Lung Ultrasound Prediction Model for Acute Respiratory Distress Syndrome A Multicenter Prospective Observational Study

ORIGINAL ARTICLE

Lung Ultrasound Prediction Model for Acute Respiratory Distress Syndrome

A Multicenter Prospective Observational Study

Marry R. Smit[1], Laura A. Hagens[1], Nanon F. L. Heijnen[2], Luigi Pisani[1,3,4], Thomas G. V. Cherpanath[1], Dave A. Dongelmans[1], Harm-Jan S. de Grooth[5], Charalampos Pierrakos[1,6], Pieter Roel Tuinman[5], Claudio Zimatore[1,7], Frederique Paulus[1], Ronny M. Schnabel[2], Marcus J. Schultz[1,3,8], Dennis C. J. J. Bergmans[2,9], and Lieuwe D. J. Bos[1]; for the DARTS Consortium

[1]Department of Intensive Care, Amsterdam University Medical Center (UMC), location University of Amsterdam, Amsterdam, the Netherlands; [2]Department of Intensive Care, Maastricht UMC+, Maastricht, the Netherlands; [3]Mahidol–Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand; [4]Department of Anesthesia and Intensive Care, Miulli Regional Hospital, Acquaviva delle Fonti, Italy; [5]Intensive Care, Amsterdam UMC, locatie Vrije Universiteit Amsterdam, Amsterdam, Nederland; [6]Department of Intensive Care, Brugmann University Hospital, Free University of Brussels, Brussels, Belgium; [7]Intensive Care Unit, Emergency and Organ Transplantation, University of Bari, Bari, Italy; [8]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom; and [9]School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, the Netherlands

ORCID IDs: 0000-0001-9938-6035 (M.R.S.); 0000-0003-4629-4465 (L.A.H.); 0000-0002-9147-891X (N.F.L.H.); 0000-0002-7499-076X (H.-J.S.d.G.).

Abstract

Rationale: Lung ultrasound (LUS) is a promising tool for diagnosis of acute respiratory distress syndrome (ARDS), but adequately sized studies with external validation are lacking.

Objectives: To develop and validate a data-driven LUS score for diagnosis of ARDS and compare its performance with that of chest radiography (CXR).

Methods: This multicenter prospective observational study included invasively ventilated ICU patients who were divided into a derivation cohort and a validation cohort. Three raters scored ARDS according to the Berlin criteria, resulting in a classification of "certain no ARDS," or "certain ARDS" when experts agreed or "uncertain ARDS" when evaluations conflicted. Uncertain cases were classified in a consensus meeting. Results of a 12-region LUS exam were used in a logistic regression model to develop the LUS-ARDS score.

Measurements and Main Results: Three hundred twenty-four (16% certain ARDS) and 129 (34% certain ARDS) patients were included in the derivation cohort and the validation cohort, respectively. With an ARDS diagnosis by the expert panel as the reference test, the LUS-ARDS score, including the left and right LUS aeration scores and anterolateral pleural line abnormalities, had an area under the receiver operating characteristic (ROC) curve of 0.90 (95% confidence interval [CI], 0.85–0.95) in certain patients of the derivation cohort and 0.80 (95% CI, 0.72–0.87) in all patients of the validation cohort. Within patients who had imaging–gold standard chest computed tomography available, diagnostic accuracy of eight independent CXR readers followed the ROC curve of the LUS-ARDS score.

Conclusions: The LUS-ARDS score can be used to accurately diagnose ARDS also after external validation. The LUS-ARDS score may be a useful adjunct to a diagnosis of ARDS after further validation, as it showed performance comparable with that of the current practice with experienced CXR readers but more objectifiable diagnostic accuracy at each cutoff.

Keywords: ARDS; diagnosis; LUS; intensive care

(Received in original form October 10, 2022; accepted in final form February 14, 2023)

A complete list of DARTS Consortium members may be found before the beginning of the References.

Supported by Health Holland (a Public–Private Partnership grant 10.2.17.181PPS) via the Dutch Lung Foundation (longfonds).

Author Contributions: All authors contributed to the conceptualization and design of the study and the interpretation of the results. M.R.S. and L.D.J.B. drafted the initial manuscript. All authors critically revised the manuscript for intellectual content. M.R.S., L.A.H., N.F.L.H., and L.D.J.B. had access to the raw data, performed the analyses, and verified the data. All authors read and approved the final manuscript.

Am J Respir Crit Care Med Vol 207, Iss 12, pp 1591–1601, Jun 15, 2023

Originally Published in Press as DOI: 10.1164/rccm.202210-1882OC on February 15, 2023
Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the Subject: Lung ultrasound (LUS) is a promising tool for diagnosis of acute respiratory distress syndrome (ARDS), but adequately sized studies with external validation are lacking. In one single-center study, an adapted ARDS definition for resource-limited settings was proposed that allowed use of bilateral B-lines or consolidations on LUS in addition to existing imaging modalities. This adapted definition was validated and modified in a high-resource, single-center study. Another report provided an expert overview of specific LUS signs that could facilitate in differentiating between ARDS and cardiogenic pulmonary edema.

What This Study Adds to the Field: We developed and externally validated the first data-driven LUS-based diagnostic approach for ARDS. The study was adequately powered, included consecutive patients, and follows current recommendations for diagnostic evaluation. The LUS-ARDS score, based on LUS aeration scores of the left and right lungs, combined with presence of anterolateral pleural line abnormalities, could accurately diagnose and exclude ARDS during external validation. Further confirmation of the scores' accuracy would allow for the use of LUS to diagnose and exclude ARDS and strengthens the potential of the LUS-ARDS score as a useful adjunct to diagnosis of ARDS.

Acute respiratory distress syndrome (ARDS) is a common cause of acute respiratory failure in the ICU and is associated with a mortality of around 40% (1). Diagnosis of ARDS is based on the Berlin definition, which consists of clinical criteria and requires the presence of bilateral opacities on chest imaging consistent with pulmonary edema (2). Chest computed tomography (CT) can be considered the gold standard for radiological assessment of pulmonary edema, but it requires the undesirable transportation of critically ill patients to the radiology department and is performed in a minority of patients (3). Chest radiography (CXR) can be performed bedside in the ICU, but the interobserver agreement for bilateral opacities is poor for CXR, resulting in conflicting diagnoses (4).

Lung ultrasound (LUS) has gained recognition as an accessible technique for the diagnosis and monitoring of ICU patients (5). LUS can estimate lung aeration by means of semiquantitative scores of artifactual and anatomical ultrasound patterns (6, 7). Although LUS has been extensively studied for the diagnosis of pneumothorax, effusions, or pneumonia, evidence for the diagnosis of ARDS is scant (8, 9). The Kigali modification of the Berlin definition, which relies on bilateral LUS abnormalities, was proposed to improve ARDS diagnosis in settings with limited access to ventilators, imaging, and other resources (10). This method was further adapted and showed reasonable diagnostic accuracy compared with the Berlin definition in ICU patients in a high-resource setting (11). The presence of pleural line abnormalities, reduced lung sliding, or "spared regions" in patients with pulmonary edema has also been suggested to be specific for ARDS in the LUS consensus statement, although these abnormalities have not yet been included in a LUS definition of ARDS, and clinical evidence is limited (12–15). None of the diagnostic tests for ARDS using LUS have been derived using data-driven methods or were externally validated, whereas these aspects are crucial for the inclusion of a LUS method in a new definition of ARDS, which is in the making.

The aim of the present study was to develop and validate a data-driven LUS score for diagnosis of ARDS using a diagnosis by an expert panel as the gold standard. We hypothesized that the developed LUS-ARDS score has high diagnostic accuracy during external validation and has a diagnostic performance that is similar to the performance of the Berlin definition with evaluation of CXR, the most frequently used concurrent diagnostic imaging technique for ARDS. Finally, we also hypothesized that the developed LUS-ARDS score could resolve uncertainty in patient classification when expert readers disagreed on the diagnosis of ARDS. Data from this study were presented at the American Thoracic Society 2022 International Conference (16) and has been published in part to answer a different research question (17).

Methods

Study Design and Ethical Concerns

The present study was a multicenter prospective observational study that was performed as part of the "Diagnosis of Acute Respiratory disTress Syndrome" (DARTS) project. The DARTS project is registered at Netherlands Trial Register (https://trialsearch.who.int/ ; identifier NL8226), and its protocol was previously published (18). The study was performed in the ICUs of two academic hospitals in the Netherlands: the Amsterdam University Medical Centers (Amsterdam UMC), location Academic Medical Center (AMC) in Amsterdam, and the Maastricht University Medical Center+ (MUMC+) in Maastricht. Patients were recruited between March 27, 2019, and February 27, 2021. The institutional review board of the Amsterdam UMC, location AMC, approved the study protocol (no. W18_311). Written deferred consent for the use of data for clinical research was obtained from the patient or the patients' representative (18). The TRIPOD guidelines were followed in the reporting of this study (19).

Patients

Patients were included in the study if they were admitted to a participating ICU and expected to be invasively ventilated for at least 24 hours. Exclusion criteria were as follows: *1*) invasive ventilation for more than 48 hours in the 7 days before screening, *2*) tracheostomy, *3*) clinical situations in which study inclusion is inappropriate

Correspondence and requests for reprints should be addressed to Marry R. Smit, Ph.D., Department of Intensive Care, Amsterdam University Medical Center, location University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands. E-mail: m.r.smit@amsterdamumc.nl.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

(e.g., withdrawal of care or highly contagious disease), and *4*) objection to participate in the study by patients or representatives. Patients included at the Amsterdam UMC, location AMC, were part of the derivation cohort for development of the LUS-ARDS score, whereas patients included at the MUMC+ acted as the validation cohort.

Study Procedure

Patients were included within 48 hours after start of invasive ventilation. The inclusion day was considered as Day 1. During inclusion, a LUS examination was performed, and patient characteristics, ventilation, and gas-exchange parameters were collected. This procedure was repeated 24 hours after inclusion (Day 2).

LUS Examination

LUS was performed with a linear transducer using a standard 12-region protocol and the clinically available ultrasound device (7, 20). Lung regions were scored with the LUS aeration score. An "A-pattern," defined as horizontal repetitions of the pleural line (A-lines), was scored as 0. "B-patterns" were scored as 1 when more than two well-spaced B-lines were present that covered less than 50% of the pleural line or as 2 when B-lines covered more than 50% of the pleural line. A "C-pattern" was defined as an anatomical image of consolidation or as complete (or near-complete) loss of aeration that is larger than 2 cm and was scored as 3. C-patterns accompanied with pleural effusion were scored as 0, as this is more suggestive of a compression atelectasis rather than an intrinsic pulmonary process (10, 11, 21). All lung regions were scored on the presence of an abnormal pleural line, subpleural consolidations, dynamic air bronchograms and pleural effusions (1 when present and 0 when absent). Use of convex or sector array probes was allowed in case lung regions could not be assessed with the linear probe. Lung regions that were unable to scan or score (i.e., because of wounds, chest drains, or subcutaneous emphysema) were complemented by the mean LUS aeration score of other lung regions of the respective hemithorax. The LUS images were scored prospectively and the sonographers were blinded from the ARDS classification. Patients with more than four missing regions were excluded from analysis. The LUS examination on Day 1 was used for the analysis, or the LUS exam on Day 2 was used if the exam on Day 1 was not available. The LUS examination and scoring are presented in more detail in the online supplement.

Expert Panel Classification of ARDS—Reference Test

Three experts from the ICU with extensive experience in diagnosing and treating ARDS independently assessed all patients on meeting the Berlin criteria for ARDS (22). The experts based their judgment on clinical parameters, blood gas analysis, and chest imaging, which were available within 72 hours after the start of invasive ventilation

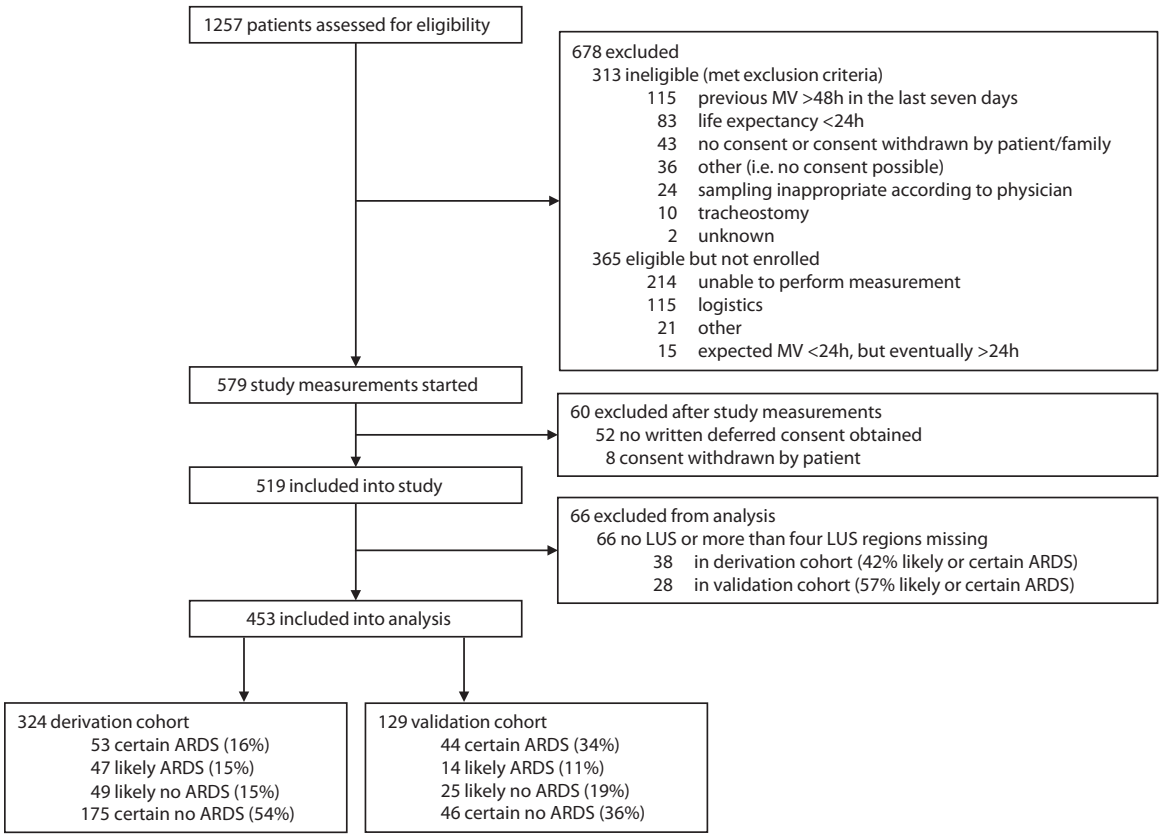

Figure 1. Study profile. A total of 519 patients were included in the Diagnosis of Acute Respiratory disTress Syndrome (DARTS) study. Sixty-six included patients who could not be analyzed in the present study because they did not receive a lung ultrasound (LUS) exam or because more than 4 of 12 lung regions were missing in the LUS exam. Patients included in the Amsterdam University Medical Centers, location Academic Medical Center, served as the derivation cohort, and patients included in the Maastricht University Medical Center+ were assigned to the validation cohort. ARDS = acute respiratory distress syndrome; MV = mechanical ventilation.

Table 1. Patient Characteristics of the Derivation Cohort

	Certain No ARDS ($n = 175$)	Uncertain Classification		Certain ARDS ($n = 53$)
		Likely No ARDS ($n = 49$)	Likely ARDS ($n = 47$)	
Age, yr, mean (SD)	62 (16)	61 (14)	63 (14)	59 (14)
Male, *n* (%)	120 (68.6)	37 (75.5)	27 (57.4)	38 (71.7)
BMI, kg/m^2, median (IQR)	25.8 (22.9, 29.9)	28.9 (24.8, 31.4)	27.0 (24.9, 30.9)	26.5 (23.7, 30.3)
APACHE II score, median (IQR)	20 (15, 26)	20 (15, 23)	20 (17, 23)	20 (15, 23)
LIPS, median (IQR)	5 (3, 6)	5 (2, 7)	6 (5, 8)	6 (6, 8)
SOFA score, median (IQR)	10 (8, 12)	10 (8, 12)	9 (7, 12)	9 (5, 12)
Global LUS score, median (IQR)	4 (1, 9)	5 (1, 7)	9 (5, 14)	14 (9, 20)
Pre-ICU LOS, days, median (IQR)	1 (0, 3)	1 (0, 2)	2 (1, 7)	2 (0, 6)
Duration MV, h, median (IQR)	20 (12, 31)	20 (13, 26)	23 (12, 26)	18 (10, 30)
Admission type, *n* (%)				
Emergency surgical	33 (18.9)	6 (12.2)	9 (19.1)	2 (3.8)
Medical	117 (66.9)	37 (75.5)	33 (70.2)	47 (88.7)
Planned surgical	25 (14.3)	6 (12.2)	5 (10.6)	4 (7.5)
Pneumonia, *n* (%)	26 (14.9)	5 (10.2)	27 (57.4)	47 (88.7)
COVID-19, *n* (%)	1 (0.6)	0 (0.0)	8 (17.0)	24 (45.3)
ARDS severity, *n* (%)				
Mild	NA	NA	8 (17.4)	3 (5.7)
Moderate	NA	NA	24 (52.2)	27 (50.9)
Severe	NA	NA	14 (30.4)	23 (43.4)
Cause of ARDS				
Nonpulmonary, *n* (%)	NA	NA	17 (36.2)	4 (7.5)
Pulmonary, *n* (%)	NA	NA	30 (63.8)	49 (92.5)
ICU LOS, d, median (IQR)	5 (3, 11)	6 (3, 9)	8 (2, 12)	9 [6, 18]
30 d mortality, *n* (%)	68 (43.3)	17 (39.5)	19 (43.2)	22 (46.8)
ICU mortality, *n* (%)	58 (34.7)	13 (28.3)	16 (34.8)	22 (44.0)
Ventilation characteristics				
Pa_{O_2}/FI_{O_2} ratio, mm Hg, median (IQR)	254 (168, 336)	174 (136, 225)	119 (96, 181)	105 (81, 144)
PEEP, cm H_2O, median (IQR)	5 (5, 8)	8 (6, 10)	8 (7, 10)	10 (8, 10)
RR, breaths/min, median (IQR)	18 (15, 23)	18 (15, 25)	20 (15, 25)	22 (17, 28)
V_T/PBW, ml/kg, median (IQR)	7.1 (6.2, 8.6)	7.1 (6.2, 8.5)	7.4 (5.8, 8.8)	7.7 (6.5, 9.2)
Ventilatory ratio, median (IQR)	1.3 (1.1, 1.6)	1.4 (1.1, 1.6)	1.5 (1.1, 1.8)	1.9 (1.7, 2.5)
Mechanical power, J/min, median (IQR)	14.9 (10.5, 20.3)	18.3 (13.3, 24.6)	17.1 (13.1, 27.2)	23.5 (17.2, 29.5)
Driving pressure, cm H_2O, median (IQR)	12 (8, 15)	12 (9, 16)	13 (9, 16)	12 (8, 17)
Compliance, ml/cm H_2O, median (IQR)	36.8 (27.4, 57.3)	41.5 (29.0, 47.0)	39.7 (26.0, 50.7)	36.1 (25.3, 52.6)

Definition of abbreviations: APACHE II = acute physiology and chronic health evaluation II; ARDS = acute respiratory distress syndrome; BMI = body mass index; COVID-19 = coronavirus disease; IQR = interquartile range; LIPS = lung injury prediction score; LOS = length of stay; LUS = lung ultrasound; MV = mechanical ventilation; NA = not applicable; PBW = predicted body weight; PEEP = positive end-expiratory pressure; RR = respiratory rate; SOFA = sequential organ failure assessment.

Data are presented as *n* (%) or median (IQR) unless indicated otherwise. Duration of MV is the time between intubation and LUS. Pa_{O_2}/FI_{O_2} ratio is defined as the lowest Pa_{O_2}/FI_{O_2} ratio in the 24 hours before LUS. Ventilation parameters were collected during LUS, and laboratory parameters were collected closest to LUS. PBW is calculated as follows: $PBW_{male} = 50 + 0.91 \cdot (\text{cm of height} - 152.4)$ and $PBW_{female} = 45.5 + 0.91 \cdot (\text{cm of height} - 152.4)$. Ventilatory ratio (VR) is calculated as follows: $VR = [\text{minute volume (ml/min)} \cdot Pa_{CO_2}]/(PBW \cdot 100 \cdot 37.5)$. Mechanical power (MP) is calculated as follows: $MP = 0.098 \cdot RR \cdot V_T \cdot [PEEP + (P_{max} - PEEP)]$. Driving pressure (DP) is calculated as $DP = P_{max} - PEEP$. Additional characteristics are presented in the online supplement.

(2, 18). On the basis of the agreement between experts, all patients were classified in the following categories: *1*) certain no ARDS when there was sufficient agreement to exclude ARDS, *2*) certain ARDS when there was sufficient agreement to diagnose ARDS, and *3*) uncertain ARDS when there were conflicting scores between experts. Patients with uncertain ARDS were subsequently discussed in a consensus meeting, resulting in the classification of either "likely ARDS" or "likely no ARDS." Consensus diagnosis was primarily based on CT if available or otherwise on CXR images. LUS exams were excluded from this analysis. The classification process is described in more detail in the online supplement.

Independent Evaluation of ARDS by Additional CXR Readers

To allow for comparison of LUS to the current standard of practice, seven individual ICU physicians and one chest radiologist outside the expert panel scored ARDS according to the Berlin criteria using CXR images in a subset of patients. This subset consisted of patients from the derivation and validation cohorts who had a CT scan available, as CT is considered the best reference standard. The seven individual ICU physicians scored the same 50 patients who were randomly selected: 35 from the derivation cohort and 15 from the validation cohort. The chest radiologist scored the first 121 included patients with a CT available in the derivation cohort. Physicians had clinical and ventilation parameters available next to CXR images.

Table 2. Patient Characteristics of the Validation Cohort

	Certain No ARDS ($n = 46$)	Uncertain Classification: Likely No ARDS ($n = 25$)	Uncertain Classification: Likely ARDS ($n = 14$)	Certain ARDS ($n = 44$)
Age, yr, mean (SD)	62 (17)	62 (13)	63 (15)	66 (12)
Male, *n* (%)	27 (58.7)	12 (48.0)	10 (71.4)	32 (72.7)
BMI, kg/m^2, median (IQR)	24.0 (22.7, 27.2)	27.3 (24.5, 29.4)	24.8 (23.6, 28.0)	26.5 (23.6, 28.6)
APACHE II score, median (IQR)	24 (16, 26)	22 (15, 25)	24 (17, 26)	16 (13, 24)
LIPS, median (IQR)	5 (4, 7)	5 (4, 6)	5 (3, 6)	6 (5, 6)
SOFA score, median (IQR)	7 (5, 10)	8 (7, 9)	8 (6, 9)	8 (7, 10)
Global LUS score, median (IQR)	5 (3, 8)	7 (3, 11)	8 (5, 13)	14 (10, 17)
Pre-ICU LOS, d, median (IQR)	2 (1, 3)	2 (1, 4)	3 (1, 9)	3 (2, 7)
Duration MV, h, median (IQR)	24 (17, 38)	24 (17, 30)	21 (17, 26)	23 (17, 34)
Admission type, *n* (%)				
Emergency surgical	11 (23.9)	2 (8.0)	2 (14.3)	0 (0.0)
Medical	27 (58.7)	20 (80.0)	10 (71.4)	44 (100.0)
Planned surgical	8 (17.4)	3 (12.0)	2 (14.3)	0 (0.0)
Pneumonia, *n* (%)	2 (4.3)	2 (8.0)	2 (14.3)	33 (75.0)
COVID-19, *n* (%)	0 (0.0)	0 (0.0)	0 (0.0)	18 (40.9)
ARDS severity, *n* (%)				
Mild	NA	NA	7 (50.0)	3 (7.0)
Moderate	NA	NA	7 (50.0)	25 (58.1)
Severe	NA	NA	0 (0.0)	15 (34.9)
Cause of ARDS				
Nonpulmonary, *n* (%)	NA	NA	12 (85.7)	3 (7.0)
Pulmonary, *n* (%)	NA	NA	2 (14.3)	37 (84.1)
ICU LOS, d, median (IQR)	6 (4, 13)	8 (4, 17)	5 (4, 9)	11 (6, 25)
30 d mortality, *n* (%)	19 (41.3)	9 (36.0)	4 (30.8)	18 (40.9)
ICU mortality, *n* (%)	16 (35.6)	7 (28.0)	2 (14.3)	18 (40.9)
Ventilation characteristics				
$Pa_{O_2}/F_{I_{O_2}}$ ratio, mm Hg, median (IQR)	307 (195, 388)	157 (128, 239)	203 (188, 234)	120 (82, 143)
PEEP, cm H_2O, median (IQR)	8 (6, 8)	8 (8, 8)	8 (7, 10)	11 (10, 12)
RR, breaths/min, median (IQR)	16 (14, 20)	17 (13, 20)	18 (15, 22)	20 (18, 24)
V_T/PBW, ml/kg, median (IQR)	7.2 (5.7, 8.6)	7.0 (6.6, 8.9)	7.3 (6.4, 9.6)	6.9 (5.9, 7.7)
Ventilatory ratio, median (IQR)	1.1 (1.0, 1.4)	1.2 (1.0, 1.4)	1.3 (1.2, 1.7)	1.4 (1.2, 1.7)
Mechanical power, J/min, median (IQR)	15.3 (10.9, 23.1)	16.3 (11.6, 21.5)	19.6 (12.1, 22.2)	24.8 (19.6, 32.2)
Driving pressure, cm H_2O, median (IQR)	15 (12, 19)	14 (10, 19)	15 (11, 19)	18 (15, 22)
Compliance, mL/cm H_2O, median (IQR)	26.6 (19.8, 34.9)	32.4 (24.9, 39.1)	28.4 (24.3, 42.2)	24.0 (19.3, 30.7)

For definition of abbreviations, *see* Table 1.
Data are presented as *n* (%) or median (IQR), unless indicated otherwise. Duration of MV is the time between intubation and LUS. The $Pa_{O_2}/F_{I_{O_2}}$ ratio is defined as the lowest $Pa_{O_2}/F_{I_{O_2}}$ ratio in the 24 hours before LUS. Ventilation parameters were collected during LUS, and laboratory parameters were collected closest to LUS. PBW is calculated as follows: $PBW_{male} = 50 + 0.91 \cdot (\text{cm of height} - 152.4)$ and $PBW_{female} = 45.5 + 0.91 \cdot (\text{cm of height} - 152.4)$. Ventilatory ratio (VR) is calculated as follows: $VR = [\text{minute volume (ml/min)} \cdot Pa_{CO_2}]/(PBW \cdot 100 \cdot 37.5)$. Mechanical power (MP) is calculated as follows: $MP = 0.098 \cdot RR \cdot V_T \cdot [PEEP + (P_{max} - PEEP)]$. Driving pressure (DP) is calculated as $DP = P_{max} - PEEP$. Additional characteristics are presented in the online supplement.

Study Outcomes

The primary outcome of the study was the diagnostic accuracy and calibration of LUS methods for ARDS as judged by the expert panel. To account for unreliability of ARDS diagnosis using CXR, a sensitivity analysis was performed that only included patients for whom a chest CT was available. To evaluate whether LUS methods could replace CXR, the diagnostic accuracy of LUS was compared with the diagnostic accuracy of individual, independent physicians on the basis of clinical data and CXR in patients for whom a chest CT was available. Last, we evaluated whether LUS provided additional diagnostic accuracy in scenarios where the expert panel was uncertain about the ARDS diagnosis.

Statistical Analysis and Development of the Index Test

The DARTS project had a sample size of at least 500 patients (18), and the statistical power of this sample size was assessed for the present study. The sample size of the derivation cohort with certain ARDS and certain no ARDS diagnoses was calculated using the *pmsampsize* package in RStudio (version 4.0.3) (23). An expected C statistic of 0.85 with a prevalence of ARDS of 10.4% resulted in a sample size of 152 patients for a model with three variables (1). The sample size of the validation cohort was pragmatically chosen as all patients who were included at the MUMC+ hospital.

Recommendations for predictive modeling were followed (24, 25). The LUS-ARDS score for diagnosis of ARDS was developed using logistic regression analysis on LUS data from patients with certain ARDS and certain no ARDS diagnoses in the derivation cohort while remaining blinded to the validation cohort. A diagnosis of ARDS by the expert panel was used as the reference test in all analyses. Variables included in logistic regression analysis were preselected on the basis of evidence from previous studies (10–13, 26, 27). The final

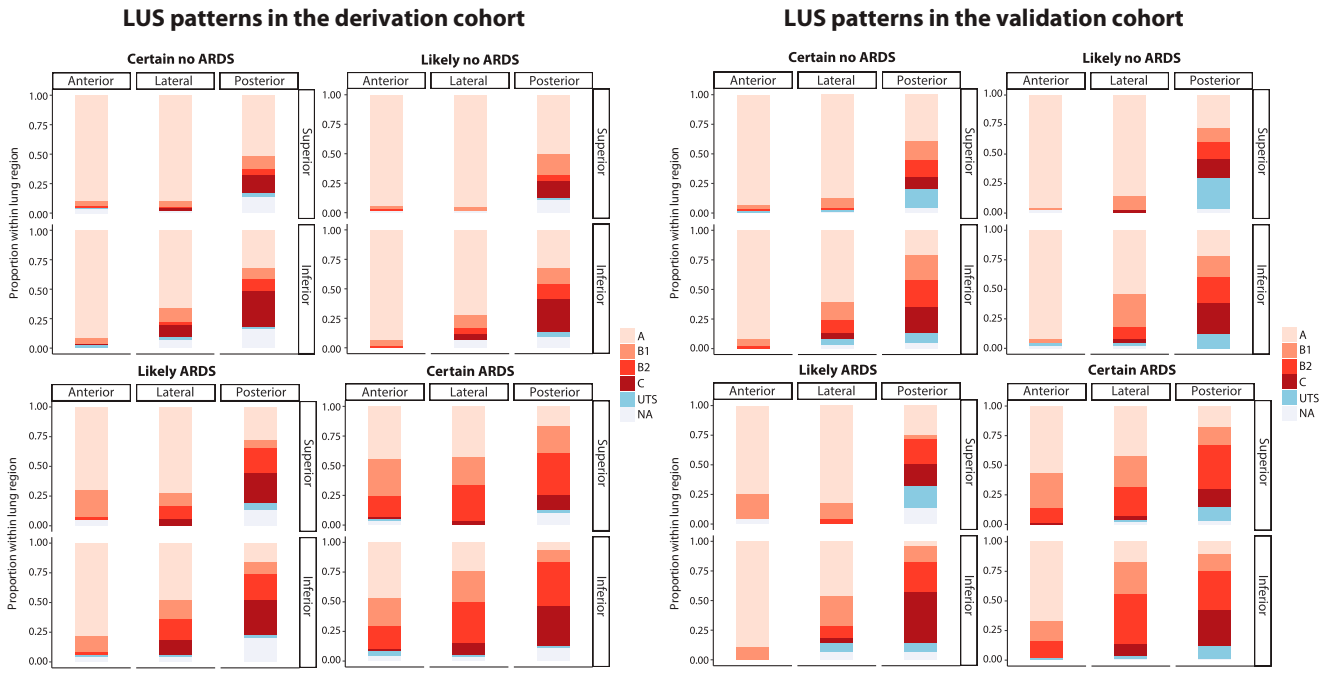

Figure 2. Distribution of lung ultrasound (LUS) patterns for different lung regions in patients from the derivation cohort ($n = 324$) and the validation cohort ($n = 129$). Patients underwent a LUS exam with six regions scanned per hemithorax. Patterns found in these regions for both lungs combined are presented in this figure and are stratified for ARDS category. ARDS = acute respiratory distress syndrome; NA = not available; UTS = unable to score.

model was determined on the basis of model discrimination using the C statistic—numerically equal to the area under the receiver operating characteristic curve (AUROCC)—and model calibration using the calibration curve and Brier score. From the final model, the LUS-ARDS score was derived. For the LUS-ARDS score, weight factors were calculated by multiplying logistic regression coefficients by 10 and rounding them to the nearest 0.5 decimal. The LUS-ARDS score was calculated for individual patients by multiplication of the variables that were finally included in the model by the corresponding weight factor. Two cutoffs for the LUS-ARDS score were chosen on the basis of the receiver operating characteristic (ROC) curve in the derivation cohort, with a low cutoff corresponding to a high sensitivity and a high cutoff corresponding to a high specificity.

Performance of the LUS-ARDS score was tested in patients from the

Table 3. Development of the Lung Ultrasound-Acute Respiratory Distress Syndrome Score

Variable	β (95% CI)	*P* Value	LUS-ARDS Points
Left LUS aeration score (0–18)	0.25 (0.10, 0.41)	0.002	+2.5 per unit
Right LUS aeration score (0–18)	0.11 (−0.06, 0.28)	0.185	+1 per unit
No. of anterolateral regions with pleural line abnormalities (0–8)	0.34 (0.07, 0.61)	0.013	+3.5 per unit
LUS-ARDS score = 2.5 × left LUS aeration score + 1 × right LUS aeration score + 3.5 × no. of anterolateral regions with an abnormal pleural line			
Examples from the present study	Patient presenting with extrapulmonary sepsis and scored as certain ARDS by the expert panel (Figure E9) • Left and right LUS aeration score were 9 and 11, respectively. • Eight anterolateral lung regions showed an abnormal pleural line. • **LUS-ARDS score: 2.5 × 9 + 1 × 11 + 3.5 × 8 = 62**		
	Patient presenting with cardiac arrest and scored as certain no ARDS by the expert panel (Figure E10) • Left and right LUS aeration scores were 0 and 4, respectively. • No anterolateral lung regions showed an abnormal pleural line. • **LUS-ARDS score: 2.5 × 0 + 1 × 4 + 3.5 × 0 = 4**		

Definition of abbreviations: ARDS = acute respiratory distress syndrome; CI = confidence interval; LUS = lung ultrasound.

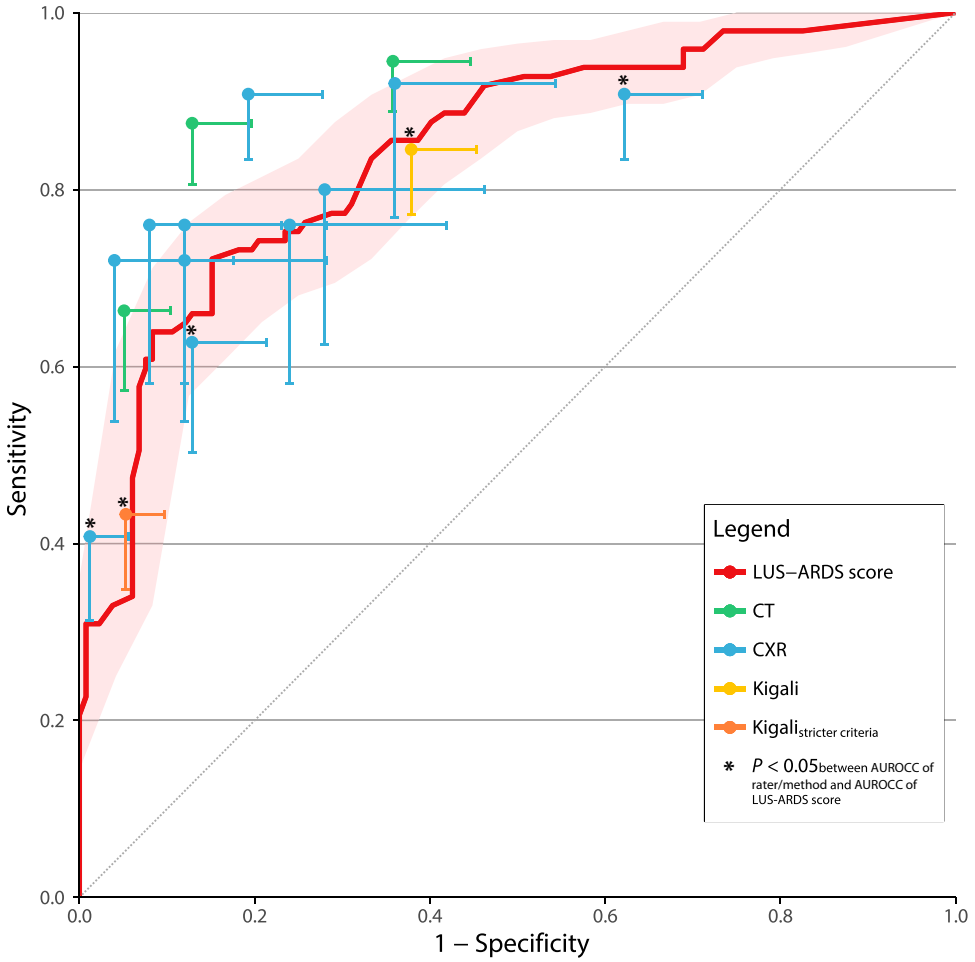

Figure 3. Diagnostic accuracy of *1*) the lung ultrasound-acute respiratory distress syndrome (LUS-ARDS) score, *2*), the current practice with single-rater assessment of the Berlin definition using chest radiography (CXR) and computed tomography (CT), and *3*) two previously published LUS methods for diagnosis of ARDS. Diagnostic accuracy was assessed in patients from both the derivation and validation cohorts who had a CT scan available, as CT is considered the most reliable imaging modality ($n = 229$). ARDS diagnosis by the expert panel was used as the reference standard. The area under the receiver operating characteristic curve (AUROCC) of the LUS-ARDS score in this group was 0.84; 95% confidence interval (CI), 0.79–0.89 (red). The three raters of the expert panel scored the Berlin definition using CT scans in all 229 patients with a CT (green). For the purpose of this study, and using CXR as the imaging modality, *1*) three raters of the expert panel scored the presence of ARDS in all 229 patients with a CT scan, *2*) seven independent ICU physicians scored the presence of ARDS in the same subset of 50 patients with a CT scan, and *3*) one chest radiologist scored the presence of ARDS in a subset of 121 patients with a CT scan from the derivation cohort. These assessments resulted in a total of 11 CXR raters (blue). The expert panel and independent raters had clinical and ventilation data available next to the images. We also assessed the diagnostic accuracy of two previous definitions of ARDS that included LUS criteria: the Kigali definition by Riviello and colleagues (10) (yellow) and the Kigali definition with stricter criteria by Vercesi and colleagues (11) (orange). The LUS criteria of these methods were compared with ARDS diagnosis by the expert panel in all 229 patients who had a CT scan available. The Kigali definition requires at least one region with a B- or C-pattern bilaterally. The Kigali definition with stricter criteria requires at least two B2- or C-patterns or three B1-patterns bilaterally. Lower CIs are presented for each rater and LUS method to assess potential overlap with the CI of the LUS-ARDS score. *$P < 0.05$, indicating significant differences between the AUROCC of a rater/method and the AUROCC of the LUS-ARDS score.

validation cohort by calculating sensitivity, specificity, and the AUROCC. Additionally, we assessed the performance of two previously published LUS methods for the diagnosis of ARDS (10, 11). Finally, the LUS-ARDS score was compared with the performance of eight experienced, independent physicians who used CXR as the imaging modality to score ARDS. Statistical differences in AUROCC between raters/methods and the LUS-ARDS score were assessed with a bootstrapping-based test (the pROC package).

Continuous data are presented as median and interquartile range (IQR). Categorical values are presented as number and percentage. Differences between groups were tested using the Mann-Whitney *U* test or Fisher's exact test. Data analysis was performed using R version 4.0.3 with the R studio interface. A P value <0.05 was considered statistically significant.

Results

Patients

A total of 519 patients were included, of whom 453 patients were analyzed (Figure 1). Of these patients, 324 were included in the

derivation (Amsterdam UMC) cohort, and 129 patients were included in the validation cohort (MUMC+) cohort (Figure 1). Patients were scanned at a median of 21 hours (IQR, 11–28) after the start of invasive ventilation in the derivation cohort and a median of 23 hours (IQR, 17–33) after the start of invasive ventilation in the validation cohort. Patient characteristics and ventilation parameters are presented in Tables 1 and 2 for the derivation and validation cohorts, respectively. Additional patient characteristics are presented in Tables E2–E8 of the online supplement. In patients with certain ARDS, the majority of patients had a chest CT available (ARDS, 73%; likely ARDS, 43%; likely no ARDS, 43%; no ARDS, 45%). An overview of the distribution of LUS patterns over the lung regions is presented in Figure 2 for the derivation and validation cohorts. Frequency of other LUS findings (e.g., pleural line abnormalities and pleural effusions) and an overview of missing regions are presented in the online supplement (Figures E1 and E2; Table E9).

Derivation of the LUS-ARDS Score

A detailed approach for model development and calibration and discrimination parameters from the logistic regression model are presented in the online supplement (Figures E3–E8).

The three variables that were included in the final logistic regression model for diagnosis of ARDS were left LUS aeration score (range = 0–18), right LUS aeration score (range = 0–18), and the number of anterolateral lung regions with an abnormal pleural line (range = 0–8). These variables and their corresponding logistic regression coefficients were transformed into the LUS-ARDS score (Table 3). The LUS-ARDS score can range from 0 to 91. Interobserver and intraobserver agreement for the LUS-ARDS score are presented in the online supplement. Examples of its practical calculation are shown in Table 3 and in Figures E9 and E10.

Diagnostic Accuracy of the LUS-ARDS Score in the Derivation Cohort

The LUS-ARDS score had high discriminative performance in diagnosing ARDS when applied to patients with "certain ARDS" labels in the derivation cohort with an AUROCC of 0.90 (95% confidence interval [95% CI], 0.85–0.95). Discriminative performance of the LUS-ARDS decreased to an AUROCC of 0.83 (95% CI, 0.77–0.88) when applied to all patients in the derivation cohort, which included an uncertain ARDS diagnosis. The low cutoff for the LUS-ARDS score was set at 8, resulting in a sensitivity of 0.94 (95% CI, 0.87–1.00) and a specificity of 0.56 (95% CI, 0.49–0.63) in patients with "certain ARDS" labels. The high cutoff for the LUS-ARDS score was set at 27, resulting in a sensitivity of 0.74 (95% CI, 0.60–0.85) and a specificity of 0.94 (95% CI, 0.90–0.97) in patients with "certain ARDS" labels (Table E10). In all patients of the derivation cohort (certain and uncertain ARDS diagnoses), 122 (38%) patients had a LUS-ARDS score between 8 and 27.

Diagnostic Accuracy of the LUS-ARDS Score in the Validation Cohort

In the validation cohort, the LUS-ARDS score had good discrimination performance in diagnosing ARDS when applied to patients with "certain ARDS" labels (AUROCC, 0.85; 95% CI, 0.77–0.93) and to all patients (AUROCC, 0.80; 95% CI, 0.72–0.87). Consistent with the derivation cohort, the low cutoff of the LUS-ARDS score resulted in a sensitivity of 0.98 (95% CI, 0.93–1.00) and a specificity of 0.39 (95% CI, 0.26–0.52), and the high cutoff resulted in a sensitivity of 0.55 (95% CI, 0.41–0.68) and a specificity of 0.91 (95% CI, 0.83–0.98) in patients with "certain ARDS" labels. Application to all patients of the validation cohort resulted in similar diagnostic characteristics (Table E10). Of all patients in the validation cohort (certain and uncertain ARDS diagnoses), 67 (52%) patients had a LUS-ARDS score between 8 and 27.

LUS-ARDS Score Is Equally Accurate for ARDS Diagnosis as Currently Available Methods

In patients from the derivation and validation cohorts who had a CT scan available ($n = 229$), which is considered to be the gold standard imaging modality, the AUROCC of the LUS-ARDS score was 0.84 (95% CI, 0.79–0.89) (Figure 3). Within patients from the validation cohort with a CT scan available ($n = 65$), the AUROCC was 0.82 (95% CI, 0.72–0.93). Sensitivities and specificities of ARDS diagnosis by 11 readers using the Berlin definition with CXR as the imaging modality followed the ROC curve of the LUS-ARDS score (Figure 3). Evaluations of eight of those readers were not included in the consensus diagnosis of ARDS, and the only reader with higher accuracy, although not significant, was involved in the expert panel and thus biased toward higher accuracy. Two previously proposed LUS methods for diagnosis of ARDS showed good performance, but there was an imbalance in sensitivity and specificity (Figure 3).

The LUS-ARDS Score Improves Diagnostic Accuracy When Diagnosis by the Expert Panel Is Uncertain

When a diagnosis of ARDS based on the expert panel was uncertain, a higher LUS-ARDS score was associated with more frequent diagnosis of ARDS in an expert panel consensus meeting ($P < 0.001$; Table E11). In these patients, a LUS-ARDS score below 8 only resulted in a consensus diagnosis of ARDS in 10 of 41 (24%) patients, whereas a LUS-ARDS score above 27 resulted in a consensus diagnosis in 18 of 20 (90%) patients. Clinical and outcome characteristics of patients with uncertain ARDS are presented in the online supplement (Table E11).

Discussion

In this multicenter observational study, we found that the LUS-ARDS score, which combined assessment of left and right LUS aeration scores together with the presence of pleural line abnormalities in anterolateral lung regions, provided good diagnostic accuracy for ARDS—also after external validation. The LUS-ARDS score showed high diagnostic accuracy in a subgroup of patients who had a chest CT scan available and had performance comparable with ARDS diagnosis by experienced physicians who used CXR as the imaging method. In patients for whom the ARDS diagnosis was uncertain because of conflicting evaluations by experts, the LUS-ARDS score was able to provide additional information and identify patients with low and high likelihoods of ARDS, as defined through a consensus meeting.

The LUS-ARDS Score

The LUS-ARDS score developed in the present study includes variables that were similar to those used previously, but the weight that was put on these variables is different and based on data rather than

expert opinion. For example, until now, the diagnosis of ARDS by LUS relied on the presence of bilateral abnormalities or specific LUS findings like pleural line abnormalities (10, 11, 21), which was common but not universally necessary for the diagnosis of ARDS in the LUS-ARDS score. Posterior pleural line abnormalities were not included in the model, because they did not improve the diagnostic accuracy, which is in line with the observations of another study (28). The reason why the left LUS aeration score attributes more to the LUS-ARDS score than the right LUS aeration score is unknown. A plausible explanation could be that right-sided consolidations are typically more prevalent in pulmonary ARDS, resulting in a lower β coefficient for the right lung (29).

The LUS-ARDS Score in Comparison with Other Imaging Methods

Assessment of the probability of a condition with a continuous score like the LUS-ARDS score is considered superior to arbitrary dichotomous methods, as dichotomization loses all the information about the certainty of diagnosis, and this information is especially crucial in syndromes that are complex to diagnose, like ARDS (24, 25). Previously described LUS methods were highly sensitive with low specificity (Kigali) or highly specific with low sensitivity (Kigali with stricter criteria) (10, 11). These diagnostic tests could therefore serve purposes similar to those of the cutoffs of the LUS-ARDS score identified here, but they fail to provide a continuous assessment of probability. Compared with the evaluation of the Berlin definition with CXR by independent readers, the LUS-ARDS score had similar or even higher diagnostic accuracy in patients who underwent the gold-standard chest CT scan. It should be appreciated, however, that the individual readers had widely varying test characteristics ranging from highly specific to highly sensitive. This is a direct consequence of the poor interrater reliability of the assessment of bilateral infiltrates on CXR. Furthermore, the approached clinicians arguably had more experience in evaluation of CXR for ARDS and were aware that their performance was evaluated, and this might have biased these results toward better accuracy.

Uncertainty in ARDS Diagnosis

Because diagnosing ARDS is challenging and observer dependent, we followed recommendations on improving an imperfect reference standard and used a panel of independent experts to classify patients as accurately as possible (30). Almost half of the patients had chest CT scans available, and this considerably decreased between-rater variability. However, a substantial group of patients was classified with an uncertain ARDS diagnosis that was due to disagreement between experts, a finding that is consistent with a previous study on disagreement in ARDS diagnosis (31). These are the patients for whom a new diagnostic test could yield the most benefit, and the LUS-ARDS score was able to correctly label patients who would be classified as having ARDS in a consensus meeting. This was achieved even though the groups were not different in terms of clinical characteristics and lung injury prediction score (32). These findings imply that the LUS-ARDS score can add diagnostic confidence with regard to patients for whom expert raters are uncertain of ARDS diagnosis.

Strengths and Limitations

Important strengths of the presented analysis include the large sample of consecutive patients and the use of a reference standard that acknowledges and corrects for high interobserver variability (31, 33). The study also has several limitations. First, we did not include all previously published LUS signs that are assumed to be typical for ARDS in the model, because we were limited in the number of variables in the model, and some of these LUS signs (such as spared regions) were not collected prospectively. Second, around 1 in 10 patients did not receive the index test because of four or more missing regions (i.e., because of wounds, chest drains, or subcutaneous emphysema), and this illustrates that a LUS-based diagnosis of ARDS is not possible in all patients. The analysis in this study was not adjusted for technical failure of LUS exams, because guidelines state that adjusting for technical failure is a choice dependent on the clinical context (34). Technical failure in 10% of the LUS exams does reflect clinical practice, and as LUS is considered an additional imaging tool, CT and CXR can still be used for diagnosis of ARDS in patients when LUS is unfeasible. Third, the validation cohort had a pragmatic sample size; namely, all patients who were included in one of the two centers. Fourth, comparing the LUS-ARDS score with the performance of experienced readers of CXR in patients from both the derivation and validation cohorts might provide an advantage for the LUS-ARDS score. We did, however, choose this approach because the number of patients with a CT scan in the validation cohort was limited, and the diagnostic accuracy of the LUS-ARDS score was only slightly lower in validation cohort patients. Fifth, the proportion of ARDS patients with nonpulmonary ARDS was low in the present study, which is in line with other ARDS studies but may limit the validity of the LUS-ARDS score in these patients (1, 35). Finally, the LUS exams in the present study were performed by three dedicated sonographers. Although the LUS technique in this study is widely used, we cannot directly imply validity and usability of the LUS-ARDS score in daily ICU practice when there is more interoperator variability.

The LUS-ARDS Score in Clinical Practice

How could the LUS-ARDS score be incorporated into clinical practice? The use of the LUS-ARDS score as a screening and diagnostic tool may improve the currently high number of underdiagnoses of ARDS in clinical practice and increase the use of appropriate treatment in these patients (1, 36). High and low LUS-ARDS scores can accurately diagnose and exclude ARDS, respectively, in a large proportion of patients. An intermediate LUS-ARDS score can provide direction on the probability of ARDS in conjunction with the clinical and physiological features of the patient. When a more certain imaging diagnosis on presence of ARDS is required, a CT scan can still be performed. LUS could also serve as the main diagnostic tool for ARDS in settings where access to CT scanners is limited, such as low-resource settings or in case of overloaded personnel or CT scanners. The major benefit of the LUS-ARDS score is that it can provide an estimate of the probability for ARDS that is consistent with the interpretation of the Berlin definition by three expert observers, a level of expertise that is usually absent in clinical practice (37). Moreover, the LUS-ARDS score showed performance similar to that of the assessment of CXR by experienced physicians, but without the variation in diagnostic accuracy between observers providing an objectifiable diagnostic accuracy at each cutoff. Indeed, the LUS-ARDS score and the LUS aeration score both have high interobserver agreement in the present cohort and

previous studies, which is an advantage over the complex interpretation of CXR (6, 33, 38–40). However, adequate LUS training remains essential before adaptation of the LUS-ARDS score in clinical practice (41–43). In the present study, the LUS-ARDS was not developed and validated with an imaging reference test but with the diagnosis of ARDS by an expert panel as the reference test to allow for the most accurate reference for ARDS diagnosis as possible. In clinical practice, the LUS-ARDS score will be integrated with other clinical parameters, and the pretest probability of ARDS will be considered before clinical decision making. Indeed, we show that a combination with Pa_{O_2}/Fi_{O_2} ratio and positive end-expiratory pressure levels already improves diagnostic accuracy further. In the future, the LUS-ARDS score may be further refined by data from centers with different ARDS prevalence and by the addition of other promising variables.

Conclusions

This study showed that a LUS score based on the left and right LUS aeration scores and the presence of an abnormal pleural line in the anterolateral lung regions can be used to accurately diagnose ARDS. The LUS-ARDS score showed performance comparable with that of ARDS diagnosis by experienced physicians using the Berlin definition with CXR, but with an objectifiable diagnostic accuracy at each cutoff. When further validated in other cohorts, the LUS-ARDS score may be considered as a useful adjunct to the diagnosis of ARDS. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

DARTS Consortium members: Lieuwe D. J. Bos, Laura A. Hagens, Marcus J. Schultz, Marry R. Smit, and Fleur L. I. M. van der Ven, Amsterdam University Medical Center, location Academic Medical Center, University of Amsterdam; Dennis C. J. J. Bergmans, Hester A. Gietema, Suzanne C. Gerretsen, Nanon F. L. Heijnen, and Ronny M. Schnabel, Maastricht University Medical Center+; and Inge Geven, Tamara M. E. Nijsen, and Alwin R. M. Verschueren, Philips Research.

References

1. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, *et al.*; LUNG SAFE Investigators; ESICM Trials Group. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* 2016;315:788–800.
2. Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, *et al.*; ARDS Definition Task Force. Acute respiratory distress syndrome: the Berlin definition. *JAMA* 2012;307:2526–2533.
3. Gattinoni L, Caironi P, Pelosi P, Goodman LR. What has computed tomography taught us about the acute respiratory distress syndrome? *Am J Respir Crit Care Med* 2001;164:1701–1711.
4. Rubenfeld GD, Caldwell E, Granton J, Hudson LD, Matthay MA. Interobserver variability in applying a radiographic definition for ARDS. *Chest* 1999;116:1347–1353.
5. Mojoli F, Bouhemad B, Mongodi S, Lichtenstein D. Lung ultrasound for critically ill patients. *Am J Respir Crit Care Med* 2019;199:701–714.
6. Chiumello D, Mongodi S, Algieri I, Vergani GL, Orlando A, Via G, *et al.* Assessment of lung aeration and recruitment by CT scan and ultrasound in acute respiratory distress syndrome patients. *Crit Care Med* 2018;46: 1761–1768.
7. Smit MR, Pisani L, de Bock EJE, van der Heijden F, Paulus F, Beenen LFM, *et al.*; Lung Ultrasound Consortium. Ultrasound versus computed tomography assessment of focal lung aeration in invasively ventilated ICU patients. *Ultrasound Med Biol* 2021;47:2589–2597.
8. Lichtenstein DA, Mezière G, Lascols N, Biderman P, Courret J-P, Gepner A, *et al.* Ultrasound diagnosis of occult pneumothorax. *Crit Care Med* 2005;33:1231–1238.
9. Mongodi S, Via G, Girard M, Rouquette I, Misset B, Braschi A, *et al.* Lung ultrasound for early diagnosis of ventilator-associated pneumonia. *Chest* 2016;149:969–980.
10. Riviello ED, Kiviri W, Twagirumugabe T, Mueller A, Banner-Goodspeed VM, Officer L, *et al.* Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition. *Am J Respir Crit Care Med* 2016;193:52–59.
11. Vercesi V, Pisani L, van Tongeren PSI, Lagrand WK, Leopold SJ, Huson MMA, *et al.*; Lung Ultrasound Consortium. External confirmation and exploration of the Kigali modification for diagnosing moderate or severe ARDS. *Intensive Care Med* 2018;44:523–524.
12. Copetti R, Soldati G, Copetti P. Chest sonography: a useful tool to differentiate acute cardiogenic pulmonary edema from acute respiratory distress syndrome. *Cardiovasc Ultrasound* 2008;6:16.
13. Volpicelli G, Elbarbary M, Blaivas M, Lichtenstein DA, Mathis G, Kirkpatrick AW, *et al.*; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. *Intensive Care Med* 2012;38:577–591.
14. Lichtenstein DA, Mezière GA. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. *Chest* 2008; 134:117–125.
15. Brusasco C, Santori G, Tavazzi G, Via G, Robba C, Gargani L, *et al.*; UCARE (Ultrasound in Critical care and Anesthesia Research Group). Second-order grey-scale texture analysis of pleural ultrasound images to differentiate acute respiratory distress syndrome and cardiogenic pulmonary edema. *J Clin Monit Comput* 2022;36:131–140.
16. Smit MR, Hagens LA, Heijnen NFL, Pisani L, Paulus F, Schnabel RM, *et al.* Lung ultrasound for diagnosis of acute respiratory distress syndrome—a prospective observational study [abstract]. *Am J Respir Crit Care Med* 2022;205:A5733.
17. Heldeweg MLA, Smit MR, Kramer-Elliott SR, Haaksma ME, Smit JM, Hagens LA, *et al.* Lung ultrasound signs to diagnose and discriminate interstitial syndromes in ICU patients: a diagnostic accuracy study in two cohorts. *Crit Care Med* 2022;50:1607–1617.
18. Hagens LA, Heijnen NFL, Smit MR, Verschueren ARM, Nijsen TME, Geven I, *et al.* Diagnosis of acute respiratory distress syndrome (DARTS) by bedside exhaled breath octane measurements in invasively ventilated patients: protocol of a multicentre observational cohort study. *Ann Transl Med* 2021;9:1262.
19. 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.
20. 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: 341–347.
21. See KC, Ong V, Tan YL, Sahagun J, Taculod J. Chest radiography versus lung ultrasound for identification of acute respiratory distress syndrome: a retrospective observational study. *Crit Care* 2018;22:203.
22. Hagens LA, Van der Ven FLIM, Heijnen NFL, Smit MR, Gietema HA, Gerretsen SC, *et al.* Improvement of an interobserver agreement of ARDS diagnosis by adding additional imaging and a confidence scale. *Front Med (Lausanne)* 2022;9:950827.
23. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
24. Steyerberg E. Clinical prediction models. New York, NY: Springer New York; 2009.
25. Leisman DE, Harhay MO, Lederer DJ, Abramson M, Adjei AA, Bakker J, *et al.* Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med* 2020;48:623–633.

26. Leopold SJ, Ghose A, Plewes KA, Mazumder S, Pisani L, Kingston HWF, *et al.* Point-of-care lung ultrasound for the detection of pulmonary manifestations of malaria and sepsis: an observational study. *PLoS One* 2018;13:e0204832.
27. Pisani L, De Nicolo A, Schiavone M, Adeniji AO, De Palma A, Di Gennaro F, *et al.* Lung ultrasound for detection of pulmonary complications in critically ill obstetric patients in a resource-limited setting. *Am J Trop Med Hyg* 2020;104:478–486.
28. Pisani L, Vercesi V, van Tongeren PSI, Lagrand WK, Leopold SJ, Huson MAM, *et al.*; Lung Ultrasound Consortium. The diagnostic accuracy for ARDS of global versus regional lung ultrasound scores—a post hoc analysis of an observational study in invasively ventilated ICU patients. *Intensive Care Med Exp* 2019;7:44.
29. Goodman LR, Fumagalli R, Tagliabue P, Tagliabue M, Ferrario M, Gattinoni L, *et al.* Adult respiratory distress syndrome due to pulmonary and extrapulmonary causes: CT, clinical, and functional correlations. *Radiology* 1999;213:545–552.
30. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009;62:797–806.
31. Sjoding MW, Hofer TP, Co I, McSparron JI, Iwashyna TJ. Differences between patients in whom physicians agree and disagree about the diagnosis of acute respiratory distress syndrome. *Ann Am Thorac Soc* 2019;16:258–264.
32. Trillo-Alvarez C, Cartin-Ceba R, Kor DJ, Kojicic M, Kashyap R, Thakur S, *et al.* Acute lung injury prediction score: derivation and validation in a population-based sample. *Eur Respir J* 2011;37:604–609.
33. Goddard SL, Rubenfeld GD, Manoharan V, Dev SP, Laffey J, Bellani G, *et al.* The randomized educational acute respiratory distress syndrome diagnosis study: a trial to improve the radiographic diagnosis of acute respiratory distress syndrome. *Crit Care Med* 2018; 46:743–748.
34. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
35. Constantin J-M, Jabaudon M, Lefrant J-Y, Jaber S, Quenot J-P, Langeron O, *et al.*; AZUREA Network. Personalised mechanical ventilation tailored to lung morphology versus low positive end-expiratory pressure for patients with acute respiratory distress syndrome in France (the LIVE study): a multicentre, single-blind, randomised controlled trial. *Lancet Respir Med* 2019;7:870–880.
36. Bellani G, Pham T, Laffey JG. Missed or delayed diagnosis of ARDS: a common and serious problem. *Intensive Care Med* 2020;46:1180–1183.
37. Sjoding MW, Hofer TP, Co I, Courey A, Cooke CR, Iwashyna TJ. Interobserver reliability of the Berlin ARDS definition and strategies to improve the reliability of ARDS diagnosis. *Chest* 2018;153:361–367.
38. Mongodi S, Bouhemad B, Orlando A, Stella A, Tavazzi G, Via G, *et al.* Modified lung ultrasound score for assessing and monitoring pulmonary aeration. *Ultraschall Med* 2017;38:530–537.
39. Smit MR, de Vos J, Pisani L, Hagens LA, Almondo C, Heijnen NFL, *et al.*; DARTS Consortium. Comparison of linear and sector array probe for handheld lung ultrasound in invasively ventilated ICU patients. *Ultrasound Med Biol* 2020;46:3249–3256.
40. Winkler MH, Touw HR, van de Ven PM, Twisk J, Tuinman PR. Diagnostic accuracy of chest radiograph, and when concomitantly studied lung ultrasound, in critically ill patients with respiratory symptoms: a systematic review and meta-analysis. *Crit Care Med* 2018;46: e707–e714.
41. See KC, Ong V, Wong SH, Leanda R, Santos J, Taculod J, *et al.* Lung ultrasound training: curriculum implementation and learning trajectory among respiratory therapists. *Intensive Care Med* 2016;42: 63–71.
42. Arbelot C, Dexheimer Neto FL, Gao Y, Brisson H, Chunyao W, Lv J, *et al.*; APECHO Study Group. Lung ultrasound in emergency and critically ill patients: number of supervised exams to reach basic competence. *Anesthesiology* 2020;132:899–907.
43. Rouby J-J, Arbelot C, Gao Y, Zhang M, Lv J, An Y, *et al.*; APECHO Study Group. Training for lung ultrasound score measurement in critically ill patients. *Am J Respir Crit Care Med* 2018; 198:398–401.