Who Comes Back?:
Machine Learning for ICU Readmission Prediction

Joseph Lattanzi
Master of Science in Applied Data Science
Bay Path University

The Readmission Crisis

20%

Medicare Patients
Readmitted within 30 days

14.67%

Average US Readmission Rate
Across all conditions

$26 Billion

Financial Burden
Annually

- Financial penalties under Hospital Readmissions Reduction Program
- Patient suffering: additional procedures, complications, mortality risk
- ICU patients face even higher readmission rates due to complexity

Research Objective & Approach

Can machine learning predict 30-day ICU readmissions using comprehensive clinical data from electronic health records?

DATA	FEATURES	MODELS	IMPACT
MIMIC-IV 545,316 ICU admissions (2008-2019) 6 core tables: Demographics, diagnoses, medications, procedures, lab results	57 predictive features engineered Clinical complexity, comorbidity indices, medication burden, healthcare utilization patterns	Three complementary approaches: • Linear Regression • Random Forest • XGBoost	Clinical translation Risk stratification, ROI analysis, resource optimization

The MIMIC-IV Dataset

Source

- Beth Israel Deaconess Medical Center (Boston, MA)
- 2008-2019 ICU Admissions
- De-identified but clinically rich
- Gold standard for critical care research

Scale & Scope

- **380,000+** unique patients
- **500,000+** hospital readmissions
- **120+ million** laboratory measurements
- **26+ million** prescription records
- **4.8+ million** diagnosis codes
- Truly longitudinal patient journeys captured

Why MIMIC-IV?

✓ **Comprehensive:** Captures entire clinical picture
✓ **Granular**: Time-stamped events, not just summaries
✓ **Validated**: Peer-reviewed, widely used in research
✓ **Realistic**: Real-world data with all its complexity

Data Structure

Table	Rows	Key Variables
Admissions	~500,000	Demographics, admission details
Patients	~380,000	Age, gender, mortality
Diagnoses	~4,800,000	ICD Codes, sequence
Prescriptions	~26,000,000	Medications, dosages
Lab Events	~120,000,000	Clinical measurements
Procedures	~650,000	Clinical procedure codes

Note: "ICD" stands for International Classification of Diseases

The Integration Challenge:

How do you meaningfully combine 120M+ records into patient-level features?

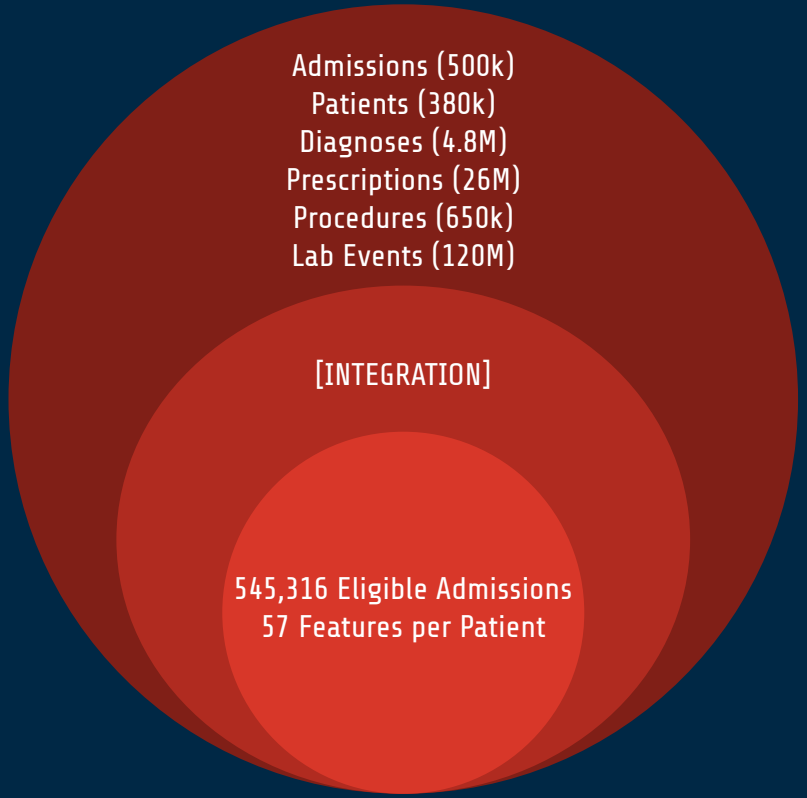
How do you capture temporal patterns?

How do you handle missingness without losing signal?

My Solution:

Aggregate clinical complexity metrics while preserving informativeness

Data Integration Pipeline



Key Integration Steps:

1. Link all tables via Subject_ID and Admission_ID
2. Filter to valid discharges *(LOS > 0, LOS < 365 days)*
3. Calculate 30-day readmission outcome
4. Aggregate clinical events per admission

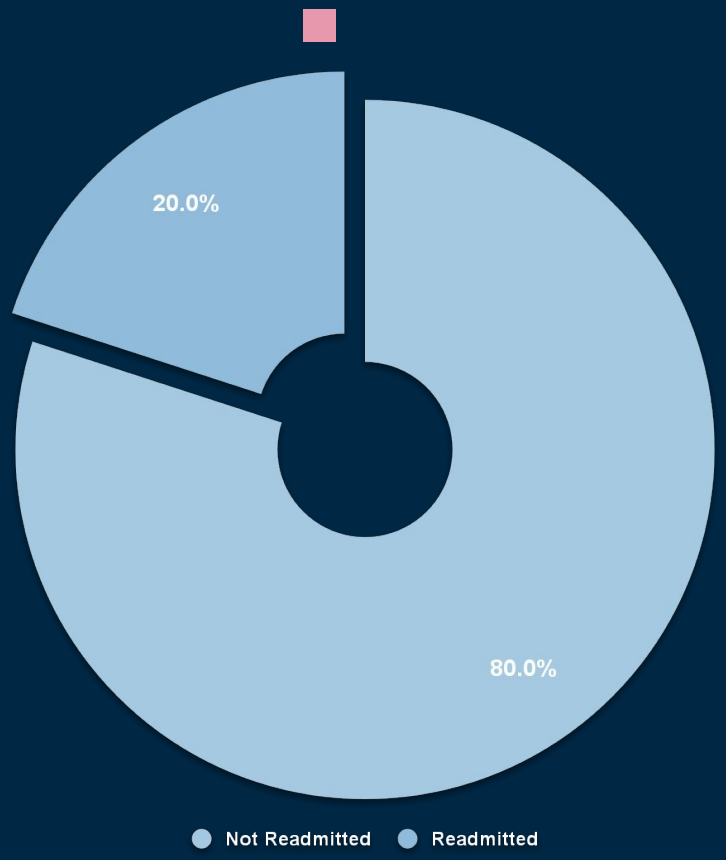
The Readmission Baseline

20.03%

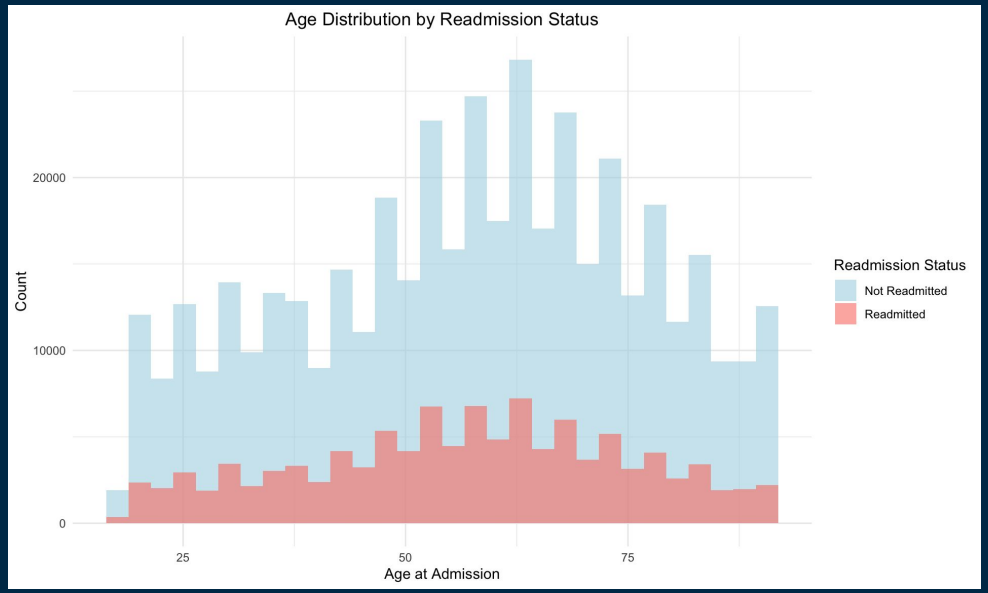
30-Day Readmission Rate
(109,269 readmissions out of 545,316 eligible discharges)

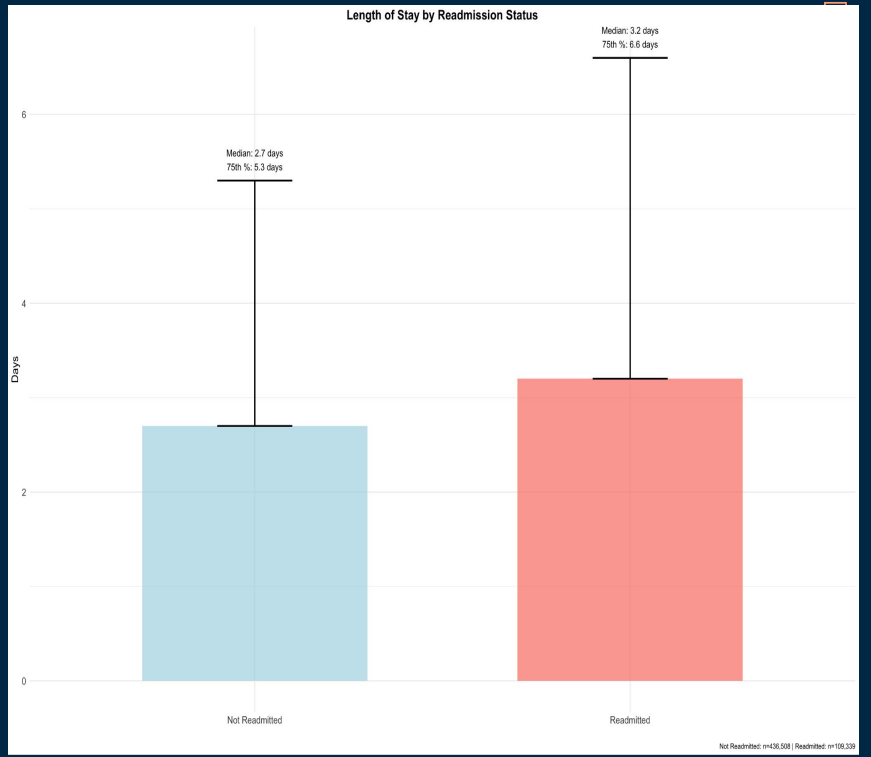
Context:

- ICU Patients: 20.03%
- National Average (All Patients): 14.67%
- Medicare Patients: 20%



Patient Demographics & Clinical Characteristics

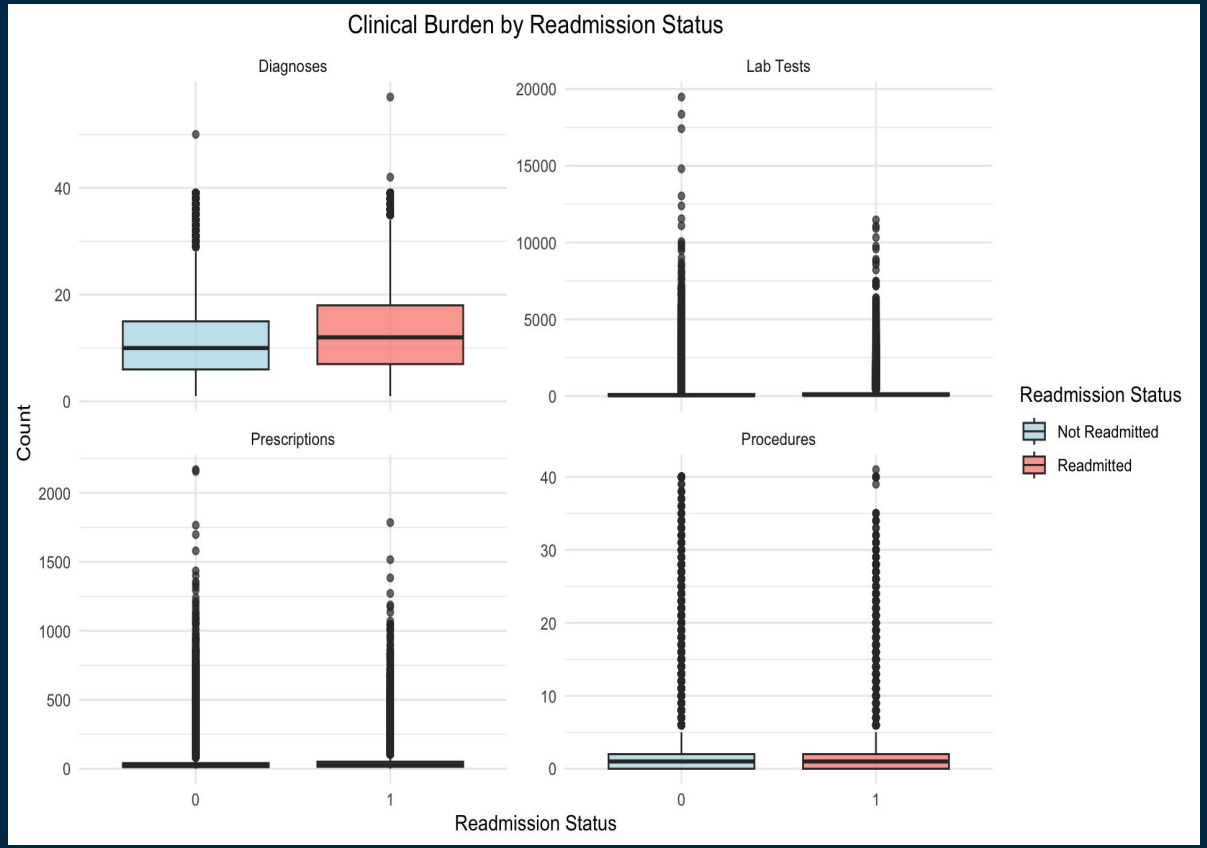




Key Findings:

- Readmitted patients: Older (median age 67 vs 64)
- Readmitted patients: Longer stays (median 6.8 vs 5.9)
- Emergency admissions: Higher readmission risk

Clinical Complexity Patterns



Diagnosis Burden:

- Total diagnoses per admission
- Readmitted: Higher (median 12 vs 10)

Medication Burden:

- Total medications per admission
- Readmitted: Higher (median 29 vs 22)

Lab Testing Intensity:

- Readmitted: Higher (median 80 vs 54)

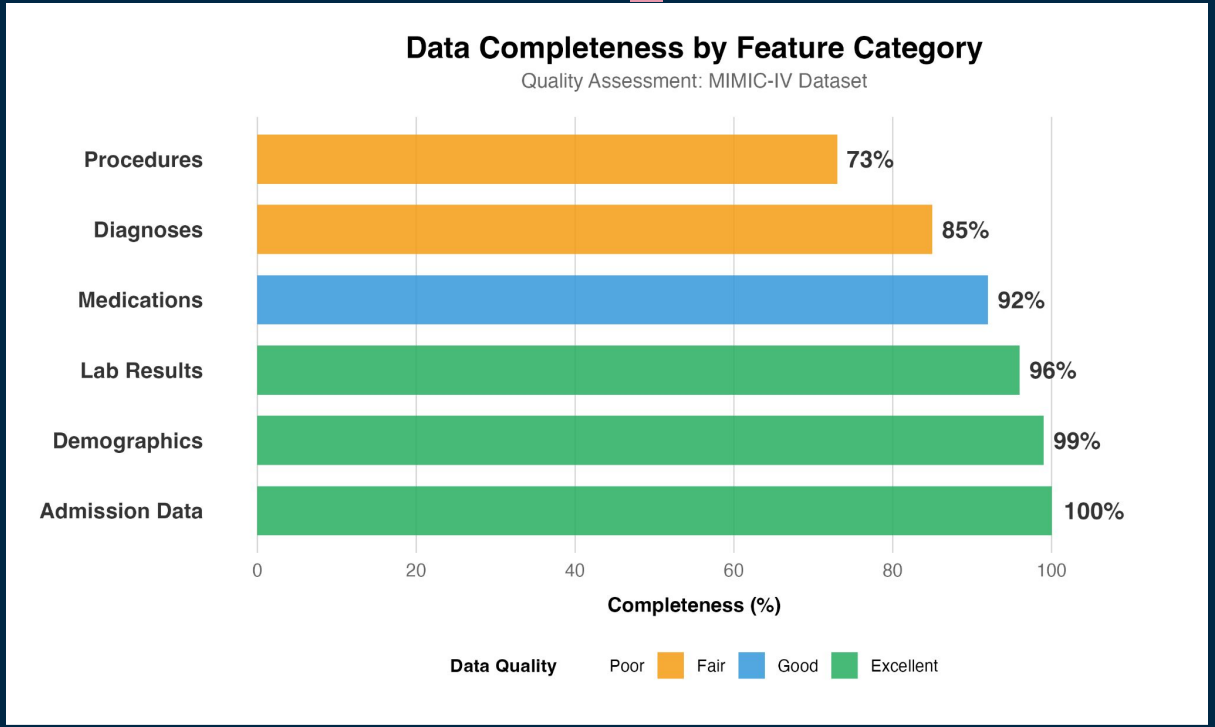
Data Quality and Completeness

MISSING DATA IS INFORMATIVE

No cardiac enzymes drawn? → Not a cardiac case
Few procedures recorded? → Less invasive care
Missingness = signal, not noise

DATA QUALITY DECISIONS

✓ Removed 19,442 patients with "UNKNOWN/DECLINED" race (data collection failures)
✓ Removed 531 patients with zero diagnosis records (likely errors)
✓ Kept patients with missing procedures/labs (absence is informative)



Feature Engineering Overview

Comorbidity Indices

Charlson Comorbidity Score
Heart failure flag
Diabetes flag
COPD flag
Pneumonia flag
Multi-morbidity indicators

Healthcare Utilization

Medication Count
Polypharmacy indicators
Lab testing intensity
Lab category diversity
Procedure complexity analysis

Medication Risk

Therapeutic category diversity
High-risk medication patterns
Elderly polypharmacy flags
Medication risk scores

Clinical Complexity

Total diagnoses count
Multi-system involvement
Diagnostic complexity score
ICU intensity patterns
Chronic complex indicators

BOTTOM LINE: 57 features | Zero missing values | All clinically validated

Feature Engineering Examples

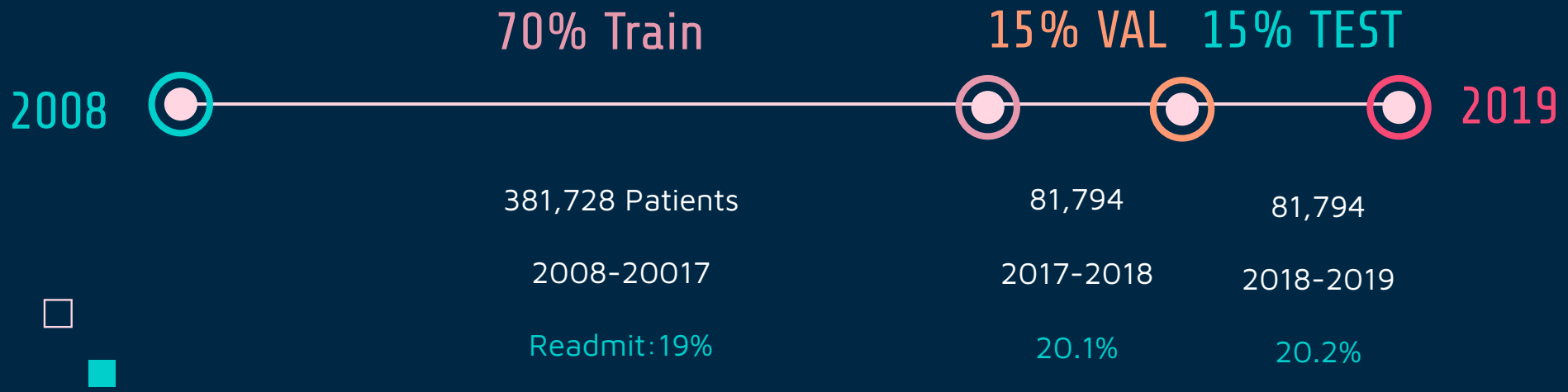
RAW DATA SOURCE	ENGINEERED FEATURE	EXAMPLE	CLINICAL RATIONALE
26M PRESCRIPTION RECORDS	MEDICATION_COUNT	28 MEDS	POLYPHARMACY = RISK
ICD-10 CODE "I50.9"	HAS_HEART_FAILURE	TRUE	HIGH-RISK CONDITION
120M LAB MEASUREMENTS	LAB_DIVERSITY_SCORE	8 CATEGORIES	TESTING BREADTH SIGNAL
AGE 72+ MEDS 15	ELDERLY_POLYPHARMACY	TRUE	AGE X MEDICATION INTERACTION
12 DX ACROSS FIVE SYSTEMS	COMPLEXITY_SCORE	14/58	MULTI-DIMENSIONAL ACUITY

- Philosophy: Aggregate clinical complexity, not raw clinical values

Data Split Strategy

Why temporal?

- ✓ Simulates real deployment: Train on history, predict future
- ✓ Prevents temporal data leakage: Treatment protocols evolve
- ✓ Conservative performance estimate: Tests stability over time
- ✓ Clinically realistic: Models must work on unseen patients



Modeling Approach

Ensemble Philosophy:

Balance interpretability with predictive power

Not Used:

- Deep Learning:
  - Insufficient temporal sequences
- SVM:
  - Computation cost with 380k patients
- Native Bayes:
  - Violated independence assumptions

	Logistic Regression	Random Forest	XGBoost
Strength	Interpretable	Non-linear	State-of-art
Healthcare Role	Clinical Standard	Interactions	Max accuracy
Interpretability	☆☆☆☆☆	☆☆☆	☆☆
Performance	☆☆☆	☆☆☆☆	☆☆☆☆☆

Training Methodology

Threshold Optimization

- Youden's Index (NOT default 0.5)
- 20% class imbalance makes 0.5 inappropriate
- Optimize to balance sensitivity/specificity
- Optimal threshold = 0.197 for XGBoost

Early Stopping

- XGBoost:
  - Stopped at interaction 407 of 500 max
- Prevents overfitting to training data

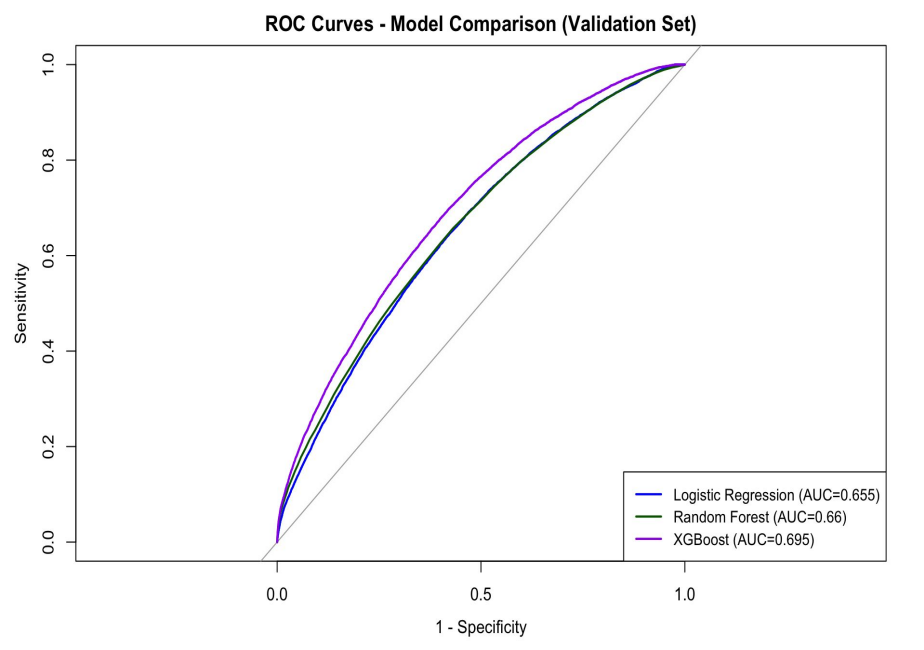
Regularization

- Logistic Regression:
  - L2 penalty (α = 0)
- XGBoost:
  - Learning rate 0.1, max depth 6, subsample 0.8

PRIMARY METRIC: AUC

AUC handles class imbalance better than accuracy

Validation Set Performance (n = 81,794)



Model	Logistic Regression	Random Forest	XGBoost
Threshold	0.1997	0.2050	0.1968
AUC	0.655	0.660	**0.695**
Sensitivity	63.74%	65.69%	**67.81%**
Specificity	58.56%	56.87%	**59.94%**
PPV	28.42%	28.22%	**30.41%**
NPV	86.22%	86.53%	**87.82%**
Accuracy	59.63%	58.68%	**61.55%**

Baseline: 20% readmission rate (random = 0.5 AUC)

Winner: XGBoost selected for final test evaluation

Final Model Performance – Held-Out Test Set

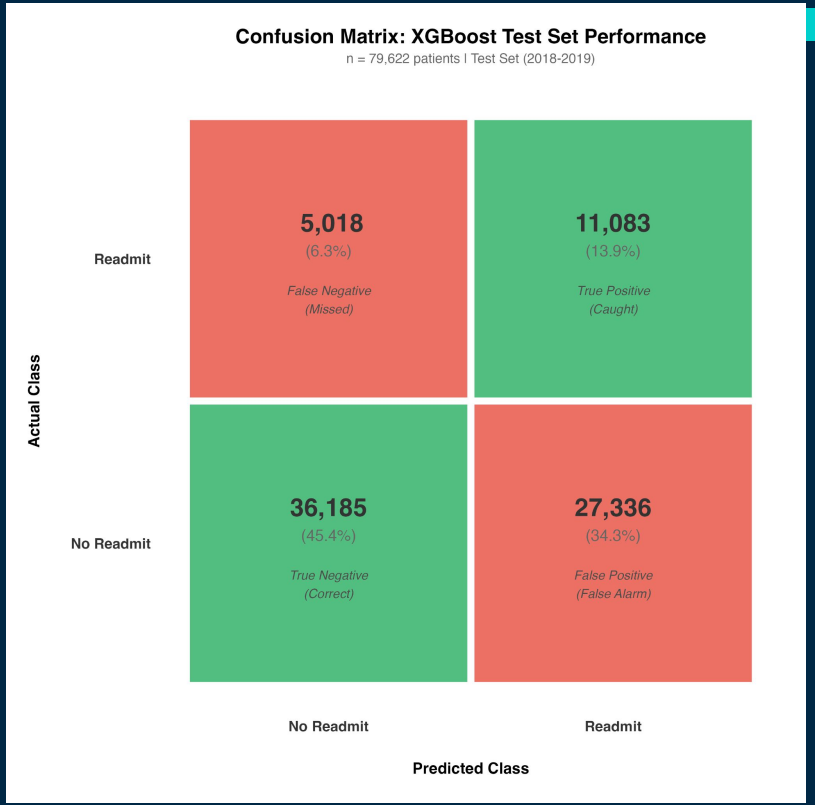
TEST SET: 81,794 patients from 2018-2019 (completely withheld)

	Validation	Test	Change
AUC	0.695	0.683	-1.28%
Sensitivity	67.8%	68.8%	+1%
Specificity	59.9%	56.9%	-3%
PPV	30.4%	29.8%	-0.6%
NPV	87.6%	87.3%	-0.3%

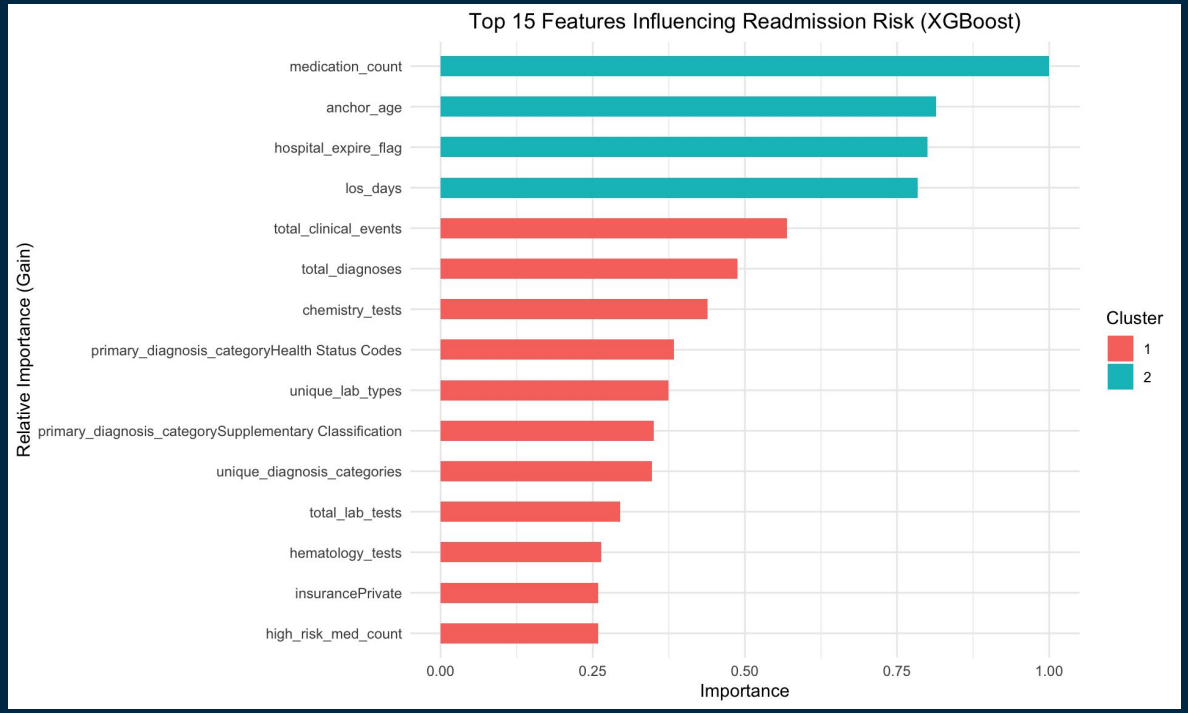
PRIMARY RESULT: AUC = 0.683

CLINICAL INTERPRETATION:

- **68.8% Sensitivity**: Identifies ~7 of 10 readmissions
- **29.8% PPV**: 49% improvement over 20% baseline
- **87.3% NPV**: When model says "low risk", correct 87% of the time



Top 15 Predictive Features (XGBoost)

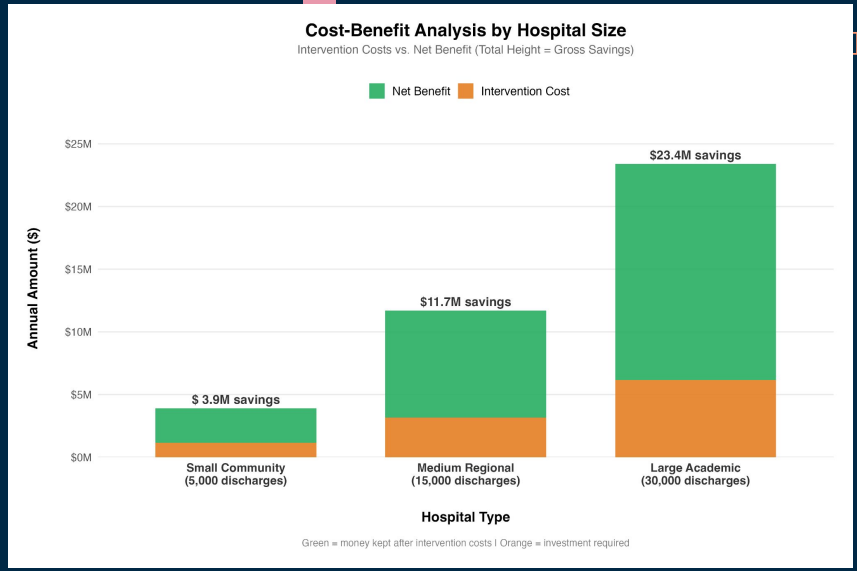


Key Insights

- Medication burden dominates (not just what it treats)
- Age & LOS capture physiologic reserve and acuity
- Multi-dimensional complexity beats single conditions
- Comorbidity indices validate healthcare standards

Financial Impact by Hospital Size

Metric	Small Community	Medium Regional	Large Academic
Annual Discharges	5,000	15,000	30,000
Readmissions Prevented	150	450	900
Gross Savings	$3,900,000	$11,700,000	$23,400,000
Net Benefit	$2,750,000	$8,550,000	$17,250,000
ROI	239.1%	271.4%	280.5%



ASSUMPTIONS:

- Model PPV: 29.8% (from test set)
- Intervention Costs: Care Transitions Intervention and nurse-led transitional care programs
- Effectiveness: 15-48% reduction based on systematic reviews (Coleman 2006; Naylor 2019)
- Base case: Uses 25% effectiveness (conservative)
- Average readmission cost: $26,000

Clinical Efficiency Analysis

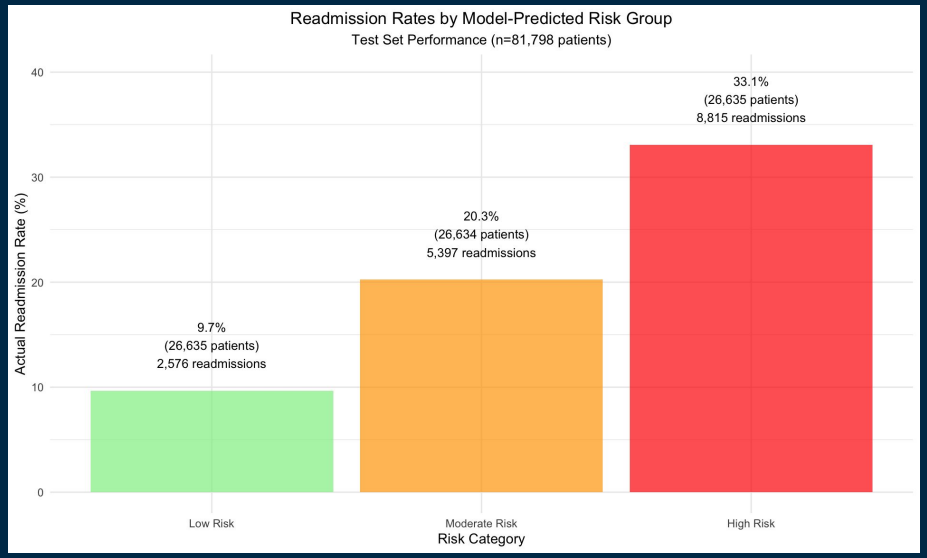
Key Insights:

- NNS = 13.4 is highly efficient for preventative medicine
- Compare to colorectal cancer screening (NNS ~300-500)
- Model guided approach 50% more efficient than random selection
- Every 13 patients screened prevents 1 costly admission

Metric	Value	Interpretation
Baseline Readmission Rate	20.0%	Without model: 1 in 5 patients readmit
Risk Among Flagged Patients (PPV)	29.8%	With model: 1 in 3 flagged readmit
Risk After Intervention (25% Reduce)	22.3%	With intervention: risk reduced to 23%
Absolute Risk Reduction	7.5%	7.5% absolute reduction in risk
Number Needed to Screen	13.4%	Screen ~13 patients to prevent 1
Number Needed to Treat	4.0	Treat 4 high-risk patients to prevent 1
Efficiency Gain vs Random Selection	1.5x	Model is 1.5x more efficient

Note: "ICD" stands for International Classification of Diseases

Clinical Translation: Match Resources to Risk



Risk Tier	High	Moderate	Low
% Cohort	33%	33%	33%
Readmission Rate	33.1%	20.3%	9.7%
Concentration	**52.5%**	32.1%	15.3%
Intervention	Intensive TCM $800-1,000 Home visits, 48hr	Standard TCM $400-600 Phone calls, appts	Minimal $100-200 Education materials

TCM: Transitional Care Management

Model Calibration Assessment

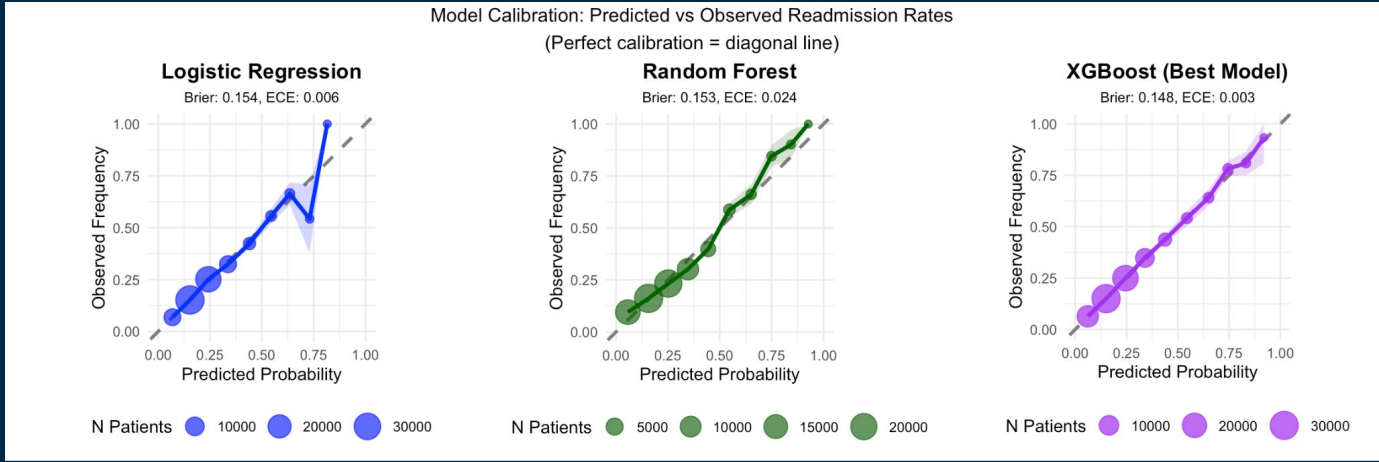
Calibration Metrics

Brier Score: 0.146 (good if <0.25)

Expected Calibration Error: 0.003 (excellent if < 0.05)

WHY THIS MATTERS:

- Clinicians can trust the risk scores
- Enables probabilistic reasoning, not just rank-ordering
- Supports resource allocation based on absolute risk



Equity Assessment: Calibration by Race/Ethnicity

Key Finding

Sensitivity range across well-defined groups: 17 percentage points

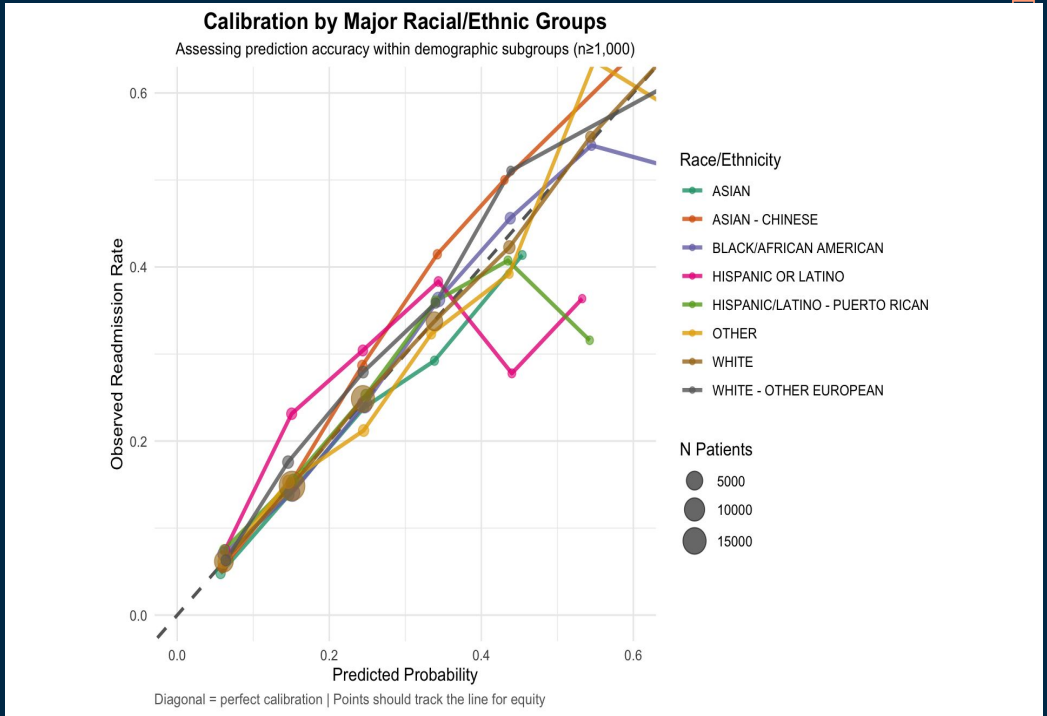
- High: Black/African American (75.5%)
- Low: White - Other European (64.3%)
    - 11.2 pp (major groups only)

Possible causes:

1. Differential readmission risk (22.2% vs 20.2%)
2. Sample size imbalance (50,193 vs 11,321)

Actions:

1. Acknowledge limitation
2. Group-specific calibration
3. Monthly monitoring



Study Limitations: Data & Scope

Single-Center Dataset

- Beth Israel Deaconess only
- Limited generalizability to rural/community hospitals
- External validation needed

Temporal Coverage

- Pre-Covid data
- Doesn't reflect telehealth expansion, current practice
- Mitigation: retrain on 2020-2024 data

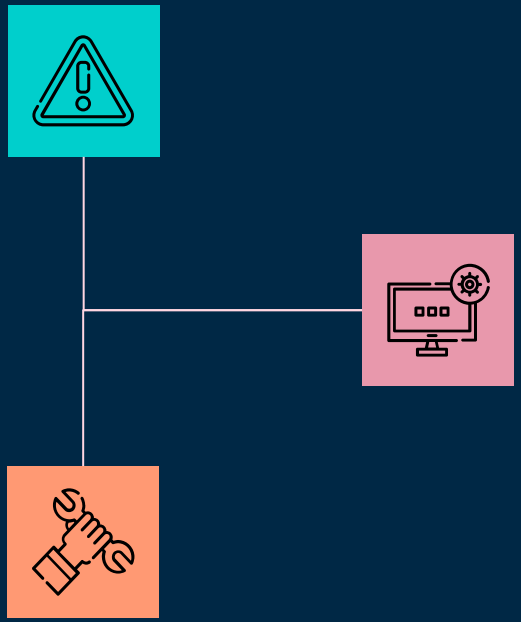
Post Discharge

- No social determinants
  - Housing, transportation, food security
- No medication adherence data
- No follow-up appointment attendance
- No caregiver support information
- These **directly** drive readmissions

Practical Deployment Challenges

Model Limitations

- Performance ceiling: 31% variance unexplained
- False positive burden: 70% of flagged don't readmit
- Interpretability: XGBoost is a black box



Mitigation Strategies

- Provide feature importance + similar patient examples
- Integrate into existing discharge workflow
- Pilot with champion physicians, show data
- Tier interventions by risk level

Deployment Challenges

- EHR Integration: Extract 57 features in real-time
- Alert fatigue: Flagging 40% of patients may overwhelm staff
- Intervention ability: Need TCM resources

What We've Accomplished

Model Performance

- XGBoost achieved 0.683 AUC on held-out data
- Excellent generalization, 1.28% drop from validation
- Sensitivity: 68.8% (catches 7 of 10 readmissions)
- PPV: 29.8% (49% improvement over 20% baseline)

Business Value

- Large Hospital: $17.25M net benefit, 280% ROI
- Medium Hospital: $8.55M net benefit, 271% ROI
- Small Hospital: $2.75M net benefit, 239% ROI

Clinical Impact

- Number Needed to Screen: 13.4 patients
- Among most efficient preventative interventions
- 1.5x more efficient than random selection
- Identifies 52.5% of readmissions in top 33% of patients

Key Predictive Features

- Clinical complexity (medication count, total clinical events)
- Physiologic reserved (age, LOS)
- Disease burden (total diagnoses, Charlson score)
- Healthcare utilization (lab tests, chemistry panels)
- Multi-dimensional complexity > single disease codes

Conclusions

Model Performance

Clinically useful accuracy (AUC 0.683, Sensitivity 68.8%)

Ready for prospective validation, not production deployment

Critical Success Factors

Validate on independent dataset (AUC > 0.65, PPV > 25%)

Integrate into clinical workflow without disruption

Robust fairness monitoring with equity protocols

Next Step

Prospective validation determines real-world impact