[Original Research]

CHEST

Diaphragm and Lung Ultrasound to Predict Weaning Outcome

Systematic Review and Meta-Analysis

Ana M. Llamas-Álvarez, MD; Eva M. Tenza-Lozano, MD, PhD; and Jaime Latour-Pérez, MD, PhD

BACKGROUND: Deciding the optimal timing for extubation in patients who are mechanically ventilated can be challenging, and traditional weaning predictor tools are not very accurate. The aim of this systematic review and meta-analysis was to assess the accuracy of lung and diaphragm ultrasound for predicting weaning outcomes in critically ill adults.

METHODS: MEDLINE, the Cochrane Library, Web of Science, Scopus, LILACS, Teseo, Tesis Doctorales en Red, and OpenGrey were searched, and the bibliographies of relevant studies were reviewed. Two researchers independently selected studies that met the inclusion criteria and assessed study quality in accordance with the Quality Assessment of Diagnostic Accuracy Studies-2 tool. The summary receiver-operating characteristic curve and pooled diagnostic OR (DOR) were estimated by using a bivariate random effects analysis. Sources of heterogeneity were explored by using predefined subgroup analyses and bivariate meta-regression.

RESULTS: Nineteen studies involving 1,071 people were included in the study. For diaphragm thickening fraction, the area under the summary receiver-operating characteristic curve was 0.87, and DOR was 21 (95% CI, 11-40). Regarding diaphragmatic excursion, pooled sensitivity was 75% (95% CI, 65-85); pooled specificity, 75% (95% CI, 60-85); and DOR, 10 (95% CI, 4-24). For lung ultrasound, the area under the summary receiver-operating characteristic curve was 0.77, and DOR was 38 (95% CI, 7-198). Based on bivariate meta-regression analysis, a significantly higher specificity for diaphragm thickening fraction and higher sensitivity for diaphragmatic excursion was detected in studies with applicability concerns.

CONCLUSIONS: Lung and diaphragm ultrasound can help predict weaning outcome, but its accuracy may vary depending on the patient subpopulation. CHEST 2017; ■(■):■-■

KEY WORDS: diaphragm ultrasound; extubation; lung ultrasound; meta-analysis; weaning

ABBREVIATIONS: AUSROC = area under the summary receiver-operating characteristic curve; DE = diaphragmatic excursion; DTF = diaphragm thickening fraction; LUS = lung ultrasound; MV = mechanical ventilation; QUADAS-2 = Quality Assessment of Diagnostic Accuracy Studies-2; ROC = receiver-operating characteristic; SBT = spontaneous breathing trial; SROC = summary receiver-operating characteristic

AFFILIATIONS: From the Intensive Care Unit (Drs Llamas-Álvarez, Tenza-Lozano, and Latour-Pérez), Elche General University Hospital, Elche, Spain; and the Department of Clinical Medicine (Dr Latour-Pérez), Miguel Hernández University, Sant Joan d'Alacant, Spain.

FUNDING/SUPPORT: This study was supported by the Department of Clinical Medicine, Miguel Hernández University. The Department of Clinical Medicine of Miguel Hernández University provided nonfinancial support supplying translation assistance.

CORRESPONDENCE TO: Ana M. Llamas-Álvarez, MD, Department of Intensive Care Medicine, Elche General University Hospital, Camino de la Almazara 11, 03203, Elche, Spain; e-mail: llamasalvarez@yahoo.es

DOI: https://doi.org/10.1016/j.chest.2017.08.028

Determining the optimal moment to extubate a critically ill patient remains a challenge, as premature removal of mechanical ventilation (MV) entails a high risk of weaning failure, prompting reintubation that exposes the patient to unnecessary hemodynamic and respiratory stress.[1] Conversely, delayed extubation increases the duration of MV and carries other risks (eg, development of ventilator-associated pneumonia, tracheal injury, barotrauma).[2] Thus, both early and delayed weaning are associated with increased mortality, stay in the ICU, and economic cost.[3]

Current guidelines for weaning[4,5] recommend the implementation of a spontaneous breathing trial (SBT) as a tool to predict weaning outcome. However, 13% to 26% of patients who are extubated following a successful SBT need to be reintubated within 48 h.[6,7] In the last few years, multiple indices and parameters have been proposed as predictors of weaning outcome, but none has shown more than modest prognostic accuracy.[8,9]

Successful liberation from MV depends on several factors, as patients must be hemodynamically stable, have an adequate ventilation-perfusion ratio, have the ability to generate a strong cough and expectorate endotracheal secretions, and generate a reliable ventilator pattern. Diverse mechanisms can influence the development of respiratory distress after extubation, such as the decrease of aeration of the pulmonary parenchyma related to heart failure induced by SBT,[10] alterations in pulmonary compliance,[11,12] and diaphragmatic dysfunction associated with MV.[13]

Ultrasound use in the ICU is an area of growing interest[14] because of its portability, speed, safety, and the encouraging results obtained for managing multiple entities.[15,16] Because ultrasound provides both morphologic and functional information in real time, it may be useful for assessing two important factors among those that can influence weaning: the status of aeration of the pulmonary parenchyma[11] and the functional status of the diaphragm,[17] eliciting clues on the probability of success when removing MV.

There are two proposed diaphragm sonographic predictors: the diaphragmatic excursion (DE),[18] which measures the distance that the diaphragm is able to move during the respiratory cycle, and the diaphragm thickening fraction (DTF),[19] which reflects variation in the thickness of the diaphragm during respiratory effort and is calculated as (thickness at end-inspiration – thickness at the end-expiration)/thickness at the end of the expiration. Ultrasound can detect the decrease in the aeration of the lung parenchyma due to cardiac, respiratory, or diaphragmatic origin. This decrease can be quantified through the so-called lung ultrasound (LUS) score,[20] a validated scale whose values range from 0 to 36 points, obtained from the sum of the grades assigned to different ventilation patterns observed in every area of the lung scan.[21,22]

The goal of the present systematic review and meta-analysis was to assess the accuracy of the lung and diaphragm ultrasound, in particular the DE, DTF, and LUS score, for predicting MV weaning outcomes in critically ill adults.

Methods

Search Strategy and Selection Criteria

In this systematic review and meta-analysis, we included any studies in participants aged ≥ 18 years, admitted to an ICU, and subjected to invasive MV for at least 24 h; patients underwent lung and/or diaphragmatic ultrasound and had data available on weaning outcome. Weaning failure was defined broadly as the need for reintubation with reconnection to invasive MV, unscheduled postextubation noninvasive MV, tracheostomy, death within the first 72 h, or SBT failure. Because weaning success and failure are mutually exclusive, weaning success was defined as the absence of criteria for failure.

Our exclusion criteria were nonprimary studies, insufficient data to calculate a 2 × 2 table for sensitivity and specificity, studies with < 20 participants, and those in which the unit of analysis was not the patient (eg, chest or lung regions).

Two researchers (A. M. L.-A. and E. M. T.-L.) independently conducted a literature search to identify potentially relevant studies in MEDLINE, the Cochrane Library, the Web of Science, Scopus, and LILACS, as well as repositories of doctoral theses (Theseus and Tesis doctorales en red) and reviews of grey literature on Open Grey. We hand-searched the bibliographies of relevant studies and in some cases attempted to contact authors of conference proceedings to obtain unpublished data or other authors to request clarifications.

The search used the terms “extubation,” “weaning,” “discontinuation of mechanical ventilation,” “disconnect of mechanical ventilation,” “ultrasound,” “ultrasonography,” and “echography.” No restrictions were imposed on the date of publication, and all papers published from database inception to November 2016 were included. Only studies conducted in humans were eligible. Despite the fact that the present review focuses on adult patients, we did not apply an age filter to the search, intending instead to manually screen search results to increase comprehensiveness. Likewise, the search was not limited according to study design or language. Google Scholar Alerts was activated to receive notifications about the publication of potentially relevant studies between November 2016 and April 2017. Discrepancies between researchers were resolved through consensus with a third expert researcher (J. L.-P.).

The results are presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations.[23]

Data Analysis

Two researchers manually extracted the data necessary to build a 2 × 2 table, using raw data from each study on true-positive, true-negative, false-negative, and false-positive findings; reported sensitivity and specificity; and/or graphics. Given the dichotomous nature of the variable "weaning outcome" and to facilitate comparison of data, we decided to build only tables regarding weaning success. Unlike with DTF and DE, high LUS scores predict weaning failure, low scores predict success, and intermediate values indicate uncertainty; that is, different cutoff points are used for predicting success and failure, and thus data for both situations were extracted from the cutoff points provided.

Two researchers (A. M. L.-A. and E. M. T.-L.) independently evaluated methodologic quality using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool,[24] a validated method for assessing risk of bias across various domains as well as factors that may affect applicability. We developed a complementary protocol with operational definitions of risk of bias in each of the explored areas (e-Table 1). Discrepancies were resolved by consensus with a third researcher (J. L.-P.).

Data were analyzed according to the European Network for Health Technology Assessment recommendations,[25] using the Mada application in the R statistical package (version 3.3.2)[26] and Review Manager 5.3,[27] developed by the Cochrane Collaboration. In the univariate analysis, a forest plot was constructed for sensitivity and specificity, and the diagnostic OR was calculated. The correlation between sensitivity and the false-positive rate was explored graphically (forest plot) and statistically by examining the Spearman correlation coefficient and its 95% CI. Heterogeneity was analyzed among studies graphically and through the I^2 statistic and the Q test. Data were presented in the receiver-operating characteristic (ROC) plane and, depending on the existence of threshold effect according to graphics and statistical evidence, we calculated the area under the curve or a pooled estimation of sensitivity and specificity. The bivariate Reitsma model was used,[28] which in the absence of covariates is equivalent to the hierarchical summary ROC of Rutter and Gatsonis.[29] In case of detecting an outlier (ie, a study with values for sensitivity and/or specificity ostensibly different from the rest), a sensitivity analysis was performed excluding that study. Possible causes of heterogeneity among studies were examined through an analysis of predefined subgroups, attending to quality and applicability criteria by using a bivariate meta-regression.

Results

The initial database search yielded a total of 3,819 references, and four additional records were identified from other sources. Fourteen subsequent references were also obtained from Google Scholar Alerts. A total of 3,837 records were thus obtained. After removing duplicates and screening titles and abstracts, we considered the full text of 55 studies, of which 36 were excluded (reasons detailed in Fig 1). Therefore, 19 references met the inclusion criteria and were included in the qualitative and quantitative synthesis.

All the selected studies have a cohort design and were conducted between 2004 and 2017. Some studies measured more than one predictor: 10 studies assessed DE; 10, DTF; and five, LUS score. The definition of "weaning failure" is not standard, covering one or more of the following items in the first 48 to 72 h: need for reintubation, nonscheduled postextubation noninvasive MV, death, tracheostomy, terminal extubation, extubation delay, and/or SBT failure (Table 1) (unpublished data, Tenza-Lozano et al, 2017).[12,30-46] Most of the studies were conducted in polyvalent ICUs, but three studies[31,36,44] took place in the respiratory ICU, and one[38] was conducted in a high dependency unit. Ultrasounds were usually conducted on patients just before or during the SBT, but four studies[12,36,40,44] only considered those individuals who successfully passed this test (Table 2).

Regarding the assessment of methodologic quality (Fig 2), the main problem we detected was the presence of aspects that could compromise applicability due to patient selection. Different studies focused exclusively or primarily on patients with COPD or whose intubation was due to respiratory causes,[36,37,44] on those with ICU-acquired weakness,[41] or on patients who had failed previous attempts at weaning.[38,39] Some studies were at risk of selection bias because they excluded specific

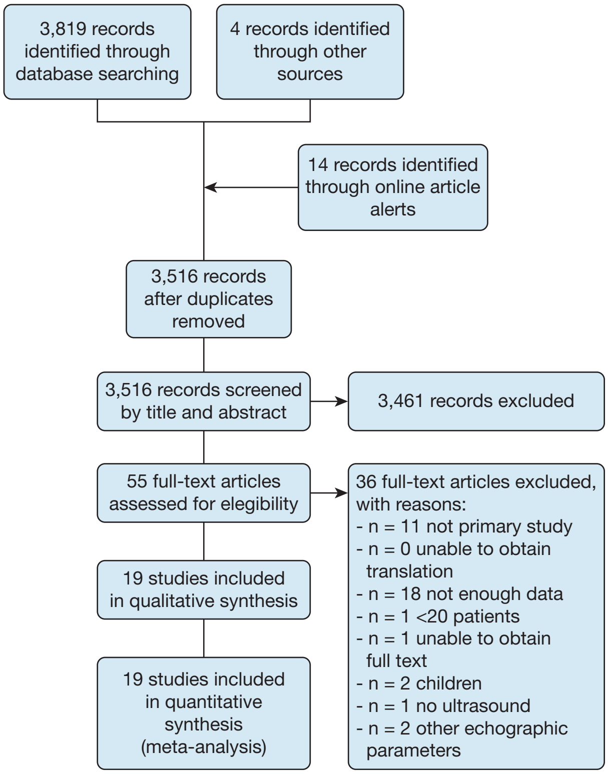

Figure 1 – *Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram for study identification and selection, with reasons for exclusion.*

TABLE 1] Characteristics of Included Studies

Study/Year	Country	Design	Measurements and Cutoff Value	Definition of Weaning Failure
Ali and Mohamad,[30] 2016	Egypt	Cohort	DTF > 30% DE > 1.5 cm	MV within 48 h of self-breathing
Baess et al,[31] 2016	Egypt	Cohort	DTF ≥ 30% DE > 1 cm	Reintubation or NIMV within 48 h
Binet et al,[32] 2014	France	Cohort	LUS score ≤ 14 (≥ 19 for failure)	Reintubation, nonscheduled NIVM, or death within 48 h
Blumhof et al,[33] 2016	United States	Cohort	DTF > 20%	Reintubation or delayed extubation (> 48 h)
Carrie et al,[34] 2017	France	Cohort	DE > 2.7 cm	SBT failure or the need for MV or death within 48 h following extubation
Dinino et al,[35] 2014	United States	Cohort	DTF ≥ 30%	Reintubation within 48 h or terminal extubation or tracheostomy
Farghaly and Hasan,[36] 2017	Egypt	Cohort	DTF ≥ 34.5% DE ≥ 1.5 cm	Inability to maintain spontaneous breathing without any ventilatory support within 48 h
Fayed et al,[37] 2016	Egypt	Cohort	DTF > 29%	Inability to maintain spontaneous breathing without any ventilatory support within 48 h
Ferrari et al,[38] 2014	Italy	Cohort	DTF > 36%	SBT failure
Flevari et al,[39] 2016	Greece	Cohort	DE ≥ 1 cm	Ventilatory support (noninvasive or invasive) within 48 h after a SBT
Jiang et al,[40] 2004	Taiwan	Cohort	DE ≥ 1.1 cm	Reintubation or NIMV within 72 h
Jung et al,[41] 2016	France	Cohort	DTF ≥ 30%	Reintubation within 72 h or tracheostomy
Kim et al,[42] 2011	Korea	Cohort	DE < 1.4 cm	MV within 48 h of self-breathing
Osman and Hashim,[43] 2017	Egypt	Cohort	DTF ≥ 28% DE ≥ 1 cm LUS score < 12	Reintubation within 48 h
Saeed et al,[44] 2016	Egypt	Cohort	DE > 1.1 cm	MV within 48 h
Shoaeir et al,[45] 2016	Egypt	Cohort	LUS score ≤ 10 (> 18 for failure)	Reintubation within 48 h
Soummer et al,[12] 2012	France	Cohort	LUS score < 13 (> 17 for failure)	Ventilatory support (either noninvasive or invasive ventilation) within 48 h after extubation
Spadaro et al,[46] 2016	Italy	Cohort	DE > 1.4 cm	SBT failure, reintubation, or NIMV within 48 h
Tenza-Lozano et al, unpublished data, 2017	Spain	Cohort	DTF > 24%	SBT failure, reintubation, or NIMV within 48 h

Measurements and cutoff values refer to weaning success prediction unless otherwise indicated. DE = diaphragmatic excursion; DTF = diaphragmatic thickness fraction; MV = mechanical ventilation; NIMV = noninvasive mechanical ventilation; LUS = lung ultrasound; SBT = spontaneous breathing trial.

populations of patients, such as those with severe ICU-acquired weakness,[12,45] or those with any admission to an ICU in the previous 12 months[30] (e-Table 2).

Overall, ultrasounds and their measurements were well conducted (e-Table 3), but one study[30] may have performed the procedure while patients were being mechanically ventilated, which would not be suitable for assessing patients' respiratory efforts.[47] In this review, there is no gold standard test for comparison, and we therefore considered the adequacy of the weaning failure definition made in each study to properly classify weaning outcome. In this sense, only one study[38] was considered to be at high risk of bias because its definition of weaning failure was just SBT failure, even though patients may require reintubation after a successful SBT. Generally, physicians responsible for deciding to withdraw MV were blinded to ultrasound results, although four studies were unclear on this point.[30,37,43,45] Concerning the flow of patients within each study, one study[30] was considered to be at high risk of bias because it excluded from the analysis patients

TABLE 2] Participant Characteristics

Study	No.	Setting	Age (y)[a]	Inclusion
Ali and Mohamad[30]	54	MD ICU	54 ± 11.23	Not specified
Baess et al[31]	30	General and respiratory ICU	59.17 ± 13.17	Patients who were planned for weaning
Binet et al[32]	48	MD ICU	59 ± 16	Ventilated > 48h
Blumhof et al[33]	52	ICU	62 ± 17	Ventilated > 24 h
Carrie et al[34]	67	1 Medical ICU 1 Medical and surgical ICU	66 (58-74)	First SBT after 48 h of MV
Dinino et al[35]	63	2 Medical ICU	66 ± 19	Ready for first SBT
Farghaly and Hasan[36]	54	Respiratory ICU	SG, 65 (55-67.8) FG, 62.5 (55-70.7)	Patients with underlying pulmonary disease causing ARF who had successfully passed the SBT
Fayed et al[37]	112	ICU	62.61 ± 12.04	COPD patients
Ferrari et al[38]	46	High dependency unit	64.6 ± 12.1	Patients with tracheostomy who had failed one or more weaning attempts
Flevari et al[39]	27	MD ICU	65 (53-75)	Difficult or prolonged weaning
Jiang et al[40]	55	Medical ICU	67 (33-84)	Patients who had successfully passed the SBT
Jung et al[41]	33	Medical and surgical ICU	58 (51-67)	Patients diagnosed for ICUAW with MV > 48 h and undergoing an SBT
Kim et al[42]	82	Medical ICU	66	Patients ventilated > 48 h and ready for an SBT
Osman and Hashim[43]	68	Different ICU (medical and surgical)	56 (45-65)	Patients ready for SBT (most postoperatively)
Saeed et al[44]	30	Respiratory ICU	59 ± 6	Patients intubated due to COPD who had successfully passed the SBT
Shoaeir et al[45]	50	ICU	SG, 47.52 ± 14.60 FG, 51.89 ± 14.58	Patients ventilated > 48 h and ready for a first SBT
Soummer et al[12]	86	2 MD ICU	61 ± 14	Patients ventilated > 48 h who successfully passed a first SBT
Spadaro et al[46]	51	ICU	65 ± 13	Patients ventilated > 48 h and ready for SBT
Tenza-Lozano et al, unpublished data	63	MD ICU	63 ± 16	Patients ventilated > 24 h and ready for SBT

ARF = acute respiratory failure; FG = failure group; ICUAW = ICU-acquired weakness; MD = multidisciplinary; SG = success group. See Table 1 legend for expansion of other abbreviations.
[a]Age is expressed according to data extracted from each study as mean ± SD or median (interquartile range).

who died but did not clarify whether death occurred during weaning.

Given that ultrasound is subject to observer interpretation, we assumed that the included studies used diverse thresholds for positivity. Thus, sensitivity and specificity data are presented separately for each study, with no global weighting.[48] In any case, no pattern of graphically negative association was observed between sensitivity and specificity in any predictor (Fig 3). The main results are shown in Table 3.

Regarding DTF, in the ROC plane, studies showed moderate spread, and the confidence ellipse tended to be projected over the SROC curve, which supported the calculation of AUC at 0.87 (Fig 4A). There was one outlier in specificity (Tenza-Lozano et al, unpublished data, 2017), and a sensitivity analysis was therefore conducted (e-Fig 1). For DE, there was no evidence of a threshold effect and low heterogeneity for sensitivity, with a pooled value of 75% (95% CI, 65-85). However, because there were important differences among studies concerning specificity, it was not appropriate to perform a weighted estimation (Fig 4B). A sensitivity analysis was performed, excluding an outlier in specificity[31] (e-Fig 2). Concerning LUS, the AUC was 0.77; however, its representation in the ROC plane involves considerable uncertainty regarding the results due to the small number of studies available (e-Fig 3).

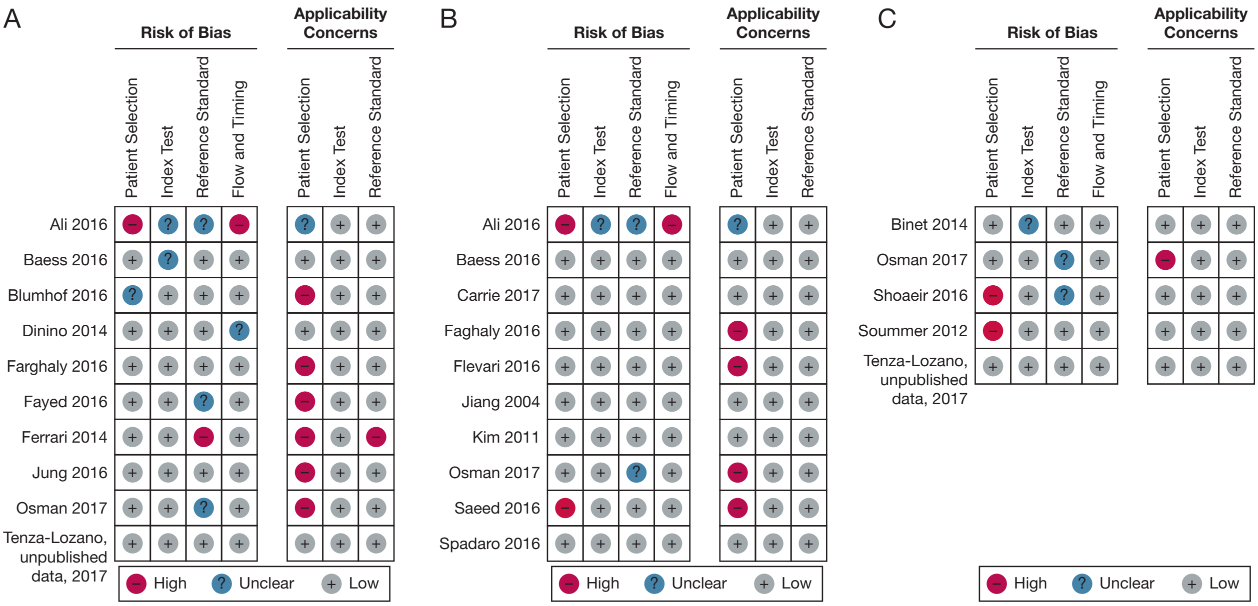

Figure 2 – *A-C, Methodological quality assessment according to the Quality Assessment of Diagnostic Accuracy Studies-2. For each study, risk of bias and applicability concerns are classified as high risk, low risk, or unclear. A, Studies measuring diaphragm thickening fraction. B, Studies measuring diaphragmatic excursion. C, Studies measuring lung ultrasound score.*

To explore potential causes of heterogeneity, a subgroup analysis and bivariate meta-regression were conducted to consider the overall quality of the studies, as well as possible problems of applicability derived from patient selection observed in the QUADAS-2 evaluation (Fig 5). This analysis was performed for DTF and DE because few studies reported on LUS. A "high-quality" subgroup comprises studies at low risk of bias in all QUADAS-2 domains. With regard to applicability, the "applicable" group comprised those studies involving a general population of critically ill patients vs a "nonapplicable" group of studies performed on more specific subpopulations of critically ill patients (ie, failed previous weaning attempts, COPD), according to our QUADAS-2 evaluation. There were no significant differences between high- and low-quality subgroups for DTF or DE. Nevertheless, in terms of applicability, there was a significantly higher specificity for DTF and higher sensitivity for DE in the "applicability concerns" subgroup. This finding suggests that DTF and DE perform better in determined subpopulations of individuals with a higher pretest probability compared with the general ICU population.

Discussion

DTF is considered a good indicator of the diaphragmatic inhalation effort,[49] and low values are associated with an increase in the duration of MV, ICU stay, and mortality.[50] Our data suggest that DTF is also a good predictor of weaning outcome, with overall consistency across studies, except for one outlier reporting lower specificity (Tenza-Lozano, unpublished data, 2017). This discordant result may be because investigators performed the ultrasound before the SBT, during a brief MV disconnection period, whereas most other studies conducted LUS during SBT. This theory would lend credence to the hypothesis that DTF varies as STB progresses due to diaphragmatic fatigue; that is, an early test would provide lower sensitivity to predict weaning failure. Currently, no available evidence supports this possibility; however, one ongoing study could eventually help to clarify this issue.[51]

DE is associated with inspiratory volume[52] but does not correlate with other indexes of respiratory effort.[49] Our data suggest a lower accuracy for DE compared with DTF in predicting weaning outcome and higher heterogeneity. Moreover, in patients undergoing MV, DTF reflects active diaphragm contraction,[53] whereas DE is derived from adding patients' respiratory effort to the pressure generated by the ventilator.[47] Thus, the use of DE is only meaningful in the absence of ventilatory support. Moreover, DE may vary depending on posture, exhibiting higher values when patients are supine vs seated, as well as when the abdominal and/or thoracic pressure is altered[54] (eg, ascites, atelectasis).

A

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ali 2016	27	4	1	22	0.96 [0.82-1.00]	0.85 [0.65-0.96]
Baess 2016	16	2	7	5	0.70 [0.47-0.87]	0.71 [0.29-0.96]
Blumhof 2016	22	6	4	20	0.85 [0.65-0.96]	0.77 [0.56-0.91]
Dinino 2014	43	4	6	10	0.88 [0.75-0.95]	0.71 [0.42-0.92]
Farghaly 2016	36	5	4	9	0.90 [0.76-0.97]	0.64 [0.35-0.87]
Fayed 2016	80	8	2	22	0.98 [0.91-1.00]	0.73 [0.54-0.88]
Ferrari 2014	24	2	5	15	0.83 [0.64-0.94]	0.88 [0.64-0.99]
Jung 2016	11	1	7	14	0.61 [0.36-0.83]	0.93 [0.68-1.00]
Osman 2017	44	0	6	18	0.88 [0.76-0.95]	1.00 [0.81-1.00]
Tenza-Lozano, unpublished data, 2017	37	13	3	10	0.93 [0.80-0.98]	0.43 [0.23-0.66]

B

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ali 2016	25	4	3	22	0.89 [0.72-0.98]	0.85 [0.65-0.96]
Baess 2016	16	6	7	1	0.70 [0.47-0.87]	0.14 [0.00-0.58]
Carrie 2017	32	9	13	13	0.71 [0.56-0.84]	0.59 [0.36-0.79]
Farghaly 2016	35	4	5	10	0.88 [0.73-0.96]	0.71 [0.42-0.92]
Flevari 2016	17	1	3	6	0.85 [0.62-0.97]	0.86 [0.42-1.00]
Jiang 2004	27	4	5	19	0.84 [0.67-0.95]	0.83 [0.61-0.95]
Kim 2011	21	22	7	32	0.75 [0.55-0.89]	0.59 [0.45-0.72]
Osman 2017	42	0	8	18	0.84 [0.71-0.93]	1.00 [0.81-1.00]
Saeed 2016	19	1	3	7	0.86 [0.65-0.97]	0.88 [0.47-1.00]
Spadaro 2016	21	2	13	15	0.62 [0.44-0.78]	0.88 [0.64-0.99]

C

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Binet 2014	39	5	0	4	1.00 [0.91-1.00]	0.44 [0.14-0.79]
Osman 2017	50	1	0	17	1.00 [0.93-1.00]	0.94 [0.73-1.00]
Shoaeir 2016	19	0	4	27	0.83 [0.61-0.95]	1.00 [0.87-1.00]
Soummer 2012	39	4	18	25	0.68 [0.55-0.80]	0.86 [0.68-0.96]
Tenza-Lozano, unpublished data, 2017	27	6	13	17	0.68 [0.51-0.81]	0.74 [0.52-0.90]

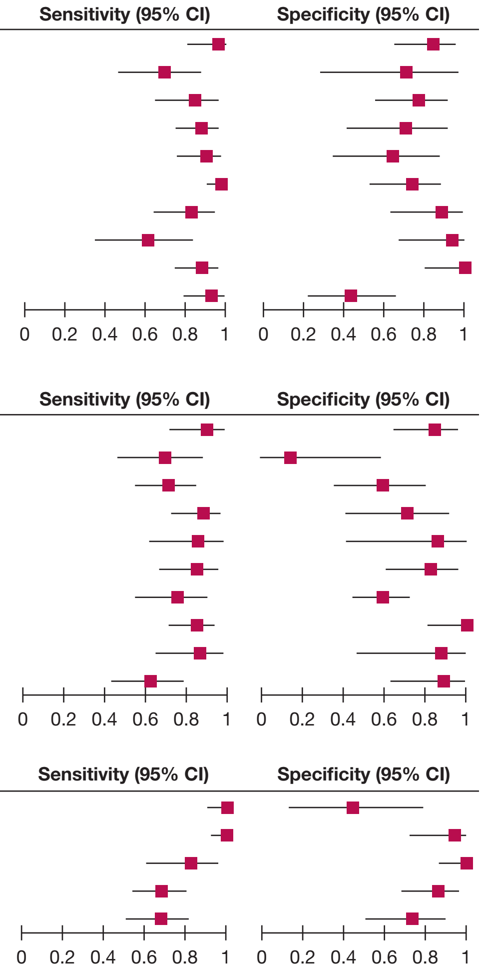

Figure 3 – A-C, Forest plot of sensitivity and specificity in studies that measure the following: (A) diaphragm thickening fraction, (B) diaphragm excursion, and (C) lung ultrasound score. FN = false negative; FP = false positive; TN = true negative; TP = true positive.

TABLE 3] Main Results Regarding Accuracy, Correlation Between Sensitivity and Specificity and Heterogeneity

	DTF	DE	LUS
Accuracy			
Pooled sensitivity (95% CI)	NA	75% (65 to 85)	NA
Pooled specificity (95% CI)	NA	NA	NA
AUSROC	0.87	NA	0.77
DOR (95% CI)	21 (11 to 40)	10.6 (5 to 24)	38 (7 to 198)
Correlation sensitivity-specificity			
Spearman rho (95% CI)	0.3 (−0.4 to 0.7)	−0.45 (−0.84 to 0.25)	0.2 (−0.8 to 0.9)
Heterogeneity			
Cochrane Q (*P* value)	9.5 (*P* = .38)	10.7 (*P* = .29)	5.1 (*P* = .27)
I^2	6%	15.8%	22%

AUSROC = area under the summary operator receiver-characteristic curve; DOR = diagnostic OR; NA = not applicable. See Table 1 legend for expansion of other abbreviations.

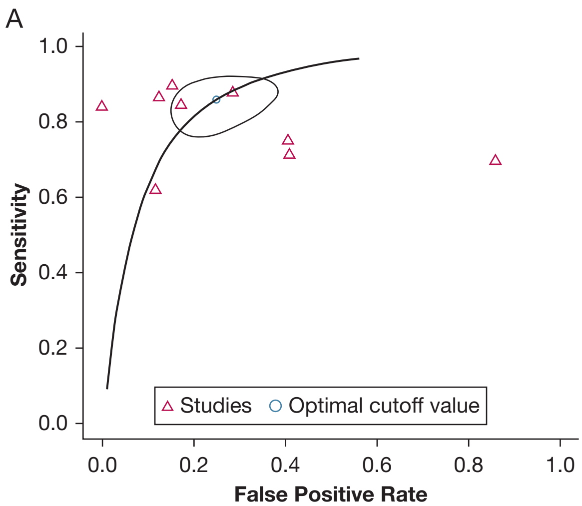

Figure 4 – *Summary receiver-operating characteristic curve of (A) diaphragm thickening fraction and (B) diaphragm excursion for weaning success prediction and confidence ellipse around the optimal cutoff value.*

Diaphragmatic ultrasound requires the use of high-frequency probes with precision to a few millimeters, meaning that slight variations in measurement between observers can substantially affect the result. Despite being an observer-dependent technique, the available evidence suggests that both DTF and DE are reproducible measures.[55,56]

In a recent systematic review, Zambon et al[57] assessed the usefulness of diaphragm ultrasound in ICU patients and, based on four studies, suggested a good performance as weaning indexes. Although overall data from this meta-analysis indicate good performance for both DTF and DE in predicting weaning outcomes, our exploration of heterogeneity and evaluation of applicability revealed a significantly higher specificity for DTF and higher sensitivity for DE in the "applicability concerns" subgroup. This novel finding contrasts with the enthusiasm shown in the current literature, suggesting that accuracy of DTF and DE may be overestimated due

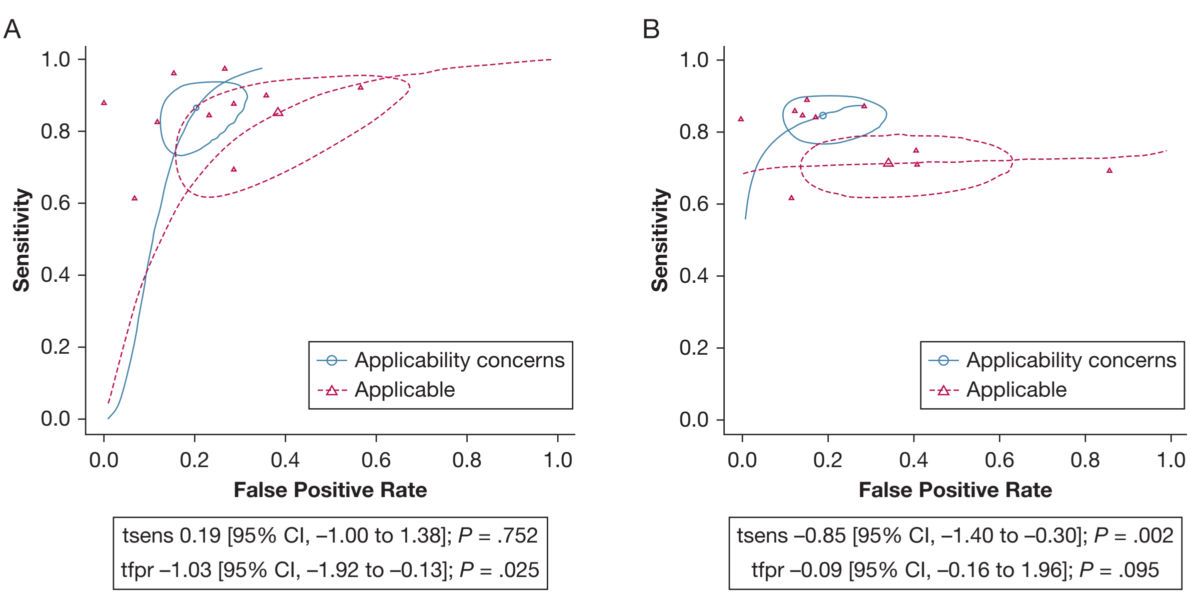

Figure 5 – *A, B, Subgroup analysis of applicability. Comparison of summary receiver-operating characteristic and confidence ellipses of applicable studies vs studies with applicability concerns, based on the analysis of methodologic quality using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. Studies were considered applicable when there was a low risk of concerns regarding applicability. A, Studies that measured diaphragm thickening fraction. B, Studies that measured diaphragm excursion. Statistical significance was set at* P $<$ *.05. tfpr = transformed false-positive rate; tsens = transformed sensitivity.*

to a large number of studies performed in subpopulations with a higher pretest probability of weaning failure (eg, in patients with COPD or those who have failed previous weaning attempts).

Other causes that may affect the weaning process involve alterations in the aeration of the pulmonary parenchyma (eg, pulmonary edema, pneumonia, atelectasis) that can be evaluated by using LUS. Our findings suggest an excellent performance for LUS in predicting weaning outcome; however, readers should exercise caution when interpreting this information because there are few available studies, and great uncertainty still remains. No studies assessing the reproducibility of the LUS score were identified.

To our knowledge, the present systematic review is the first to include a meta-analysis that assesses the accuracy of the lung and diaphragm ultrasound to predict weaning outcome, providing novel data derived from an applicability assessment. Another strength of this study is its comprehensiveness because the bibliographic search was not restricted by language, date of publication, or participant age. We made a major effort to contact different authors to obtain unpublished data. However, the main limitation of this study is the small number of studies available on each ultrasound predictor, which is understandable because these are incipient applications of ultrasound in the critical patient. This fact should be taken into account when interpreting subgroup analyses, as they are imprecise (although indicative).

Our findings suggest that DTF, LUS score, and, to a lesser extent, DE can provide valuable information for predicting weaning outcome, but taken alone, these factors may not perform as well as individual studies suggest. Diaphragmatic dysfunction and pulmonary aeration loss are two of the main potential causes of weaning failure, but they are not the only ones, and it is essential to contextualize the information obtained from ultrasound with clinical and laboratory data, as well as information derived from other imaging techniques such as echocardiography. Another aspect that should be taken into account is that all the reviewed studies implemented ultrasound only in patients previously classified as “ready to wean” according to traditional assessment; however, it is unknown how many patients would meet this criterion according to ultrasound alone, while failing to fulfil traditional parameters. More high-quality studies that rigorously assess the ultimate role of lung and diaphragm ultrasound in critically ill patients, and not only its accuracy and applicability, are needed.

Conclusions

Our data suggest that DTF is by itself a modest predictor of weaning outcome in the general population of critically ill patients. We do not support the use of DE because its accuracy is lower, and its measurement and interpretation entail several pitfalls. The LUS score seems to be an accurate predictor, but more studies are needed to reduce uncertainty.

Acknowledgments

Author contributions: A. M. L.-A. had full access to all of the data in the study; takes responsibility for the integrity of the data and the accuracy of the data analysis; is guarantor for the entire manuscript; and contributed to the study concept and design, the literature search, data collection and analysis, methodologic quality assessment, figures, translation and drafting of the manuscript. E. M. T.-L. contributed to the literature search, data collection, and methodologic quality assessment. J. L.-P. contributed to the study design, statistical analysis and revision of the manuscript. All authors approved the final version of the manuscript.

Financial/nonfinancial disclosures: None declared.

Other contributions: The authors appreciate the assistance of Meggan Harris, MSc, for her help with the edition and review of the manuscript, and the assistance of Pascual Gascón-Sansano for his help with the full-text article obtention.

Role of the sponsor: The sponsor did not participate in the study design, execution, or the preparation of the manuscript.

Additional information: The e-Figures and e-Tables can be found in the Supplemental Materials section of the online article.

References

1. Lemaire F, Teboul JL, Cinotti L, et al. Acute left ventricular dysfunction during unsuccessful weaning from mechanical ventilation. *Anesthesiology*. 1988;69(2):171-179.
2. Caroleo S, Agnello F, Abdallah K, Santangelo E, Amantea B. Weaning from mechanical ventilation: an open issue. *Minerva Anestesiol*. 2007;73(7-8):417-427.
3. Epstein SK, Ciubotaru RL, Wong JB. Effect of failed extubation on the outcome of mechanical ventilation. *Chest*. 1997;112(1):186-192.
4. Ouellette DR, Patel S, Girard TD, et al. Liberation from mechanical ventilation in critically ill adults: an Official American College of Chest Physicians/American Thoracic Society clinical practice guideline: inspiratory pressure augmentation during spontaneous breathing trials, protocols minimizing sedation, and noninvasive ventilation immediately after extubation. *Chest*. 2017;151(1):166-180.
5. Boles JM, Bion J, Connors A, et al. Weaning from mechanical ventilation. *Eur Respir J*. 2007;29(5):1033-1056.
6. Frutos-Vivar F, Ferguson ND, Esteban A, et al. Risk factors for extubation failure in patients following a successful spontaneous breathing trial. *Chest*. 2006;130(6):1664-1671.
7. Ouanes-Besbes L, Dachraoui F, Ouanes I, et al. NT-proBNP levels at spontaneous

breathing trial help in the prediction of post-extubation respiratory distress. *Intensive Care Med*. 2012;38(5):788-795.

8. Meade M, Guyatt G, Cook D, et al. Predicting success in weaning from mechanical ventilation. *Chest*. 2001;120(suppl 6):400S-424S.
9. Conti G, Montini L, Pennisi MA, et al. A prospective, blinded evaluation of indexes proposed to predict weaning from mechanical ventilation. *Intensive Care Med*. 2004;30(5):830-836.
10. Teboul J, Monnet X, Richard C. Weaning failure of cardiac origin: recent advances. *Critical Care*. 2010;14(2):211.
11. Bouhemad B, Brisson H, Le-Guen M, Arbelot C, Lu Q, Rouby JJ. Bedside ultrasound assessment of positive end-expiratory pressure-induced lung recruitment. *Am J Respir Crit Care Med*. 2011;183(3):341-347.
12. Soummer A, Perbet S, Brisson H, et al. Ultrasound assessment of lung aeration loss during a successful weaning trial predicts postextubation distress. *Crit Care Med*. 2012;40(7):2064-2072.
13. Martin AD, Smith B, Gabrielli A. Mechanical ventilation, diaphragm weakness and weaning: a rehabilitation perspective. *Respir Physiol Neurobiol*. 2013;189(2):377-383.
14. Mayo P, Arntfield R, Balik M, et al. The ICM research agenda on critical care ultrasonography. *Intensive Care Med*. 2017;43(9):1257-1269.
15. Lichtenstein DA. BLUE-protocol and FALLS-protocol: two applications of lung ultrasound in the critically ill. *Chest*. 2015;147(6):1659-1670.
16. Llamas-Álvarez AM, Tenza-Lozano EM, Latour-Pérez J. Accuracy of lung ultrasonography in the diagnosis of pneumonia in adults: systematic review and meta-analysis. *Chest*. 2017;151(2): 374-382.
17. Dube BP, Dres M. Diaphragm dysfunction: diagnostic approaches and management strategies. *J Clin Med*. 2016;5(12):E113.
18. Boussuges A, Gole Y, Blanc P. Diaphragmatic motion studied by M-mode ultrasonography: methods, reproducibility, and normal values. *Chest*. 2009;135(2):391-400.
19. Summerhill EM, El-Sameed YA, Glidden TJ, McCool FD. Monitoring recovery from diaphragm paralysis with ultrasound. *Chest*. 2008;133(3):737-743.
20. Bouhemad B, Liu ZH, Arbelot C, et al. Ultrasound assessment of antibiotic-induced pulmonary reaeration in ventilator-associated pneumonia. *Crit Care Med*. 2010;38(1):84-92.
21. Via G, Storti E, Gulati G, Neri L, Mojoli F, Braschi A. Lung ultrasound in the ICU: from diagnostic instrument to respiratory monitoring tool. *Minerva Anestesiol*. 2012;78(11):1282-1296.
22. Mayo P, Volpicelli G, Lerolle N, Schreiber A, Doelken P, Vieillard-Baron A. Ultrasonography evaluation during the weaning process: the heart, the diaphragm, the pleura and the lung. *Intensive Care Med*. 2016;42(7):1107-1117.
23. Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med*. 2009;6(7):e1000097.
24. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8): 529-536.
25. European Network for Health Technology Assessment, Eunethta. Methodological guideline on “Meta-analysis of diagnostic test accuracy studies.” Available at http://www.eunethta.eu/outputs/methodological-guideline-meta-analysis-diagnostic-test-accuracy-studies. Accessed March 13, 2017.
26. R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed May 22, 2017.
27. *Review Manager (RevMan), version 5.3*. Copenhagen, Denmark: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
28. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
29. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865-2884.
30. Ali ER, Mohamad AM. Diaphragm ultrasound as a new functional and morphological index of outcome, prognosis and discontinuation from mechanical ventilation in critically ill patients and evaluating the possible protective indices against VIDD. *J Chest Dis Tuberc*. 2016;66:339-351.
31. Baess AI, Abdallah TH, Emara DM, Hassan M. Diaphragmatic ultrasound as a predictor of successful extubation from mechanical ventilation: thickness, displacement, or both? *Egypt J Bronchol*. 2016;10:162-166.
32. Binet C, Neron L, Rochart N, et al. Validation d’un indice échographique prédictif d’échec de sevrage de la ventilation mécanique. *Annales Françaises D Anesthesie Et De Reanimation*. 2014;33. A383-A383.
33. 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.
34. Carrie C, Gisbert-Mora C, Bonnardel E, et al. Ultrasonographic diaphragmatic excursion is inaccurate and not better than the MRC score for predicting weaning-failure in mechanically ventilated patients. *Anaesth Crit Care Pain Med*. 2017;36(1): 9-14.
35. Dinino E, Gartman EJ, Sethi JM, McCool FD. Diaphragm ultrasound as a predictor of successful extubation from mechanical ventilation. *Thorax*. 2014;69(5):423-427.
36. 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.
37. Fayed AM, Abd El Hady MA, Shaaban MS, Fikry DM. Use of ultrasound to assess diaphragmatic thickness as a weaning parameter in invasively ventilated chronic obstructive pulmonary disease patients. *J Am Sci*. 2016;12(6): 96-105.
38. 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.
39. Flevari A, Lignos M, Konstantonis D, Armaganidis A. Diaphragmatic ultrasonography as an adjunct predictor tool of weaning success in patients with difficult and prolonged weaning. *Minerva Anestesiol*. 2016;82(11):1149-1157.
40. Jiang JR, Tsai TH, Jerng JS, Yu CJ, Wu HD, Yang PC. Ultrasonographic evaluation of liver/spleen movements and extubation outcome. *Chest*. 2004;126(1): 179-185.
41. Jung B, Moury PH, Mahul M, et al. Diaphragmatic dysfunction in patients with ICU-acquired weakness and its impact on extubation failure. *Intensive Care Med*. 2016;42(5):853-861.
42. 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.
43. Osman AM, Hashim RM. Diaphragmatic and lung ultrasound application as new predictive indices for the weaning process in ICU patients. *Egypt J Radiol Nucl Med*. 2017;48(1):61-66.
44. Saeed AM, El Assal GI, Ali TM, Hendawy MM. Role of ultrasound in assessment of diaphragmatic function in chronic obstructive pulmonary disease patients during weaning from mechanical ventilation. *Egypt J Bronchol*. 2016;10: 167-172.
45. Shoaeir M, Noeam K, Mahrous A, Alaa A. Lung aeration loss as a predictor of reintubation using lung ultrasound in mechanically ventilated patients. *Biolife*. 2016;4(3):514-520.
46. Spadaro S, Grasso S, Mauri T, et al. Can diaphragmatic ultrasonography performed during the T-tube trial predict weaning failure? The role of diaphragmatic rapid shallow breathing index. *Critical Care*. 2016;20(1):305.
47. Matamis D, Soilemezi E, Tsagourias M, et al. Sonographic evaluation of the

diaphragm in critically ill patients. Technique and clinical applications. *Intensive Care Med*. 2013;39(5):801-810.

48. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. United Kingdom: The Cochrane Collaboration; 2010.
49. Umbrello M, Formenti P, Longhi D, et al. Diaphragm ultrasound as indicator of respiratory effort in critically ill patients undergoing assisted mechanical ventilation: a pilot clinical study. *Crit Care*. 2015;19:161.
50. Dubé BP, Dres M, Mayaux J, Demiri S, Similowski T, Demoule A. Ultrasound evaluation of diaphragm function in mechanically ventilated patients: comparison to phrenic stimulation and prognostic implications. *Thorax*. 2017;72(9):811-818.
51. Zhou P, Zhang Z, Hong Y, et al. The predictive value of serial changes in diaphragm function during the spontaneous breathing trial for weaning outcome: a study protocol. *BMJ Open*. 2017;7:e015043.
52. Cohen E, Mier A, Heywood P, Murphy K, Boultbee J, Guz A. Excursion-volume relation of the right hemidiaphragm measured by ultrasonography and respiratory airflow measurements. *Thorax*. 1994;49(9):885-889.
53. Goligher EC, Fan E, Herridge MS, et al. Evolution of diaphragm thickness during mechanical ventilation: impact of inspiratory effort. *Am J Respir Crit Care Med*. 2015;192(9):1080-1088.
54. Haji K, Royse A, Green C, Botha J, Canty D, Royse C. Interpreting diaphragmatic movement with bedside imaging, review article. *J Crit Care*. 2016;34:56-65.
55. Francis CA, Hoffer JA, Reynolds S. Ultrasonographic evaluation of diaphragm thickness during mechanical ventilation in intensive care patients. *Am J Crit Care*. 2016;25(1):e1-e8.
56. Bobbia X, Clement A, Claret PG, et al. Diaphragmatic excursion measurement in emergency patients with acute dyspnea: toward a new diagnostic tool? *Am J Emerg Med*. 2016;34(8):1653-1657.
57. Zambon M, Greco M, Bocchino S, et al. Assessment of diaphragmatic dysfunction in the critically ill patient with ultrasound: a systematic review. *Intensive Care Med*. 2017;43(1):29-38.