Figure 1. Study workflow

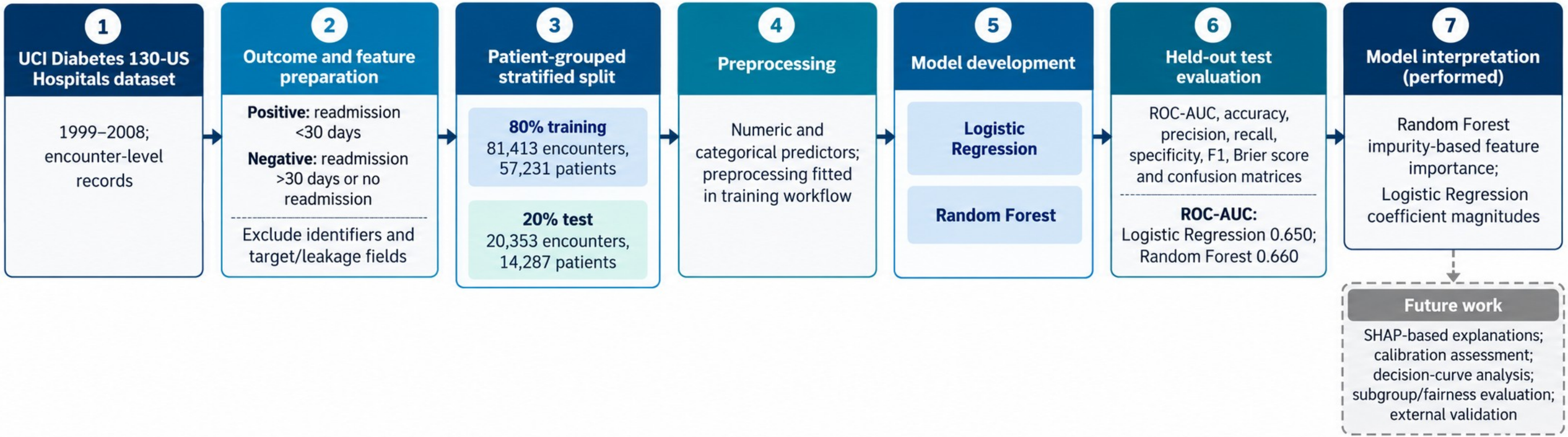


Figure 2. Calibration Analysis

Agreement Between Predicted and Observed 30-Day Readmission Risk

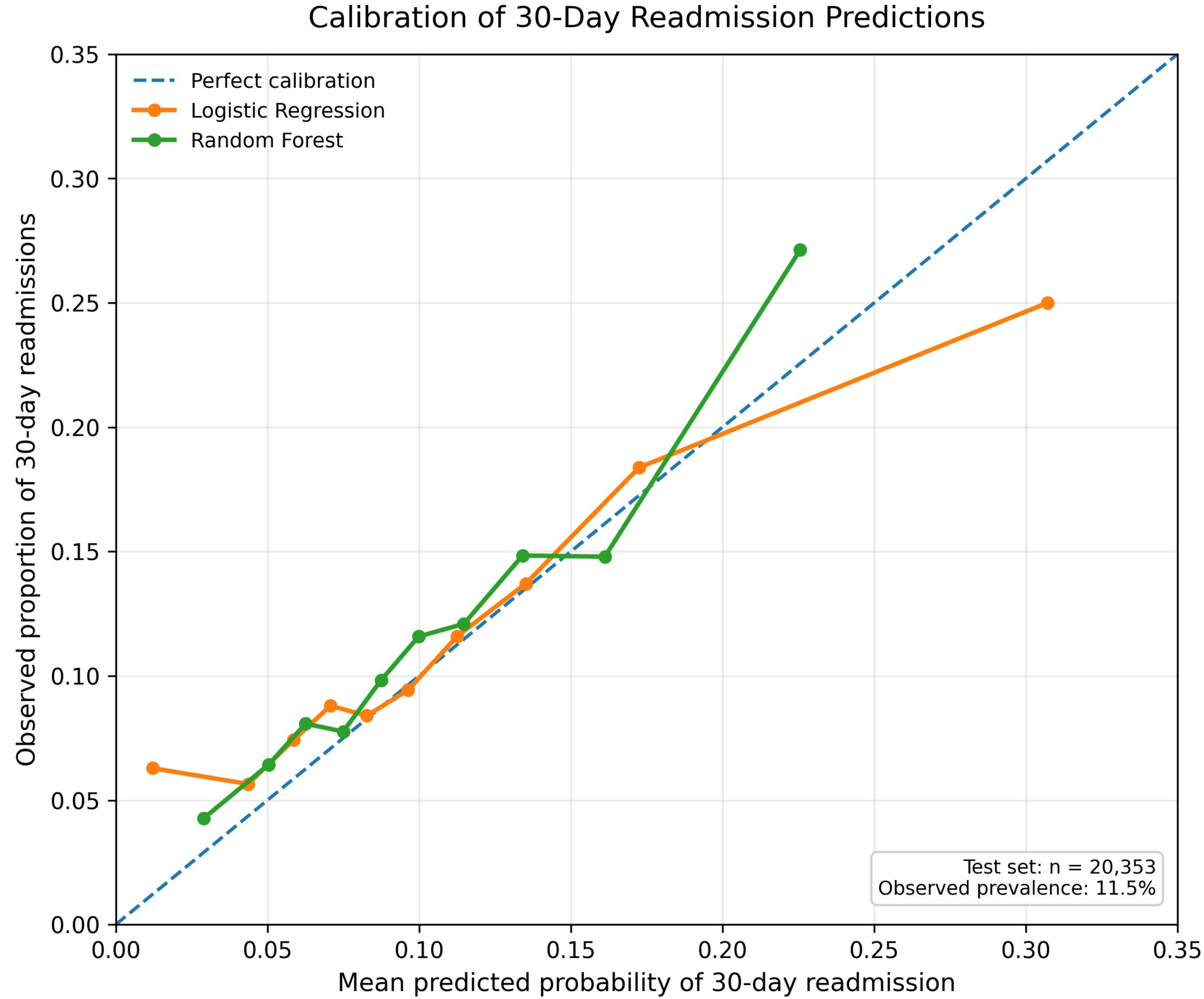


Figure 3. Proposed Subgroup Fairness Evaluation

Conceptual framework for future evaluation; not performed in the current study

1 Input data and trained models

Saved prediction models
- Logistic Regression
- Random Forest

Held-out test data
- Features
- True 30-day readmission outcome
- Predicted probabilities and/or class labels

2 Define candidate subgroup dimensions

- **Age group** (e.g., 0–10, 10–20, …)
- **Sex / gender** (e.g., male, female)
- **Ethnicity** (e.g., race/ethnic groups)
- **Socioeconomic status** (e.g., income, deprivation index)
- **Insurance status** (e.g., Medicare, Private, Self-pay)
- **Disease category** (e.g., primary diagnosis groups, comorbidity groups)

Candidate subgroup dimensions (subject to data availability)

3 Evaluate model performance within each subgroup

Calculate performance metrics separately for each subgroup

Candidate performance metrics (within each subgroup)
- Discrimination: AUROC
- Sensitivity (Recall)
- Specificity
- Precision (Positive Predictive Value)
- F1-score

Compare performance across subgroups to identify potential variation.

4 Compare errors and assess fairness

Assess potential disparities in model performance across subgroups

Candidate fairness measures (across subgroups)
- Demographic parity difference
- Equal opportunity difference
- Equalized odds difference

Evaluate differences in error rates (false positives and false negatives) across subgroups.

Investigate subgroups with notably different performance for further analysis.

5 Report and interpret for future work

Summarize findings and assess potential disparities

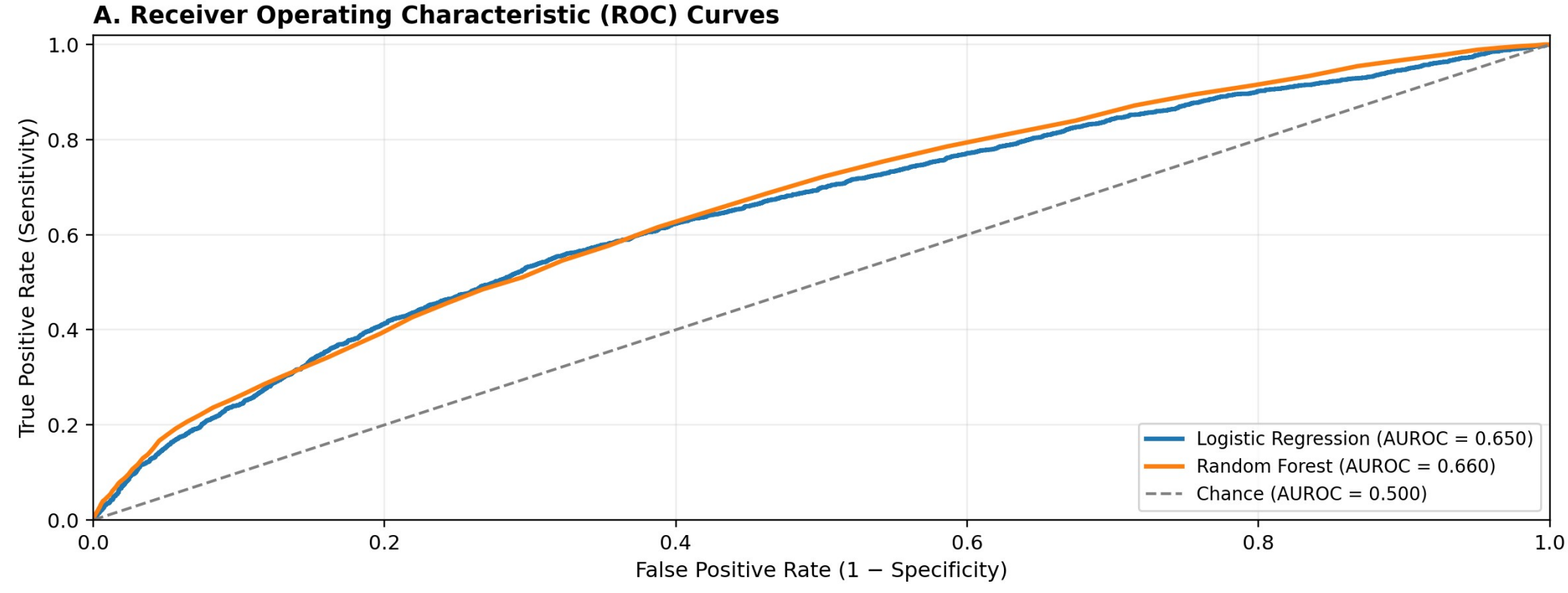
Key aspects to report (in future analysis)
- Subgroup sample sizes and event rates
- Performance metrics with confidence intervals
- Fairness metrics with confidence intervals
- Interpretation of any observed disparities
- Limitations (e.g., small subgroup sizes, missing data)
- Implications for clinical use and future validation

Status in current study

Not performed on the real dataset; proposed future work only.
Subgroup, fairness, and clinical utility analyses are planned for future research using the saved models and test predictions.

Figure 4. Comparative Model Performance for 30-Day Unplanned Hospital Readmission Prediction

ROC curves use predicted probabilities; confusion matrices use saved class predictions at the default threshold



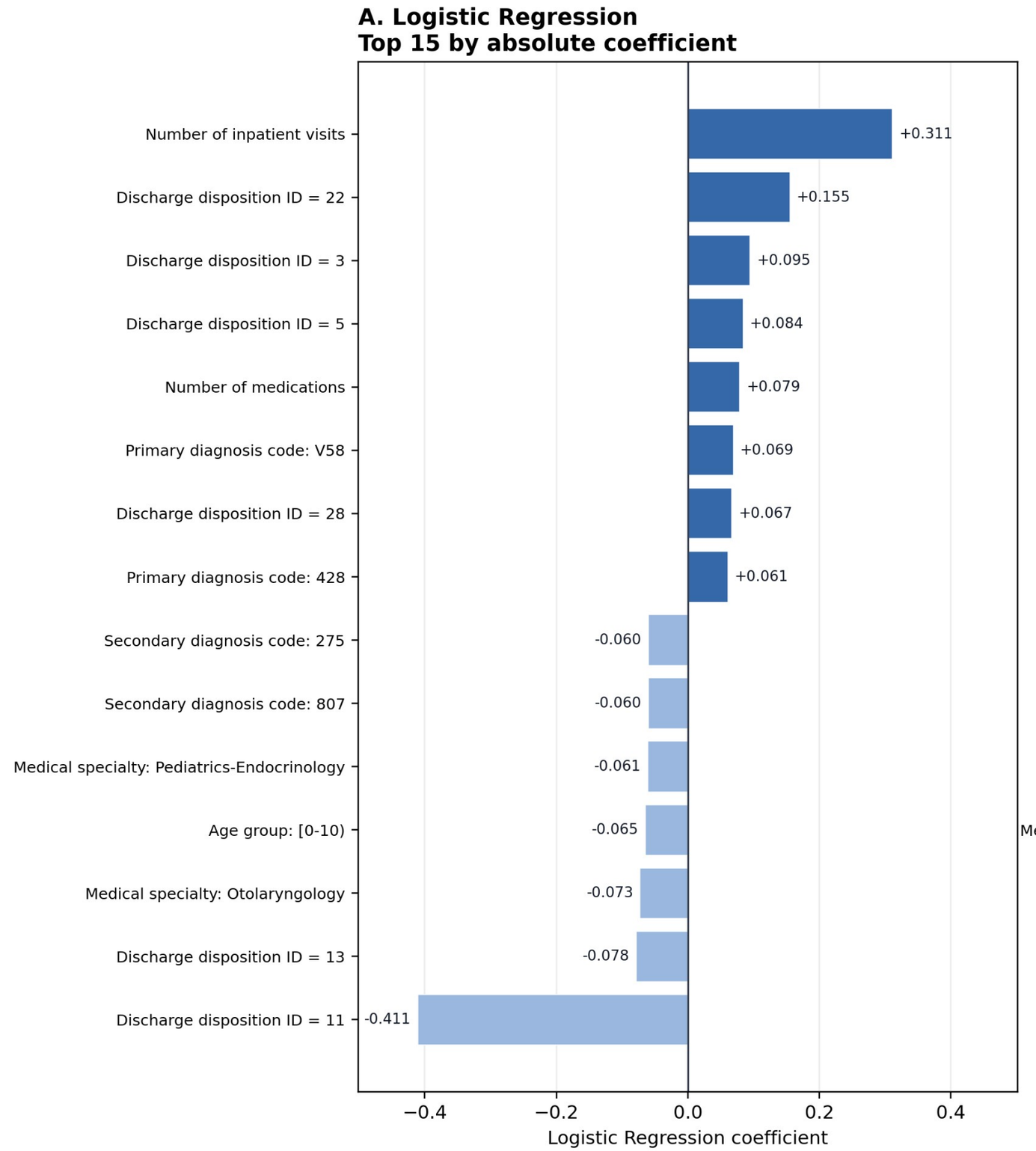
Test set: n = 20,353; readmissions = 2,334 (11.47%). ROC discrimination and threshold-based classification are distinct; the Random Forest classification result should be interpreted cautiously.

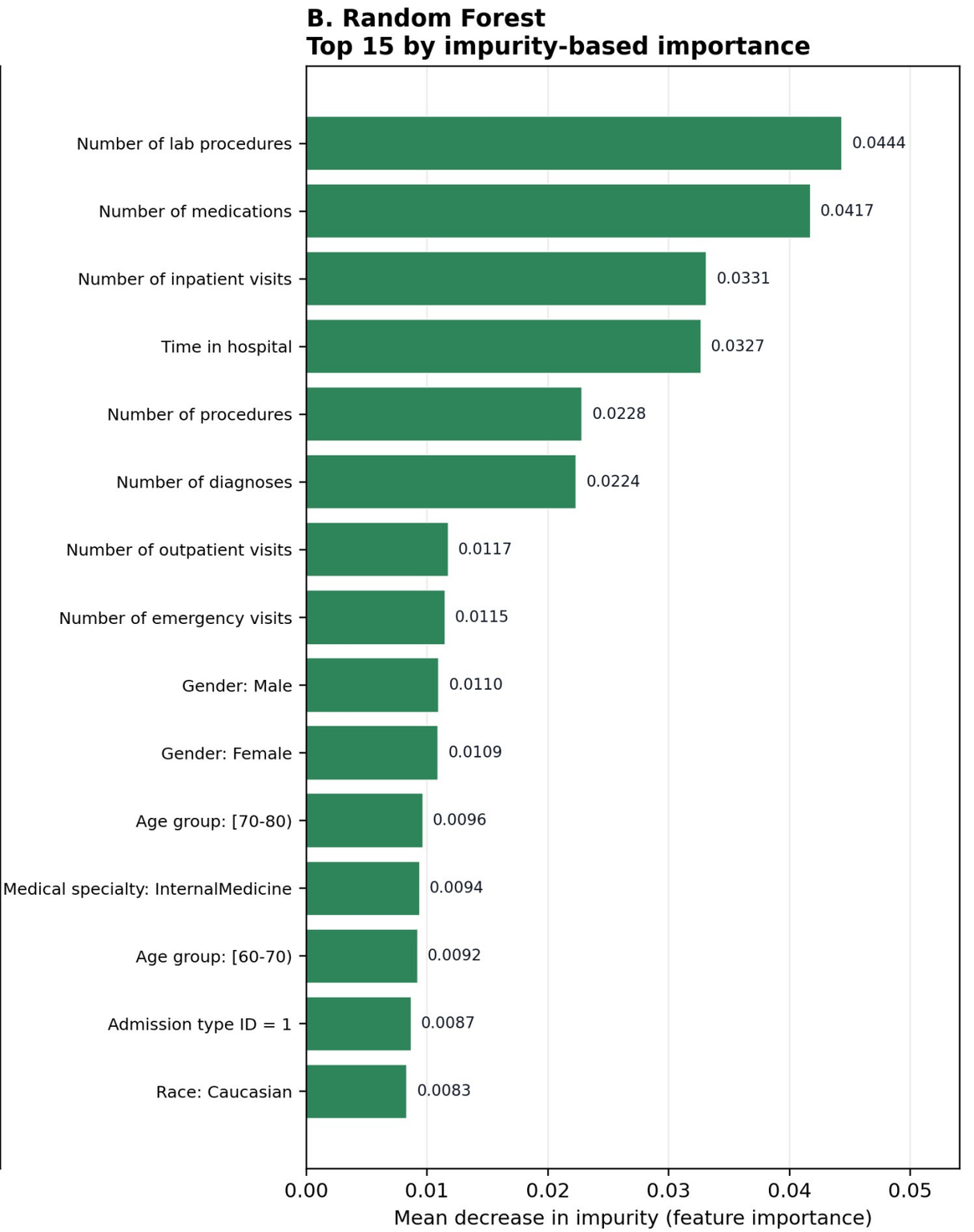
At the saved default threshold: TP = 45; FN = 2,289.

Default classification rule predicted every test encounter as "No readmission" (sensitivity = 0; TP = 0).

Figure 5. Predictor Importance in Logistic Regression and Random Forest Models

Verified coefficients and impurity-based feature importances from the saved fitted models





Feature labels preserve the encoded model-input categories where applicable. Coefficients and impurity-based importance are model-specific and should not be interpreted as causal effects.

Positive coefficients indicate higher model log-odds;
negative coefficients indicate lower model log-odds, holding other inputs constant.

Figure 6. Proposed Explainability Analysis for 30-Day Readmission Prediction

