Section: Anaesthesia

Original Research Article

DIAPHRAGM ULTRASOUND FOR PREDICTING EXTUBATION FAILURE IN MECHANICALLY VENTILATED PATIENTS: A PROSPECTIVE DIAGNOSTIC ACCURACY STUDY

Sujeet Bhamre[1], Atul Sangale[2], Vinaykumar Rathod[3]

[1,2,3]Assistant Professor, Department of Anaesthesia, SSNK Memorial Medical College and Hospital, Kopargaon Taluka, Ahilyanagar (Ahmednagar) District, Maharashtra, India.

Received *: 04/09/2026*
Received in revised form *: 02/10/2026*
Accepted *: 15/10/2026*

Corresponding Author:
Dr. Sujeet Bhamre,
Assistant Professor, Department of Anaesthesia, SSNK Memorial Medical College and Hospital, Kopargaon Taluka, Ahilyanagar (Ahmednagar) District, Maharashtra, India.
Email: sujeet17may@gmail.com

DOI: 10.70034/ijmedph.2026.4.182

Source of Support: Nil,
Conflict of Interest: None declared

Int J Med Pub Health
2026; 16 (4); 1196-1203

ABSTRACT

Background: Extubation fails in a clinically important minority of intensive care patients and is associated with longer stay and higher morbidity. Once conventional readiness criteria, including a rapid shallow breathing index (RSBI) below 105 breaths/min/L, have already been met, that index has less room to discriminate. Diaphragm ultrasound measures muscle function directly. We estimated the accuracy of diaphragm thickening fraction (DTF), diaphragm excursion (DE), RSBI, and a twelve-zone lung aeration score for extubation failure.

Materials and Methods: This prospective observational diagnostic accuracy study was conducted in the intensive critical care unit of a teaching hospital and is reported to STARD 2015. Between August 2025 and February 2026, 50 adults who had been ventilated for more than 48 hours, and who met prespecified readiness criteria, underwent a 30- to 60-minute spontaneous breathing trial. At 30 minutes, RSBI, right-sided DE, DTF, and the lung aeration score were recorded. The team deciding on extubation was blinded to the ultrasound measurements. The reference standard was reintubation, or death, within 48 hours. Optimal cut-offs maximised the Youden index.

Results: Extubation succeeded in 44 patients (88.0%) and failed in 6 (12.0%). Failure was associated with longer ventilation before extubation (7.8 ± 3.2 vs 4.4 ± 1.6 days), higher RSBI (81.2 ± 9.4 vs 52.4 ± 8.6 breaths/min/L), a higher lung aeration score (14.8 ± 2.5 vs 8.5 ± 3.6), lower DE (1.18 ± 0.21 vs 1.65 ± 0.24 cm), and lower DTF (25.4 ± 4.5% vs 42.8 ± 6.4%) (all $p \leq 0.002$). For prediction of failure, DTF ≤ 32% gave sensitivity 83.3%, specificity 90.9%, positive predictive value 55.6%, negative predictive value 97.6%, and area under the curve 0.898 (95% CI 0.772–1.000). The areas under the curve were 0.875 for DE, 0.864 for RSBI, and 0.848 for the lung aeration score. Confidence intervals overlapped.

Conclusion: Among patients who already met readiness-to-wean criteria, a DTF of 32% or less during the spontaneous breathing trial was the strongest point-estimate marker of extubation failure. A DTF above 32% argued against reintubation within 48 hours, but positive predictive values were modest and the failure group was small. The measurement should inform, not replace, the extubation decision.

Keywords: Diaphragm Ultrasound, Diaphragm Thickening Fraction, Diaphragm Excursion, Lung Ultrasound, Extubation Failure, Mechanical Ventilation, Spontaneous Breathing Trial.

INTRODUCTION

Removal of the endotracheal tube is the last step in liberation from invasive mechanical ventilation. Extubation fails, and the patient is reintubated, in about 10% to 20% of planned extubations.[1] Reintubation prolongs intensive care and hospital stay and raises the risk of ventilator-associated pneumonia and death.[2–3] Deciding early enough to avoid diaphragm injury, but not so early that the airway must be restored, remains one of the less certain judgements in critical care.

Weaning should be considered as soon as the reason for respiratory failure has improved.[4] A spontaneous breathing trial (SBT), lasting 30 to 120 minutes and conducted on a T-piece or on a low level of pressure support, is the usual test of whether the tube can be removed.[4] The rapid shallow breathing index (RSBI), the ratio of respiratory rate to tidal volume, is the numeric adjunct most often added to that test.[5] An RSBI below 105 breaths/min/L, adequate oxygenation on modest inspired oxygen and positive end-expiratory pressure, a tolerable cough, and haemodynamic stability are widely accepted readiness criteria.[4,6] Those criteria still misclassify patients. Some who meet every one of them are reintubated within two days.

Diaphragm weakness explains part of this gap. A few days of controlled ventilation are enough to produce fibre atrophy, and contractile injury follows when the muscle is loaded again.[7–8] Ultrasound studies show that diaphragm thickness commonly falls when inspiratory effort is very low.[9] Weakness acquired in the intensive care unit is easy to miss at the bedside and is associated with prolonged weaning and extubation failure.[10–11]

Two ultrasound measurements describe the diaphragm directly. In the zone of apposition, the thickening fraction (DTF) is the percentage increase in thickness from end-expiration to end-inspiration. From a subcostal window, M-mode records excursion (DE) of the hemidiaphragm.[12] DiNino and colleagues found that thickening of at least 30% predicted freedom from ventilatory support for 48 hours after extubation, with an area under the receiver operating characteristic curve of 0.79.[13] Later single-centre studies reported thickening-fraction thresholds from 20% to 36%, the exact value depending on the level of pressure support and on whether the outcome was success of the breathing trial or success of extubation, and excursion thresholds around 1.0 to 1.4 cm.[14–17] In a systematic review, the summary area under the curve for DTF was approximately 0.87, and accuracy for excursion was lower and more heterogeneous.[18]

Lung ultrasound addresses a different reason for failure. A twelve-zone aeration score summarises loss of A-lines, accumulation of B-lines, and consolidation.[19–20] An increase in that score during an SBT that the patient otherwise passes predicts respiratory distress after the tube is removed.[19] Diaphragm function and lung aeration can therefore fail separately, and a study that reports only one of them leaves the other mechanism unmeasured.

Cut-offs are not interchangeable across studies. A threshold chosen to predict failure of the SBT will not match a threshold chosen to predict reintubation among patients who have already passed the trial. Few reports give the full two-by-two table for both readings of the same cut-off, even though sensitivity for failure is specificity for success, and the predictive values change with which event is called positive. We conducted a prospective diagnostic accuracy study, with the extubation decision blinded to ultrasound, in adults who had already met institutional readiness criteria. The index tests were DTF, DE, RSBI, and a twelve-zone lung aeration score. The reference standard was extubation failure within 48 hours.

MATERIALS AND METHODS

Design, setting, and ethics: This was a prospective observational diagnostic accuracy study in the intensive critical care unit of the Department of Anaesthesia, SSNK Memorial Medical College and Hospital, Kopargaon, Ahilyanagar (Ahmednagar), Maharashtra, India. Patients were enrolled from August 2025 through February 2026. The report follows the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015.[21] The Institutional Ethics Committee approved the protocol before enrolment (approval number: to be inserted). Written informed consent was obtained from the legally authorised representative of each patient. The study received no external funding. Protocol registration: [insert Clinical Trials Registry–India number, if applicable].

Participants and sample size

The sample-size calculation used standard methods for a single receiver operating characteristic curve.[22] Against a null area under the curve of 0.50, an expected area of at least 0.85 for DTF, an extubation-failure incidence of 12% to 15%, a two-sided alpha of 0.05, and power of 80% indicated a minimum of 50 patients. During the study window, 58 consecutive adults were screened when the clinical team judged them potentially ready to wean. Eight were excluded, and 50 were enrolled and analysed. [Figure 1]

Inclusion required age of 18 years or older, invasive ventilation through an endotracheal tube for more than 48 hours, resolution or clear improvement of the condition that had caused acute respiratory failure, and fulfilment of every institutional readiness criterion [Table 1, panel A] before the first planned SBT.

Exclusion criteria were (1) pre-existing neuromuscular disease, including myasthenia gravis, Guillain–Barré syndrome, and motor neuron disease; (2) known phrenic-nerve injury or hemidiaphragmatic paralysis; (3) tracheostomy before the weaning assessment; (4) recent

thoracotomy, chest drain, or upper abdominal dressing that prevented an adequate right subcostal or intercostal window; (5) haemodynamic instability needing escalation of vasopressors; (6) pregnancy; and (7) refusal of consent.

Table 1: Institutional readiness-to-wean criteria and the twelve-zone lung aeration score.

A. Criteria required before the spontaneous breathing trial

Arterial blood gas	Respiratory rate and mechanics	Clinical and metabolic stability
PaO2 > 60 mmHg	Respiratory rate < 35 breaths/min	Tidal volume > 5 mL/kg predicted body weight
PaCO2 < 50 mmHg (or the patient's baseline)	RSBI < 105 breaths/min/L	Vital capacity > 10 mL/kg
FiO2 < 0.50	Minute ventilation 4–10 L/min	Afebrile (temperature < 38.0 °C)
PaO2/FiO2 ratio > 150 mmHg	Maximal inspiratory pressure ≤ −20 cmH2O	No important electrolyte or acid–base disturbance
PEEP ≤ 5 cmH2O	Intact cough and minimal secretions	Haemodynamically stable (no, or low-dose, vasopressor)

PaO2, partial pressure of arterial oxygen; PaCO2, partial pressure of arterial carbon dioxide; FiO2, fraction of inspired oxygen; PEEP, positive end-expiratory pressure; RSBI, rapid shallow breathing index. Maximal inspiratory pressure ≤ −20 cmH2O denotes a value at least as negative as −20 cmH2O.

B. Score applied to each of 12 thoracic zones (total 0–36)

Points	Aeration	Ultrasonographic pattern	Profile
0	Normal	Horizontal A-lines with lung sliding, or ≤ 2 isolated B-lines	A
1	Moderate loss	≥ 3 well-spaced, regular or irregular B-lines	B1
2	Severe loss	Coalescent B-lines (white lung)	B2
3	Complete loss	Tissue-like consolidation with dynamic or static air bronchograms	C

The global score is the sum of the worst pattern in each of six zones per hemithorax (upper and lower anterior, lateral, and posterior), bounded by the parasternal, anterior axillary, and posterior axillary lines.

Spontaneous breathing trial and ultrasound

Patients who met all readiness criteria underwent one standardised SBT of 30 to 60 minutes, either on a T-piece or on pressure support of 8 cmH2O or less, with positive end-expiratory pressure of 5 cmH2O or less and inspired oxygen fraction of 0.40 or less. At 30 minutes the RSBI was recorded as respiratory frequency divided by tidal volume in litres. Ultrasound was performed at the same time, with the patient semi-recumbent (head of bed 30° to 45°), by one anaesthesiologist trained in critical-care ultrasound. Images were acquired with the patient breathing quietly. The technique followed described windows for excursion and for thickness in the zone of apposition.[12,23] Diaphragm excursion. A 2–5 MHz curvilinear transducer was placed on the right subcostal margin, between the mid-clavicular and anterior axillary lines, using the liver as the window. The beam was aimed medially, cephalad, and dorsally so that it met the posterior third of the right hemidiaphragm at a near-right angle. The diaphragm was identified in B-mode as the bright line on the liver surface. M-mode then displayed the respiratory trace. Excursion, in centimetres, was the vertical distance from the end-expiratory baseline to the inspiratory peak, averaged over three consecutive quiet breaths. Diaphragm thickening fraction. A 7–12 MHz linear transducer was held perpendicular to the chest wall in the zone of apposition, in the 8th to 10th intercostal space on the right mid-axillary or anterior axillary line, 0.5 to 2 cm below the costophrenic recess. At a depth of 1.5 to 3 cm the diaphragm is a three-layer structure: a hypoechoic muscle between the echogenic pleural line and the echogenic peritoneal line. In M-mode, thickness was measured from the inner pleural line to the inner peritoneal line at end-inspiration (TEI), when the muscle is thickest, and at end-expiration (TEE), when it is thinnest. DTF was calculated as DTF (%) = [(TEI − TEE) / TEE] × 100. Lung aeration score. A 2–5 MHz convex transducer was used in B-mode to examine 12 zones, six on each side. Each zone received the score of its worst pattern [Table 1, panel B]. The global score ranges from 0, normal aeration, to 36, complete loss of aeration. The four index tests were therefore RSBI, DE, DTF, and the lung aeration score, all obtained at the 30-minute mark of the SBT and before the extubation decision. No index result was revised after the outcome was known. The reference standard was not applied at the time of ultrasound.

Reference standard and blinding

After the ultrasound examination and the remainder of the SBT, the attending intensive care team decided whether to extubate. That team was blinded to DE, DTF, and the lung aeration score. RSBI formed part of the institutional readiness screen [Table 1] and was therefore known, as a pass–fail criterion below 105 breaths/min/L, before enrolment; the numeric RSBI used in the accuracy analysis was the value recorded at 30 minutes of the SBT. Post-extubation non-invasive ventilation or high-flow oxygen was used when the institutional protocol indicated it. The primary reference standard was extubation failure: inability to sustain spontaneous breathing, requiring reintubation, or death, within 48 hours of extubation. Extubation success was freedom from reintubation for at least 48 hours. All enrolled patients were

extubated and were followed to that time point. [Figure 1]

Statistical analysis

Data were entered in Microsoft Excel and analysed in IBM SPSS Statistics, version 26.0 (IBM Corp., Armonk, NY, USA). Distribution was examined with the Kolmogorov–Smirnov and Shapiro–Wilk tests. Continuous variables are reported as mean ± standard deviation and were compared with the two-sample Student t test on 48 degrees of freedom. Categorical variables are reported as number and percentage and were compared with the chi-square test, or with Fisher's exact test when an expected cell count was below 5. A receiver operating characteristic curve was constructed for each index test against extubation failure.[24] The area under the curve is reported with its 95% confidence interval. The cut-off was the value that maximised the Youden index, J = sensitivity + specificity − 1.[25] Because the same threshold can be read as a test for failure or as a test for success, both orientations are reported from the same two-by-two counts. For a failure-oriented reading, a true positive is a patient who failed and was classified as high risk. For a success-oriented reading, a true positive is a patient who succeeded and was classified as low risk. Sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios, and overall accuracy (true positives plus true negatives, divided by 50) were calculated from those counts. A two-sided p value below 0.05 was regarded as statistically significant. No pairwise comparison of correlated areas under the curve was prespecified, and none was performed. Cut-offs were derived in the same patients in whom accuracy was estimated.

RESULTS

Study population

Of 58 patients screened, 8 met exclusion criteria and 50 were enrolled [Figure 1] All 50 were extubated after the SBT. Forty-four (88.0%) remained extubated for 48 hours, and 6 (12.0%) were reintubated within 48 hours. The six failures were reintubations. Prophylactic non-invasive ventilation was used after extubation in 34 of 50 patients (68.0%). Age did not differ significantly between patients whose extubation failed and those whose extubation succeeded (65.5 ± 7.4 vs 59.8 ± 10.2 years; t = 1.317; p = 0.194). Men predominated in both groups: 30 of 44 (68.2%) in the success group and 5 of 6 (83.3%) in the failure group (Fisher's exact test, p = 0.654). [Table 2]

The clinical course differed. Ventilation before extubation was longer when extubation later failed (7.8 ± 3.2 vs 4.4 ± 1.6 days; t = 4.262; p < 0.001). Subsequent intensive-care stay (11.5 ± 4.6 vs 6.4 ± 3.5 days; t = 3.228; p = 0.002) and hospital stay (15.3 ± 2.4 vs 10.2 ± 3.8 days; t = 3.185; p = 0.003) were also longer. Stay after extubation is a consequence of the outcome as well as a marker of a more severe preceding illness; it was not treated as a predictor.

Weaning mechanics and ultrasound

At 30 minutes of the SBT, every index test differed between the outcome groups (Table 2). RSBI was higher when extubation failed (81.2 ± 9.4 vs 52.4 ± 8.6 breaths/min/L; t = 7.618; p < 0.001). The lung aeration score was higher (14.8 ± 2.5 vs 8.5 ± 3.6; t = 4.133; p < 0.001). Right hemidiaphragm excursion was lower (1.18 ± 0.21 vs 1.65 ± 0.24 cm; t = 4.555; p < 0.001), and DTF was lower (25.4 ± 4.5% vs 42.8 ± 6.4%; t = 6.419; p < 0.001).

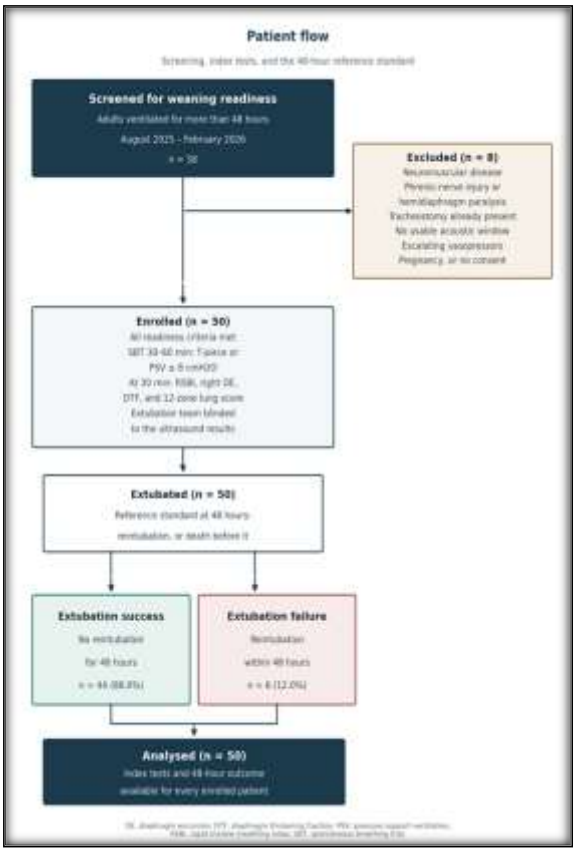


Figure 1: Flow of patients from screening to the 48-hour extubation outcome. Index ultrasound was performed at 30 minutes of the spontaneous breathing trial. The extubation decision was blinded to diaphragm and lung ultrasound measurements

Table 2: Demographic, clinical, and ultrasonographic variables by extubation outcome (N = 50)

Variable	Success (n = 44)	Failure (n = 6)	Statistic	p value
Age, years	59.8 ± 10.2	65.5 ± 7.4	t = 1.317	0.194
Women, n (%)	14 (31.8)	1 (16.7)	Fisher exact	0.654
Men, n (%)	30 (68.2)	5 (83.3)	—	—
Ventilation before extubation, days	4.4 ± 1.6	7.8 ± 3.2	t = 4.262	<0.001
Intensive-care stay, days	6.4 ± 3.5	11.5 ± 4.6	t = 3.228	0.002
Hospital stay, days	10.2 ± 3.8	15.3 ± 2.4	t = 3.185	0.003
RSBI, breaths/min/L	52.4 ± 8.6	81.2 ± 9.4	t = 7.618	<0.001

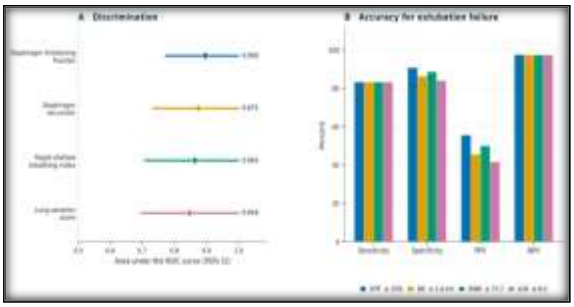
Lung aeration score	8.5 ± 3.6	14.8 ± 2.5	t = 4.133	<0.001
Diaphragm excursion, cm	1.65 ± 0.24	1.18 ± 0.21	t = 4.555	<0.001
Diaphragm thickening fraction, %	42.8 ± 6.4	25.4 ± 4.5	t = 6.419	<0.001

Values are mean ± standard deviation unless noted. Continuous variables were compared by the two-sample t test (df = 48). Sex was compared by Fisher's exact test. RSBI, rapid shallow breathing index. A p value below 0.05 was taken as statistically significant.

Diagnostic accuracy

Discrimination is summarised in [Table 3 and Figure 2]. DTF had the highest area under the curve, 0.898 (95% CI 0.772–1.000; p = 0.002). At the Youden cut-off of 32%, a DTF of 32% or less predicted extubation failure with sensitivity 83.3% (5/6), specificity 90.9% (40/44), positive predictive value 55.6% (5/9), negative predictive value 97.6% (40/41), positive likelihood ratio 9.17, negative likelihood ratio 0.18, and Youden index 0.74. Overall accuracy was 90.0% (45/50). Read the other way, DTF above 32% predicted successful extubation with sensitivity 90.9%, specificity 83.3%, positive predictive value 97.6%, and negative predictive value 55.6%. DE of 1.4 cm or less predicted failure with area under the curve 0.875 (95% CI 0.735–1.000; p = 0.003), sensitivity 83.3% (5/6), specificity 86.4% (38/44), positive predictive value 45.5% (5/11), negative predictive value 97.4% (38/39), positive likelihood ratio 6.11, negative likelihood ratio 0.19, and Youden index 0.70. Accuracy was 86.0% (43/50). RSBI of 77.7 breaths/min/L or higher predicted failure with area under the curve 0.864 (95% CI 0.708–1.000; p = 0.004), sensitivity 83.3% (5/6), specificity 88.6% (39/44), positive predictive value 50.0% (5/10), negative predictive value 97.5% (39/40), positive likelihood ratio 7.33, negative likelihood ratio 0.19, and Youden index 0.72. Accuracy was 88.0% (44/50). A lung aeration score of 8.2 or higher predicted failure with area under the curve 0.848 (95% CI 0.695–1.000; p = 0.006), sensitivity 83.3% (5/6), specificity 84.1% (37/44), positive predictive value 41.7% (5/12), negative predictive value 97.4% (37/38), positive likelihood ratio 5.24, negative likelihood ratio 0.20, and Youden index 0.67. Accuracy was 84.0% (42/50). The score is an integer, so a cut-off of 8.2 classifies a score of 9 or more as the high-risk result. Sensitivity was 83.3% for every index because each correctly classified five of the six failures.

The four 95% confidence intervals for the area under the curve all include values well below 0.90 and all reach 1.000. The point estimates favour DTF, then DE, then RSBI, then the lung aeration score, but this order was not tested by a comparison of correlated curves, and it should not be read as a demonstrated ranking.



[Figure 2] Receiver operating characteristic analysis of the four index tests (N = 50). Panel A shows the area under the curve and 95% confidence interval for prediction of extubation failure. Panel B shows sensitivity, specificity, and predictive values at the Youden cut-off, with extubation failure as the positive class. Numerical values are those in [Table 3].

Table 3: Diagnostic accuracy of the four index tests for extubation failure and, at the same cut-off, for extubation success (N = 50)

Panel A takes extubation failure as the positive class (n = 6). Panel B takes extubation success as the positive class (n = 44). Counts, predictive values, and likelihood ratios change with that choice. The area under the curve does not.

Index test	Cut-off	TP, FP, FN, TN	Sens. %	Spec. %	PPV %	NPV %	LR+	LR−	Youden	AUC (95% CI)	p
A. Positive class = extubation failure											
RSBI, breaths/min/L	≥ 77.7	5, 5, 1, 39	83.3	88.6	50.0	97.5	7.33	0.19	0.72	0.864 (0.708–1.000)	0.004
Lung aeration score	≥ 8.2	5, 7, 1, 37	83.3	84.1	41.7	97.4	5.24	0.20	0.67	0.848 (0.695–1.000)	0.006
Diaphragm excursion, cm	≤ 1.4	5, 6, 1, 38	83.3	86.4	45.5	97.4	6.11	0.19	0.70	0.875 (0.735–1.000)	0.003
Thickening fraction, %	≤ 32.0	5, 4, 1, 40	83.3	90.9	55.6	97.6	9.17	0.18	0.74	0.898 (0.772–1.000)	0.002
B. Positive class = extubation success											
RSBI, breaths/min/L	< 77.7	39, 1, 5, 5	88.6	83.3	97.5	50.0	5.32	0.14	0.72	0.864 (0.708–1.000)	0.004
Lung aeration score	< 8.2	37, 1, 7, 5	84.1	83.3	97.4	41.7	5.05	0.19	0.67	0.848 (0.695–1.000)	0.006
Diaphragm excursion, cm	> 1.4	38, 1, 6, 5	86.4	83.3	97.4	45.5	5.18	0.16	0.70	0.875 (0.735–1.000)	0.003
Thickening fraction, %	> 32.0	40, 1, 4, 5	90.9	83.3	97.6	55.6	5.45	0.11	0.74	0.898 (0.772–1.000)	0.002

TP, true positive; FP, false positive; FN, false negative; TN, true negative; Sens., sensitivity; Spec., specificity; PPV, positive predictive value; NPV, negative predictive value; LR+ and LR−, positive and negative likelihood ratios, calculated from the counts as sensitivity/(1 − specificity) and (1 − sensitivity)/specificity; AUC, area under the receiver operating characteristic curve; CI, confidence interval; RSBI, rapid shallow breathing index. Overall accuracy, (TP + TN)/50, was 88.0% for RSBI, 84.0% for the lung aeration score, 86.0% for excursion, and 90.0% for thickening fraction. All four areas under the curve differed from 0.50 at the p values shown. Likelihood ratios are reported to two decimal places from the exact counts.

DTF, diaphragm thickening fraction; DE, diaphragm excursion; RSBI, rapid shallow breathing index; LUS, lung aeration score; PPV, positive predictive value; NPV, negative predictive value. Bars in panel A are 95% confidence intervals. Every upper limit was 1.000, and the four intervals overlap. Cut-offs in the legend are those for predicting extubation failure.

DISCUSSION

In 50 adults who had already satisfied a structured readiness screen, extubation failed within 48 hours in 12%. Diaphragm thickening fraction, measured at 30 minutes of the spontaneous breathing trial, had the highest point estimate of discrimination. A thickening fraction of 32% or less identified five of the six failures and mislabelled four of the 44 successes. Excursion, RSBI, and the lung aeration score separated the groups in the expected direction, with areas under the curve of 0.875, 0.864, and 0.848. Those intervals overlap the interval for thickening fraction, and no formal comparison of the curves was done. The practical signal is therefore not that thickening fraction has been proved superior, but that a low thickening fraction, a small excursion, a high RSBI inside the range still allowed by the readiness screen, and a higher aeration score all marked the patients who were reintubated.

The 32% threshold sits close to published thickening-fraction cut-offs, with an important difference in the reference standard. DiNino and colleagues reported that thickening of at least 30% predicted extubation success with sensitivity 88%, specificity 71%, and an area under the curve of 0.79.13 Ferrari and colleagues used a threshold above 36%, but the outcome was success of the breathing trial rather than freedom from reintubation.[14] Farghaly and Hasan, studying patients who had already passed an SBT, found that a thickening fraction of at least 34.2% predicted successful extubation with sensitivity 90% and specificity 64.3%.[15] Their failure rate was 25.9%, about twice ours, so the predictive values are not directly comparable. Blumhof and colleagues found that thickening above 20%, measured on modest pressure support, predicted extubation success, and that the same measurement lost accuracy at higher levels of support.[16] The number attached to a “positive” thickening fraction therefore moves with the level of assistance and with whether the event of interest is trial failure or reintubation. Our patients were assessed on a T-piece or on pressure support of 8 cmH2O or less, and the event was reintubation within 48 hours. Against that background, a summary area under the curve of about 0.87 for thickening fraction in the review by Llamas-Álvarez and colleagues is consistent with our point estimate of 0.898.[18] Consistency of the point estimate should not be mistaken for precision: our confidence interval runs from 0.772 to 1.000.

Excursion of 1.4 cm or less was the second index by point estimate. It is a displacement, not a normalised change, and it varies with the M-mode angle, abdominal pressure, posture, and how hard the patient is trying.[12] Farghaly and Hasan reported a success threshold of 1.05 cm or more.[15] Kim and colleagues defined ultrasonographic diaphragm dysfunction as excursion below 1.0 cm, or as paradoxical motion, and found more weaning failure in patients who met that definition.[17] The windows used here are those for which reproducibility and normal values have been described.[23] Thickening fraction, being a ratio of two thicknesses along the same cursor, is less exposed to some of those errors, which is a reason to prefer it when only one diaphragm measurement can be made. We recorded only the right hemidiaphragm, the side with the better hepatic window. Unilateral left dysfunction would have been missed.

The RSBI result needs to be read in light of how patients entered the study. Every enrolled patient already had an RSBI below 105 breaths/min/L, because that was a readiness criterion. Inside a population truncated by the classical Yang and Tobin threshold,[5] the cut-off that best separated later extubation failure was 77.7 breaths/min/L. This does not replace 105 as a screen for attempting a breathing trial. It says that, after patients have passed that screen, a value still below 105 can sit high enough to mark risk. Specificity of the classical threshold is expected to look poor, or even unevaluable, once patients above 105 have been kept out. Reporting our Youden cut-off as if it were a new universal RSBI threshold would overstate what a pre-screened cohort can show.

The lung aeration score was higher when extubation failed (mean 14.8 vs 8.5). Soummer and colleagues found that the score at the end of a successful breathing trial predicted post-extubation distress, with an area under the curve of 0.86, and that the score rose during the trial only in patients who later developed distress.[19] Their absolute scores were higher than ours, and they analysed the change as well as the final value. We made a single measurement at 30 minutes and did not record a pre-trial score, so we cannot say whether aeration was worsening during the trial. A Youden cut-off of 8.2 on an integer scale is the boundary between a score of 8 and a score of 9. It should be applied as “9 or higher”, not as a fractional score. Lung ultrasound and diaphragm ultrasound are not substitutes. One

describes the parenchyma; the other describes the pump. With six failure events we did not fit a combined model, and a combined model derived here would not have been trustworthy.

Predictive values, rather than sensitivity, should guide the bedside reading. The pretest probability of failure in this cohort was 12%. A positive likelihood ratio of 9.17, the largest we observed, raised that probability only to 56%: of nine patients with thickening fraction at or below 32%, five failed and four succeeded. A negative likelihood ratio of 0.18 lowered the probability of failure to about 2%, which is the negative predictive value of 97.6%. The safer use of a high thickening fraction is reassurance, not a guarantee. The safer use of a low thickening fraction is a reason to pause: to review secretions, fluid balance, cardiac function, and cough, and to consider a further period of ventilation or a planned period of non-invasive ventilation, rather than an automatic decision against extubation. The same restraint applies to excursion of 1.4 cm or less, an RSBI of 77.7 or higher, and an aeration score of 9 or higher. Ultrasound did not, in this protocol, override a clinical team that was kept unaware of it.

Patients who failed had been ventilated for nearly twice as long before the breathing trial. That association is biologically plausible. Diaphragm thickness can change within the first week of ventilation, and low inspiratory effort is one driver.[9] It is also confounded. A longer ventilation course marks a more severe illness, more sedation, and more weakness of other muscles, any of which can cause reintubation with or without a low thickening fraction. We did not adjust the ultrasound comparisons for days of ventilation or for diagnosis. Longer intensive-care and hospital stays in the failure group should be read as part of the outcome, not as predictors available at the moment of extubation.

The design has several protections against bias in a diagnostic-accuracy study. Enrolment was consecutive among patients who reached a weaning assessment. Readiness criteria were specified. Index tests were obtained during the breathing trial, not on full ventilatory support, when thickening is difficult to interpret.[16] The clinicians who extubated the patient did not know the ultrasound results, so the reference standard was not assigned with knowledge of DTF, DE, or the aeration score. Every enrolled patient was extubated and was classified at 48 hours, so there is no partial verification in the analysed set. Both orientations of each cut-off are shown, with the counts from which they come.

The limitations are equally specific. The study was done in one intensive care unit, by one operator, and inter-observer and intra-observer variability were not measured. Only the right hemidiaphragm was imaged. Six failures mean that sensitivity moves in steps of one patient in six, or 16.7 percentage points, and every index shared the same sensitivity because each missed one failure. The confidence intervals of the areas under the curve are wide, reach 1.0, and overlap; a test of correlated areas was not performed.[26] Student t tests in a group of six are sensitive to single observations even when a normality test has been applied. Admission diagnoses, cough strength beyond a binary “intact” judgement, delirium, fluid balance, and cardiac function were not reported, although each is a recognised cause of extubation failure in a patient whose diaphragm thickens normally.[1,3] Prophylactic non-invasive ventilation was used in 34 of 50 patients and may have prevented some reintubations, which would shrink the failure group and could bias accuracy either way. The sample size was calculated to distinguish an area under the curve of 0.85 from 0.50, not to compare thickening fraction with RSBI and not to put a narrow interval around sensitivity. Finally, the cut-offs were chosen and then evaluated in the same 50 patients. Accuracy estimated that way is biased upward. The 32% threshold is a hypothesis for the next study, not a rule that has been validated.

A useful next study would prespecify thickening fraction of 32% as the threshold, enrol enough patients to observe several dozen extubation failures, use at least two blinded readers, record both hemidiaphragms, and state in advance how post-extubation non-invasive ventilation will be handled in the reference standard. Until that confirmation exists, diaphragm ultrasound during the breathing trial is best treated as one more piece of physiology at the bedside: reassuring when thickening is above 32%, and a reason for a second look when it is not.

CONCLUSION

In mechanically ventilated adults who already met readiness-to-wean criteria, diaphragm thickening fraction during the spontaneous breathing trial had the highest point estimate of accuracy for extubation failure within 48 hours, ahead of diaphragm excursion, the rapid shallow breathing index, and a twelve-zone lung aeration score. Thickening above 32% was associated with a high negative predictive value for reintubation. Positive predictive values were only moderate, the confidence intervals were wide, and the cut-offs were not externally validated. The ultrasound findings should be added to clinical judgement, not substituted for it.

Acknowledgements

The authors thank the nursing staff of the intensive critical care unit and the attending team, who remained blinded to the diaphragm and lung ultrasound measurements while making extubation decisions.

REFERENCES

1. Thille AW, Richard JCM, Brochard L. The decision to extubate in the intensive care unit. Am J Respir Crit Care Med. 2013;187(12):1294-1302.
2. Epstein SK, Ciubotaru RL, Wong JB. Effect of failed extubation on the outcome of mechanical ventilation. Chest. 1997;112(1):186-192.
3. Thille AW, Harrois A, Schortgen F, Brun-Buisson C, Brochard L. Outcomes of extubation failure in medical

intensive care unit patients. Crit Care Med. 2011;39(12):2612-2618.
4. Boles JM, Bion J, Connors A, Herridge M, Marsh B, Melot C, et al. Weaning from mechanical ventilation. Eur Respir J. 2007;29(5):1033-1056.
5. Yang KL, Tobin MJ. A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. N Engl J Med. 1991;324(21):1445-1450.
6. MacIntyre NR, Cook DJ, Ely EW Jr, Epstein SK, Fink JB, Heffner JE, et al. Evidence-based guidelines for weaning and discontinuing ventilatory support. Chest. 2001;120(6 Suppl):375S-395S.
7. Levine S, Nguyen T, Taylor N, Friscia ME, Budak MT, Rothenberg P, et al. Rapid disuse atrophy of diaphragm fibers in mechanically ventilated humans. N Engl J Med. 2008;358(13):1327-1335.
8. Jaber S, Petrof BJ, Jung B, Chanques G, Berthet JP, Rabuel C, et al. Rapidly progressive diaphragmatic weakness and injury during mechanical ventilation in humans. Am J Respir Crit Care Med. 2011;183(3):364-371.
9. Goligher EC, Fan E, Herridge MS, Murray A, Vorona S, Brace D, et al. Evolution of diaphragm thickness during mechanical ventilation. Impact of inspiratory effort. Am J Respir Crit Care Med. 2015;192(9):1080-1088.
10. Dres M, Goligher EC, Heunks LMA, Brochard LJ. Critical illness-associated diaphragm weakness. Intensive Care Med. 2017;43(10):1441-1452.
11. Jung B, Moury PH, Mahul M, de Jong A, Galia F, Prades A, et al. Diaphragmatic dysfunction in patients with ICU-acquired weakness and its impact on extubation failure. Intensive Care Med. 2016;42(5):853-861.
12. Matamis D, Soilemezi E, Tsagourias M, Akoumianaki E, Dimassi S, Boroli F, et al. Sonographic evaluation of the diaphragm in critically ill patients. Technique and clinical applications. Intensive Care Med. 2013;39(5):801-810.
13. DiNino E, Gartman EJ, Sethi JM, McCool FD. Diaphragm ultrasound as a predictor of successful extubation from mechanical ventilation. Thorax. 2014;69(5):431-435.
14. Ferrari G, de Filippi G, Elia F, Panero F, Volpicelli G, Aprà F. Diaphragm ultrasound as a new index of discontinuation from mechanical ventilation. Crit Ultrasound J. 2014;6(1):8.
15. Farghaly S, Hasan AA. Diaphragm ultrasound as a new method to predict extubation outcome in mechanically ventilated patients. Aust Crit Care. 2017;30(1):37-43.
16. Blumhof S, Wheeler D, Thomas K, McCool FD, Mora J. Change in diaphragmatic thickness during the respiratory cycle predicts extubation success at various levels of pressure support ventilation. Lung. 2016;194(4):519-525.
17. Kim WY, Suh HJ, Hong SB, Koh Y, Lim CM. Diaphragm dysfunction assessed by ultrasonography: influence on weaning from mechanical ventilation. Crit Care Med. 2011;39(12):2627-2630.
18. Llamas-Álvarez AM, Tenza-Lozano EM, Latour-Pérez J. Diaphragm and lung ultrasound to predict weaning outcome: systematic review and meta-analysis. Chest. 2017;152(6):1140-1150.
19. Soummer A, Perbet S, Brisson H, Arbelot C, Constantin JM, Lu Q, et al. Ultrasound assessment of lung aeration loss during a successful weaning trial predicts postextubation distress. Crit Care Med. 2012;40(7):2064-2072.
20. Bouhemad B, Mongodi S, Via G, Rouquette I. Ultrasound for "lung monitoring" of ventilated patients. Anesthesiology. 2015;122(2):437-447.
21. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527.
22. Obuchowski NA. Sample size calculations in studies of test accuracy. Stat Methods Med Res. 1998;7(4):371-392.
23. Boussuges A, Gole Y, Blanc P. Diaphragmatic motion studied by M-mode ultrasonography: methods, reproducibility, and normal values. Chest. 2009;135(2):391-400.
24. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.
25. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-35.
26. 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(3):837-845.