Predicting Risk of Endocarditis Using a Clinical Tool (PREDICT): Scoring System to Guide Use of Echocardiography in the Management of *Staphylococcus aureus* Bacteremia

Supplementary Material

Description of Statistical Methods

Model Derivation

Descriptive statistics on study variables are presented as median (range or interquartile range), mean (standard deviation [SD]), or frequency count (percentage) as appropriate. The associations between candidate risk factors and diagnosis of endocarditis were measured using logistic regression and summarized with odds ratios (OR) and 95% confidence intervals (CI). Each factor was screened for an association with IE via univariate logistic regression, but only those with at an alpha level of 0.1 or less were carried forward in multivariable analyses. These screened variables were then entered in a multivariable logistic model, which was reduced to the most important risk factors using stepwise variable selection with backwards elimination (alpha level of 0.05). Two-way interactions between candidate risk factors were also tested for importance. To provide clinicians with a risk tool applicable on the first day ("Day 1") a patient is diagnosed with SAB, a scoring system was derived based on the modeling steps above using only factors known up until that point. Excluding a small minority of patients who were lost to short-term follow-up, The algorithm was then repeated to derive a second scoring system that incorporated the information regarding persistently positive blood cultures on day 3 of admission and this served as a risk stratification tool on "Day 5" (because it would typically take up to 2 days before the culture result is ascertained).

Model Validation

To internally validate the fit and performance of both final models, bootstrap resampling was used. For each of 400 bootstrap resamples (selected via random sampling with replacement from the original set of patients, each of an equal sample size), the modeling procedure was repeated with the same stepwise criteria so that variability and optimism bias in the automated variable selection could be assessed. In particular, the stepwise model selected in each resample was tested on the original sample, with the difference in performance between these two models averaged across all resamples to estimate the optimism bias due to overfitting. Bias-corrected measures of predictive accuracy were then obtained by subtracting the optimism bias estimate from the model performance index derived on the original sample. In addition, the frequency of "selected" variables was summarized as a percentage across all bootstrap resamples, and only predictors consistently retained in the modeling (i.e., in at least 70% of resamples) were included in the risk score.

To assess calibration, which is the model's ability to predict accurately the absolute level of risk that is subsequently observed, the accuracy of the model predicted values relative to observed proportions of subjects with IE was plotted according to decile risk groups, with a non-linear calibration curve estimated using a nonparametric loess smooth.

Risk Stratification

From the final selected models, regression coefficients were used to derive the two separate risk scores for predicting IE. In particular, points were assigned for the presence of each risk factor in the model and weighted approximately by the respective regression coefficients. For optimal scoring, regression coefficients were re-scaled by first dividing by the minimum absolute value among all coefficients and then multiplying each re-scaled coefficient by a constant (such as 2 or 3, choosing the one producing the most optimal weighting scheme), and finally rounding the re-

scaled values to the nearest integer. Using these point values, a subject's risk score was simply computed as the aggregate number of points from their risk factor profile, with higher scores corresponding to increased risk of IE.

Application of Statistical Methods

Model derivation

Two scoring systems were derived for the clinical purpose of risk stratification at two time points in reference to initial assessment; Day 1 score (SAB diagnosis day) and Day 5 score (when results of day 3 blood cultures are known). Among the baseline factors (available at the time of SAB diagnosis) identified from univariable screening and carried forward into multivariable modeling, the following were selected in the original model fit and corresponded to increased risk of IE: onset of SAB (non-nosocomial), presence of CIED, prior intravenous drug abuse (IVDA), a recent *S. aureus* infection, and absence of a skin or surgical site as a source of infection (Table 1). However, based on internal validation of the model fit via bootstrap resampling, only onset of SAB and presence of CIED were consistently selected across the bootstrap resamples (≥90% for both; <60% for all the others) and deemed robust predictors of IE. From the fit of this final reduced model on the original sample, the *c*-statistic was 0.723 indicating fairly good discrimination (**Table 2 – in the manuscript**). However, the estimated optimism bias due to variability in the automated process that selected this model was 0.030, and thus, a bias-corrected estimate of 0.693 would provide a reasonable approximation of the model c-statistic in a future independent validation. Using a smoothing estimator relating observed IE outcomes to predicted values, the calibration curve of the Day 1 model appeared fairly proximal to the 45 degree line of identity representing ideal calibration (**Figure 1A**).

To derive a prediction model that could be applied on Day 5, the effect of prolonged bacteremia beyond 3 days, along with those from the previous set of baseline variables, was assessed for a total of 662 subjects with SAB (7 subjects who died within 5 days of an SAB diagnosis and 9 subjects lacking a negative culture whose duration of bacteremia was therefore unknown were excluded). Baseline risk factors retained in the initial stepwise selection for the Day 5 model were similar those retained in the Day 1 modeling, except for prior IDU (not selected) and prosthetic valve (selected), but as seen previously only the onset of SAB (community vs. healthcare-associated vs. nosocomial) and presence of CIED were consistently retained across the bootstrap resamples. Prolonged bacteremia beyond 3 days was found to be predictive of IE in the original sample and was internally validated as a predictor after being selected in 100% of the bootstrap models. The final Day 5 model therefore included onset of SAB, presence of CIED and sustained bacteremia, and the new information yielded by the addition of the latter predictor led to a significant gain in model discrimination. The *c*-statistic for this model fit on the original data was 0.792, though this was overoptimistic by 0.031 (bias-corrected c-statistic=0.761). From visual inspection of the calibration curve in **Figure 1B**, the data suggest that the Day 5 model is adequately calibrated with reasonable accuracy over its entire range of predictions.

Risk stratification

On the basis of final Day 1 and Day 5 models, regression coefficients were used to derive two respective risk scores for predicting IE in patients hospitalized for SAB. Points were assigned for the presence of each risk factor, weighted in magnitude by the corresponding regression coefficients, and summed together to define an individual's risk score (Table 3). For example, a

patient who had an ICD and acquired SAB in a healthcare setting would have a Day 1 risk score of 3 out of 5 possible points; if this patient's bacteremia persisted for at least 72 hours, then their Day 5 risk score would be 5 of 7 possible points. From a logistic model with risk score included as a continuous predictor, an ROC curve was derived and superimposed over the ROC curve from the corresponding multivariable model for comparison. In **Figure 2A**, the ROC curves from the Day 1 predictor and risk score models are similar (c-statistic, or area under the ROC curve: 0.723 vs. 0.720, respectively) showing that the risk score adequately summarizes the multivariable model. Likewise, the Day 5 multivariable and risk score models have very comparable ROC curves as illustrated in **Figure 2B** (c-statistic: 0.792 vs. 0.794, respectively). Bubble plots in Figures 3A and 3B display the range of Day 1 and Day 5 scores, respectively, in relation to absolute risk of IE, with the size of the bubble proportional to the frequency observed in the analysis set.

The test performance characteristics of the two risk scores are summarized for all possible dichotomous cut-points in **Table 3**, under the assumption that echocardiography performed at the respective time points would accurately reflect the final diagnosis of IE.

Figure 1A. Calibration Plot of the Day 1 (Baseline) Risk Model

Figure 1B. Calibration Plot of the Day 5 Risk Model

Figure 2A. Receiver operating characteristic curve plot of the Day1 predictor and risk score models

Figure 2B. Receiver operating characteristic curve of the Day5 predictor and risk score models

Figure 3A. Bubble plot of infective endocarditis risk across the range of Day 1 risk scores

Figure 3B. Bubble plot of infective endocarditis risk across the range of Day 5 risk scores

Figure 1A: Calibration Plot of the Day 1 (Baseline) Risk Model

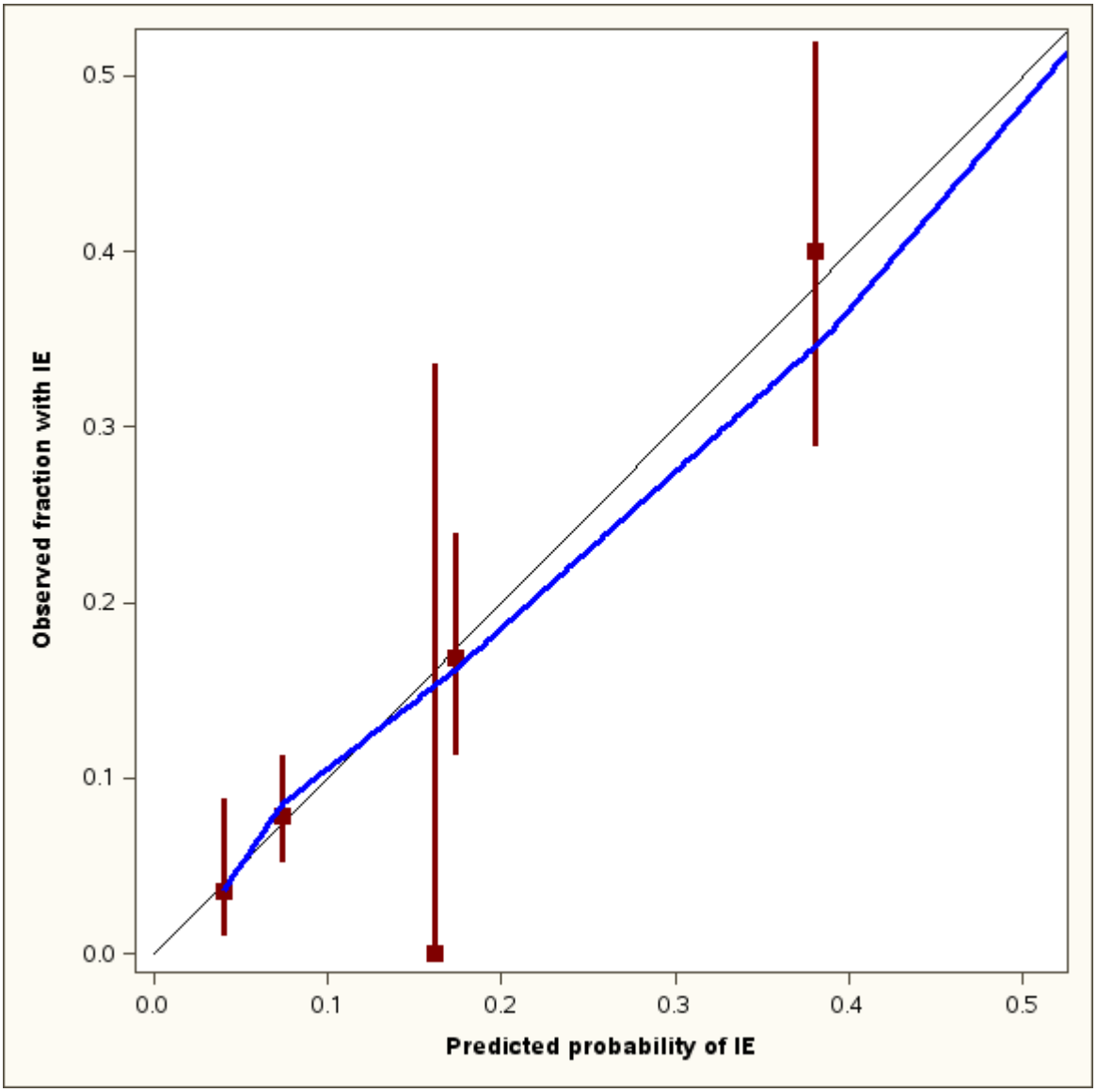

Figure 1A: Calibration Plot of the Day 1 (Baseline) Risk Model

Using a smoothing estimator relating observed IE outcomes to predicted values, the calibration curve of the Day 1 model appeared fairly proximal to the 45 degree line of identity representing ideal calibration

Figure 1B: Calibration Plot of the Day 5 Risk Model

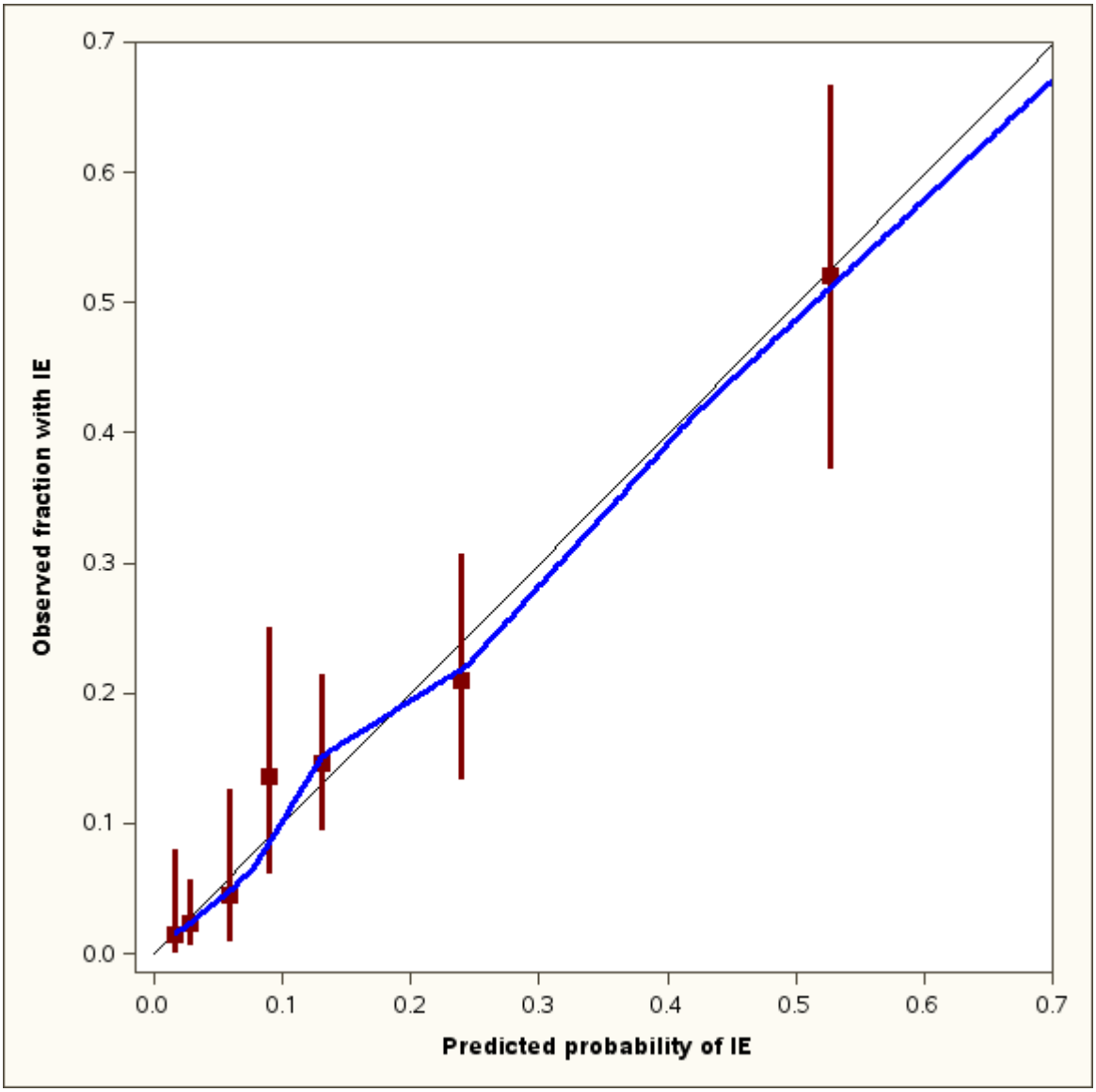

Figure 1B: Calibration Plot of the Day 5 Risk Model
From visual inspection of the calibration curve in, the data suggest that the Day 5 model is adequately calibrated with reasonable accuracy over its entire range of predictions.

Figure 2A: Receiver operating characteristic curve plot of the Day 1 predictor and risk score models

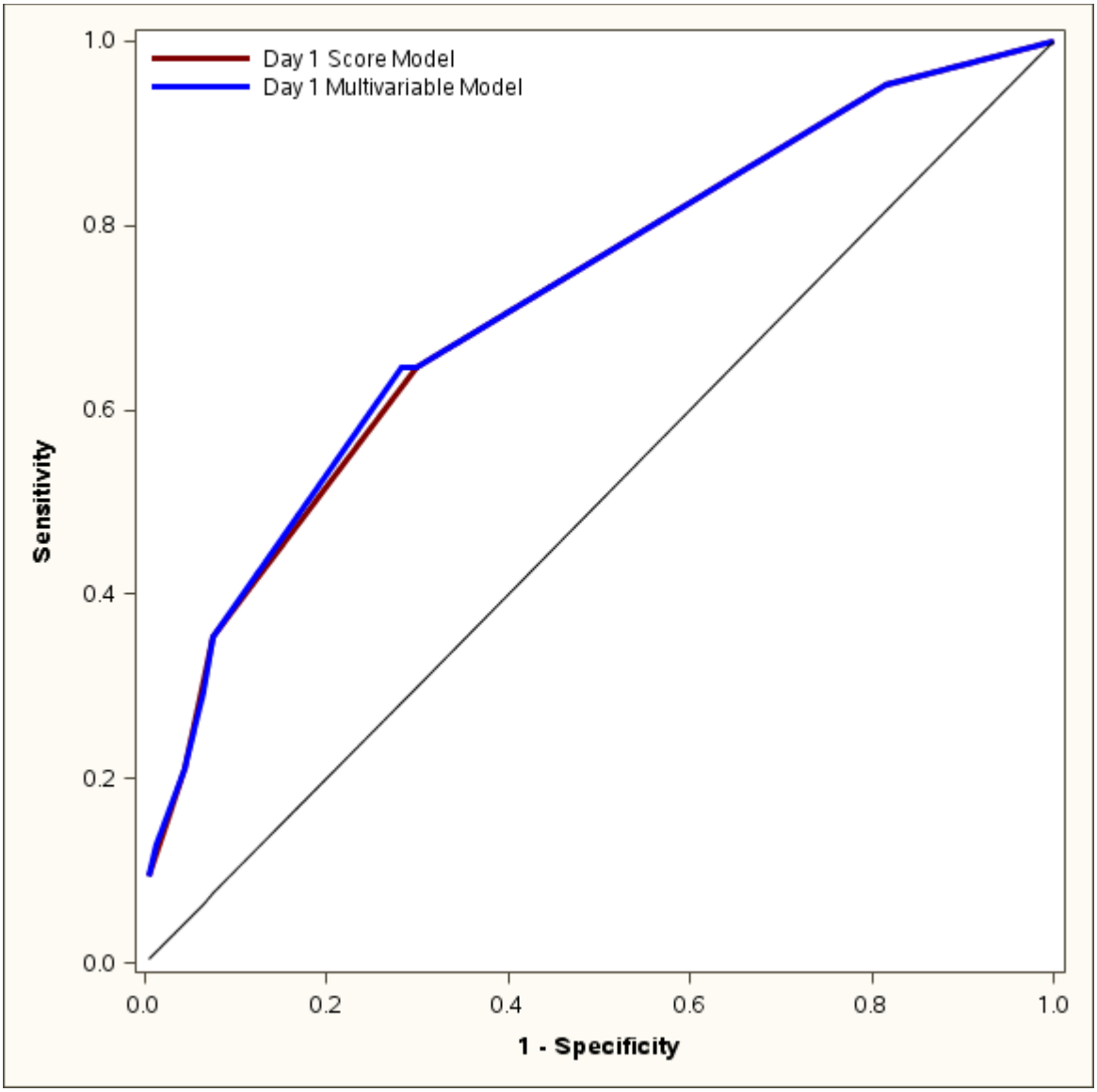

Figure 2A: Receiver operating characteristic curve plot of the Day 1 predictor and risk score models

The ROC curves from the Day 1 predictor and risk score models are similar (c-statistic, or area under the ROC curve: 0.723 vs. 0.720, respectively) showing that the risk score adequately summarizes the multivariable model.

Figure 2B: Receiver operating characteristic curve of the Day 5 predictor and risk score models

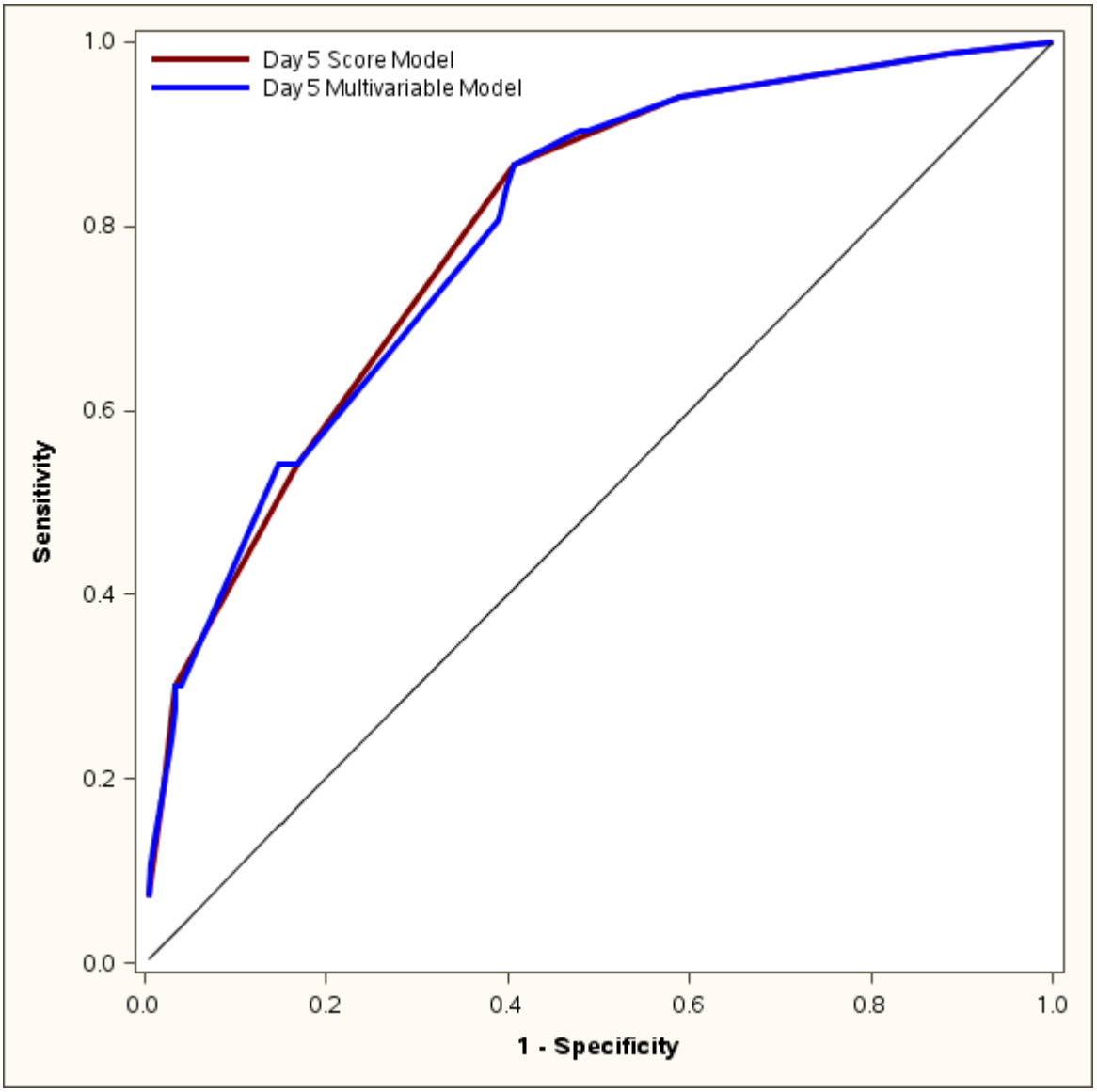

Figure 2B: Receiver operating characteristic curve of the Day 5 predictor and risk score models
the Day 5 multivariable and risk score models have very comparable ROC curves (c-statistic: 0.792 vs. 0.794, respectively).

Figure 3A: Bubble plot of infective endocarditis risk across the range of Day 1 risk scores

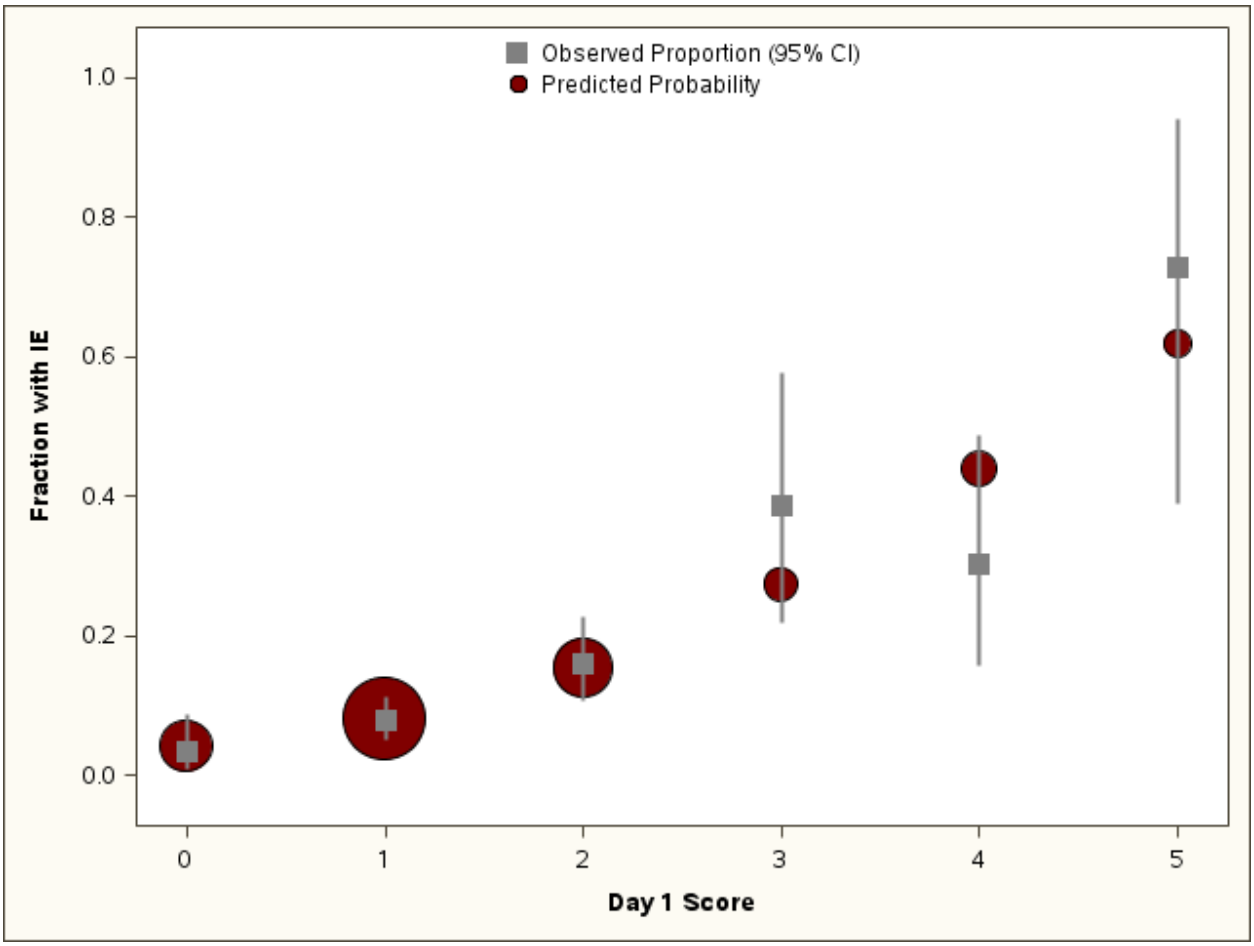

Figure 3A: Bubble plot of infective endocarditis risk across the range of Day 1 risk scores
Bubble plots in 3A display the range of Day 1 scores in relation to absolute risk of IE, with the size of the bubble proportional to the frequency observed in the analysis set.

Figure 3B: Bubble plot of infective endocarditis risk across the range of Day 5 risk scores

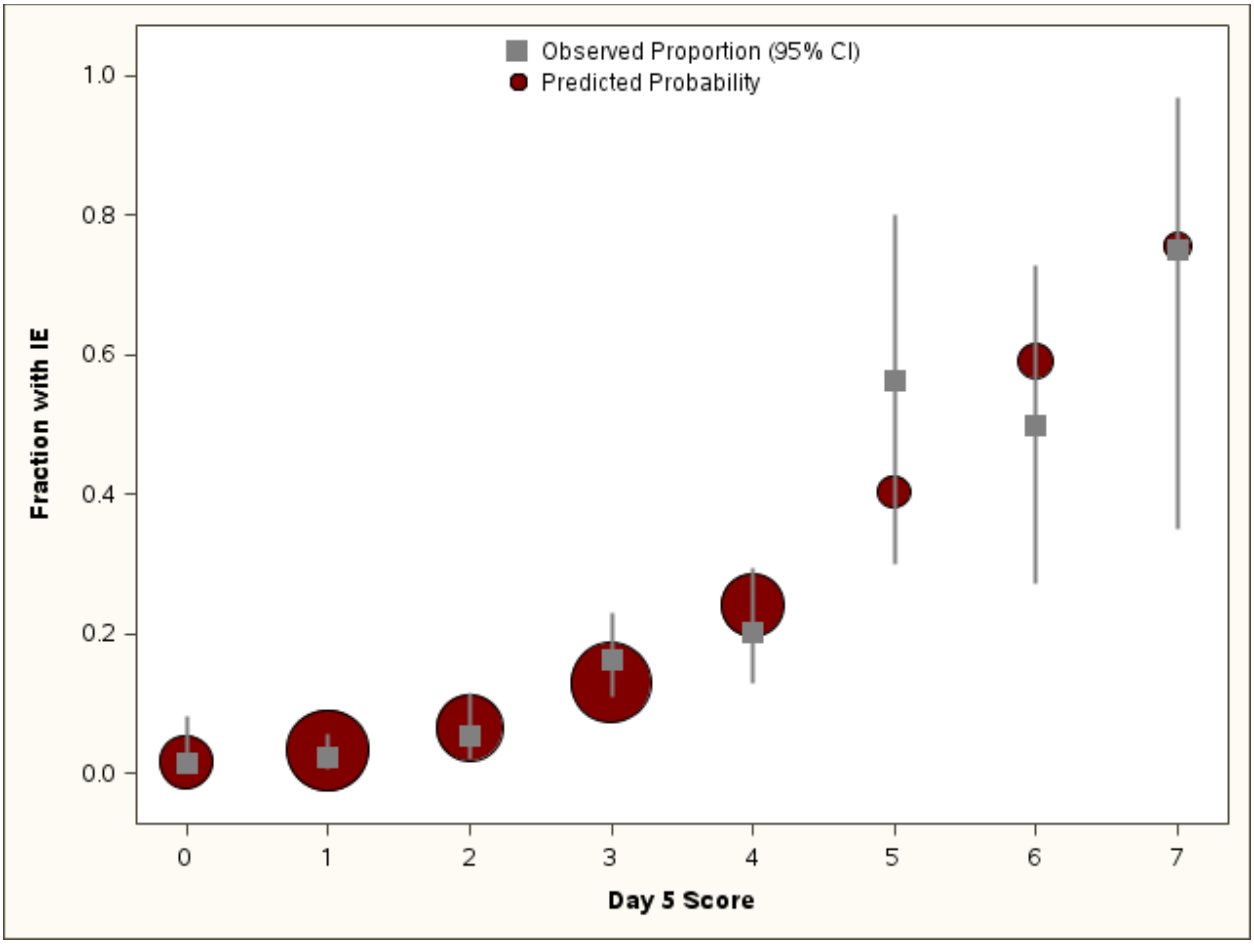

Figure 3B: Bubble plot of infective endocarditis risk across the range of Day 5 risk scores

Bubble plots in 3B display the range of Day 5 scores in relation to absolute risk of IE, with the size of the bubble proportional to the frequency observed in the analysis set.