Supplementary Online Content

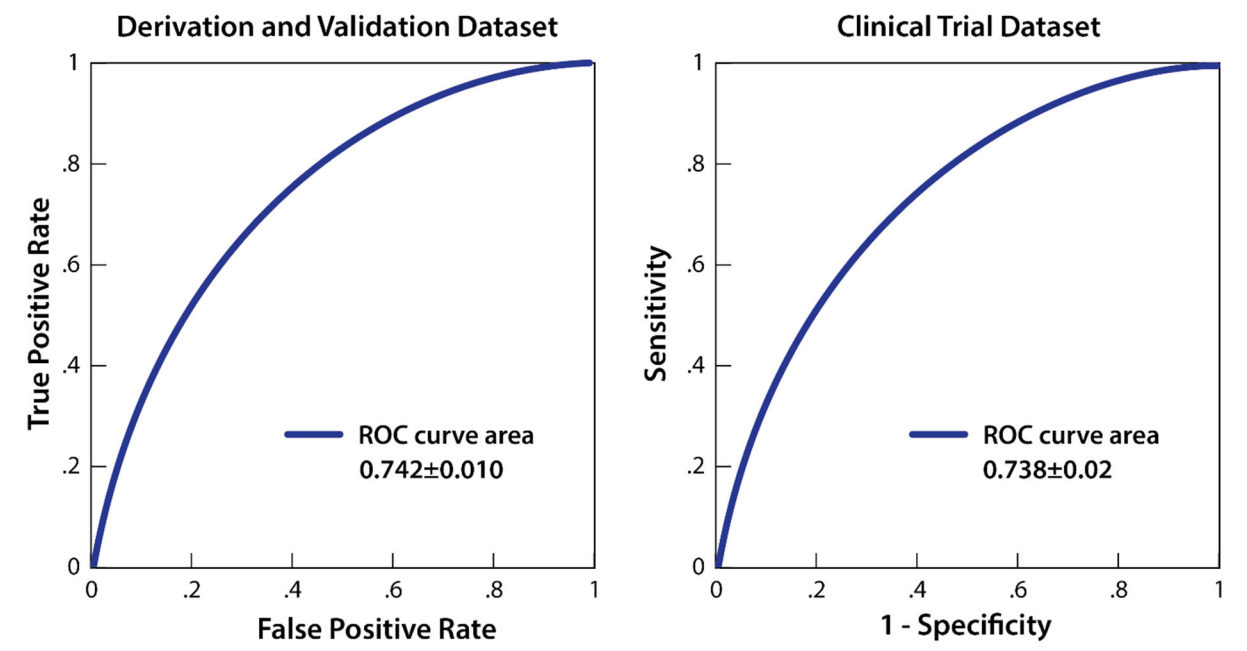
Ahmad T, Desai NR, Yamamoto Y, et al. Alerting clinicians to 1-year mortality risk in patients hospitalized with heart failure: the REVEAL-HF randomized clinical trial. *JAMA Cardiol*. Published online August 10, 2022. doi:10.1001/jamacardio.2022.2496

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Creation and Performance of Mortality Risk Score

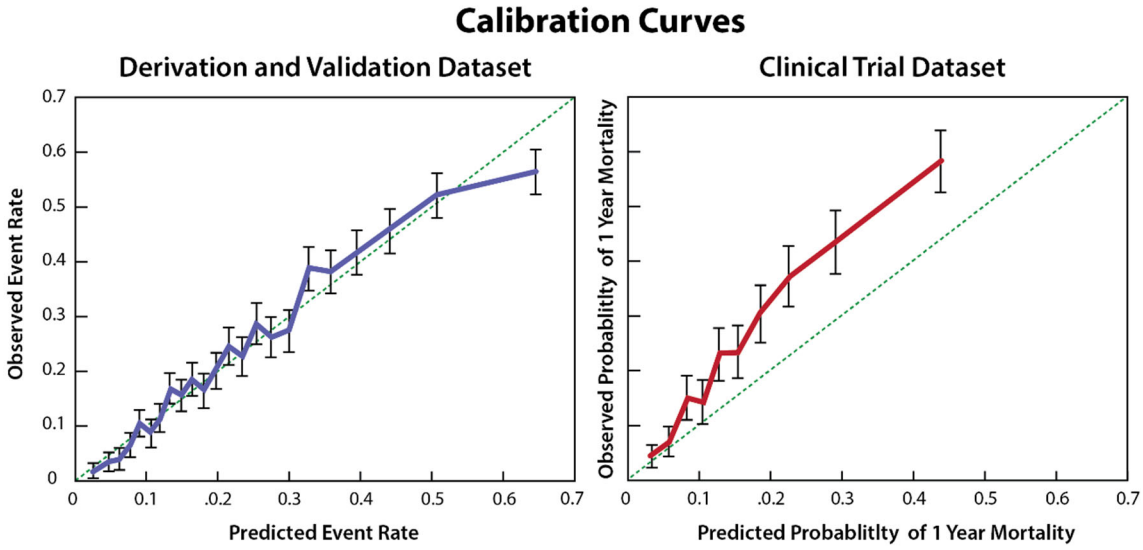
Our derivation and validation datasets were based on patients from our health system to maximize generalizability. We analyzed data on patients admitted to the Yale New Haven Hospital between January 1, 2014, and April 14, 2018 that met the inclusion and exclusion criteria and we were not under comfort measures only or died within 36 hours of admission. This yielded 7376 unique patients on whom we performed feature selection using population-based incremental learning. First, we randomly split the data into a training set and a validation set with balanced outcomes. Second, we preprocessed the data by imputing training set medians, center and scale numeric data by training set means and SDs and clip all values at ±10 for numeric stability. Finally, we fit a logistic regression model on the training set and evaluate the area under the curve on the validation set to gauge the performance of the feature set. We set the number of features considered to 15 for ease of implementation. We choose the 15 features with the highest probabilities after 500 iterations using a population size of 1,000. The final model was a logistic regression fit on the full dataset after performing similar preprocessing as before, except using statistics from the full dataset. The model's coefficients were finally transformed to work with the original feature centers and scales. The variables included in the final risk score were as follows: age, weight, systolic blood pressure, red cell distribution width, blood urea nitrogen, monocyte count, lymphocyte percentage, blood urea nitrogen-to-creatinine ratio, troponin, NT-proBNP, mean corpuscular volume, intensive care unit admission, and measurement of arterial pH. A variable for patients who were on comfort measures was in the initial risk score "codecomfort" and was removed, leaving 14 variables in total. The model achieved an area

under the curve of 0.742 ± 0.01 on the full dataset. The model achieved an area under the curve of 0.738 ± 0.02 on the clinical trial dataset which was similar to its discrimination performance in the derivation and validation dataset.



To measure the calibration of our risk score, we partitioned the space of predicted event rates/probabilities into 23 consecutive bins, constructed such that each bin contains just over 500 of the model's predicted probabilities. Using the mean of observed events within each bin, we estimate the true within-bin event rates and construct 95% confident intervals using the normal approximation to the binomial distribution. The plot below compares these estimates (y-axis) against the model's within-bin mean predicted probabilities, demonstrating excellent calibration. During the study period, we observed persistently higher rates of mortality than predicted by the algorithm during the clinical trial period, resulting in worse calibration compared with the derivation dataset. Excluding COVID+ patients did not meaningfully change either the discrimination of calibration of our model. However, given the higher rates of adverse outcomes during this time period, we hypothesized that our findings might be related to changes in cohort

characteristics during the COVID pandemic and the associated excess burden of heart failure deaths.



eTable 1. Composite Outcome by Hospital

Hospital	Alert Group	Usual Care	OR
YNNH	300/769 (39.0%)	314/742 (42.3%)	0.87 [0.71, 1.07]
SRC	160/389 (41.1%)	133/360 (36.9%)	1.19 [0.89, 1.60]
BH	106/279 (38.0%)	112/283 (39.6%)	0.94 [0.67, 1.31]
GH	53/153 (34.6%)	44/149 (29.5%)	1.26 [0.78, 2.05]

eTable 2. Treatment Decisions According to Alert and Risk Levels

	Alert	No Alert	Risk difference (CI)
ACE-I/ARB/ARNI	766/1590 (48.18%)	738/1534 (48.11%)	0.1 (-3.4 to 3.6)
Very Low Risk	321/425 (75.53%)	283/360 (78.61%)	
Low Risk	349/752 (46.41%)	348/770 (45.19%)	
Medium Risk	77/307 (25.08%)	88/319 (27.59%)	
High Risk	17/96 (17.71%)	19/77 (24.68%)	
Very High Risk	2/10 (20%)	0/8 (0%)	
Beta Blocker	1307/1590 (82.2%)	1264/1534 (82.4%)	-0.2 (-2.9 to 2.4)
Very Low Risk	380/425 (89.41%)	320/360 (88.89%)	
Low Risk	616/752 (81.91%)	634/770 (82.34%)	
Medium Risk	224/307 (72.96%)	250/319 (78.37%)	
High Risk	78/96 (81.25%)	56/77 (72.73%)	
Very High Risk	9/10 (90%)	4/8 (50%)	
MRA	376/1590 (23.65%)	385/1534 (25.1%)	-1.4 (-4.4 to 1.6)
Very Low Risk	140/425 (32.94%)	126/360 (35%)	
Low Risk	167/752 (22.21%)	172/770 (22.34%)	
Medium Risk	49/307 (15.96%)	76/319 (23.82%)	
High Risk	19/96 (19.79%)	10/77 (12.99%)	
Very High Risk	1/10 (10%)	1/8 (12.5%)	
SGLT2i	138/1590 (8.68%)	134/1534 (8.74%)	-0.7 (-2.6 to 1.2)
Very Low Risk	65/425 (15.29%)	59/360 (16.39%)	
Low Risk	54/752 (7.18%)	53/770 (6.88%)	
Medium Risk	14/307 (4.56%)	18/319 (5.64%)	
High Risk	4/96 (4.17%)	3/77 (3.9%)	
Very High Risk	1/10 (10%)	1/8 (12.5%)	
ICD Implantation	19/1590 (1.19%)	21/1534 (1.37%)	-0.2 (-0.9 to 0.6)
Very Low Risk	2/425 (0.47%)	7/360 (1.94%)	
Low Risk	11/752 (1.46%)	11/770 (1.43%)	
Medium Risk	4/307 (1.3%)	2/319 (0.63%)	
High Risk	2/96 (2.08%)	1/77 (1.3%)	
Very High Risk	0/10 (0%)	0/8 (0%)	
Palliative Care Referral	163/1590 (10.25%)	164/1534 (10.69%)	0.02 (-2.1 to 2.1)

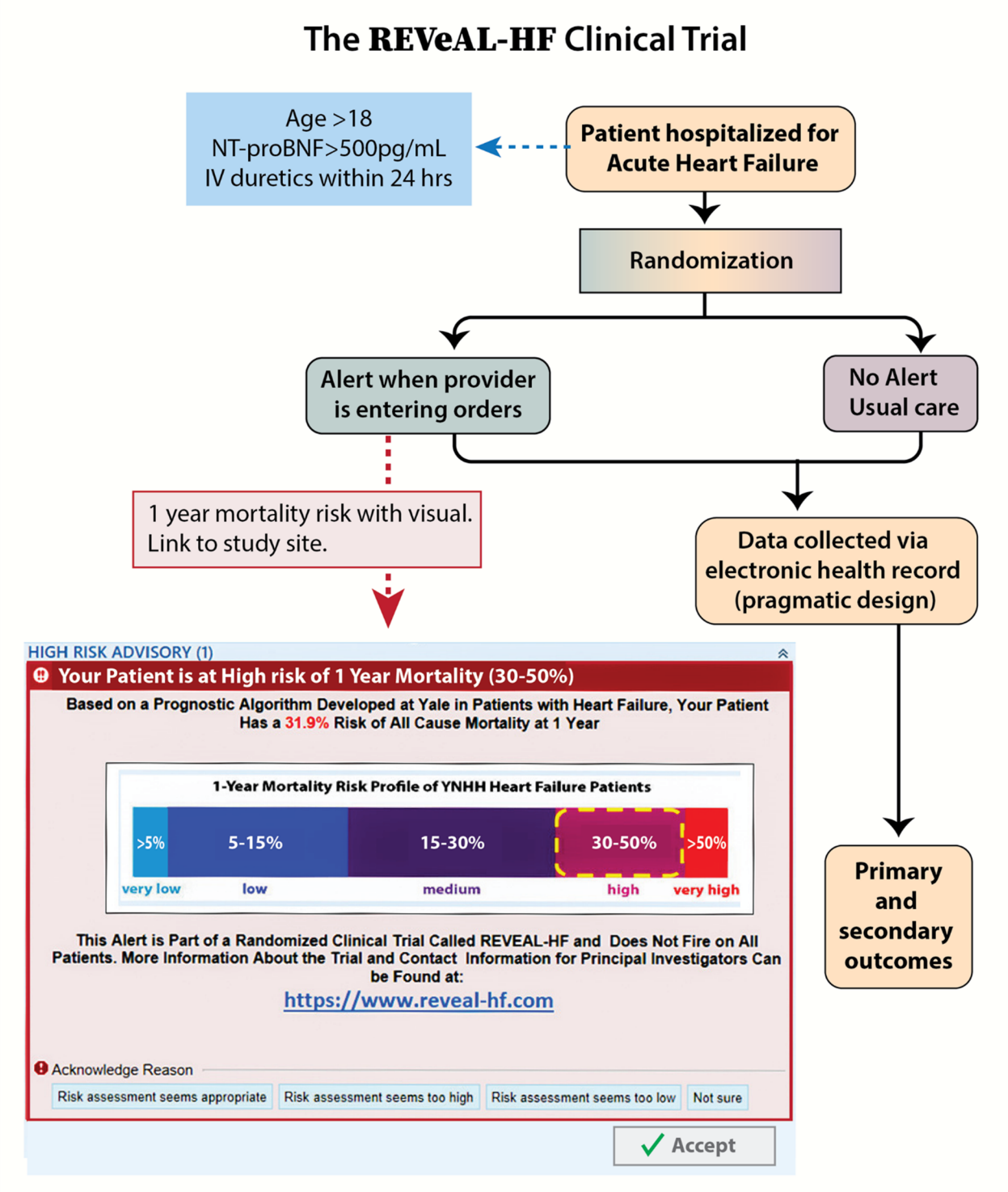
Very Low Risk	14/425 (3.29%)	8/360 (2.22%)	
Low Risk	71/752 (9.44%)	72/770 (9.35%)	
Medium Risk	59/307 (19.22%)	62/319 (19.44%)	
High Risk	15/96 (15.63%)	18/77 (23.38%)	
Very High Risk	4/10 (40%)	4/8 (50%)	

eTable 3. Clinical Opinion of Estimated Risk and Observed Mortality

Risk Level	Clinician's Opinion	Result
Very Low Risk	Risk assessment seems too low	8.9% [5.6, 13.3]
Very Low Risk	Risk assessment seems appropriate	8.3% [6.8, 10.0]
Very Low Risk	Risk assessment seems too high	13.0% [5.4, 24.9]
Low Risk	Risk assessment seems too low	31.6% [27.0, 36.5]
Low Risk	Risk assessment seems appropriate	26.7% [24.9, 28.5]
Low Risk	Risk assessment seems too high	25.0% [17.8, 33.4]
Medium Risk	Risk assessment seems too low	46.8% [40.4, 53.4]
Medium Risk	Risk assessment seems appropriate	48.2% [45.4, 51.0]
Medium Risk	Risk assessment seems too high	37.3% [27.0, 48.7]
High Risk	Risk assessment seems too low	81.4% [70.3, 89.7]
High Risk	Risk assessment seems appropriate	68.7% [64.1, 73.0]
High Risk	Risk assessment seems too high	64.5% [45.4, 80.1]
Very High Risk	Risk assessment seems too low	100% [66.4, 100]
Very High Risk	Risk assessment seems appropriate	88.5% [77.8, 95.3]
Very High Risk	Risk assessment seems too high	66.7% [9.4, 99.2]

eFigure 1. The REVEAL-HF Clinical Trial

The study included all adults ≥18 years who had an NT-proBNP levels of >500 pg/mL and received intravenous loop diuretics within 24 hours of admission. In the intervention arm, an alert displaying the predicted 1-year mortality rate, as well as other relevant information, was displayed to clinicians when they opened the "order entry" portion of the medical record



eFigure 2. Interaction Between Predicted Risk Categories, Observed Rates of Mortality, and Clinician Assessment of Risk
As predicted risk increases (X-axis), observed mortality increases. Within each predicted risk category, we assessed the impact of physician input to evaluate whether physician intuition accurately sub-stratified individuals within predicted risk categories.