

ASA DataFest 2026 · Duke University

From Patterns to Prevention: How Social Determinants Shape Type 2 Diabetes Care Pathways

The Carolina Correlations:
Eric Hu, Ruitong Liu, Mia Zhou, Gloria Yang

Markov Chain Prediction Network

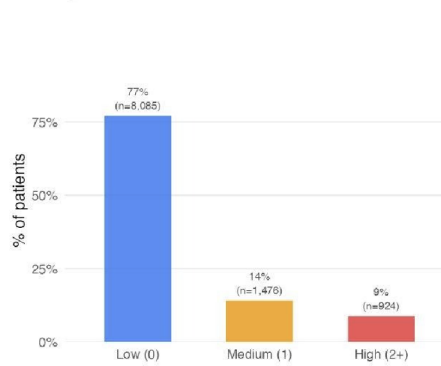
Among top 20 most frequent diseases, we focused on the smaller comorbidity network consisting of Type 2 diabetes and four other related diseases



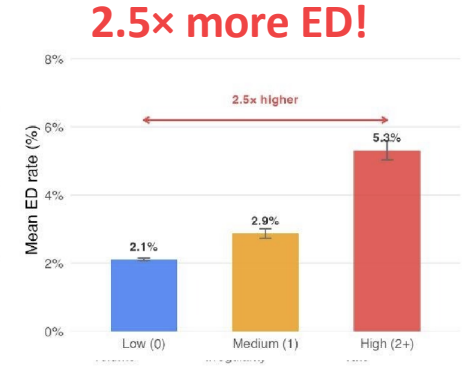
Social Determinants Shape Patient Journey Archetypes in Diabetic Care

High-SDOH patients visit more, visit irregularly, use ED 2.5x more, and concentrate in the ED-Dependent archetype

A Patient distribution
by SDOH risk tier



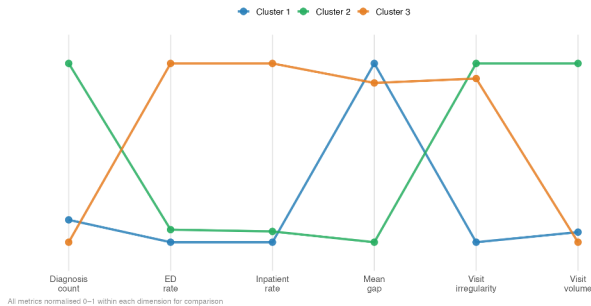
B ED use — headline finding
High SDOH = 2.5x more ED visits



K-Mean clustering (k=3)

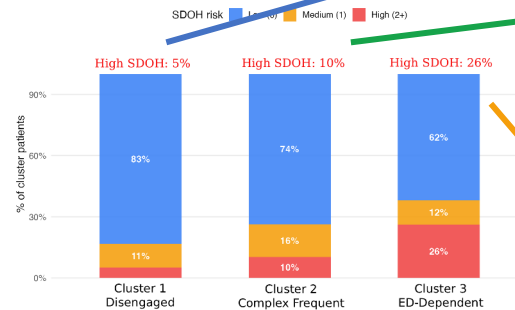
Three Distinct Patient Journey Archetypes

Each line = one cluster. Higher = more extreme on that dimension



High-SDOH Patients Concentrate in ED-Dependent Archetype

SDOH risk tier composition within each patient journey cluster (among patients with SDOH survey data)



Cluster 1 (Disengaged)

Longest mean gap, lowest ED rate, fewest visits -> healthier or simply don't engage often

Cluster 2 (Complex Frequent)

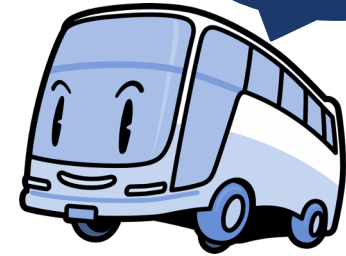
Shortest mean gap, high diagnosis count, most irregular rhythm -> complex conditions and are actively using the system

Cluster 3 (ED dependent)

Highest ED rate, moderate visit volume, moderate irregularity -> social barriers delay care & rely on emergency services

From call to action: Mobile Shuttles to lower ED rate

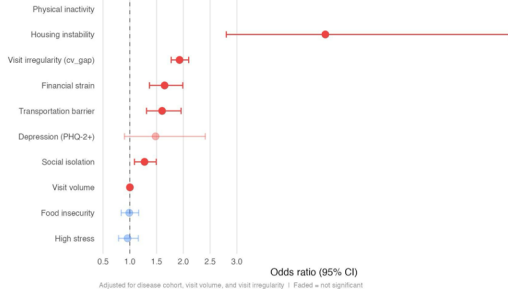
Click me for the shuttle dashboard!



Which Social Factors Most Predict ED-Dependent Care?

Odds ratios from logistic regression — outcome: ED visit rate in top quartile
Values > 1 = increases risk of ED-dependent pathway

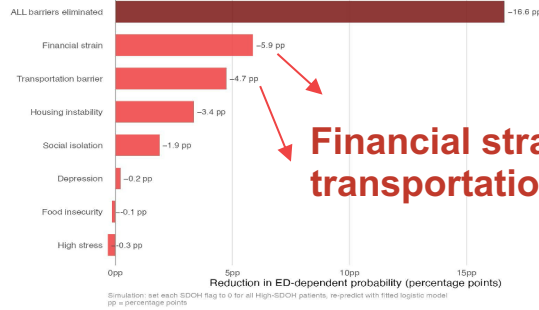
● p ≥ 0.05 ● p < 0.05 ● NA ● Decreases ED risk ● Increases ED risk ● NA



Simulated Impact of Eliminating Each Social Barrier

Among High-SDOH patients (baseline ED-dependent risk: 50.5%)

Bar = predicted reduction in ED-dependent probability if barrier removed



Financial strain (-5.9%) & transportation limit (-4.7%)!

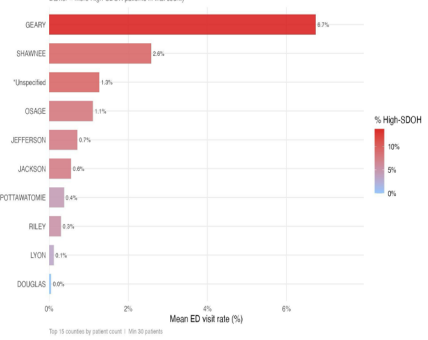
Counties with More Social Burden Have Higher ED Rates

Each dot = one county | Dot size = patient count | r = correlation



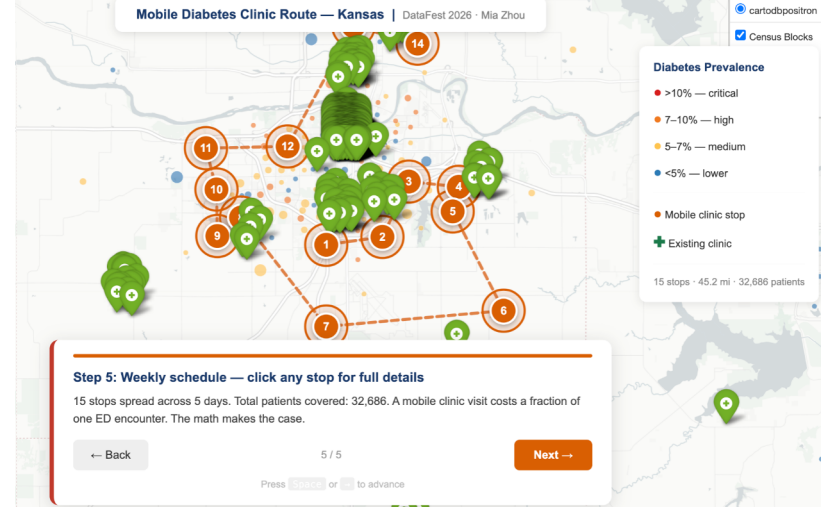
ED Use Varies Substantially Across Counties

Color = more High-SDOH patients in that county



Geary County: 6.7% ED rate (highest SDOH burden)

Nearest-neighbor algorithm based on prevalence, patient density, and access (15 prioritised nodes)



45 miles, covering over 32,000 patients

From risk to action: ML-Driven Early Alert for Care Pathway Disruption

PATIENTS
24,213

ENCOUNTERS
317,764

BASE RISK RATE
20.1%

TOP 10% CAPTURE
19.8%

TOP 10% PRECISION
24.8%

How reliably can we flag upcoming care breakdown risk?

For each extra 5% outreach capacity, how many future events do we gain?

Which clusters show continuity breakdown and emergency pressure together?

Where does cluster-level burden concentrate?

Which neighborhoods have the biggest risk gap?

Which individual patient story should we review next?

Open Patient Drill-Down

Who should we contact first under limited outreach capacity?

Logistic Regression selects 6,008 encounters (10%). This captures 19.8% of observed future-risk events.

PATIENT	ENCOUNTER	DATE	RISK SCORE	OBSERVED RISK	ED ≤ 90D	GAP > 120D
4600183	145697166	2025-08-04	0.881	0	0	0
117972	141585649	2025-04-10	0.870	1	0	1
2600158	140305929	2025-03-06	0.861	0	0	0
3843121	139430709	2025-02-10	0.857	0	0	0
8684235	136254022	2025-02-26	0.856	0	0	0

CASE REVIEW

Patient Drill-Down Timeline

Patient 117972 | 6 encounters from 01/14/22 to 10/13/25

What does the patient timeline show?

- Each point is one encounter time point.
- Y-axis is gap days since the previous encounter.
- Higher points mean longer continuity breaks.
- Red dashed line is a 120-day threshold; points above it suggest follow-up risk.
- Repeated high points indicate persistent follow-up disruption risk.

Patient ID: [Load Patient](#)

TOTAL ENCOUNTERS: 6 **ED ENCOUNTERS**: 0 **ED RATIO**: 0.0% **MEAN GAP**: 273.6 days **FOLLOW-UP SPAN**: 1,368 days

Time Point vs Gap Days

Encounter List

DATE	TYPE	VISIT TYPE	PRIMARY DIAGNOSIS KEY	ED	RISK SCORE (LR)	RISK SCORE (MLP)
01/14/22	Office Visit	NEW PATIENT	91249	0	0.000	0.000
04/14/22	Office Visit	FOLLOW UP	91249	0	0.000	0.000
10/10/24	Lab Visit	LAB	91249	0	0.000	0.000
04/10/25	Office Visit	FOLLOW UP	406269	0	0.784	0.581
04/10/25	Lab Visit	LAB	406269	0	0.970	0.798
10/13/25	Lab Visit	LAB	406269	0	0.000	0.000

Problem: Access barriers → silent disengagement → preventable ED crisis

Composite target: 90-day ED risk + visit gap > 120 days

Scoring: Every encounter ranked → priority contact list

Incremental capture: Each +5% outreach → marginal burden captured vs. random

Patient profile: Repeated 120-day crossings → proactive outreach trigger

Population guidance: Cluster & neighborhood panels → cohort-level deployment

Bottom line: Risk score + thresholds + contact list = executable early warning

THANK YOU!!

**The Carolina Correlations:
Eric Hu, Ruitong Liu, Mia Zhou, Gloria Yang**