Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

One-Hour Diaphragmatic Thickening Fraction and the HACOR Score, Alone and Combined (HACOR-D), for Predicting Non-invasive Ventilation Failure in the Emergency Department

Mohammad KHALID
Department of Emergency Medicine, Kasturba Medical College, Manipal Academy of Higher Education
Mohammed IMAAD KHATEEB
Department of Emergency Medicine, Ramaiah Medical College, Ramaiah University of Applied Sciences
Sachin Nayak SUJIR
sachin.sujir@manipal.edu
Department of Emergency Medicine, Kasturba Medical College, Manipal Academy of Higher Education
Udit Umesh UMRANI
Department of Emergency Medicine, Kasturba Medical College, Manipal Academy of Higher Education
Pranav PRAKASH
Department of Emergency Medicine, Kasturba Medical College, Manipal Academy of Higher Education
Vimal KRISHNAN S
Department of Emergency Medicine, Kasturba Medical College, Manipal Academy of Higher Education

Research Article

Keywords: Point-of-care ultrasound, diaphragmatic thickening fraction, HACOR score, non-invasive ventilation, acute respiratory distress

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11245150/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Non-invasive ventilation (NIV) is a cornerstone of management for acute respiratory distress in the emergency department, but failure requiring intubation remains common. The HACOR score is an established predictor of NIV failure and diaphragmatic thickening fraction (DTF) a promising one. This study aimed to determine, among patients who remained on NIV to the 1-hour landmark, whether HACOR provides incremental predictive value beyond DTF at 1 hour for subsequent NIV failure.

Methods

This secondary analysis used a prospective single-center cohort of adults with acute respiratory distress treated with NIV in an emergency department. Because DTF at 1 hour was a principal predictor, the analysis was landmarked at 1 hour and included only patients who remained on NIV to that reassessment; the model therefore estimates risk conditional on 1-hour NIV tolerance, not from the point of NIV initiation. HACOR was calculated at NIV initiation; DTF was measured at baseline and 1 hour. The primary outcome was NIV failure (endotracheal intubation) between 1 hour and 24 hours after NIV initiation. The primary comparison was the ΔAUC between DTF at 1 hour alone and a combined logistic model adding HACOR (HACOR-D). Bootstrap confidence intervals characterized uncertainty; likelihood-ratio testing assessed model fit.

Results

Among 126 patients, 36 (28.6%) experienced NIV failure. DTF at 1 hour had an AUC of 0.802 (95% CI 0.708 to 0.896), compared with 0.725 (95% CI 0.632 to 0.818) for HACOR. The combined model (HACOR-D) had an AUC of 0.812. The ΔAUC versus DTF at 1 hour alone was 0.010 (bootstrap 95% CI − 0.011 to 0.070; DeLong 95% CI − 0.028 to 0.048), with no significant improvement in discrimination. Adding HACOR significantly improved likelihood-based model fit (likelihood-ratio χ^2=4.52; P = 0.033). Both retained statistically significant coefficients in the mutually adjusted model.

Conclusions

In patients who remained on NIV to the 1-hour landmark, DTF at 1 hour demonstrated an AUC of 0.802 for subsequent NIV failure, conditional on 1-hour NIV tolerance. Adding baseline HACOR (HACOR-D) significantly improved likelihood-based model fit but did not demonstrate additional AUC-based discrimination. The magnitude of any incremental benefit remains uncertain and requires validation in larger, multicenter cohorts.

Background

Non-invasive ventilation (NIV) is a cornerstone of management for acute respiratory distress in the emergency department (ED), allowing many patients to avoid the risks of endotracheal intubation and invasive mechanical ventilation. However, NIV fails in a substantial proportion of patients, and failure requiring intubation is associated with worse outcomes than either successful NIV or primary intubation, including longer intensive care stay and higher mortality, particularly when recognition is delayed [1–4]. Early, reliable bedside identification of patients at risk of NIV failure is therefore essential to guide timely escalation of care, especially in resource-limited ED settings where continuous invasive monitoring is not always feasible. The HACOR score (heart rate, acidosis, consciousness, oxygenation, respiratory rate) is a validated bedside predictor of NIV failure that has been evaluated in emergency department populations and across a broader range of acute respiratory failure phenotypes, including acute cardiogenic pulmonary oedema, non-COPD chronic respiratory failure, and COVID-19 pneumonia managed both on general wards and in critical care [5–10]. In a recent prospective ED cohort, HACOR measured 1 hour after NIV initiation, and its early change, discriminated NIV failure better than baseline HACOR [11]. Diaphragmatic thickening fraction (DTF), measured by point-of-care ultrasound (POCUS), has more recently emerged as a complementary bedside tool: it is rapid, repeatable, and directly assesses the mechanical reserve of the primary respiratory muscle, and has shown promise across several acute respiratory failure populations [12–15]. In a previously published cohort of 126 ED patients, DTF measured 1 hour after NIV initiation was independently associated with NIV failure (AUC 0.802) and also correlated with duration of invasive ventilation, ICU stay, and in-hospital mortality [16].

HACOR and DTF assess different physiological domains: HACOR summarizes global cardiorespiratory and neurologic severity, whereas DTF reflects diaphragmatic mechanical reserve. A systematic review by Kheir et al. explicitly identified this as a gap in the field, noting that further work was needed to establish whether a combination of lung ultrasound, DTF, and the HACOR score could best predict NIV outcomes [15]. Prior work has suggested that combining clinical scores with ultrasound measures may improve prediction, but formal evaluation of HACOR added to DTF in undifferentiated ED patients with acute respiratory distress remains limited [15, 17–20]. Determining whether HACOR adds meaningful information beyond dynamic DTF is clinically relevant because it could establish whether diaphragm ultrasound alone is sufficient for risk stratification or whether a validated clinical score provides additional information. This distinction has direct implications for clinical workflow: if DTF alone is sufficient, risk stratification could rely on a single bedside ultrasound measurement without additional laboratory testing or score calculation, whereas if HACOR adds meaningful information, combined assessment would be warranted despite the added complexity.

This study aimed to determine whether baseline HACOR provides incremental predictive value beyond DTF measured 1 hour after NIV initiation for predicting subsequent NIV failure (intubation between 1 and 24 hours after NIV initiation) among ED patients with acute respiratory distress who remained on NIV without intubation at the 1-hour landmark. DTF and HACOR were also evaluated individually, and the primary comparison was the combined HACOR–DTF model (HACOR-D) versus DTF at 1 hour alone. We

hypothesized that combining the measures would improve discrimination. To our knowledge, few studies have formally evaluated the incremental predictive value of baseline HACOR when added to 1-hour DTF, with quantified discrimination, calibration, and internal validation, in an undifferentiated ED population with acute respiratory distress.

Methods

Study Design and Setting

This was a secondary analysis, using the same dataset, of a single-center, prospective observational study conducted in the Emergency Medicine Department of a tertiary care teaching hospital in India, involving adult patients presenting with acute respiratory distress between May 2023 and May 2025. The parent study has been described in detail previously and reported the association between DTF alone and NIV failure [16]. The present manuscript addresses a distinct research question that was not evaluated in the parent publication – namely, whether the HACOR score provides incremental predictive information when added to DTF at 1 hour, and whether a formally modeled combined predictor improves discrimination over either measure alone. No results pertaining to the HACOR score or the combined model have been reported elsewhere. The sample size for this secondary analysis was determined by the size of the parent cohort rather than by a priori power calculations for prediction-model development or for detection of a prespecified incremental AUC difference; this analysis should accordingly be regarded as an exploratory prediction-model analysis rather than definitive validation. The study was approved by the Institutional Ethics Committee (IEC2: 135/2023), registered with the Clinical Trials Registry of India (CTRI/2023/06/053740, dated 9th June 2023), adhered to the principles of the Declaration of Helsinki, and followed the STROBE reporting guidelines for observational studies [21] and the TRIPOD reporting guideline for multivariable prediction models [22].

Participants

Adult patients (> 18 years) presenting with acute respiratory distress – defined as respiratory rate > 25/min with signs of increased work of breathing, with or without hypoxaemia (SpO_2 <90% on room air or PaO_2/FiO_2 <300) – and initiated on NIV were included. We use the term “acute respiratory distress” because inclusion was based on clinical signs of increased respiratory effort and did not require gas-exchange criteria for acute respiratory failure; the cohort is the same as that described in the parent study [16]. Patients requiring immediate intubation or intubation before completion of the 1-hour NIV trial, and those with interstitial lung disease, chest wall deformity, pregnancy, trauma, prior diaphragmatic/thoracic surgery, chronic home NIV use, or who declined consent or left against medical advice, were excluded. Because early-intubation patients were excluded at this eligibility stage rather than after enrollment, the analytic cohort of 126 patients already represents the landmark-eligible population, with no further post-hoc exclusions applied for this analysis. Full inclusion/exclusion criteria and NIV protocol are reported in the parent publication [16].

Because DTF at 1 hour was a principal predictor in this analysis, the study population was necessarily restricted to patients who remained on NIV long enough to undergo reassessment at approximately 1 hour – a landmark analysis design. The prediction time origin for the primary analysis was therefore 1 hour after NIV initiation, not the point of ED presentation. Patients who required intubation before this landmark timepoint were excluded, as DTF at 1 hour could not be measured in them. Accordingly, the target population for inference was patients who remained on NIV without intubation until the 1-hour landmark: the model estimates the conditional risk of subsequent NIV failure in this landmark-eligible population and should not be interpreted as a prediction model applicable at NIV initiation.

HACOR Score

The HACOR score was calculated at NIV initiation using heart rate, arterial pH (acidosis), Glasgow Coma Scale (consciousness), SpO_2/FiO_2 ratio (oxygenation), and respiratory rate, following the original scoring system of Duan et al., with a maximum possible score of 25 [5, 8]. Because arterial pH and Glasgow Coma Scale were recorded only at baseline in this cohort, HACOR could not be recalculated at the 1-hour timepoint; the HACOR score used throughout this analysis therefore reflects the value at NIV initiation.

Diaphragm Ultrasound

DTF was measured by POCUS as previously described [16], using a GE Versana Active™ ultrasound machine with a high-frequency linear probe (7–12 MHz). Patients were positioned semi-recumbent at 45° or in their position of maximal comfort. The right hemidiaphragm was imaged at the zone of apposition, between the 7th and 9th intercostal spaces and between the anterior and mid-axillary lines, where it appears as a hypoechoic muscle layer between hyperechoic pleural and peritoneal lines. Diaphragm thickness (inner-to-inner) was measured in M-mode at peak inspiration (TEI) and end-expiration (TEE) during ongoing NIV, without interruption of ventilatory support; at least three consistent measurements were obtained and averaged. Measurements were performed within 30 minutes of NIV initiation (baseline) and after 1 hour of NIV, with a skin marker placed after the first assessment to ensure consistent probe placement. DTF was calculated as DTF (%) = (TEI − TEE) / TEE × 100. All examinations were performed by a single emergency medicine resident who had completed at least 25 supervised diaphragm ultrasound examinations in accordance with American College of Emergency Physicians training guidelines; images and video clips were stored electronically. The operator was unaware of HACOR values at the time of DTF acquisition, and DTF measurements were completed before the outcome occurred.

Outcome

The primary outcome was NIV failure, defined as endotracheal intubation occurring between the 1-hour landmark and 24 hours after NIV initiation. Intubation was guided by institutional criteria. Major criteria were cardiac or respiratory arrest; PaO_2/FiO_2 <100 mmHg despite NIV; need for airway protection

(Glasgow Coma Scale < 8, active seizures, or excessive tracheal secretions); and haemodynamic instability unresponsive to fluid resuscitation or vasoactive agents (systolic blood pressure < 90 mmHg or mean arterial pressure < 65 mmHg). Minor criteria were PaO_2/FiO_2 <150 mmHg on NIV, respiratory rate > 35/min, persistent signs of respiratory muscle fatigue, and pH < 7.25 with rising $PaCO_2$. Intubation was advised when at least one major or more than two minor criteria were met, but the final decision rested with the attending physician. Treating physicians were blinded to diaphragm ultrasound measurements, which were not part of routine clinical care and were not disclosed to the clinical team. Treating physicians could not be blinded to the individual components comprising the HACOR score – heart rate, arterial pH, Glasgow Coma Scale, oxygenation and respiratory rate – as these are standard vital signs and laboratory values monitored as part of routine ED care; the possibility that these variables influenced the decision to intubate, independent of their aggregation into a HACOR score, cannot therefore be excluded and is discussed as a limitation.

Statistical Analysis

There were no missing values for any predictor or the outcome; all 126 patients were included. Baseline characteristics (Table 1) are presented descriptively; candidate predictors were limited to HACOR and DTF, as defined by the research question, and no data-driven variable selection was performed. Individual predictive performance of baseline DTF, DTF at 1 hour, and HACOR was assessed using receiver operating characteristic (ROC) analysis, with area under the curve (AUC), data-derived operating points, sensitivity, specificity, predictive values, and likelihood ratios reported for each. 95% confidence intervals for individual AUCs were calculated using the DeLong method [23]. For descriptive purposes, operating points were selected as the point on each ROC curve closest to perfect classification (sensitivity and specificity of 1); they were considered exploratory and were not treated as clinically actionable thresholds. Linearity of HACOR and DTF at 1 hour on the logit scale was assessed using the Box–Tidwell approach. A combined logistic regression model (HACOR-D) included HACOR and DTF at 1 hour, and its predicted probability was evaluated as the HACOR-D risk estimate. HACOR-D is a logistic prediction model, not a point-based clinical score. The primary comparison was HACOR-D versus DTF at 1 hour alone. The ΔAUC and its 95% confidence interval were estimated with paired bootstrap resampling (2000 replications, percentile method); DeLong's test for correlated ROC curves [23] was reported as a sensitivity analysis, given concerns about its use for nested models [24]. The 95% confidence band around the HACOR-D ROC curve was estimated using the same 2000 bootstrap resamples, with sensitivity linearly interpolated at fixed values of 1-specificity in each resample and the pointwise 2.5th and 97.5th percentiles plotted; these resamples were generated separately from those used for optimism correction, calibration, and Brier score uncertainty (below). A likelihood-ratio test comparing the nested DTF-at-1-hour model with the full model was performed as a secondary analysis of incremental likelihood-based model fit. The analytical hierarchy was therefore: primary, bootstrap ΔAUC; sensitivity, DeLong comparison; secondary, likelihood-ratio test of incremental model fit, which assesses model fit rather than discrimination. Other pairwise AUC comparisons were exploratory, and their P values were not adjusted for multiplicity.

Because HACOR-D was developed and evaluated in the same dataset, internal validation used 2000 bootstrap replications to estimate optimism-corrected AUC, calibration slope, calibration intercept, and Brier score, following established approaches to bootstrap internal validation [25]. Calibration was assessed graphically across 5 risk groups (quintiles of predicted risk), consistent with contemporary calibration frameworks [26] and with the Brier score; grouped calibration was used because there were only 36 outcome events. Exploratory decision-curve analysis (Additional file 1), using the decision-curve framework [27], estimated net clinical benefit of HACOR-D, DTF at 1 hour, and HACOR relative to treating all or no patients across a range of threshold probabilities (1%–80%), selected to span the range of model-predicted probabilities and plausible intubation decision points, extending well beyond the cohort's observed NIV failure prevalence (28.6%). For DTF at 1 hour and HACOR individually, predicted probabilities were obtained from separate univariate logistic regression models before inclusion in the decision-curve analysis. Two-sided $P < 0.05$ was considered statistically significant for hypothesis tests. Analyses were performed in Python (v3.11) using scikit-learn and statsmodels and cross-checked against the original SPSS (v27) analysis of individual predictors from the parent study.

Results

Of 135 patients enrolled, 9 were excluded (discharge against medical advice or transfer to another facility), leaving 126 patients who met the eligibility criteria and remained on NIV to the 1-hour landmark; the full patient flow is shown in the parent publication [16]. Of these 126 patients, 36 (28.6%) experienced NIV failure between the 1-hour landmark and 24 hours after NIV initiation and required endotracheal intubation. Baseline demographic, clinical, and laboratory characteristics are summarized in Table 1. The median baseline HACOR score was higher among patients who failed NIV than among those who did not (median 8.5 vs 5; $P < 0.001$). The underlying diagnoses were heterogeneous (Additional file 1).

Table 1
Baseline demographic, clinical, and laboratory characteristics of the cohort, by NIV outcome (n = 126)

Characteristic	NIV success (n = 90)	NIV failure (n = 36)	P value
Age, years, median (IQR)	67 (55.75–75)	63.5 (47.75–72)	0.216
Male sex, n (%)	60 (66.7)	23 (63.9)	0.766
Hypertension, n (%)	68 (75.6)	17 (47.2)	0.002*
Diabetes mellitus, n (%)	52 (57.8)	16 (44.4)	0.175
Chronic kidney disease, n (%)	24 (26.7)	4 (11.1)	0.058
Ischemic heart disease, n (%)	23 (25.6)	4 (11.1)	0.074
Bronchial asthma/COPD, n (%)	25 (27.8)	12 (33.3)	0.536
Respiratory rate @baseline, /min	31 (30–36)	32 (30–36)	0.730
Respiratory rate @1hr, /min	26 (24–30)	28 (26–31.5)	0.022*
Heart rate @baseline, /min	106.5 (95.75–120.75)	120 (99.25–129.5)	0.056
Heart rate @1hr, /min	100 (85–110)	111 (94.25–120)	0.038*
Systolic BP @baseline, mmHg	140 (120–180)	110 (90–137.5)	< 0.001*
SpO_2 @baseline, %	85 (80–90)	80 (70–85.75)	0.004*
Glasgow Coma Scale	15 (15–15)	15 (13–15)	< 0.001*
FiO_2 @baseline, %	40 (40–60)	60 (40–80)	0.003*
pH	7.32 (7.26–7.42)	7.30 (7.20–7.36)	0.027*
PaO_2/FiO_2 ratio	165.5 (129.5–217.75)	141.5 (101.75–172)	0.004*
Lactate, mg/dL	21 (13–37.5)	44 (16–84.5)	0.007*
DTF @baseline, %	39.05 (29.1–55.25)	30.6 (20.68–44.15)	0.018*
DTF @1hr, %	44.25 (31.25–59.85)	21.2 (16.58–37.98)	< 0.001*
HACOR score	5 (3.75–8)	8.5 (6.25–11)	< 0.001*

*Continuous variables reported as median (IQR); compared using Mann-Whitney U test. Categorical variables compared using chi-square or Fisher's exact test. BP – blood pressure; DTF – diaphragmatic thickening fraction; HACOR – heart rate, acidosis, consciousness, oxygenation, respiratory rate. *Statistically significant.*

Individual Predictive Performance

DTF at 1 hour had the highest observed AUC among the individual predictors of subsequent NIV failure (AUC 0.802, 95% CI 0.708 to 0.896), followed by HACOR (AUC 0.725, 95% CI 0.632 to 0.818) and baseline DTF (AUC 0.635, 95% CI 0.528 to 0.742) (Table 2; Fig. 1). DTF at 1 hour had a numerically higher AUC than HACOR, although the difference was not statistically demonstrable (ΔAUC 0.077; P = 0.169). Data-derived operating points are reported for descriptive purposes only and should not be interpreted as validated clinical thresholds; predictive values and accuracy depend on the 28.6% event prevalence of this cohort and are less transportable than AUC and likelihood ratios.

Table 2
Predictive performance of DTF, HACOR score, and HACOR-D for NIV failure (n = 126; NIV-failure prevalence 28.6%)

Predictor	AUC (95% CI)	Operating point	Sens (%)	Spec (%)	PPV (%)	NPV (%)	LR+	LR−	Acc (%)
DTF @ baseline	0.635 (0.528–0.742)	≤ 34.65%	61.1	60.0	37.9	79.4	1.53	0.65	60.3
DTF @ 1 hour	0.802 (0.708–0.896)	≤ 32.35%	75.0	74.4	54.0	88.2	2.93	0.34	74.6
HACOR score	0.725 (0.632–0.818)	≥ 7.5	58.3	74.4	47.7	81.7	2.28	0.56	69.8
HACOR-D (HACOR + DTF@1hr)	0.812 (0.725–0.900)	prob ≥ 0.39	75.0	85.6	67.5	89.5	5.19	0.29	82.5

DTF – diaphragmatic thickening fraction; HACOR – heart rate, acidosis, consciousness, oxygenation, respiratory rate; AUC – area under the curve; PPV – positive predictive value; NPV – negative predictive value; LR+ – positive likelihood ratio; LR− – negative likelihood ratio; Acc – overall accuracy. Operating points are data-derived and exploratory, not validated clinical thresholds. PPV, NPV, and accuracy depend on cohort prevalence.

Combined Model (HACOR-D)

A multivariable logistic regression model combining baseline HACOR and DTF at 1 hour was constructed as a landmark model because DTF at 1 hour could only be assessed in patients who remained on NIV to that timepoint. Both HACOR (OR 1.16 per point, 95% CI 1.01 to 1.33; P = 0.041) and DTF at 1 hour (OR 0.94 per percentage point, 95% CI 0.91 to 0.97; P < 0.001) remained significantly associated with NIV failure after mutual adjustment (Table 3). The fitted equation is shown below for reproducibility; it is a development-cohort model and not a validated clinical score.

Table 3
Multivariable regression model and nested model comparison for HACOR-D (HACOR + DTF@1hr) (A) Regression coefficients

Variable	Coefficient (B)	Odds Ratio	95% CI	P value
HACOR score (per point)	0.145	1.16	1.01–1.33	0.041*
DTF @ 1 hour (per %)	−0.059	0.94	0.91–0.97	< 0.001*
Constant	0.175	–	–	0.844

(B) Nested model comparison – DTF@1hr alone versus DTF@1hr + HACOR

Model	Log-likelihood	AIC	BIC	LR χ^2 (df = 1)	P value
DTF @1hr alone	−60.72	125.45	131.12	–	–
DTF @1hr + HACOR	−58.46	122.92	131.43	4.52	0.033*

*HACOR – heart rate, acidosis, consciousness, oxygenation, respiratory rate; DTF – diaphragmatic thickening fraction; CI – confidence interval; AIC – Akaike information criterion; BIC – Bayesian information criterion; LR – likelihood ratio. *Statistically significant.*

$$\text{logit}[P(\text{NIV failure})] = 0.175 + 0.145 \times \text{HACOR} - 0.059 \times \text{DTF@1hr(\%)}$$

$$P(\text{NIV failure}) = 1 / [1 + \exp(-\text{logit})]$$

HACOR-D had an apparent AUC of 0.812 (95% CI 0.725 to 0.900). Its data-derived operating point (predicted probability ≥ 0.39) had 85.6% specificity, but this threshold was derived and evaluated in the same dataset and should be considered hypothesis-generating rather than clinically validated.

Linearity Assessment

Box–Tidwell testing did not indicate departure from linearity for HACOR (P = 0.486) but suggested possible nonlinearity for DTF at 1 hour (P = 0.006). A restricted cubic-spline sensitivity model modestly improved fit (AIC 119.5 vs 122.9; likelihood-ratio $\chi 2 = 5.47$, P = 0.019 for the nonlinear term) but produced only a small increase in discrimination (AUC 0.816 vs 0.812 for the linear model). Given the limited number of outcome events, the additional degrees of freedom required by the spline, and the exploratory objective, the linear specification was retained as the primary model to limit overfitting, and the nonlinear model was treated as a sensitivity analysis rather than as the primary model. The spline findings require confirmation in larger, externally validated cohorts.

Model Fit and Internal Validation

For the primary comparison, HACOR-D versus DTF at 1 hour alone had a ΔAUC of 0.010 (bootstrap 95% CI − 0.011 to 0.070); the corresponding DeLong 95% CI was − 0.028 to 0.048 (P = 0.618). Both intervals included zero, providing no statistically demonstrable improvement in discrimination. As a secondary

analysis, adding HACOR significantly improved likelihood-based model fit (likelihood-ratio χ^2=4.52, df = 1; P = 0.033), with lower AIC but marginally higher BIC for HACOR-D. Thus, although adding HACOR significantly improved likelihood-based model fit, this did not translate into a demonstrable improvement in rank discrimination, as reflected by the primary ΔAUC analysis. Bootstrap internal validation yielded an optimism-corrected AUC of 0.803 versus an apparent AUC of 0.812, a calibration slope of 0.96, a calibration intercept of − 0.02 (95% CI − 0.47 to 0.57), and an optimism-corrected Brier score of 0.158 (95% CI 0.125 to 0.199) versus an apparent Brier score of 0.149.(Table 4)These findings indicate limited apparent optimism but do not establish external validity. Grouped calibration showed no obvious systematic over- or underprediction (Fig. 2).

Table 4
Internal validation (bootstrap, 2000 replications) for HACOR-D

Metric	Apparent	Optimism-corrected
AUC (95% CI)	0.812 (0.725–0.900)	0.803 (0.722–0.888)
Calibration slope	1.00 (by construction)	0.96
Calibration intercept	0 (by construction)	−0.02 (95% CI − 0.47 to 0.57)
Brier score	0.149	0.158 (95% CI 0.125–0.199)

AUC – area under the curve; CI – confidence interval.

Comparison of AUCs (correlated ROC analysis)

In exploratory secondary comparisons, HACOR-D had a higher AUC than HACOR alone (P = 0.034) and baseline DTF (P < 0.001), and DTF at 1 hour had a higher AUC than baseline DTF (P < 0.001) (Table 5; Fig. 1). These P values were not adjusted for multiplicity. HACOR-D was not superior to DTF at 1 hour alone in the primary comparison, and DTF at 1 hour was not superior to HACOR alone.

Table 5
Exploratory pairwise comparisons of correlated ROC curves

Comparison	ΔAUC	95% CI	P value
HACOR-D vs. HACOR alone	0.087	0.006 to 0.167	0.034*
HACOR-D vs. DTF @ 1 hour alone	0.010	−0.028 to 0.048	0.618
DTF @ 1 hour vs. HACOR alone	0.077	−0.033 to 0.187	0.169
HACOR-D vs. DTF @ baseline	0.177	0.097 to 0.258	< 0.001*
DTF @ 1 hour vs. DTF @ baseline	0.168	0.082 to 0.254	< 0.001*

*AUC – area under the curve; DTF – diaphragmatic thickening fraction; HACOR – heart rate, acidosis, consciousness, oxygenation, respiratory rate. *Statistically significant. The HACOR-D-versus-DTF@1hr*

comparison is the primary discrimination endpoint. Secondary pairwise AUC comparisons are exploratory, and P values were not adjusted for multiplicity.

Decision Curve Analysis

Exploratory decision-curve analysis is presented in Additional file 1. At threshold probabilities below approximately 10%, all approaches offered net benefit similar to intubating all patients. Between approximately 10% and 46%, the apparent net benefit of HACOR-D was equal to or slightly higher than that of DTF at 1 hour alone (difference 0 to 0.04), whereas between approximately 47% and 61% DTF at 1 hour alone was higher (difference up to 0.06). The curves therefore crossed, and no approach showed a consistent advantage across the threshold range. HACOR alone had the lowest net benefit at most thresholds above 24%. These apparent, non-optimism-corrected estimates from a development cohort with 36 events are descriptive and do not constitute evidence of clinical utility.

Discussion

The principal finding was that, among patients who remained on NIV at the 1-hour landmark, DTF measured at that timepoint provided substantial discrimination for subsequent NIV failure (AUC 0.802), consistent with the earlier report [16], and that adding baseline HACOR did not produce a demonstrable improvement in discrimination (ΔAUC 0.010; bootstrap 95% CI − 0.011 to 0.070; DeLong P = 0.618). The confidence interval remains compatible with a modest gain or loss in discrimination, so the study neither establishes nor excludes clinically meaningful incremental value. Threshold-dependent differences should be regarded as exploratory. These estimates are conditional on having remained on NIV without intubation to the 1-hour landmark and do not apply to patients requiring earlier intubation.

Association, model fit, and discrimination should be distinguished. HACOR remained independently associated with NIV failure after mutual adjustment (OR 1.16 per point) and significantly improved likelihood-based model fit (likelihood-ratio χ^2=4.52; P = 0.033). Although HACOR therefore contributed statistically significant incremental information to the fitted model, this did not translate into a demonstrable improvement in rank discrimination (HACOR-D AUC 0.812 versus 0.802 for DTF at 1 hour alone). A significant odds ratio or likelihood-ratio test indicates that a variable carries information conditional on the others, not that it improves the model's ability to separate patients who will and will not fail NIV. Whether this information improves clinical decisions remains unresolved. Exploratory decision-curve analysis likewise showed no consistent net-benefit advantage of HACOR-D over DTF at 1 hour alone, with crossing curves across the threshold range (Additional file 1); these development-cohort estimates require external validation.

Direct comparisons between diaphragm ultrasound measures and HACOR remain limited. In a small cohort with acute exacerbation of chronic obstructive pulmonary disease, a diaphragm thickness index and HACOR were combined into a decision rule without a combined-model AUC or formal comparison, and the rule's sensitivity was lower than either measure alone [17]. Comparisons of HACOR with other ultrasound- or clinical-index-based predictors in the ED have shown broadly comparable discrimination

[28, 29]. The present analysis extends this work by formally modelling the combination in an undifferentiated ED population.

HACOR and DTF assess different domains of respiratory failure physiology: HACOR summarizes cardiorespiratory and neurologic severity, whereas DTF reflects diaphragmatic mechanical reserve. Because they were measured at different timepoints, HACOR-D integrates baseline clinical severity with a 1-hour dynamic diaphragmatic response; a measure of response after 1 hour of NIV may already capture much of the prognostic information contained in baseline severity, which could explain the limited discriminative gain. Consistent with this, a recent prospective ED cohort found that baseline HACOR did not significantly discriminate NIV failure, whereas HACOR measured at 1 hour (AUC 0.760) and its early change (AUC 0.798) did [11]. As HACOR was available only at baseline in the present cohort, our analysis compares a baseline clinical score with a 1-hour dynamic measure and does not address whether a 1-hour HACOR would perform comparably to DTF at 1 hour. Because NIV failure was defined by the treating team's decision to intubate, HACOR should be interpreted as a predictor of a clinician-determined endpoint rather than a purely biological outcome.

Clinically, these findings support prospective evaluation of DTF at 1 hour as a dynamic risk-stratification measure among patients who remain on NIV to that landmark. They do not establish a validated DTF threshold or show that HACOR is unnecessary. Bootstrap internal validation showed limited apparent optimism in HACOR-D, but the analysis was not externally validated. Data-derived cutoffs and decision-curve estimates therefore should not be used as clinical thresholds without prospective validation.

Performance should also be interpreted in the context of population, timing, and outcome. DTF achieved a higher AUC (0.940) in a COPD cohort with a two-hour NIV trial and a 48-hour outcome [30] but was a weak predictor (AUC 0.572) in an ARDS extubation cohort [18]. Performance in other settings is likely to vary with the underlying cause of respiratory failure, NIV mode and interface, local NIV protocols, clinicians' thresholds for intubation, patient selection, and operator expertise in diaphragm ultrasound, which is particularly relevant because DTF measurement is operator dependent. External validation in undifferentiated ED populations is therefore required.

Limitations

This study has several limitations. First, it is a secondary, hypothesis-generating analysis of a single-centre cohort with 36 failure events; the ΔAUC confidence interval remains compatible with both modest benefit and modest harm from adding HACOR, so the study cannot exclude clinically relevant incremental discrimination. Second, treating physicians were blinded to DTF but not to the routine vital signs and laboratory values comprising HACOR, several of which also feature in the intubation criteria; these variables may therefore have influenced the intubation decision and introduce partial incorporation bias and management bias for HACOR specifically, though not for DTF. Third, this is a landmark analysis: by design, the analytic cohort is conditioned on being event-free (not yet intubated) at the 1-hour landmark, because DTF at 1 hour can only be measured in patients who survive event-free to that

timepoint. This conditioning is the standard and statistically appropriate way to estimate risk from a landmark-measured predictor, and is distinct from immortal-time bias, which arises when person-time before a time-varying exposure is misclassified as exposed time; no such misclassification occurs here, since DTF at 1 hour is only defined from the landmark forward. Nonetheless, the practical consequence is the same point raised above: the model estimates NIV-failure risk conditional on having already tolerated 1 hour of NIV and does not address immediate intubation decisions at presentation or patients intubated before reassessment; it should not be applied, and was not intended to be applied, at the point of NIV initiation. Fourth, DTF at 1 hour showed possible nonlinearity, but the spline sensitivity analysis produced only a small change in AUC; the linear specification was retained because of the limited number of events and the exploratory design. Fifth, HACOR could not be recalculated at 1 hour because pH and Glasgow Coma Scale were not repeated; serial HACOR and DTF may better capture illness trajectory. Sixth, the model was internally but not externally validated, and all operating thresholds and decision-curve estimates were data-derived. Seventh, DTF was measured by a single trained operator (a supervised emergency medicine resident meeting ACEP training standards) without formal interobserver or intraobserver reliability assessment, limiting assessment of reproducibility. Finally, the heterogeneous case mix may enhance real-world relevance but precluded adequately powered disease-specific validation.

Conclusions

Among adults with acute respiratory distress who remained on NIV without intubation to the 1-hour landmark, DTF measured at that timepoint demonstrated substantial discrimination for subsequent NIV failure. Adding baseline HACOR (HACOR-D) significantly improved likelihood-based model fit but did not demonstrate a statistically significant improvement in discrimination over DTF alone (ΔAUC 0.010, 95% CI − 0.011 to 0.070). The incremental value of baseline HACOR beyond 1-hour DTF therefore remains uncertain, and confirmation is required in larger, multicenter cohorts with standardized intubation criteria, repeated HACOR assessment at the 1-hour landmark, and formal assessment of DTF measurement reliability.

Abbreviations

ADHF
Acute decompensated heart failure
AE
Acute exacerbation
AIC
Akaike information criterion
ARDS
Acute respiratory distress syndrome
AUC
Area under the curve

BIC
Bayesian information criterion
BP
Blood pressure
CI
Confidence interval
COPD
Chronic obstructive pulmonary disease
CTRI
Clinical Trials Registry of India
DTF
Diaphragmatic thickening fraction
ED
Emergency department
ESRD
End-stage renal disease
FiO_2
Fraction of inspired oxygen
HACOR
Heart rate, Acidosis, Consciousness, Oxygenation, Respiratory rate
HACOR-D
Combined HACOR and one-hour DTF model
ICU
Intensive care unit
IEC
Institutional Ethics Committee
IQR
Interquartile range
LR
Likelihood ratio
LR+
Positive likelihood ratio
LR−
Negative likelihood ratio
LRTI
Lower respiratory tract infection
MODS
Multiple organ dysfunction syndrome
NIV
Non-invasive ventilation

NPV
Negative predictive value
OR
Odds ratio
$PaCO_2$
Partial pressure of arterial carbon dioxide
PaO_2
Partial pressure of arterial oxygen
POCUS
Point-of-care ultrasound
PPV
Positive predictive value
ROC
Receiver operating characteristic
SCAPE
Sympathetic crashing acute pulmonary oedema
SpO_2
Peripheral oxygen saturation
SPSS
Statistical Package for the Social Sciences
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology
TEE
Thickness at end-expiration
TEI
Thickness at end-inspiration
TRIPOD
Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis

Declarations

Ethics approval and consent to participate

The study protocol was approved by the institutional ethics committee (Kasturba Medical College and Kasturba Hospital Institutional Ethics Committee, IEC2: 135/2023) and is registered with the Clinical Trials Registry of India CTRI/2023/06/053740, dated 9th June 2023. The study adhered to the principles of the Declaration of Helsinki. This manuscript reports a analysis of the same prospectively collected cohort as a previously published parent study, which addressed the association between DTF alone and NIV failure [16]. The current analysis addresses the distinct question of whether baseline HACOR provides incremental predictive information beyond DTF measured at 1 hour. Informed consent was obtained from patients or their legal representatives as part of the parent study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Author Contribution

Conceptualisation – MIK, MK, SNS, VK. Data curation – MIK, SNS. Formal analysis – MIK, MK, SNS, PP, VK. Investigation – MIK, MK, SNS, VK. Methodology – MIK, MK, SNS, UUU, VK. Supervision – MK, SNS, VK. Writing – original draft – MIK, MK, PP, UUU, VK. Writing – review & editing – MIK, MK, SNS, PP, UUU, VK

Acknowledgement

We thank the staff of the Department of Emergency Medicine, Kasturba Medical College, Manipal, for their support during this study.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Rochwerg B, Brochard L, Elliott MW, Hess D, Hill NS, Nava S, et al. Official ERS/ATS clinical practice guidelines: non-invasive ventilation for acute respiratory failure. Eur Respir J. 2017;50:1602426.
2. RENOVATE Investigators and the BRICNet Authors. High-Flow Nasal Oxygen vs Non-invasive Ventilation in Patients With Acute Respiratory Failure: The RENOVATE Randomized Clinical Trial. JAMA. 2025;333:875–90.
3. Demoule A, Girou E, Richard J-C, Taille S, Brochard L. Benefits and risks of success or failure of noninvasive ventilation. Intensive Care Med. 2006;32:1756–65.

4. Thille AW, Contou D, Fragnoli C, Córdoba-Izquierdo A, Boissier F, Brun-Buisson C. Noninvasive ventilation for acute hypoxemic respiratory failure: intubation rate and risk factors. Crit Care. 2013;17:R269.
5. Duan J, Han X, Bai L, Zhou L, Huang S. Assessment of heart rate, acidosis, consciousness, oxygenation, and respiratory rate to predict noninvasive ventilation failure in hypoxemic patients. Intensive Care Med. 2017;43:192–9.
6. Duan J, Chen L, Liu X, Bozbay S, Liu Y, Wang K, et al. An updated HACOR score for predicting the failure of noninvasive ventilation: a multicenter prospective observational study. Crit Care. 2022;26:196.
7. Chong CY, Bustam A, Noor Azhar M, Abdul Latif AK, Ismail R, Poh K. Evaluation of HACOR scale as a predictor of noninvasive ventilation failure in acute cardiogenic pulmonary edema patients: a prospective observational study. Am J Emerg Med. 2024;79:19–24.
8. da Silva MPN, Lunardi AC. Comparison of the clinimetric properties of the two versions of the HACOR scale for predicting noninvasive ventilation failure in Brazilian patients. Acute Crit Care. 2025;40:322–9.
9. Ding M, Han X, Bai L, Huang S, Duan J. Impact of HACOR score on noninvasive ventilation failure in non-COPD patients with acute-on-chronic respiratory failure. Can Respir J. 2021;2021:9960667.
10. Pepe F, Fabiani G, Cogozzo C, Mazzilli V, Villanti M, Pini R, et al. HACOR score to predict NIV failure in patients with COVID-related hypoxemic respiratory failure managed in the ordinary ward and in the critical care setting. Intern Emerg Med. 2024;19:2319–29.
11. Doğan B, Yazla M, Kaya B, Çelikel E, Kablan A, Erdem AB. Time-dependent predictive value of the HACOR score for non-invasive ventilation failure in the emergency department: a prospective cohort study. BMC Emerg Med. 2026;26:248.
12. Antenora F, Fantini R, Iattoni A, Castaniere I, Sdanganelli A, Livrieri F, et al. Prevalence and outcomes of diaphragmatic dysfunction assessed by ultrasound technology during acute exacerbation of COPD: A pilot study. Respirology. 2017;22:338–44.
13. Marchioni A, Castaniere I, Tonelli R, Fantini R, Fontana M, Tabbì L, et al. Ultrasound-assessed diaphragmatic impairment is a predictor of outcomes in patients with acute exacerbation of chronic obstructive pulmonary disease undergoing noninvasive ventilation. Crit Care. 2018;22:109.
14. Corradi F, Vetrugno L, Orso D, Bove T, Schreiber A, Boero E, et al. Diaphragmatic thickening fraction as a potential predictor of response to continuous positive airway pressure ventilation in Covid-19 pneumonia: A single-center pilot study. Respir Physiol Neurobiol. 2021;284:103585.
15. Kheir M, Dong V, Roselli V, Mina B. The role of ultrasound in predicting noninvasive ventilation outcomes: a systematic review. Front Med. 2023;10:1233518.
16. Khateeb MI, Khalid M, Sujir SN, Naik S, Prakash P, Krishnan SV. Association between diaphragm thickening fraction and non-invasive ventilation failure in patients with acute respiratory failure in the Emergency Department. Eur J Emerg Med. 2026 Apr;22. 10.1097/MEJ.0000000000001346.

17. Gomaa SAI, Radwan WA, Nassar YS, Elhenawy YH, Elansary MG. Diaphragmatic dysfunction by ultrasound and HACOR score for early prediction of noninvasive ventilation failure in patients with acute exacerbation of chronic obstructive pulmonary disease. Egypt J Radiol Nucl Med. 2025;56:202.
18. Liu Y, Zhou Y, Liu P, Ying W, Wu H, Dong Z. Combined lung and diaphragm ultrasound predicts extubation outcomes in ARDS: a prospective study. Eur J Med Res. 2024;29:510.
19. Song J, Luo Q, Lai X, Hu W, Yu Y, Wang M, et al. Combined cardiac, lung, and diaphragm ultrasound for predicting weaning failure during spontaneous breathing trial. Ann Intensive Care. 2024;14:60.
20. Parada-Gereda HM, Tibaduiza AL, Rico-Mendoza A, Molano-Franco D, Nieto VH, Arias-Ortiz WA, et al. Effectiveness of diaphragmatic ultrasound as a predictor of successful weaning from mechanical ventilation: a systematic review and meta-analysis. Crit Care. 2023;27:174.
21. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–8.
22. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–45.
24. Demler OV, Pencina MJ, D'Agostino RB, Sr. Misuse of DeLong test to compare AUCs for nested models. Stat Med. 2012;31:2577–87.
25. Steyerberg EW, Harrell FE Jr, Borsboom GJO, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54:774–81.
26. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol. 2016;74:167–76.
27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26:565–74.
28. Rathish M, Renuka MK, Prasant N, Sailaja B. Lung ultrasound score versus HACOR score as a predictor for weaning outcome in patients at high risk for extubation failure. Lung India. 2025;42:211–17.
29. Praphruetkit N, Boonchana N, Monsomboon A, Ruangsomboon O. ROX index versus HACOR scale in predicting success and failure of high-flow nasal cannula in the emergency department for patients with acute hypoxemic respiratory failure: a prospective observational study. Int J Emerg Med. 2023;16:3.
30. Suttapanit K, Lerdpaisarn P, Charoensuksombun C, Sanguanwit P, Supatanakij P. Diaphragmatic ultrasonographic evaluation as an assessment guide for predicting noninvasive ventilation failure in acute exacerbation of chronic obstructive pulmonary disease. Am J Emerg Med. 2025;93:13–20.

Figures

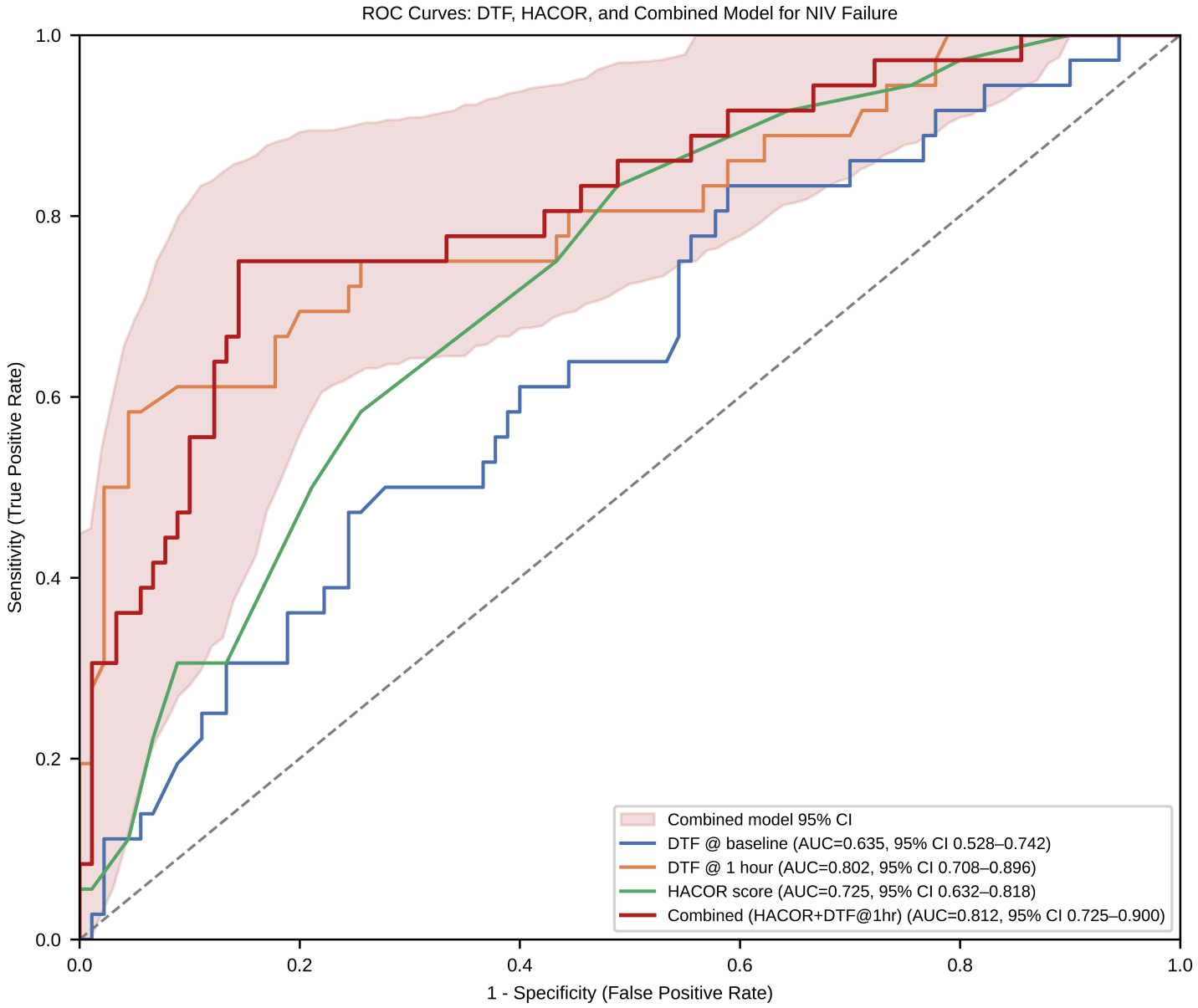


Figure 1

ROC curves of DTF at baseline, DTF at 1 hour, HACOR score, and the combined HACOR–DTF model (HACOR-D; labelled “Combined”) for prediction of NIV failure, with bootstrap 95% confidence band shown for HACOR-D.

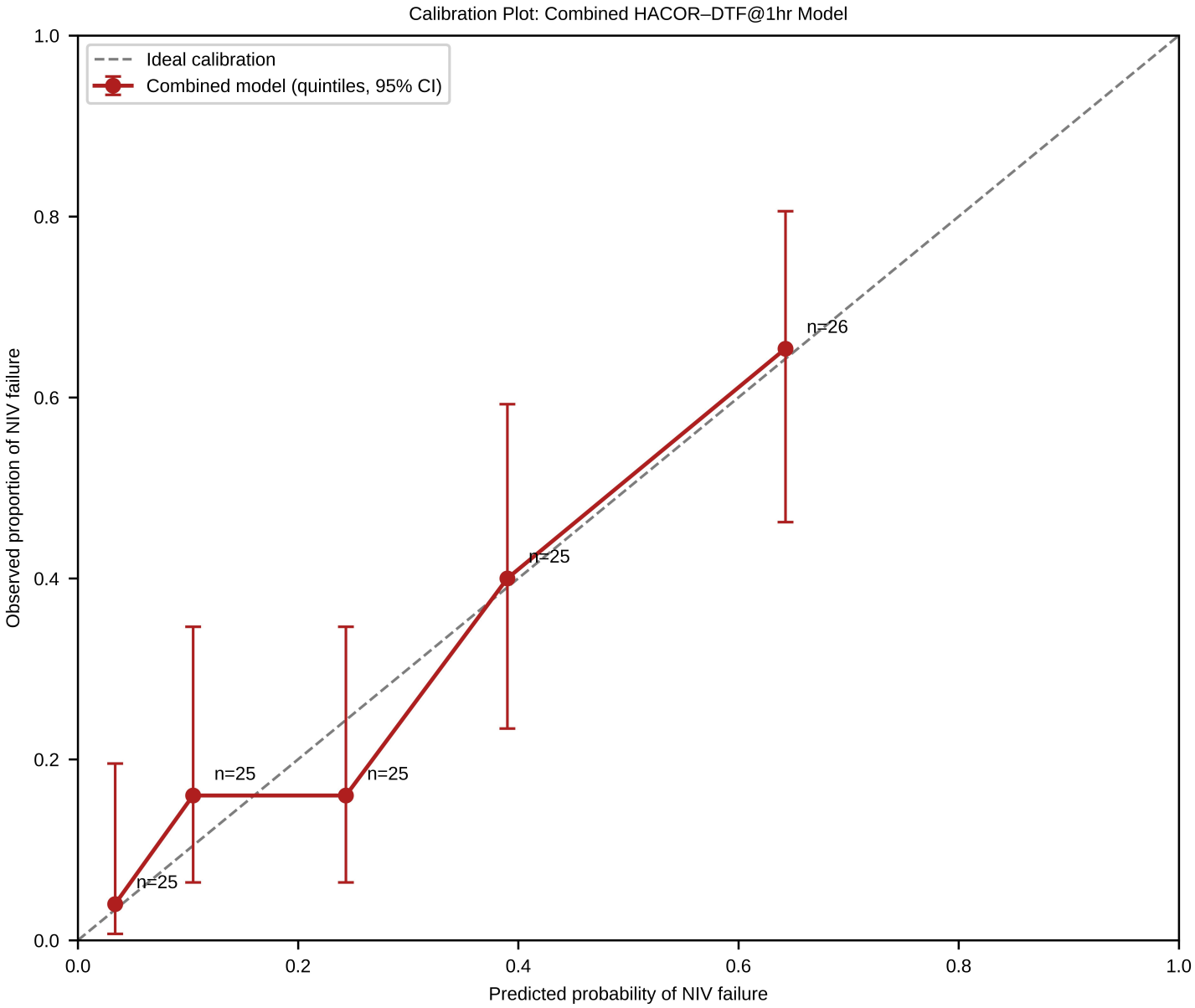


Figure 2

Calibration plot for HACOR-D (HACOR + DTF@1hr) (observed versus predicted probability of NIV failure, by quintile of predicted risk, with Wilson 95% confidence intervals around observed proportions).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx