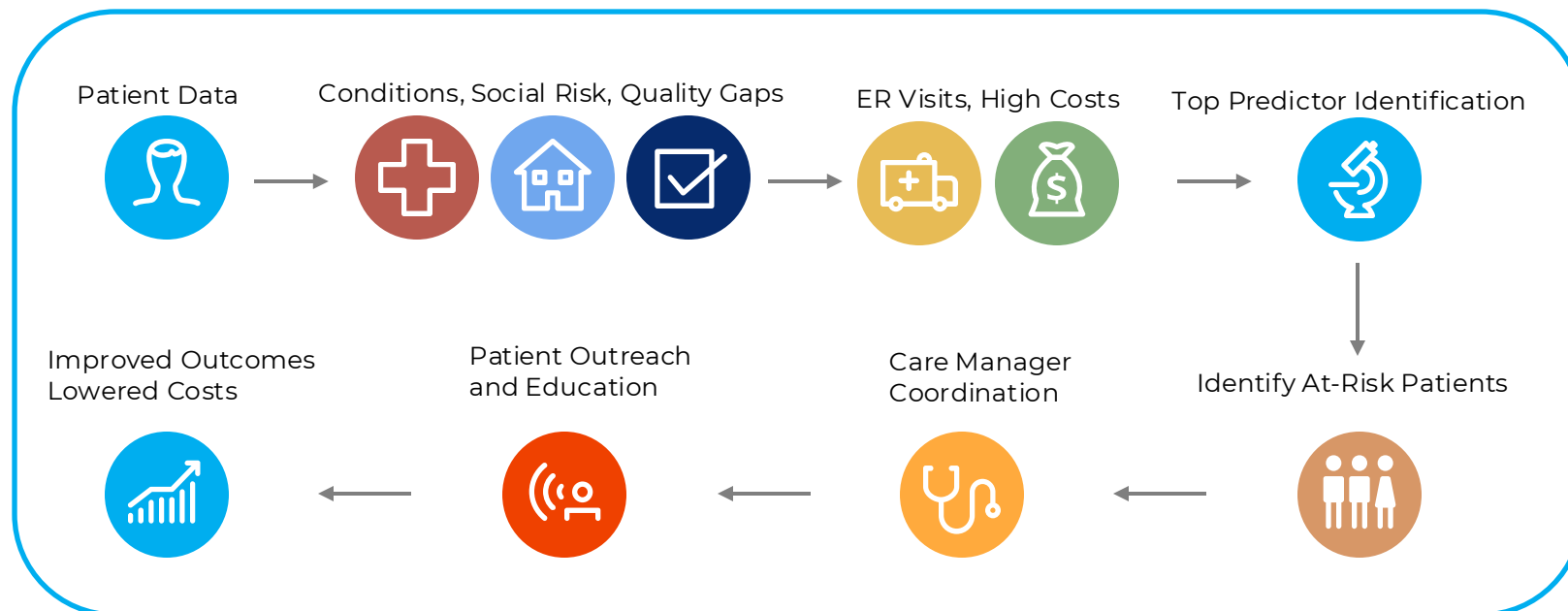Author: Milinda Polisetty | Population Health Intern | Netrin Health

Background

Behind every preventable hospital visit is a person facing more than a diagnosis, often living with chronic conditions while navigating financial stress, housing insecurity, or limited health access.

This study explores the root causes of healthcare overutilization using claims data integrated with modeled social risk indicators from Socially Determined's SocialScape®, the platform that aggregates socioeconomic data to estimate social risks. These indicators reflect SDOH barriers, modeled using ZIP+4 population algorithms and individual data to approximate patient vulnerability.

By applying predictive modeling, we aim to equip care management teams with actionable insights beyond clinical flags to enable earlier intervention and improve the outcomes of patients at higher risk, simultaneously reducing the cost of care.

Objective

- Identify Key Risk Drivers: Analyze clinical and social factors driving high healthcare utilization.
- Stratify High-Risk Patients: Use machine learning to segment patients based on the risks.
- Strategic Decisions: Suggest data-driven strategies and support care managers in targeted outreach.

Patient Data → Conditions, Social Risk, Quality Gaps → ER Visits, High Costs → Top Predictor Identification → Identify At-Risk Patients → Care Manager Coordination → Patient Outreach and Education → Improved Outcomes Lowered Costs

SOCIALLYDETERMINED

Netrin Health

Data

Claims Data
Patient Information, Quality Gaps, Chronic Health Conditions, Medical and Pharmacy risk scores, and Healthcare Costs.

Individual-Level Social Risk Scores
Financial Strain, Housing Instability, Transportation Barriers, Food Insecurity, and Health Literacy Challenges.
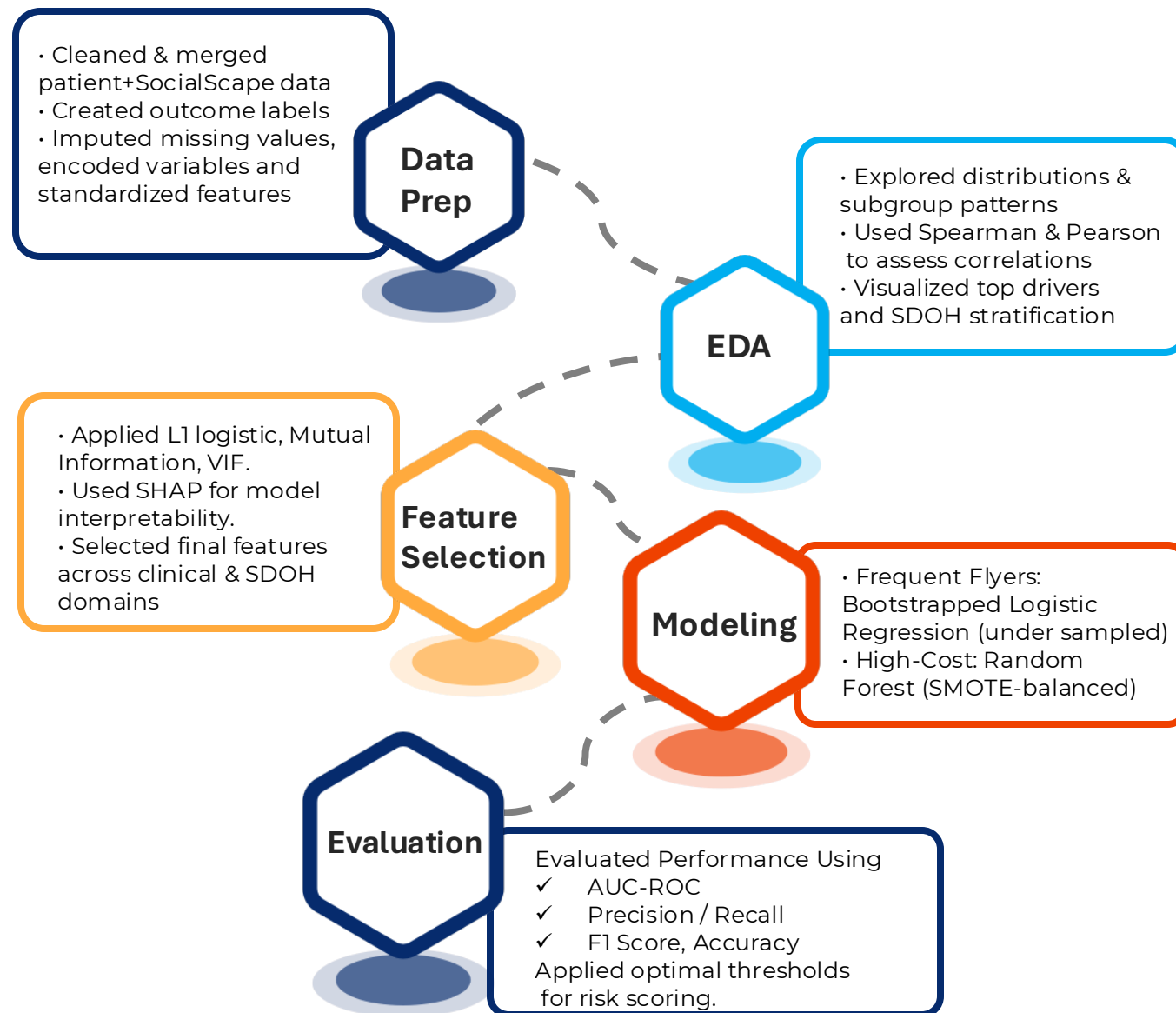
12,998 unique patients

Outcome Variables
- Frequent Flyers: ≥3 ER visits/1 year → 144 patients (1.1%)
- High-Cost: Annual cost > $8,100/1 year→ 2,092 patients (16.1%)

30+ Data Features Including:
- Clinical: Age, Gender, Medical Risk, Illness Band, Pharmacy Risk, Conditions like HTN, Diabetes, Heart Disease, Oncology, etc.

- SDOH: Financial, Housing, Literacy, Food, Transportation risks.

- Subdividers Of Risks: Liabilities, Assets, Demographics, Education, Housing Stability/Quality, Culture, Food Quality, etc.

Data Prep
· Cleaned & merged patient+SocialScape data
· Created outcome labels
· Imputed missing values, encoded variables and standardized features

EDA
· Explored distributions & subgroup patterns
· Used Spearman & Pearson to assess correlations
· Visualized top drivers and SDOH stratification

Feature Selection
· Applied L1 logistic, Mutual Information, VIF.
· Used SHAP for model interpretability.
· Selected final features across clinical & SDOH domains

Modeling
· Frequent Flyers: Bootstrapped Logistic Regression (under sampled)
· High-Cost: Random Forest (SMOTE-balanced)

Evaluation
Evaluated Performance Using
✓ AUC-ROC
✓ Precision / Recall
✓ F1 Score, Accuracy
Applied optimal thresholds for risk scoring.

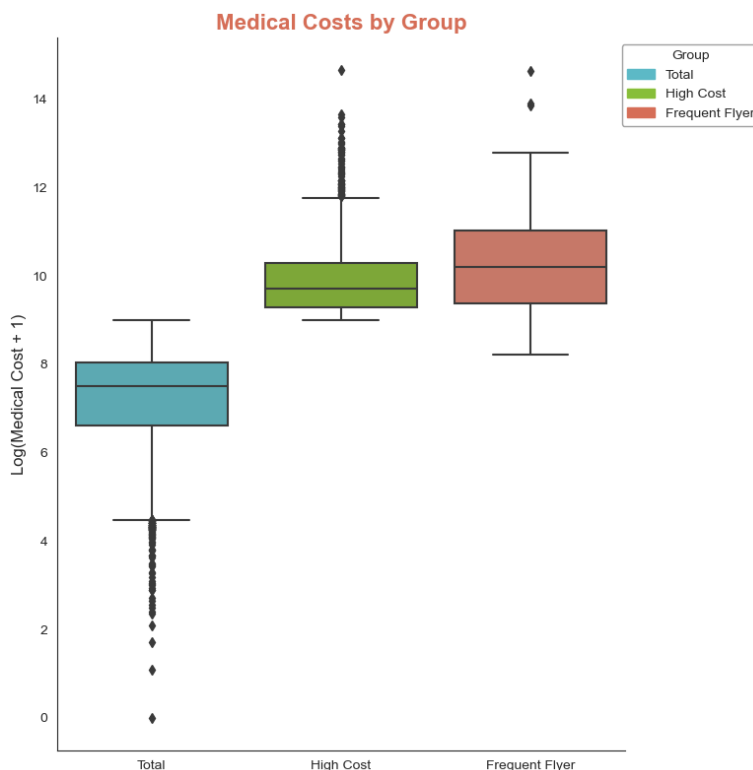# Findings

Medical Costs by Group

Fig 1 High-cost and "frequent flyers" exhibit significantly higher medical costs compared to the overall population, highlighting disproportionate spending.
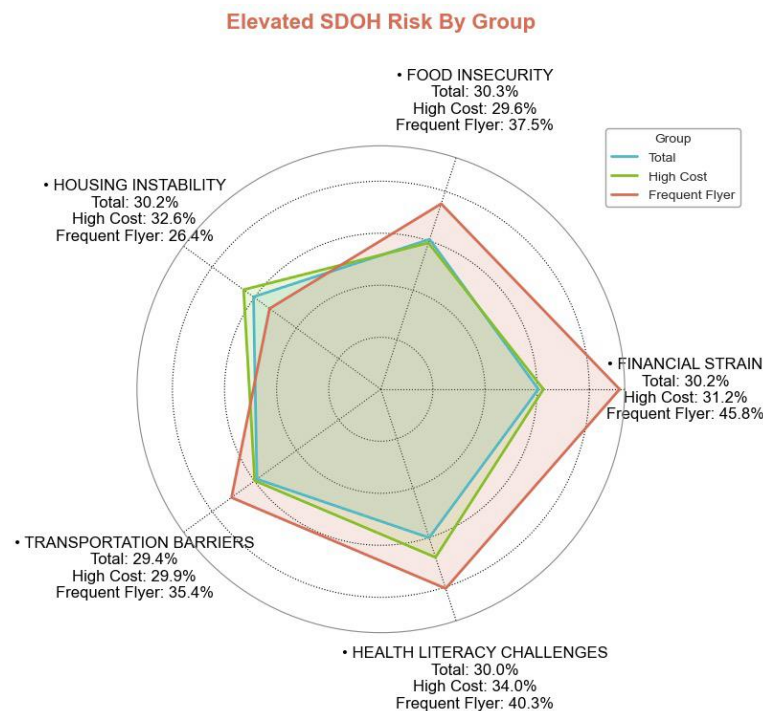

Elevated SDOH Risk By Group

• FOOD INSECURITY
Total: 30.3%
High Cost: 29.6%
Frequent Flyer: 37.5%

• HOUSING INSTABILITY
Total: 30.2%
High Cost: 32.6%
Frequent Flyer: 26.4%

• FINANCIAL STRAIN
Total: 30.2%
High Cost: 31.2%
Frequent Flyer: 45.8%

• TRANSPORTATION BARRIERS
Total: 29.4%
High Cost: 29.9%
Frequent Flyer: 35.4%

• HEALTH LITERACY CHALLENGES
Total: 30.0%
High Cost: 34.0%
Frequent Flyer: 40.3%

Fig 2 Frequent flyers show elevated social risk across most social risks and High-cost patients show higher risks than the overall population.

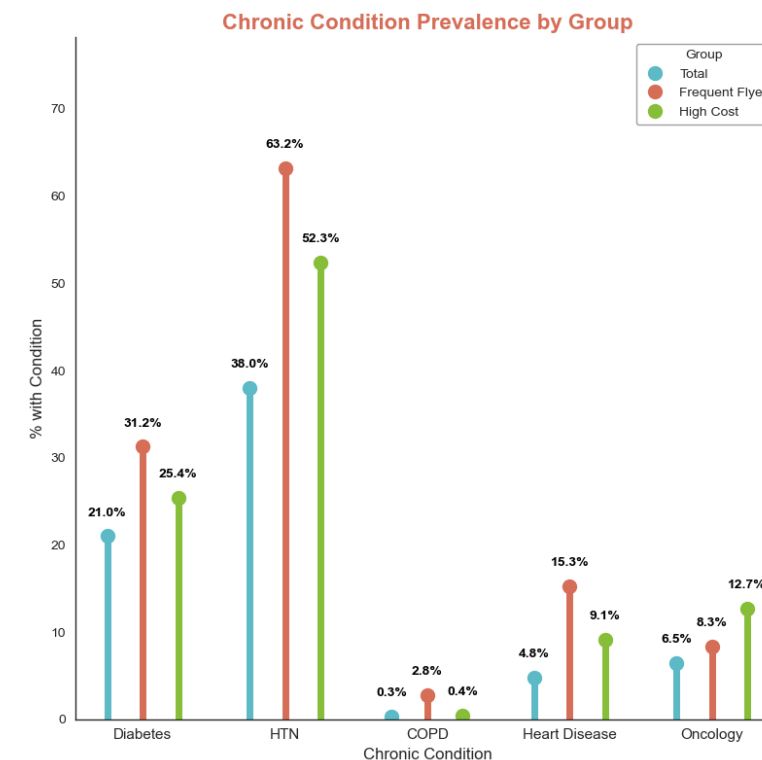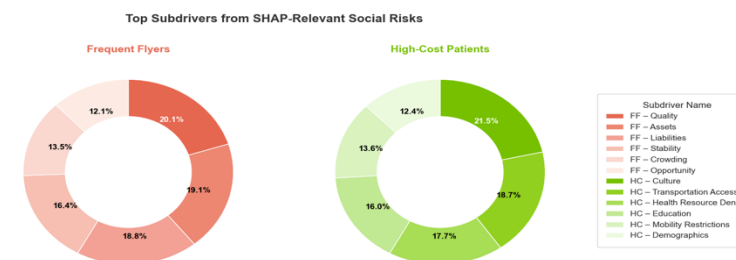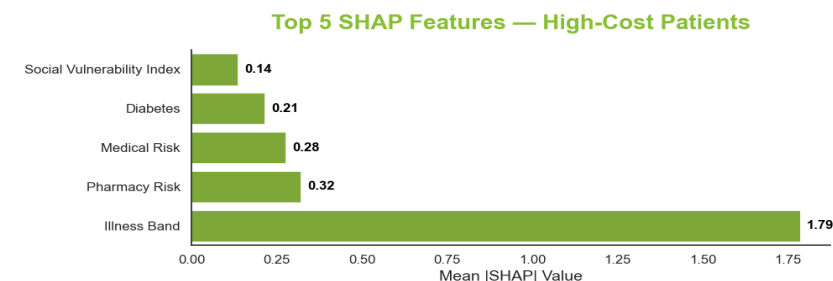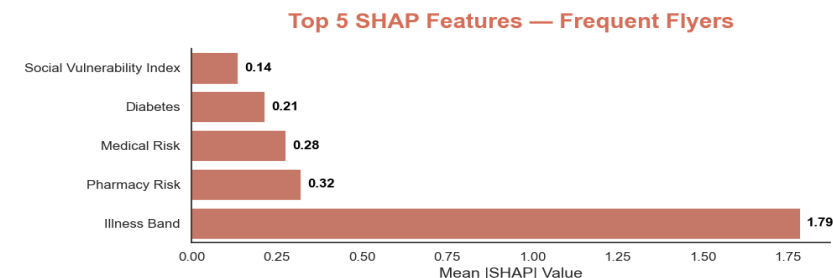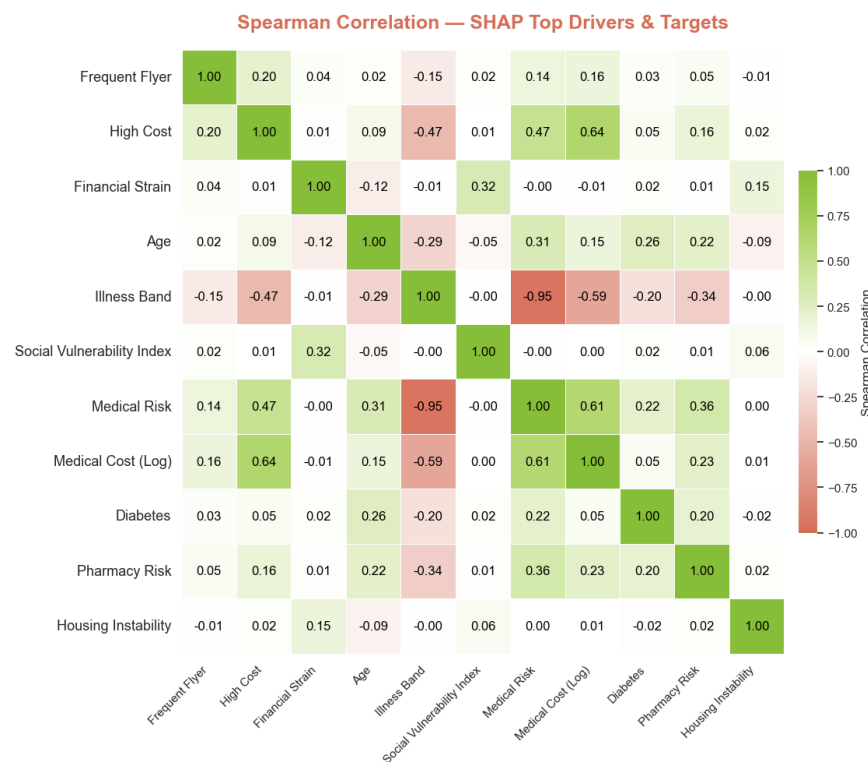**Chronic Condition Prevalence by Group**

Fig 3 Frequent flyers have higher prevalence of chronic diseases, especially hypertension and diabetes, compared to the general and high-cost populations.

Findings

Spearman Correlation — SHAP Top Drivers & Targets

Top 5 SHAP Features — Frequent Flyers

Top 5 SHAP Features — High-Cost Patients

Top Subdrivers from SHAP-Relevant Social Risks

Model Results (Fig 4,5,6,7)
Clinical conditions and social risks were key predictors of both high-cost and frequent flyer patients.

Frequent flyers were heavily impacted by social vulnerabilities, while high-cost patients showed a combination of clinical burden and care quality gaps.

Discussion

- For this study, it is evident that healthcare overutilization is not solely a clinical issue but is equally driven by social risks.

- The predictive models performed with high accuracy (AUC: 0.97 for frequent flyers, 0.96 for high-cost patients), affirming the reliability of machine learning in identifying risk.

- SHAP analysis of the risk drivers showed that while both groups share similar predictors, their relative influence varies significantly, highlighting the need for tailored interventions.

- Frequent flyers were predominantly impacted by social risk, while high-cost patients were also affected by clinical burden and healthcare quality gaps.

- The correlation matrix revealed that social risks often cluster and compound, intensifying their effect on healthcare use.

- This layered insight helps care teams move beyond generic outreach and target support where it matters most.

- Overall, this study supports integrating predictive insights into care coordination workflows to reduce preventable costs and improve outcomes among at-risk patients.

Limitations

- The dataset lacked relevant and important clinical and demographic details, restricting deeper analysis.

- SocialScape® scores were derived using undisclosed models that apply ZIP+4-level socioeconomic data to estimate individual-level social risks and additional proprietary data sources, rather than direct survey responses.

Future Scope

- **Patient Outreach:** 70 frequent flyers and 1,540 high-cost patients were identified through model-driven stratification for targeted engagement and social risk screening.

- **Ground-Level Insights:** Care managers are collecting patient-level social needs data to validate and enrich SocialScape's community-level insights.

- **Personalized Support:** Insights from patient outreach will guide us to create personalized educational materials and help connect at-risk individuals with targeted social support resources, based on their identified needs.