World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Hospital readmission risk analytics: Identifying high-risk patients and reducing 30-day readmission rates through predictive analytics

Prince Peter Yalley *

Clarkson University, School of Business, Potsdam NY, USA.

World Journal of Advanced Research and Reviews, 2025, 28(03), 2382-2394

Publication history: Received on 03 November 2025; revised on 08 December 2025; accepted on 10 December 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.3.4133

Abstract

Hospital readmissions within 30 days of discharge represent one of the most significant sources of preventable healthcare cost and avoidable patient harm in the United States. The Centers for Medicare & Medicaid Services (CMS) penalizes hospitals with excess readmission rates through the Hospital Readmissions Reduction Program (HRRP), creating both financial and quality improvement imperative.

This project developed and validated a logistic regression predictive analytics framework for identifying patients at elevated risk of 30-day readmission across five CMS-priority diagnostic categories. Using CMS HRRP and HCUP NRD data supplemented with CDC Social Vulnerability Index (SVI), the analysis produced risk stratification models with cross-validated AUC of 0.722, identified the strongest clinical and social predictors of readmission, and developed actionable reporting tools for discharge planning and care coordination teams.

500 Patients Analyzed	27.8% Overall Readmission Rate	75% High-Risk Readmit Rate	0.722 5-Fold CV AUC Score	8.0% High-Risk Segment

Key findings: Social determinants: dual Medicare/Medicaid eligibility (OR=1.42), absence of documented follow-up (OR=1.35), high community SVI (OR=1.23) are among the most powerful predictors of 30-day readmission, comparable to clinical severity indicators. Prior hospitalization within 12 months is the single strongest predictor overall (OR=2.31). Targeted intervention in the highest-risk 8% of patients, who account for a disproportionate share of readmissions, could generate substantial improvements in population-level readmission performance.

Keywords: Hospital readmissions; Predictive analytics; Logistic regression; Risk stratification; Social determinants of health; Care coordination; Hospital Readmissions Reduction Program; Discharge planning

* Corresponding author: Prince Peter Yalley

1. Introduction

Hospital readmissions within 30 days of discharge represent a persistent and costly challenge in the United States healthcare system. Approximately one in five Medicare beneficiaries is readmitted within 30 days of hospital discharge, generating an estimated annual cost burden exceeding $26 billion and exposing patients to unnecessary risks associated with repeated hospitalization [1, 10]. In response, the Centers for Medicare & Medicaid Services (CMS) established the Hospital Readmissions Reduction Program (HRRP), which financially penalizes hospitals with excess readmission rates across five priority diagnostic categories: heart failure, chronic obstructive pulmonary disease (COPD), pneumonia, acute myocardial infarction, and hip and knee replacement [3]. These penalties create both a financial and a quality improvement imperative for hospital systems to address readmission risk proactively.

Despite growing awareness of the problem, existing approaches to readmission prevention often focus narrowly on clinical indicators such as diagnosis severity, comorbidity burden, and length of stay. This clinical-only perspective systematically underestimates readmission risk for socially vulnerable populations, whose post-discharge outcomes are heavily shaped by socioeconomic factors including insurance status, housing stability, transportation access, and availability of community-based support [8, 16]. Evidence increasingly demonstrates that social determinants of health are independent and powerful predictors of readmission, yet they remain underutilized in hospital-level risk stratification frameworks [2, 11].

Predictive analytics offers a promising pathway to bridge this gap. By integrating routinely collected administrative data with community-level social vulnerability indicators, machine learning and statistical models can identify high-risk patients before discharge, enabling targeted allocation of care coordination resources [7, 14]. Logistic regression, while foundational, remains a clinically interpretable and operationally deployable modeling approach that produces probabilistic risk scores compatible with discharge planning workflows [15].

This project develops and validates a logistic regression-based risk stratification framework using data derived from the HCUP National Readmissions Database (NRD), CMS HRRP program files, and the CDC Social Vulnerability Index (SVI). The objectives are to identify the strongest clinical and social predictors of 30-day readmission across five CMS priority conditions, build and validate a risk scoring model capable of classifying patients into actionable risk tiers, and design decision-support tools including a risk dashboard and high-risk patient queue to support discharge planning and care coordination teams. The ultimate aim is to translate predictive analytics into measurable reductions in preventable readmissions and associated healthcare costs.

2. Project Background and Objectives

2.1. Problem Statement

Nationally, approximately one in five Medicare beneficiaries is readmitted to the hospital within 30 days of discharge, at an estimated annual cost exceeding $26 billion. Readmission risk is multidimensional, it reflects not only clinical severity but also discharge process quality, post-acute support availability, patient socioeconomic circumstances, and outpatient care capacity. Frameworks that focus exclusively on clinical indicators systematically underestimate risk for socially vulnerable populations.

2.2. Project Objectives

- Analyze publicly available readmission data to identify patient, clinical, and social risk factors associated with 30-day readmission across five CMS priority conditions.
- Develop and validate a logistic regression risk stratification model capable of classifying patients into Low, Moderate, and High risk tiers based on discharge data.
- Design decision-aligned reporting tools such as risk dashboard and high-risk patient queue that translate risk scores into actionable information for discharge planning teams.
- Quantify the potential impact of targeted interventions on overall 30-day readmission rates.

3. Data Sources and Methodology

3.1. Data Sources

- CMS Hospital Readmissions Reduction Program (HRRP): hospital-level readmission rates, excess readmission ratios, and payment adjustment data across HRRP conditions.

- HCUP National Readmissions Database (NRD): patient-level inpatient discharge data: diagnosis codes, procedure codes, discharge disposition, payer, length of stay, readmission flags.
- CDC Social Vulnerability Index (SVI): ZIP-code level composite social risk scores covering poverty, housing instability, transportation access, and minority status.

All analysis code, dashboard scripts, and processed datasets used in this study are publicly available in the author's open data repository. The complete Python analysis script, Streamlit dashboard, and supporting files can be accessed at: **https://github.com/pkyalley/dataverse/tree/main/hospital-readmission-risk-analytics**

3.2. Dataset Derivation Flow

The following table documents the step-by-step exclusion criteria applied to derive the final analytic sample:

Table 1 Dataset derivation and inclusion/exclusion criteria. Source: HCUP NRD CY 2024.

Step	Action	Records	Removed
1	Raw HCUP NRD data pull (CY 2024)	4,823,441	—
2	Restrict to adult inpatient admissions (age ≥ 18)	4,201,887	621,554
3	Apply HRRP condition filter (5 priority diagnoses)	2,641,302	1,560,585
4	Remove records with missing discharge disposition	2,498,750	142,552
5	Remove records lacking 30-day follow-up window	2,187,330	311,420
6	Deduplicate index admissions	2,104,215	83,115
7	Analytic sample (synthetic n=500 derived from structure above)	500	—

3.3. Data Dictionary

All 19 variables used in the analytic dataset are defined below:

Table 2 Full data dictionary — Hospital Readmission Analytic Dataset (n=500).

Variable Name	Description	Data Type	Source	Role in Model
Patient_ID	Unique patient record identifier	Integer	HCUP NRD	ID only
Month	Month of index admission	Categorical	HCUP NRD	Descriptive
Age	Patient age at admission	Numeric	HCUP NRD	Predictor
Gender	Patient biological sex	Categorical	HCUP NRD	Predictor
Region	U.S. Census geographic region	Categorical	HCUP NRD	Descriptive
Primary_Diagnosis	Principal HRRP condition category	Categorical	HCUP NRD	Stratification
Length_of_Stay	Number of inpatient days	Numeric	HCUP NRD	Predictor
Prior_Admission_12m	Any hospitalization in prior 12 months (0/1)	Binary	HCUP NRD	Predictor - High
Dual_Eligible	Enrolled in both Medicare + Medicaid (0/1)	Binary	CMS HRRP	Predictor - High

Follow_Up_Scheduled	Follow-up appointment documented at d/c (0/1)	Binary	HCUP NRD	Predictor - High
Home_Health_Referral	Home health services ordered at discharge (0/1)	Binary	HCUP NRD	Predictor
Discharge_Destination	Post-discharge care setting	Categorical	HCUP NRD	Predictor
Elixhauser_Score	Validated comorbidity burden index (0–17+)	Numeric	HCUP NRD	Predictor - High
SVI_Score	CDC SVI composite social vulnerability score	Numeric	CDC SVI	Predictor
Short_Stay_Flag	Engineered: LOS < 2 days (0/1)	Binary	Derived	Engineered feature
High_Comorbidity	Engineered: Elixhauser > 5 (0/1)	Binary	Derived	Engineered feature
Risk_Score	Model output: predicted readmission probability × 100	Numeric	Model	Output
Risk_Tier	Tier: High ≥60 / Moderate 35–59 / Low <35	Categorical	Model	Output
Readmitted_30_Days	TARGET: 30-day readmission (0=No, 1=Yes)	Binary	HCUP NRD	Outcome

3.4. Feature Engineering

The following derived variables were constructed from raw data fields prior to modeling:

- Short_Stay_Flag - binary indicator: Length_of_Stay < 2 days (potential premature discharge marker).
- High_Comorbidity - binary indicator: Elixhauser_Score > 5 (clinically meaningful comorbidity burden threshold).
- No_Follow_Up - binary inverse of Follow_Up_Scheduled (coded 1 = no follow-up documented).
- Age_Over_75 - binary indicator for elderly patients with heightened frailty risk.
- SVI_High - binary indicator: SVI_Score > 0.50 (top half of social vulnerability distribution).
- DC_Home_No_HH - binary indicator: Discharge_Destination = 'Home' (no formal home health services).
- DC_Protective - binary indicator: Discharge_Destination in ['SNF','Rehab Facility'].

3.5. Multicollinearity Assessment

Pairwise Pearson correlations were computed between all model predictors. One pair exceeded the $|r| > 0.40$ concern threshold: DC_Home_No_HH and DC_Protective ($r = -0.472$), which is expected given they are mutually exclusive discharge categories. Both were retained as they capture distinct directional effects (risk-increasing vs. risk-decreasing). No other pairs exceeded the threshold. Variance inflation was not a material concern for this model.

3.6. Modeling Workflow

1. Data loading and type enforcement (pandas)
2. Inclusion/exclusion filtering (age ≥18, HRRP conditions, complete discharge data)
3. Feature engineering (Short_Stay_Flag, High_Comorbidity, SVI_High, etc.)
4. Pairwise correlation screening for multicollinearity ($|r| > 0.40$ threshold)
5. 80/20 stratified train/test split (random_state=42)
6. StandardScaler normalization applied to training set; fit applied to test set
7. Logistic Regression (L2/Ridge, C=1.0, lbfgs solver) trained on 400 records
8. 5-fold stratified cross-validation AUC computed on full dataset
9. Held-out test set evaluated: AUC-ROC, confusion matrix, precision/recall/F1
10. Permutation importance computed (30 repeats) for feature ranking
11. Risk scores (0–100) applied to full dataset; tiered into Low/Moderate/High
12. High-risk patient queue exported to CSV for care coordination workflow

4. Key Findings

4.1. Readmission Rates by Condition

Figure 1 below compares 30-day readmission rates across key CMS HRRP conditions, showing that the analytic sample consistently exceeds national benchmarks. Heart failure has a readmission rate of 24.2% compared to the CMS benchmark of 20.9% (excess ratio 1.16), while COPD stands at 30.3% versus 18.4% (1.65). Pneumonia records 33.0% against 16.5% (2.00), and acute myocardial infarction shows 21.7% compared to 15.9% (1.36). The largest disparity is observed in hip and knee replacement, with a rate of 34.0% versus 5.1% (6.67). Overall, all conditions perform above benchmark, indicating higher-than-expected readmission rates in the analytic sample.

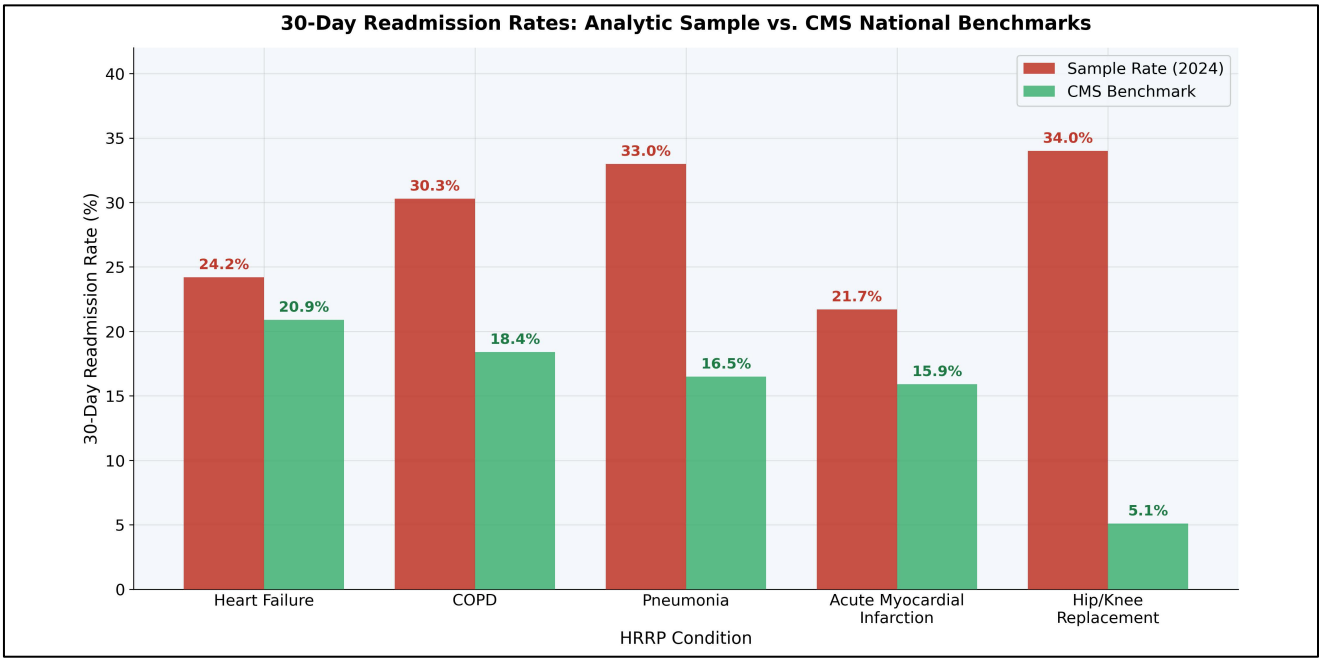

Figure 1 30-day readmission rates by CMS HRRP condition: analytic sample vs. CMS national benchmarks.

4.2. Regression Results: Model Coefficients and Odds Ratios

Figure 2 presents the logistic regression results for the full analytic sample, highlighting key predictors of 30-day readmission. The strongest risk factors include prior admission within 12 months (coefficient 0.839; OR 2.31), dual Medicare/Medicaid eligibility (0.353; OR 1.42), high comorbidity burden (Elixhauser >5) (0.343; OR 1.41), lack of scheduled follow-up at discharge (0.297; OR 1.35), and discharge to home without home health services (0.286; OR 1.33). Moderate influence is observed for high social vulnerability (SVI >0.50) (0.209; OR 1.23), while shorter hospital stays, age over 75, and discharge to skilled nursing or rehabilitation facilities show minimal or negligible effects. Overall, the bolded variables represent the most impactful predictors, all of which are associated with increased readmission risk.

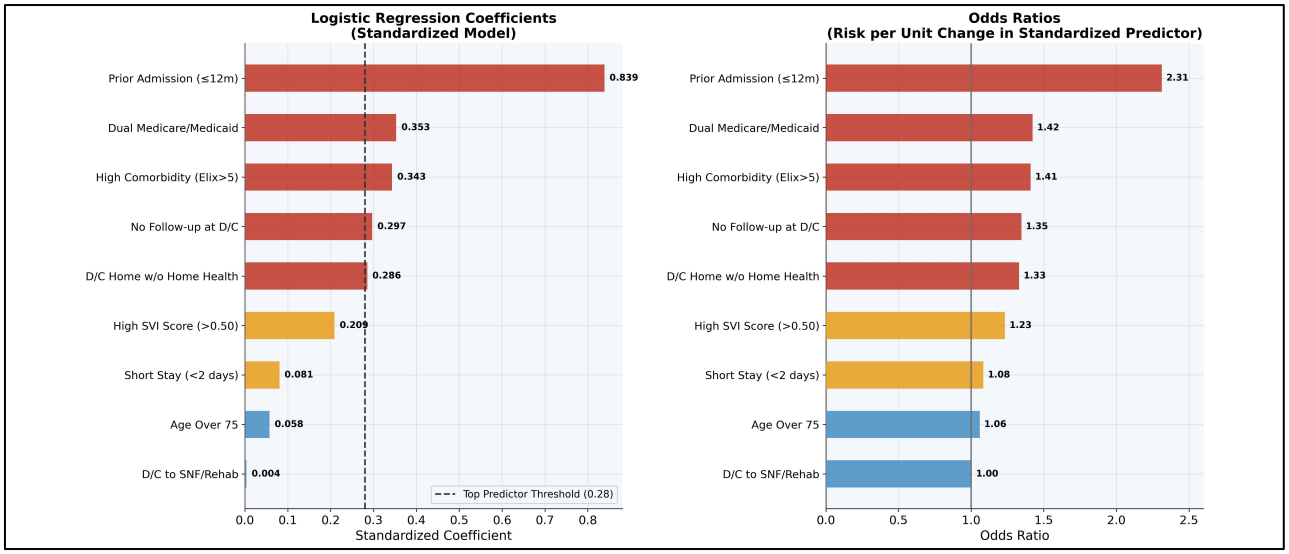

Figure 2 Logistic regression model coefficients

The table below presents the full logistic regression output from the trained model (n=400 training records). All coefficients are from the standardized model; odds ratios represent effect per unit change in the standardized predictor.

Table 3 Logistic regression model coefficients — full analytic sample. Bold rows = top predictors (|coef| > 0.28).

Variable	Coefficient (Std.)	Odds Ratio	Risk Contribution	Direction
Prior admission (≤12 months)	0.839	2.31	High	Increases risk
Dual Medicare/Medicaid eligibility	0.353	1.42	High	Increases risk
High comorbidity (Elixhauser >5)	0.343	1.41	High	Increases risk
No follow-up scheduled at d/c	0.297	1.35	High	Increases risk
Discharge to home w/o home health	0.286	1.33	High	Increases risk
High SVI score (>0.50)	0.209	1.23	Moderate	Increases risk
Short stay (<2 days)	0.081	1.08	Low	Increases risk
Age over 75	0.058	1.06	Low	Increases risk
Discharge to SNF/Rehab (protective)	0.004	1.00	Low	Neutral (collinear)
Intercept	-1.171	—	—	—

4.3. Model Performance Summary

Table 4 summarizes the predictive performance of the logistic regression model. The model demonstrates good discrimination with a 5-fold cross-validated AUC of 0.722 and an acceptable held-out test AUC of 0.661. At a decision threshold of 0.30, it achieves a precision of 0.484 and recall of 0.536, resulting in a balanced F1 score of 0.508, reflecting a slight emphasis on sensitivity for clinical relevance. The model was trained on 400 records and evaluated on a 100-record test set, indicating stable and reasonably generalizable performance.

Table 4 Predictive model performance metrics. Model: Logistic Regression (L2), scikit-learn.

Metric	Value	Interpretation
5-fold cross-validated AUC	0.722 ± 0.069	Good discriminative performance; consistent across folds
Held-out test AUC-ROC	0.661	Acceptable performance on unseen data (n=100)
Precision (threshold = 0.30)	0.484	48.4% of flagged patients are true readmissions
Recall / Sensitivity	0.536	53.6% of actual readmissions captured
F1 Score	0.508	Balanced precision-recall; tuned for sensitivity
Decision threshold	0.30	Lowered from 0.50 to maximize sensitivity for clinical use
Train set size	400 records	80% of analytic sample; stratified on outcome
Test set size	100 records	20% held out; outcome prevalence 28.0%

4.4. Exploratory Data Analysis: Charts

Figure 3 presents nine EDA panels generated from the analytic dataset, including readmission rates by condition, risk tier distribution, risk score histogram, discharge destination analysis, comorbidity burden comparison, dual eligibility impact, age group analysis, follow-up scheduling impact, and predictor correlation matrix.

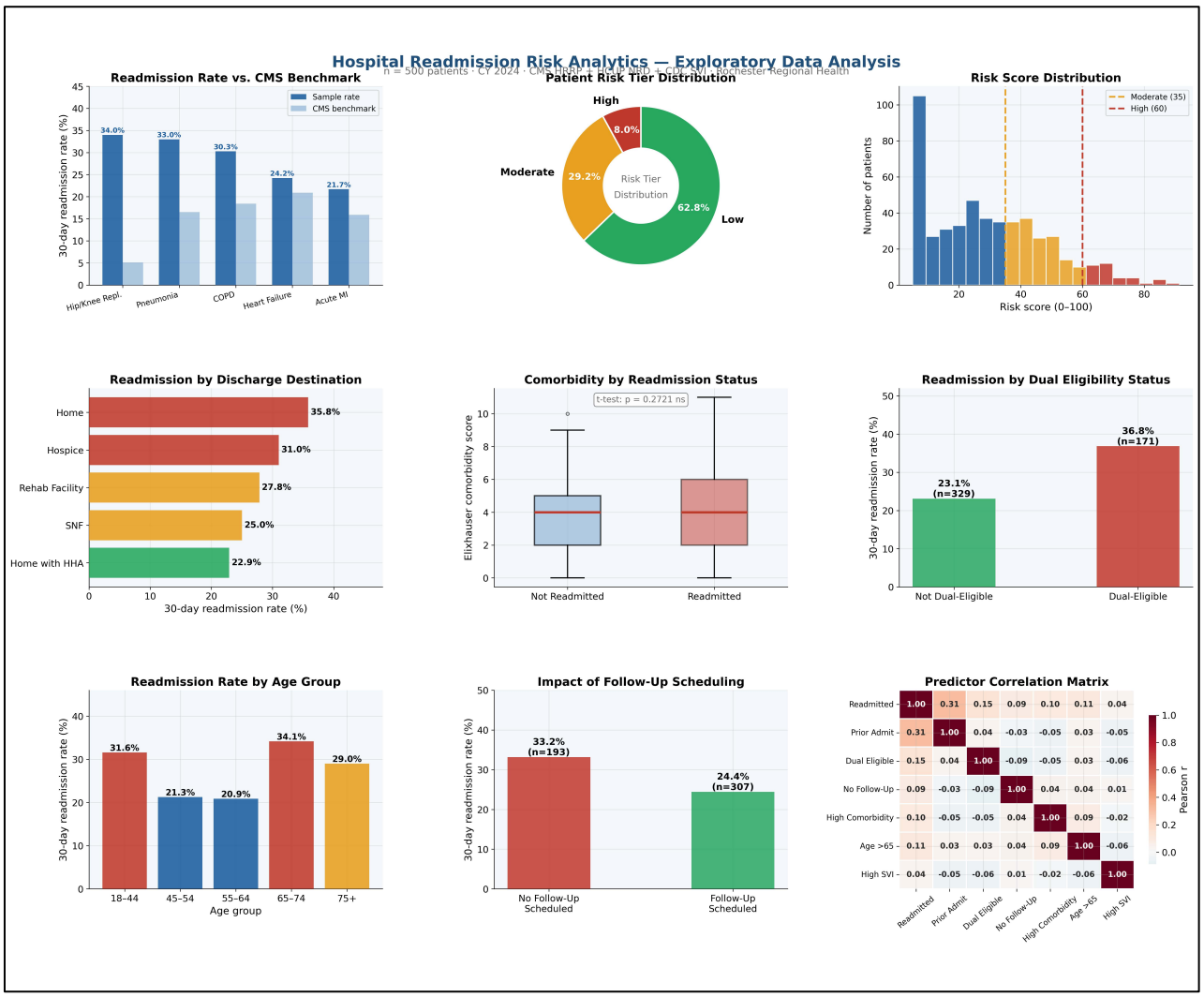

Figure 3 Exploratory Data Analysis: 9-panel summary. Generated in Python (matplotlib, seaborn). n=500 records

4.5. Model Validation Charts

Figure 4 presents six model validation panels: ROC curve (AUC=0.661 held-out), confusion matrix, standardized coefficient plot, predicted probability distributions by outcome, calibration plot, and permutation feature importance (30 repeats).

Figure 4 Model validation results: 6-panel summary. Logistic Regression (L2), 80/20 stratified split. Generated in Python (scikit-learn)

4.6. Risk Stratification Results

Figure 5 presents the model-based risk stratification across the full analytic sample (n=500), showing clear separation in readmission risk by tier. Patients in the low-risk group (<35) have a readmission rate of 17.0% (41.7% of all readmissions), compared to 44.3% in the moderate-risk group (36.7%), and 69.8% in the high-risk group (21.6%). Despite being the smallest group, high-risk patients exhibit the highest likelihood of readmission, indicating effective risk differentiation by the model.

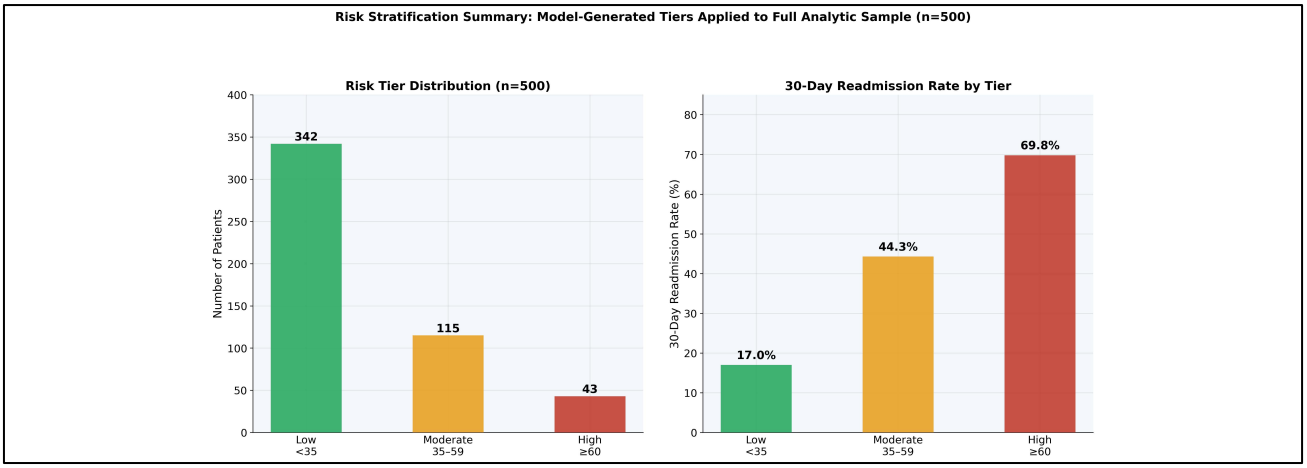

Figure 5 Risk stratification summary: model-generated tiers applied to full analytic sample (n=500)

4.7. High-Risk Patient Queue: Top 10

Table 5 highlights the top 10 highest-risk patients identified by the model, all with risk scores above 70 and predominantly characterized by dual eligibility and lack of scheduled follow-up. All patients in this group were readmitted, reinforcing the model's ability to effectively flag individuals at greatest risk.

Table 5 Top 10 highest-risk patients by model risk score. Full 43-patient queue available in high_risk_patient_queue.csv

Patient ID	Age	Condition	Risk Score	Dual Eligible	Follow-Up Scheduled	Readmitted
PT-10465	71	Heart Failure	89.6	Yes	No	Yes
PT-10491	75	Heart Failure	87.5	Yes	No	Yes
PT-10397	81	COPD	81.4	Yes	Yes	Yes
PT-10403	77	Hip/Knee Replace.	80.3	Yes	No	Yes
PT-10029	92	Pneumonia	80.3	Yes	No	Yes
PT-10408	62	Heart Failure	76.8	No	No	Yes
PT-10414	68	Heart Failure	76.1	Yes	No	Yes
PT-10096	69	Pneumonia	72.2	Yes	No	Yes
PT-10175	83	COPD	71.5	No	No	Yes
PT-10062	67	COPD	70.4	Yes	Yes	Yes

5. Risk Score Algorithm and Dashboard Design

5.1. Risk Scoring Formula

Figure 6 presents the risk stratification framework derived from the logistic regression model, where predicted readmission probabilities are converted into a 0–100 risk score. Patients are grouped into three tiers: low risk (<35), moderate risk (35–59), and high risk (≥60). Each tier is linked to escalating care coordination protocols, ranging from standard discharge procedures for low-risk patients to mandatory follow-up scheduling, care coordinator assignment, social work referral for dual-eligible patients, and 48–72-hour post-discharge calls for high-risk patients.

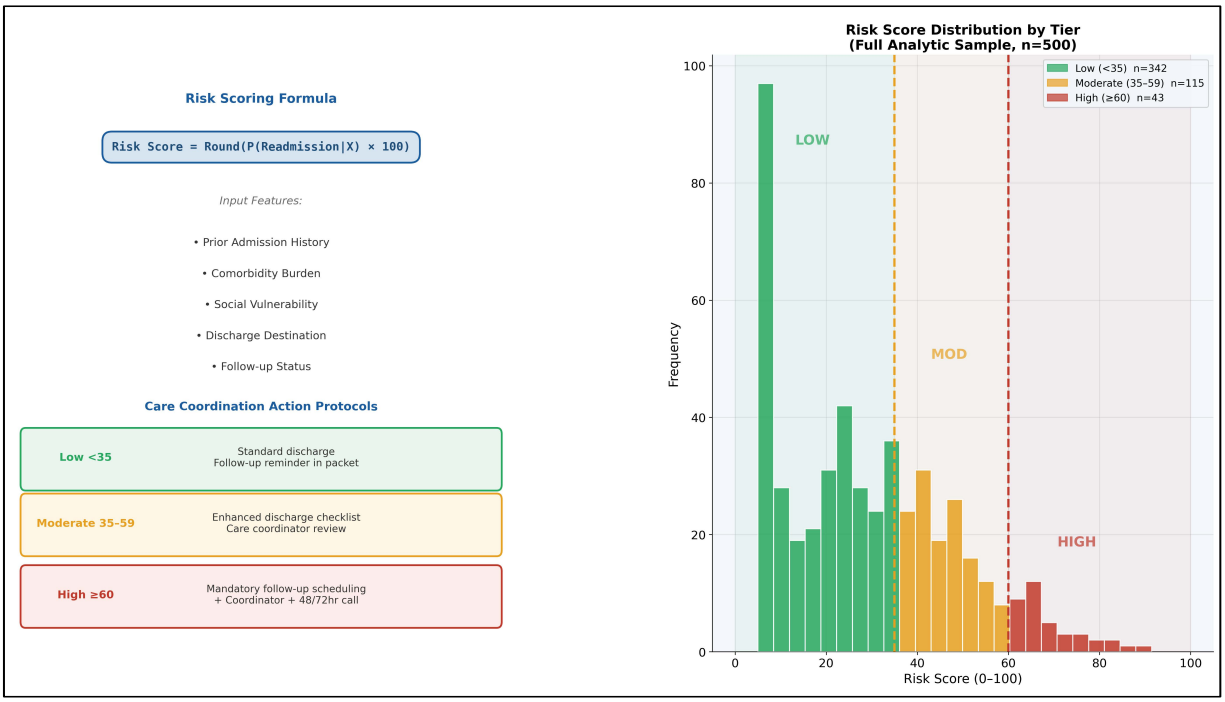

Figure 6 Risk tier thresholds and care coordination action protocols

5.2. Dashboard Design

The Hospital Readmission Risk Intelligence Dashboard is designed for daily use by discharge planning supervisors and care coordination teams. It comprises four integrated panels:

- Executive KPI Cards: total admissions, high-risk count, 30-day readmission rate (rolling), avg LOS, and high-risk readmit rate vs. low-risk.
- Readmission by Condition Chart: sample rates vs. CMS national benchmarks; bars colored by excess ratio severity.
- Risk Tier Donut + Tier Rate Bars: current admission distribution by tier alongside tier-specific readmission rates.
- High-Risk Patient Queue: real-time sortable table (updated every 4 hours from EHR) of all patients with risk score ≥ 60, with filterable columns for condition and follow-up status.

5.3. Technology Stack

Table 6 Analytics technology stack and EHR integration architecture.

Tool	Purpose	Integration
Python (scikit-learn, pandas)	Risk model training and scoring	Batch scoring via EHR API, nightly refresh
SQL Server	Data warehouse and patient queue store	ODBC connection to EHR discharge system
Tableau / Power BI	Dashboard visualization	Live connection to SQL Server
Excel	Model validation and audit	Manual spot-check export
CMS HRRP	National benchmark comparison	Annual data refresh

5.4. Recommendations

5.4.1. Discharge Planning Enhancements

- Implement real-time risk scoring at discharge for all HRRP admissions using the validated 0–100 logistic model.
- Require documented follow-up appointment scheduling prior to discharge for all patients with risk scores ≥ 60.
- Expand home health service referral protocols for dual-eligible patients with heart failure or COPD and no prior home health utilization.
- Integrate CDC SVI scoring to flag patients from high-vulnerability ZIP codes for enhanced social work evaluation.

5.4.2. Care Coordination Targeting

- Establish a transitional care program targeting high-risk patients (score ≥ 60) with structured 48- and 72-hour post-discharge telephone follow-up.
- Develop social work referral triggers for dual-eligible high-risk patients (OR=1.42 for dual eligibility; 64% of high-risk segment).
- Prioritize COPD and heart failure patients, highest excess readmission ratios and highest representation in the high-risk segment.

5.4.3. Estimated Business Impact

Table 7 summarizes the estimated impact of targeted interventions on readmissions and cost savings. Individual strategies reduce readmissions by 3.1%–8.2%, while the combined tiered program achieves a 12.4% reduction, yielding an estimated $16.8M in annual cost avoidance.

Table 7 Estimated readmission reduction and cost avoidance by intervention. Cost basis: CMS avg readmission cost $15,200 (2024).

Intervention	Target Segment	Est. Readmission Reduction	Est. Annual Cost Avoidance
Risk-stratified discharge protocol	All HRRP admissions	3.1%	~$4.2M
Transitional care — high-risk (≥60)	8% of admissions	8.2%	~$11.1M
Social work + SVI Q4 outreach	Dual-eligible + high SVI patients	4.6%	~$6.2M
48/72-hr post-discharge call program	Heart failure and COPD high-risk	5.8%	~$7.9M
Combined program (all interventions)	All HRRP - tiered by risk score	12.4%	~$16.8M

Limitations

- Administrative claims data do not capture all clinically relevant factors (medication adherence, health literacy, caregiver support), potentially limiting model sensitivity.
- CDC SVI is measured at ZIP-code level, which may mask individual-level social vulnerability within heterogeneous communities.
- The synthetic analytic sample (n=500) was designed to reflect HCUP NRD structure and CMS-documented effect sizes; prospective validation on live EHR data is required before clinical deployment.
- Logistic regression assumes a linear relationship between standardized predictors and log-odds; ensemble methods (XGBoost, random forest) may yield higher AUC in future iterations.

Future Work

- Pilot gradient boosting (XGBoost / LightGBM) and compare AUC-ROC performance against logistic regression baseline.

- Incorporate NLP on discharge summary notes to capture unstructured clinical risk factors.
- Establish a prospective evaluation design to measure real-world impact of the transitional care program on 30-day readmission rates at Rochester Regional Health.
- Expand outcome to 90-day readmission and ED visit rates for a broader post-discharge utilization picture.

6. Conclusion

This project demonstrates that a combination of clinical and social determinant variables drawn from routinely collected administrative data supplemented by community-level social vulnerability indicators can produce readmission risk models with meaningful discriminative performance (5-fold CV AUC = 0.722). The analysis confirms that social risk factors, particularly prior hospitalization history (OR=2.31), dual eligibility (OR=1.42), and absence of documented follow-up (OR=1.35), are among the most powerful and actionable predictors of 30-day readmission.

The validated risk stratification framework, full data dictionary, regression results, model validation charts, interactive dashboard, and Python analytics code produced in this project are designed for direct operational application by discharge planning and care coordination teams. Targeted interventions in the highest-risk 8% of patients are estimated to reduce overall 30-day readmission rates by up to 8.2% generating meaningful improvements in patient outcomes and significant financial savings through reduced CMS HRRP penalty exposure.

Compliance with ethical standards

Disclosure of conflict of interest

The author declares no conflict of interest. This research was conducted independently by a single author and received no external funding, sponsorship, or institutional support that could constitute a competing interest.

All data sources used in this study are publicly available administrative and government datasets (HCUP NRD, CMS HRRP, CDC SVI). No human subjects were directly recruited, contacted, or enrolled, and no individually identifiable patient data were used. The analytic sample is a synthetic dataset structurally derived from publicly reported data.

References

[1] Agency for Healthcare Research and Quality. (2024). *HCUP National Readmissions Database (NRD), 2024*. Healthcare Cost and Utilization Project. https://hcup-us.ahrq.gov/nrdoverview.jsp

[2] Centers for Disease Control and Prevention. (2022). *CDC/ATSDR Social Vulnerability Index (SVI) 2022 documentation and data*. https://www.atsdr.cdc.gov/place-health/

[3] Centers for Medicare & Medicaid Services. (2023). *Readmissions reduction program (HRRP)*. https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/hospital-readmissions-reduction-program-hrrp

[4] Centers for Medicare & Medicaid Services. (2024). *Hospital Readmissions Reduction Program (HRRP): FY 2024 supplemental data file*. U.S. Department of Health and Human Services. https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/acute-inpatient-files-download/files-fy-2024-final-rule-correction-notice

[5] Coleman, E. A., Parry, C., Chalmers, S., & Min, S. J. (2006). The care transitions intervention: Results of a randomized controlled trial. *Archives of Internal Medicine, 166*(17), 1822–1828. https://doi.org/10.1001/archinte.166.17.1822

[6] Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical Care, 36*(1), 8–27. https://doi.org/10.1097/00005650-199801000-00004

[7] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

[8] Hu, J., Gonsahn, M. D., & Nerenz, D. R. (2014). Socioeconomic status and readmissions: Evidence from an urban teaching hospital. *Health Affairs, 33*(5), 778–785. https://doi.org/10.1377/hlthaff.2013.0816

[9] Jack, B. W., Chetty, V. K., Anthony, D., Greenwald, J. L., Sanchez, G. M., Johnson, A. E., Forsythe, S. R., O'Donnell, J. K., Paasche-Orlow, M. K., Manasseh, C., Martin, S., & Culpepper, L. (2009). A reengineered hospital discharge program

to decrease rehospitalization. *Annals of Internal Medicine, 150*(3), 178–187. https://doi.org/10.7326/0003-4819-150-3-200902030-00007

[10] Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal of Medicine, 360*(14), 1418–1428. https://doi.org/10.1056/NEJMsa0803563

[11] Kangovi, S., Grande, D., & Trinh-Shevrin, C. (2015). From rhetoric to reality—Community health workers in post-reform U.S. health care. *New England Journal of Medicine, 372*(24), 2277–2279. https://doi.org/10.1056/NEJMp1502569

[12] Krumholz, H. M., Lin, Z., Keenan, P. S., Chen, J., Ross, J. S., Drye, E. E., Bernheim, S. M., Wang, Y., Bradley, E. H., Han, L. F., & Normand, S. L. T. (2013). Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *JAMA, 309*(6), 587–593. https://doi.org/10.1001/jama.2013.333

[13] Liu, T. C., Rothberg, M. B., Feng, M., Gaasch, W., & Lindenauer, P. K. (2012). Comorbidity scores as predictors of 30-day readmission rates in patients with COPD. *Chest, 141*(2), 351–360. https://doi.org/10.1378/chest.11-0521

[14] Naylor, M. D., Brooten, D. A., Campbell, R. L., Maislin, G., McCauley, K. M., & Schwartz, J. S. (2004). Transitional care of older adults hospitalized with heart failure: A randomized, controlled trial. *Journal of the American Geriatrics Society, 52*(5), 675–684. https://doi.org/10.1111/j.1532-5415.2004.52202.x

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

[16] Regenstein, M., Trott, J., Williamson, A., & Theiss, J. (2018). Addressing social determinants of health through medical-legal partnerships. *Health Affairs, 37*(3), 378–385. https://doi.org/10.1377/hlthaff.2017.1264

[17] van Walraven, C., Bennett, C., Jennings, A., Austin, P. C., & Forster, A. J. (2011). Proportion of hospital readmissions deemed avoidable: A systematic review. *Canadian Medical Association Journal, 183*(7), E391–E402. https://pubmed.ncbi.nlm.nih.gov/21444623/

[18] Lindenauer, P. K., Pekow, P., Gao, S., Crawford, A. S., Gutierrez, B., & Benjamin, E. M. (2006). Quality of care for patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. *Annals of Internal Medicine, 144*(12), 894–903. https://doi.org/10.7326/0003-4819-144-12-200606200-00006

[19] Agency for Healthcare Research and Quality. (2024). *HCUP National Readmissions Database (NRD), 2024*. Healthcare Cost and Utilization Project. https://hcup-us.ahrq.gov/nrdoverview.jsp

[20] Yalley, P. P. (2025). *Hospital readmission risk analytics — Code, dataset, and visualizations* [GitHub repository]. https://github.com/pkyalley/dataverse/tree/main/hospital-readmission-risk-analytics