Model Fairness Assessment

Governance Committee Report

Test Model - Chart Verification v1.0.0
2026-01-08

⚠ ADVISORY ASSESSMENT

This is an advisory assessment based on statistical analysis of the test dataset.
Final deployment decisions rest with the governance committee.

Executive Summary

Issues Near Threshold

7 metric(s) near threshold

0	7	0
WITHIN THRESHOLD	NEAR THRESHOLD	EXCEEDED THRESHOLD

Key Findings

The model's ability to distinguish between outcomes **requires review**.

Model calibration is **acceptable** but could be improved.

7 fairness metric(s) detected near threshold.

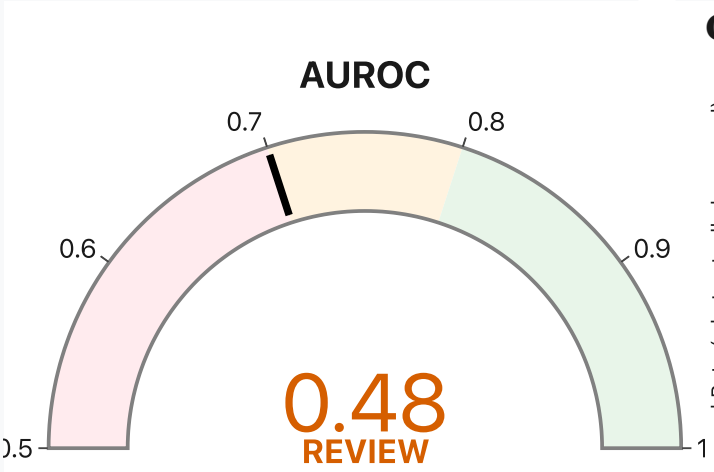
Sample size (N = 400) is limited; results should be interpreted with caution.

This assessment provides advisory guidance based on the CHAI RAIC framework. Final deployment decisions rest with the governance committee and clinical stakeholders.

Overall Model Performance

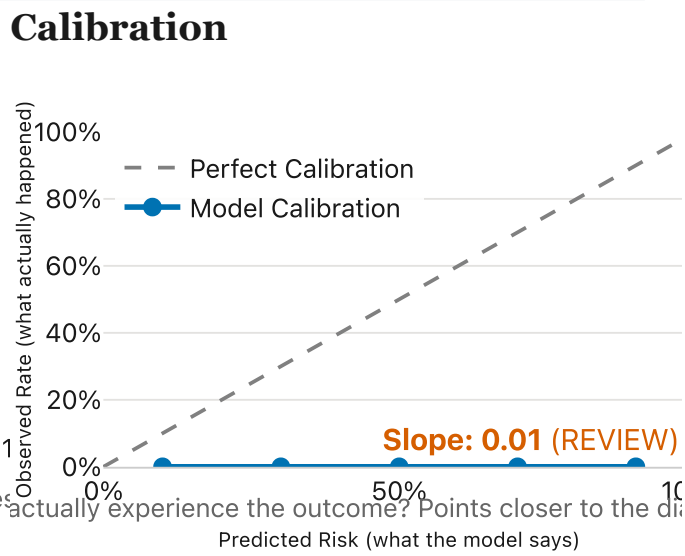
Four key metrics evaluating model discrimination and calibration:

AUROC



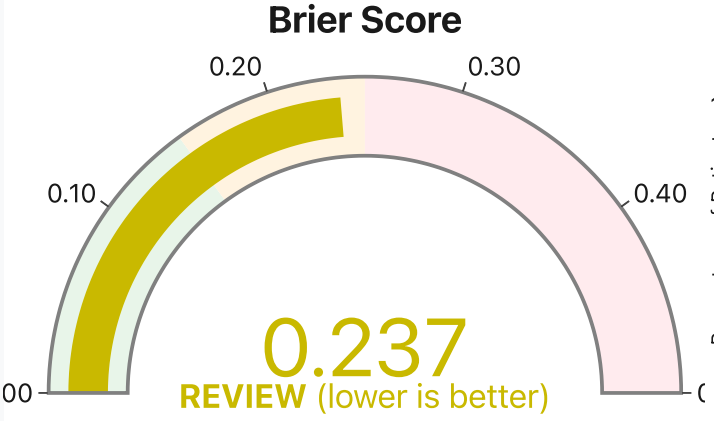
bility to rank patients correctly. Score of 0.5 = random gues

Calibration



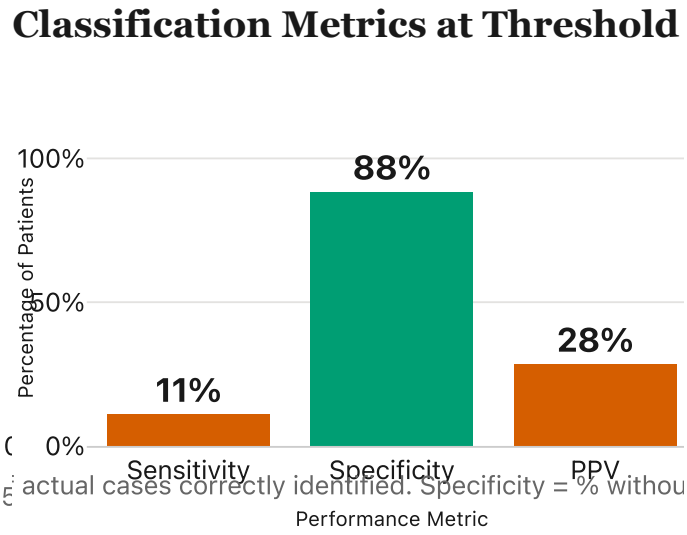
actually experience the outcome? Points closer to the di

Brier Score



ter. Think of it as the 'error' in risk predictions. Score <0.15

Classification



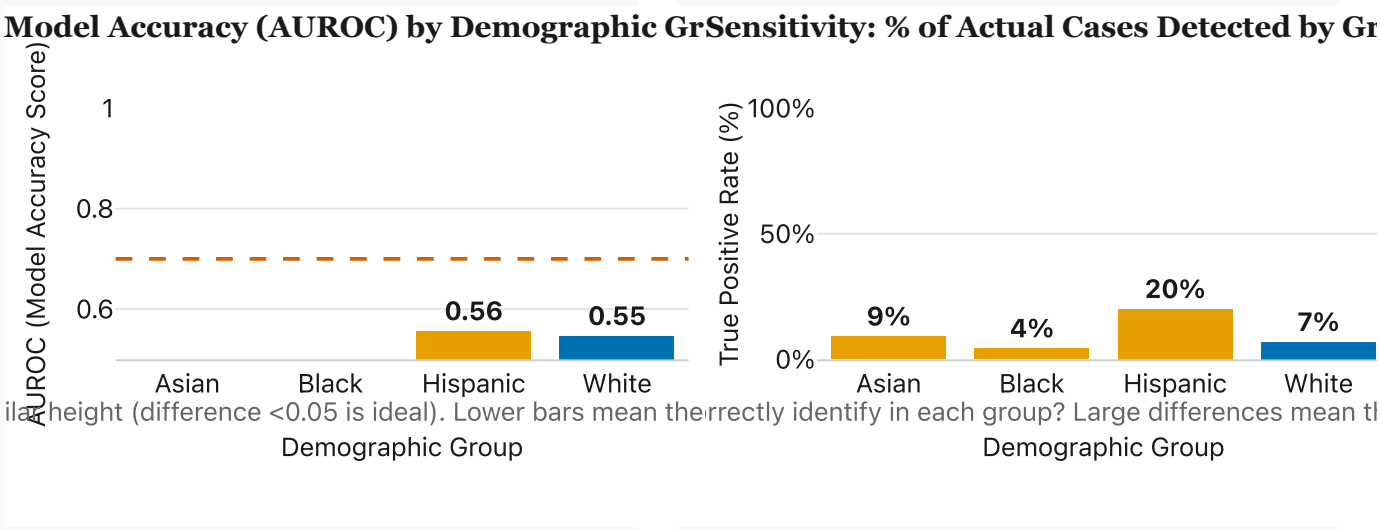
actual cases correctly identified. Specificity = % withou

Subgroup Performance

Performance comparison across demographic groups defined by the data scientist:

Race

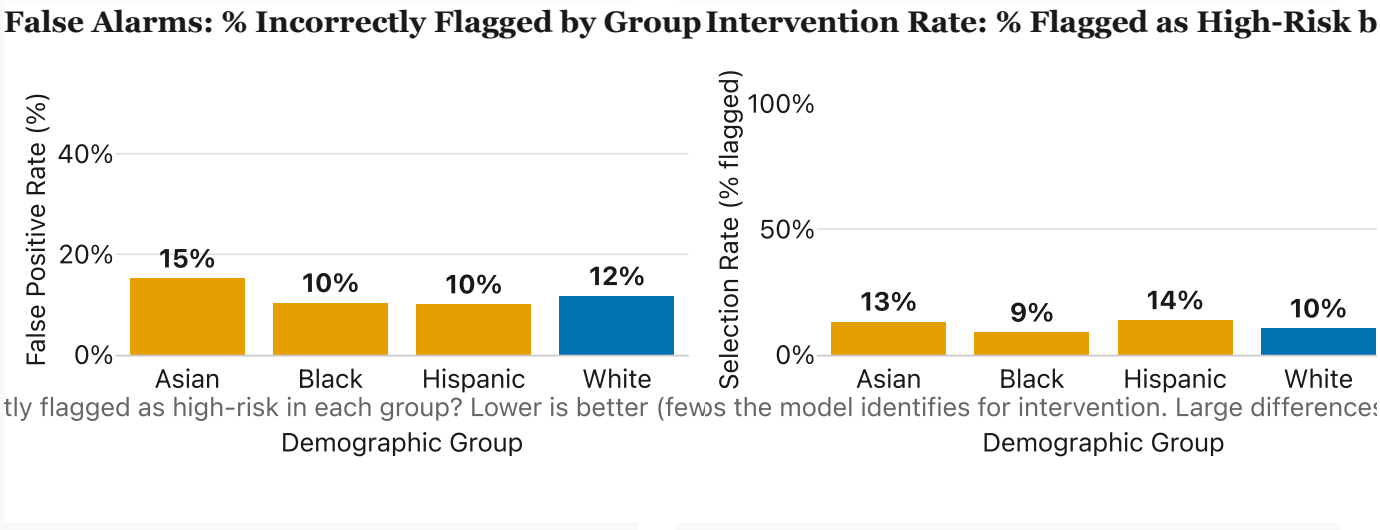
AUROC by Subgroup



Sensitivity by Subgroup

ilar height (difference <0.05 is ideal). Lower bars mean the rrectly identify in each group? Large differences mean tl

FPR by Subgroup



Selection Rate by Subgroup

tly flagged as high-risk in each group? Lower is better (few s the model identifies for intervention. Large differences

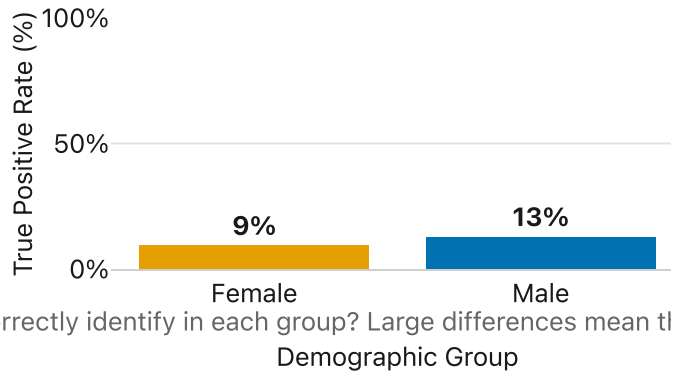
Sex

AUROC by Subgroup

Model Accuracy (AUROC) by Demographic Gr

Sensitivity by Subgroup

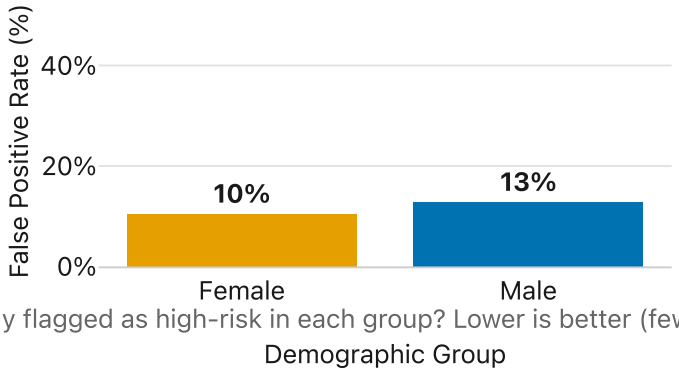
Sensitivity: % of Actual Cases Detected by Gr



ilar height (difference <0.05 is ideal). Lower bars mean the rrectly identify in each group? Large differences mean tl

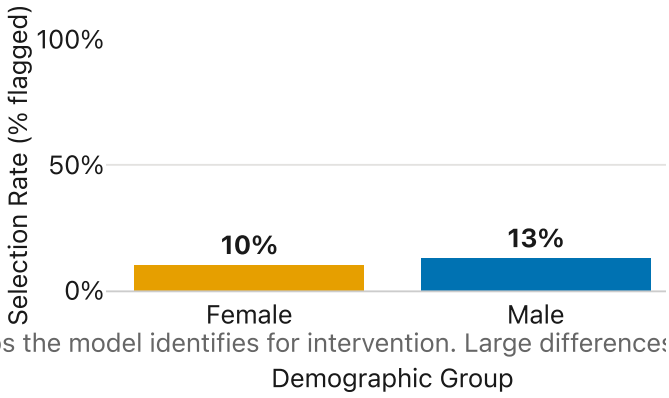
FPR by Subgroup

False Alarms: % Incorrectly Flagged by Group



Selection Rate by Subgroup

Intervention Rate: % Flagged as High-Risk b



tly flagged as high-risk in each group? Lower is better (few s the model identifies for intervention. Large differences

Governance Decision

Model: Test Model - Chart Verification v1.0.0

Intended Use: Not specified

Intended Population: Not specified

Decision

☐ APPROVE for deployment
☐ APPROVE with conditions
☐ DEFER pending additional review
☐ DO NOT APPROVE

Comments

Signatures

_____

Governance Committee Chair

_____

Date