Lung Ultrasound Score to Predict Surfactant Administration in Premature Neonates with Respiratory Distress Syndrome: A Systematic Review and Meta-analysis

Mengke Xie [1], Biying Deng [1] and Song Li [1, *]

[1]Department of Neonatology, Dongguan Children's Hospital Affiliated to Guangdong Medical University, Dongguan, China

Corresponding author: Department of Neonatology, Dongguan Children's Hospital Affiliated to Guangdong Medical University, Dongguan, China. Email: lisong@ba-hospital.com

Abstract

Context: The present systematic review aimed to investigate whether the lung ultrasound score (LUS) can accurately predict surfactant administration in premature neonates with respiratory distress syndrome (RDS).

Methods: The systematic review was conducted according to the Cochrane collaboration or the preferred reporting items for systematic review and meta-analyses (PRISMA) guidelines. English language databases included PubMed, Ovid, Cochrane Library, Embase, Web of Science, and Scopus. The coverage date of this review was from the inception of each database to the end of May 2022. Data were extracted independently by two authors, assessed for quality using the quality assessment of diagnostic accuracy studies-2 (QUADAS-2) tool, and analyzed for heterogeneity using MetaDisc1.4 software.

Results: Seven eligible articles and 653 neonates were finally selected from 176 studies for meta-analysis. Considering the LUS to predict the surfactant need in premature neonates, we extracted the variable from the studies. Also, we plotted the summary receiver operating characteristic curve with an area under the curve (AUC) of 0.94 (95% confidence interval (CI): 0.92 - 0.95). Forest plots of the LUS showed a pooled sensitivity of 0.86 (95% CI: 0.82 - 0.90) and a pooled specificity of 0.79 (95% CI: 0.75 - 0.82). The Cochrane-Q test, chi-square test, and inconsistency index confirmed the heterogeneity of the non-threshold effect ($I^2 > 50\%$ or $P < 0.05$). The meta-regression analysis showed that the relative diagnostic odds ratio for the number of recruited cases (> 100 vs. < 100) was 0.65 (95% CI: 0.33 - 0.98, P-value < 0.05). The comparison of the cut-offs of the 4 - 6 score vs. the 8 - 12 score yielded a Z of 21.44 (P < 0.001).

Conclusions: The LUS accurately predicts the onset of RDS in premature neonates and can guide surfactant administration but is subject to the cut-off effect. Variation in cut-offs is related to gestational age and disease severity.

Keywords: Lung Ultrasound Score, Surfactant, Meta-analysis, Heterogeneity, Neonatal Respiratory Distress Syndrome

1. Context

Neonatal respiratory distress syndrome (nRDS) is one of the most common life-threatening diseases in the neonatal intensive care unit (NICU) for premature neonates. According to Berman's study, the incidence of respiratory distress syndrome (RDS) in premature neonates ranges from 86% at 24 weeks to less than 1% at 39 weeks (1). The diagnosis of RDS is based on clinical symptoms or symptoms with chest X-ray (CXR) and arterial blood gas findings (2), and CXR is considered the first choice for diagnosing RDS. However, premature neonates in long-term care units with repeated exposure to CXR are susceptible to ion damage or have an increased risk of malignancy (3). Apart from CXR, phospholipid maturation of amniotic fluid can be measured to anticipate the use of surfactant, such as the lecithin/sphingomyelin (L/S) ratio or lamellar body count, which is derived from fetal alveolar type 2 epithelial cells and is measured similar to platelet calculation (4). However, these semi-quantitative methods can be helpful in modern respiratory management, nursing care, and treatment (5-7).

Exogenous surfactant is considered the cornerstone of the nRDS treatment, although there is no complete agreement on the best criteria for using this life-saving medication (8). Patients inappropriately treated with surfactant may experience side effects, including

bradycardia, decreased oxygen saturation, and pulmonary hemorrhage (9, 10). Early treatment with continuous positive airway pressure (CPAP) and selective surfactant in the setting of nRDS in premature neonates has been shown to reduce mortality and chronic lung disease (11). Thus, the European Society of Pediatric Research developed criteria for selective surfactant use in premature neonates treated with CPAP to prevent lung injury due to tracheal intubation and high oxygen concentrations (12). The consensus of European clinicians using surfactant is that it is recommended when naturally breathing neonates require a fraction of inspired oxygen (FiO$_2$) > 0.3 to remain in the normal saturation range with nasal CPAP (nCPAP), 6 cm with H$_2$O support. Certainly, in some countries, such as India and China, the predictive diagnostic criteria for RDS are different from those in Europe, where traditional methods such as clinical signs coupled with CXR are still in use (13). Usually, CXR is graded to show changes from 1 to 4; however, clinically, if a patient has progressive dyspnea, they may have ground-glass appearance changes on CXR, and an experienced clinician may treat the lung using surfactant or mechanical ventilation, but this predictive criterion proves somewhat vague.

In any case, the criteria for determining surfactant replacement therapy present a challenge (14). Currently, surfactant use is based on the changes in blood oxygen saturation (FiO$_2$), but the recommended FiO$_2$ cut-off does not strictly reflect the pathophysiological changes in the lung and ultimately ends up missing the optimal treatment window (15). Hence, it would be useful to have an early predictive tool for surfactant need to ensure timely surfactant administration. Some reports have already highlighted the usefulness of lung ultrasound score (LUS) in predicting NICU admission or the need for mechanical ventilation or nCPAP failure (16-18). Many scores based on LUS findings have proved their reliability in adult critical care (6, 19). Interestingly, comparing the merits of CXR with LUS is often preferred in the NICU (20). Perri et al. (21) concluded that LUS had higher sensitivity and specificity than CXR and better positive and negative predictive values. Despite the current consensus on ultrasound as a reliable tool for predicting the need for surfactant replacement, some issues still need to be resolved. In order to advance knowledge based on the existing evidence and to address the limitations of the previous reviews, we undertook a systematic review focused on evaluating the following questions:

(1) What is the LUS cut-off that best predicts the need for surfactant, and when is the optimal time window for the LUS performance?

(2) Is LUS superior in terms of accuracy, sensitivity, and specificity based on systematic evaluation?

(3) Is there some room for improvement in the current LUS?

2. Objectives

We seek to synthesize previous research under a unified framework that we hypothesize will be highly heterogeneous, will inform the development of new, evidence-based strategies for LUS trials, and will guide policy choices about how to direct future research and resources.

3. Materials and Method

3.1. Search Strategy

This systematic review and meta-analysis were conducted and reported in accordance with the guidelines from preferred reporting items for systematic reviews and meta-analyses (PRISMA) and diagnostic test accuracy guidelines (22). The systematic review protocol was registered in PROSPERO (the International Prospective Register of Systematic Reviews; registration number: CRD42022307223; PROSPERO). Six databases were searched: PubMed MEDLINE, Ovid Medline, Embase, Scopus, Web of Science, and the Cochrane Library. The coverage dates for this review began from the inception of each database (Medline, 1946; Embase, 1947; Scopus, 1966; Ovid Medline, 1946; and Cochrane Library, 1995; Web of Science, 1970) and ended in April 2022. Controlled vocabulary (i.e., Medical Subject Headings [MeSH] terms) and keywords were identified. According to the different search criteria of these databases, different search formulas were developed for each search using the Excel form.

3.2. Type of Participants and Diagnostic Test

All neonates admitted to the NICU requiring non-invasive ventilation for nRDS were eligible for recruitment. The diagnosis of nRDS was made by the attending neonatologist based on clinical signs. Neonates with significant chromosomal abnormalities, congenital anomalies, and those requiring surfactant therapy and intubation in the delivery room were excluded from the study. The surfactant replacement protocol was based on European guidelines: Surfactant was administered whenever FiO$_2$ was greater than 0.3 or 0.4 in neonates with a gestational age of less than or greater than 28 weeks. Otherwise, clinical signs, CXR, or blood tests were considered.

The LUS was performed according to the method proposed by Brat et al. (11). Each lung was divided into

three areas (upper anterior, lower anterior, and lateral) and examined with a linear microprobe through both transverse and longitudinal scans. Each lung area was scored from 0 to 3, with a total score of 0 - 18 (0 = A-lines and lung sliding; 1 = more than 3 B-lines, A-lines may be present; 2 = compact B-lines without consolidation, absence of A-lines; 3 = confluent B-lines or white lung with subpleural or focal consolidation).

Clinical signs include shallow breathing, tachypnea, grunting, nasal flaring, and intercostal retractions. The CXR findings include small lung volumes and diffuse reticulogranular ground glass with air bronchograms. Oxygen saturation/fraction of inspired oxygen ratio (SF ratio) and oxygenation index (OI) were calculated by transcutaneous oxygen saturation or blood gas analysis.

3.3. Study Selection

Two investigators (M.K.X. and B.Y.D.) independently screened the search results for potentially eligible studies. Before identifying the literature, the inclusion and exclusion criteria were defined to increase validity and reproducibility. Any discrepancies were resolved by discussion with the senior author (S. L.). Studies were included if they were diagnostic accuracy studies or prospective or retrospective cohorts. Studies were required to report one of the pre-specified outcomes described in this review. Case reports, case series (< 10 patients), conference abstracts, and letters to the editor were considered ineligible. Exclusion criteria included: (1) non-English language articles; (2) articles on patients with meta-analysis, duplicate literature, and abstracts; and (3) non-cohort or case-control studies.

3.4. Methodological Quality Assessment

Two authors (M.K.X and B.Y.D) used the quality assessment of diagnostic accuracy studies-2 (QUADAS-2) tool to assess the quality of the collected data (23). RevMan V.5.3.5 software was used to implement the QUADAS-2 tool, consisting of four domains: Patient selection, index test, reference standard, and flow and timing, each assessed for the risk of bias. The first three domains were also assessed for applicability concerns. The risk of bias was rated as low, high, or unclear. If all signal, questions in a domain are answered 'yes,' the risk of bias may be low. If any of the signal questions are answered, 'no,' bias is indicated. If there is not enough data to make a judgment, the risk of bias is reported as unclear but still considered high.

3.5. Statistical Analysis

Diagnostic accuracy parameters (true positive/negative and false positive/negative) and cut-offs were obtained from the original material and analyzed using MetaDiSc1.4 or STATA 15.0 software. In order to assess the appropriateness of the statistical summary of accuracy estimates across studies, summary receiver operating characteristic curves were plotted, and the area under the curve (AUC) was calculated, ranging from 0.5 to 1.0, with higher values indicating better diagnostic performance. Pooled sensitivity, pooled specificity, pooled likelihood ratio, and the diagnostic odds ratio were summarized from a random effects model.

The I^2 statistic values varied from 0% to 100% and were divided according to cut-offs of 25%, 50%, and 75%, representing low, medium, and high heterogeneity, respectively. Heterogeneity can arise from the cut-off effect, non-cut-off effect, publication bias, etc., and requires the design of subgroups to explore meta-regression or funnel plots (Deeks' funnel). In addition, sensitivity analysis can identify data significantly impacting the meta-analysis results and analyze the reasons for their heterogeneity (24, 25).

4. Results

4.1. Search Results

A total of 176 articles (653 cases) were identified from the six database searches, and two additional studies were identified from other reference lists (ClinicalTrials.gov). After removing reviews, animal studies, duplicates, and articles that were inconsistent with the study purpose, 77 full-text articles remained. Then, we excluded articles with insufficient data, unfinished clinical trials, cases, etc., and finally screened seven eligible articles (Figure 1). The literature in the databases was screened following the requirements of the PRISMA statement. The flow chart was divided into four stages (search, primary screening, inclusion, and synthesis), with M. K. X. and B. Y. D. conducting the search and primary screening and S. L. conducting the final synthesis results to identify the final selected articles.

4.2. Study Characteristics

As seen in Table 1, all studies, except one published in 2015, were observational or prospective cohort studies published or updated until 2021. Observational criteria included neonates with 26 to 36 weeks of age, but those with < 34 weeks were preferred. Five papers chose the European guidelines as a criterion to predict surfactant use (21, 26-29). Two papers (8, 11) showed criteria like the European guidelines, but not very explicit, choosing FiO_2 > 0.3 or 0.4 as a cut-off. Four papers (8, 21, 27, 28) chose CXR as a diagnostic tool to assess the severity of
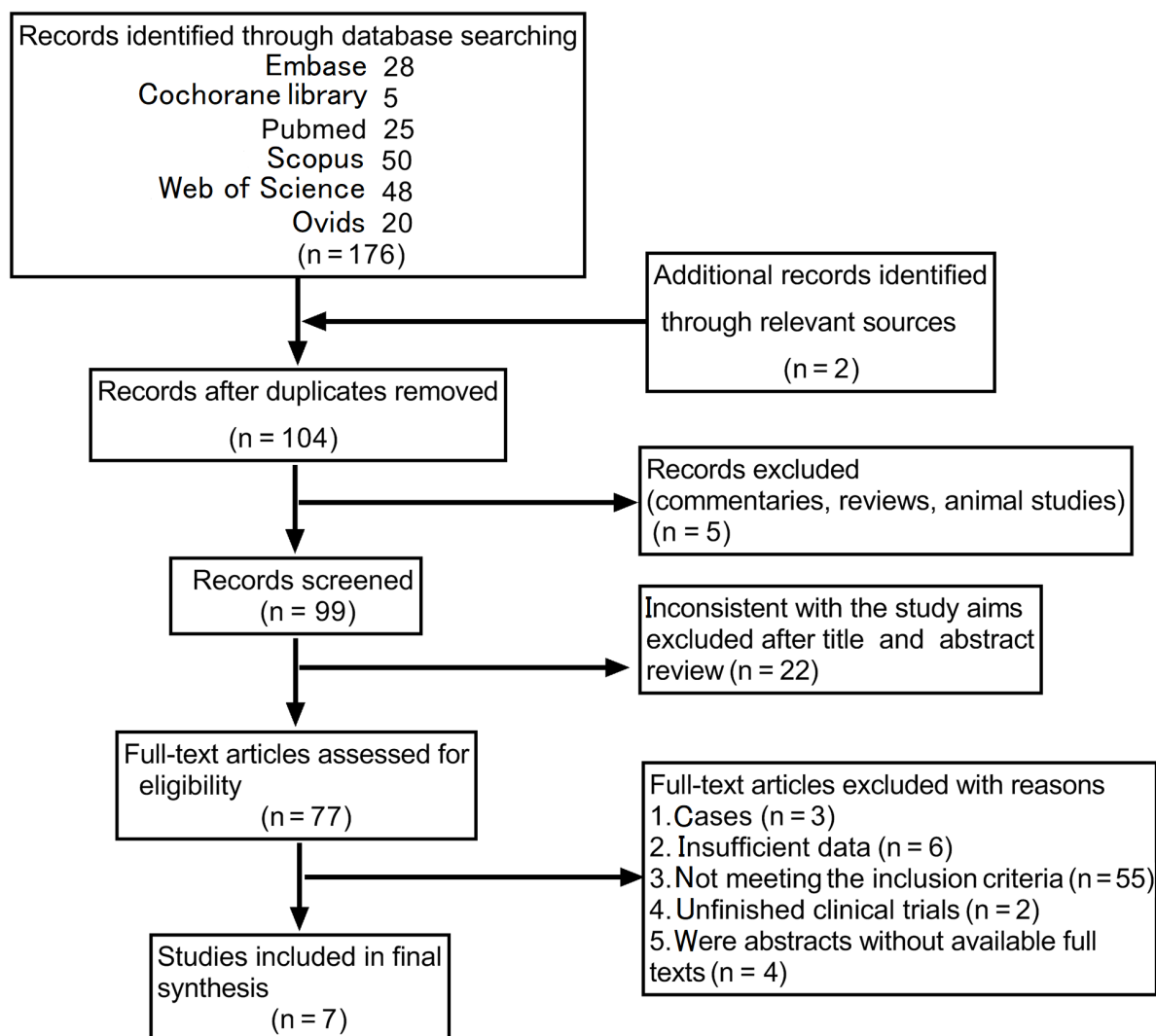
Figure 1. The study selection flowchart. Flowchart of the research process based on the preferred reporting items for systematic review and meta-analyses statement

RDS, all incorporating blood gas tests and clinical signs to determine whether surfactant should be used. Four studies (11, 21, 26, 27) examined the cut-offs of 4 - 6 score as a predictive cut-off for LUS for surfactant use, while three studies (8, 28, 29) examined those of 8 - 12 score as a predictive cut-off. As seen in Table 2, the selected articles were all clinical trials conducted with blinded, single-blinded, or double-blinded methods, and the raw data were consecutive. The evaluation of lung ultrasound was transthoracic, and almost all used a high-frequency probe (10 - 15 MHz), except for one article (11), which was unclear. Ultrasound physicians were either from the ultrasound ward or from the NICU, the latter requiring at

least 1 - 3 months of training to perform LUS. The results' characteristics have four studies (8, 11, 28, 29), with LUS performed 1 - 2 hours after birth and two studies (21, 27) 1 - 2 hours after CPAP support.

4.3. Quality Assessment and Risk of Bias

As shown in Figure 2, methodological quality was assessed according to the criteria listed in the QUADAS-2 tool. In general, the selected articles showed high quality and reliability, with only a few deviations and inapplicability (21, 26, 27). Specifically, the risk of bias in our quality assessment came from two sources: One study (21) could not avoid being written as a case-control study,

Table 1. The Main Data Extracted from the Included Studies and the Diagnostic Methods

Authors	Year	Origin	Study Types	Types Participants/Diagnostic test	Case (n)	Gestational Age, Mean ± SD (week)	Cut-offs	TP	FP	FN	TN
Aldecoa-Bilbao et al. (28)	2021	Spain	Prospective cohort	RDS/EG; CS; BT; CXR	50	28.7 ± 2.2	8	13	1	1	35
Brat et al. (11)	2015	France	Prospective cohort	RDS/EG unclear; CS; BT	65	< 34, 30 ± 2; > 34, 36 ± 2	4	31	56	1	42
Gregorio-Hernandez et al. (29)	2020	Spain	Observational studies	RDS/EG; CS; BT	64	29, median (26 - 32); 32, median (31 - 33)	12	15	4	1	44
De Martino et al. (26)	2018	France	Prospective cohort	RDS/EG; CS	133	28 ± 2	6	117	18	19	112
Perri et al. (21)	2018	Italy	Case-control	RDS/EG; CS; X-ray; BT	56	31 ± 3	5	19	4	3	29
Raimondi et al. (8)	2021	Italy	Prospective cohort	RDS/CS; BT, CXR; EG unclear	240	26.6 ± 6; 29.4 ± 6; 32.4 ± 6	9	85	22	23	110
Vardar et al. (27)	2021	Turkey	Prospective cohort	RDS/EG; CS; CXR; BT	45	30, median (27 - 32)	4	24	0	1	20

Abbreviations: RDS, respiratory distress syndrome; CS, clinical signs; CXR, chest X-ray; EG, Europe guideline; BT, blood test; wk, week; FP, false positive; FN, false negative; SD, standard deviation.

Table 2. Characteristics of the Included Studies

Authors	Year	Blinded Method	Sampling Method	LUS Operator	LUS Technique	LUS Timing	LUS Model	LUS Equipment
Aldecoa-Bilbao et al. (28)	2021	Single-blinded	Consecutive	NR	Transthoracic	1 - 2 h after birth	A semi-quantitative LUS score (0 - 18) [a]	A linear probe (VF 13 - 15 MHz)
Brat et al. (11)	2015	Double-blinded	Consecutive	Physicians	Transthoracic	1 h after birth	A semi-quantitative LUS score (0 - 18)	NR
Gregorio-Hernandez et al. (29)	2020	Double-blinded	Consecutive	Neonatologists	Transthoracic	2.5 h median (1.5 - 4.8)	A semi-quantitative LUS score (6 - 18) [b]	A linear probe (VF 15 MHz)
De Martino et al. (26)	2018	Single-blinded	Consecutive	Attending, senior fellows	Transthoracic	NR	A semi-quantitative LUS score (0 - 18)	A micro-linear probe (15 MHz)
Perri et al. (21)	2018	Single-blinded	Consecutive	Physicians	Transthoracic	2 h initiation of nCPAp	LUS score (0 - 18)	A linear probe (12 MHz)
Raimondi et al. (8)	2021	Single-blinded	Consecutive	Ultrasound physicians	Transthoracic	2 h after birth	A semi-quantitative LUS score (0 - 18)	A linear or micro-linear probe (10 - 15 MHz)
Vardar et al. (27)	2021	Double-blinded	Consecutive	Physicians	Transthoracic	After 1 hour of CPAP support	A semi-quantitative LUS score (0 - 18)	A linear probe $\geq$ 7.5 MHz

Abbreviations: LUS, lung ultrasound score; CPAP, continuous positive airway pressure; nCPAP, nasal continuous positive airway pressure; NR, not reported.

[a] Each lung region was scored on a scale of 0 - 3 with a total score of 0 - 18.

[b] Each normal lung area was given 1 point, and all normal lung areas were rated from 6 points.

and another28 did not include all cases. Second, and most importantly, one study (26) had an inappropriate time interval between the gold standard and the LUS. According to the European diagnostic criteria, postnatal treatment with a surfactant or CPAP support effectively prevents RDS in premature neonates (12). Based on this criterion, most articles suggest that the use of LUS screening 1 - 2 hours after birth effectively predicts the development of RDS and can guide the use of surfactants (30-32).

4.4. Data Synthesis and Heterogeneity

After summarizing the seven included studies, we pooled the data and then analyzed the sources of heterogeneity. Lung ultrasound score predicted the need for surfactant in premature neonates with a pooled sensitivity, pooled specificity, pooled positive likelihood ratio, pooled negative likelihood ratio, and diagnostic odds ratio of 0.86 (95% confidence interval (CI): 0.82 - 0.90; I^2 = 53.9%), 0.79 (95% CI: 0.75 - 0.82; I^2 = 93.6%), 7.25 (95% CI: 2.68 - 19.60; I^2 = 95.5%), 0.16 (95% CI: 0.11 - 0.24; I^2 = 34.9%), and 45.99 (95% CI: 21.30 - 99.30; I^2 = 50.1%), respectively

(Figure 3A-E). Significant heterogeneity was observed in the inconsistency index for both specificity and positive likelihood ratios, with $I^2 > 90\%$, whereas the Cochrane-Q or chi-square test yielded a p-value of < 0.05. Deeks' funnel plot asymmetry test suggested a publication bias for the studies with a P-value of 0.04 (Figure 4). Summary receiver operating characteristic curves were plotted using the random effects method (DerSimonian-Laird), yielding an AUC of 0.94 (95% CI: 0.92 - 0.95) (Figure 5). There was a correlation between sensitivity and 1-specificity with a Spearman correlation coefficient of -0.21 and a P-value of 0.65.

4.5. Meta-regression and Influence Analysis

Given the amount of heterogeneity found in the above analysis, meta-regression and grouping studies were used to explore the origin of heterogeneity. As shown in Table 3, co-variables were compared in subgroups according to the year of publication (2015 - 2018 vs. 2019 - 2021), number of cases (< 100 vs. > 100), study design (prospective vs. case-control) and neonate gestational age (< 34weeks vs. >
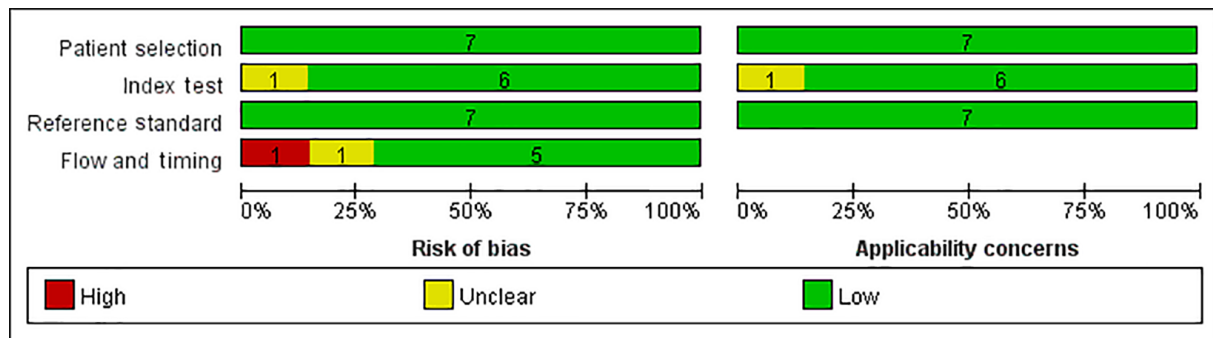
Figure 2. Risk of bias and applicability concerns diagram. The methodological quality of the articles was assessed using the quality assessment of diagnostic accuracy studies-2 tool and expressed as a percentage for the included studies.

34weeks). As a result, the 'cases' group showed differences with a p-value of 0.04 and relative diagnostic odds ratios (0.65, 95% CI: 0.33 - 0.98). As there were inconsistencies in the cut-offs of the LUS, comparing the cut-offs of the 4 - 6 score and 8 - 12 score groupings resulted in a Z value of 21.44 with a P-value < 0.01, indicating the presence of heterogeneity (Figure 6). Additional sensitivity analysis of the data, shown in Figure 7, revealed that outlier 2 had a relatively large impact on the robustness of the combined scoring results due to the excessive sensitivity of the original data (11).

5. Discussion

This systematic review aimed to assess the role of LUS in predicting surfactant therapy in premature neonates with RDS. The included studies were based on cohort studies of clinical trials where the bedside LUS was more typically used for evaluation, moving from an optional to a routine examination (33). Most of the included articles chose the European guidelines to evaluate the use of surfactant after the occurrence of RDS and then used lung ultrasound as the investigative metric for evaluation. However, some essential parameters, such as blood gas analysis, CXR, and clinical signs, cannot be overlooked. As experienced clinicians, they combined all the data to assess the severity of RDS. We summarized these studies and found that the LUS was good at evaluating pulmonary changes in the illness and accurately predicted the timing of surfactant use. Certainly, this does not mean that the LUS is free of controversy, so we would like to discuss the advantages and limitations of the LUS in detail.

As shown in Tables 1 and 2, they placed the LUS study in a tertiary NICU, but the ultrasound operation was not necessarily accomplished by the ultrasound physicians. Indeed, five studies (11, 21, 26, 27, 29)

chose to train their NICU physicians as ultrasound operators, and only Raimondi et al. (8) chose ultrasound physicians. Neonatologists receive at least 1 - 3 months of theoretical and practical training before the study, undergo testing, and perform ultrasound examinations under the supervision of senior physicians. To obtain clear ultrasound images of the lungs, the operators choose high-frequency hockey-stick linear probes for study in neonates. Theoretically, all neonates < 36 weeks are eligible for the study, except those suffering from congenital disease. In practice, neonates < 34 weeks are more likely to have RDS, while RDS in neonates > 34 weeks is not exclusively caused by RDS, as it may be caused by wet lung or pneumonia. Moreover, neonates > 34 weeks have relatively mature alveolar development and do not necessarily require surfactant treatment (34). Considering Tables 1 and 2, it can be seen that the cut-offs for LUS are not consistent and may produce heterogeneity.

We further evaluated the recruited studies against the four domains (patient selection, index test, reference standard, flow, and timing) and analyzed the results using the QUADAS-2 tool (Figure 2). Most patients received the same reference standard, allowing for the correct classification of the target disease. A single-blinded or double-blinded method was used to ensure the reliability of the test. A pre-January cohort study was used to ensure high article quality for the study type of articles. Overall, the seven articles selected were of high quality and data reliability. Only a few articles had minor problems with the selection of time intervals. For example, appropriate intervals between the gold standard and new methodological measurements were not reported, which may result in a high risk of bias. Our results show that prediction with LUS is best performed within 1 - 2 hours after birth or with CPAP support and not later than 24 hours because the LUS pattern in neonates can mutate,
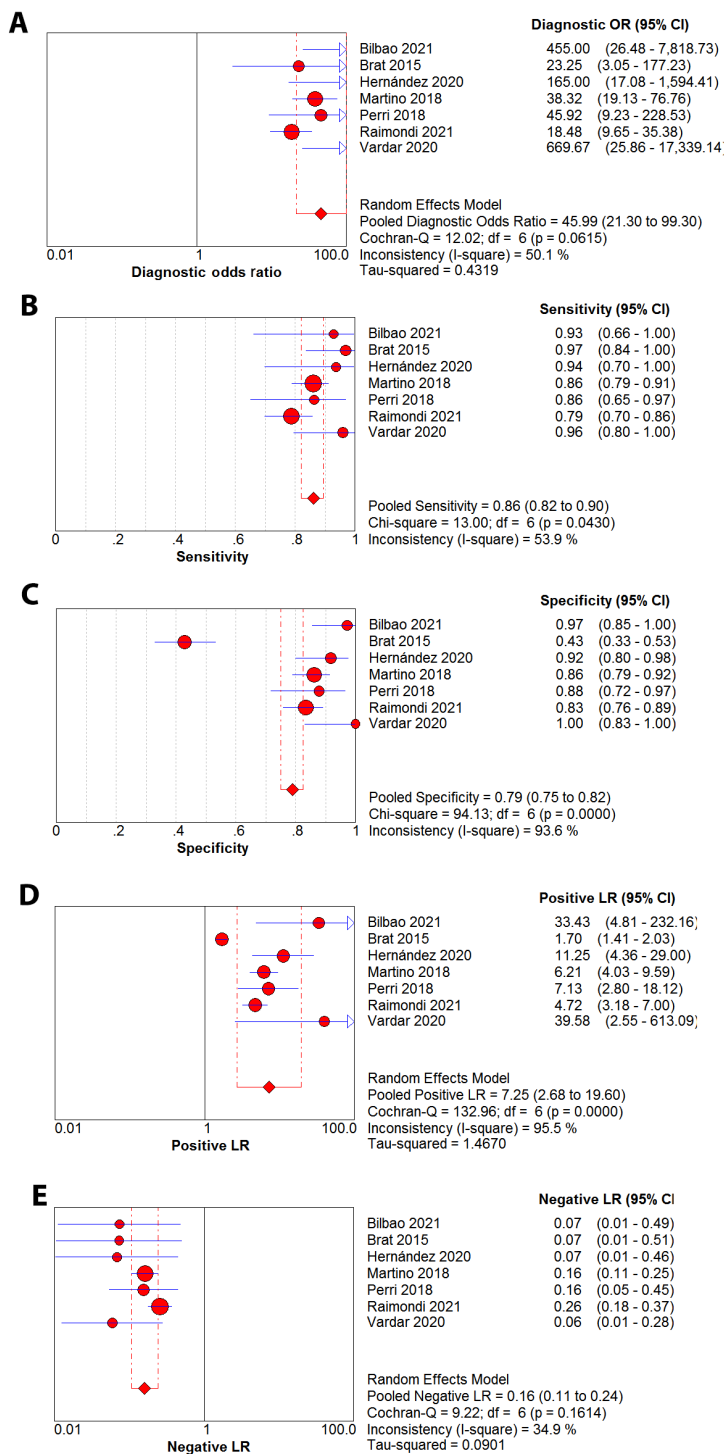
Figure 3. Diagnostic forest plot. A-E, forest diagrams of evaluation of lung ultrasound score, showing diagnostic odds ratio (DOR), pooled sensitivity, pooled specificity, pooled positive likelihood ratio (LR), and pooled negative LR
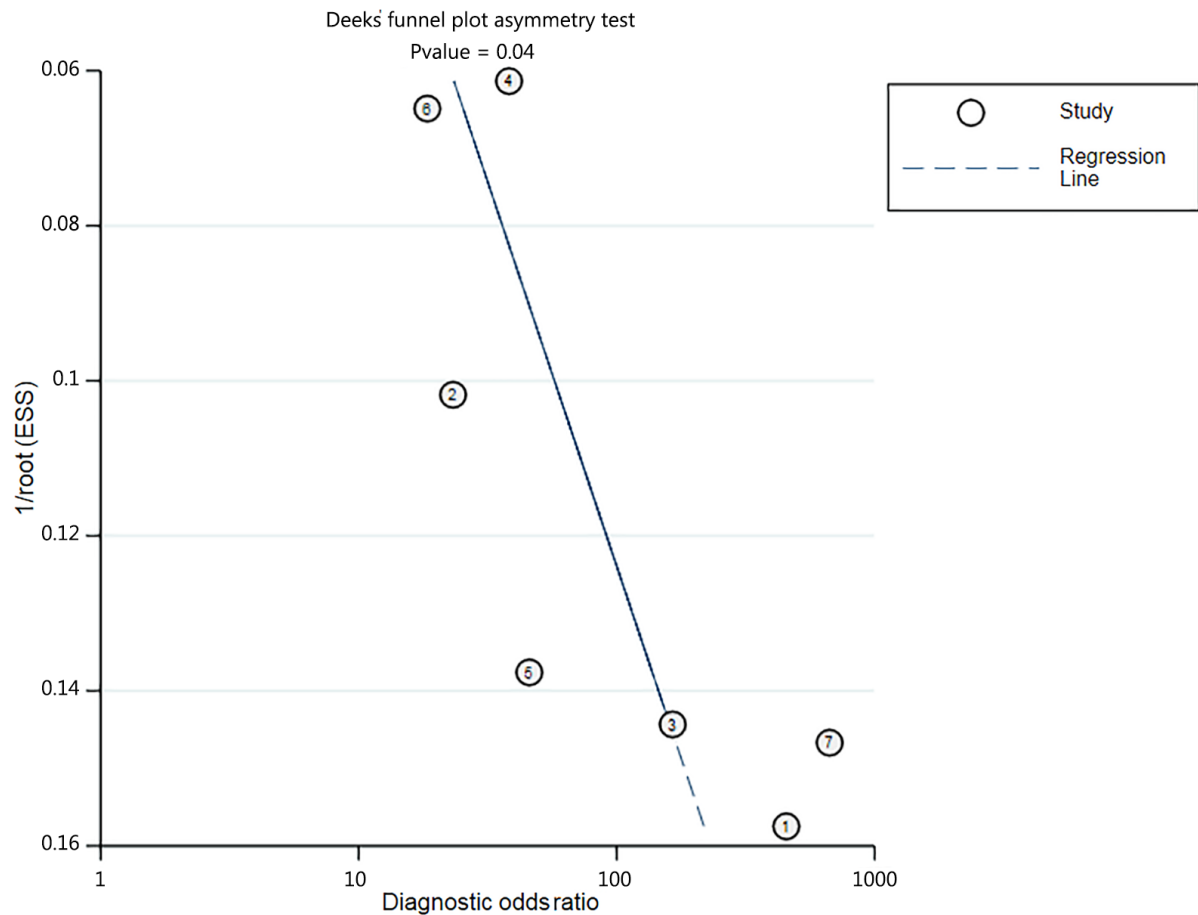
Figure 4. Deeks' funnel plot. Deeks' funnel plot displays the linear regression of the square root of the effective sample size (1/root (ESS)) on the diagnostic odds ratio. P = 0.04 indicates an asymmetric funnel shape, indicating publication bias.

Table 3. The Results of Meta-regression Analysis for Subgroup Comparisons

Co-variables	P-Value	RDOR	95% CI
Year of publication (2015 - 2018)/(2019 - 2021)	0.59	1.86	0.10 - 35.46
Number of cases (< 100 vs. > 100)	0.04 [a]	0.65	0.33 - 0.98
Study design (prospective vs. case-control)	0.79	0.53	0.01 - 41.9
GA (wk) (< 34 vs. > 34)/(< 28 vs. > 28)	0.92	0.80	0.00 - 231.46

Abbreviations: GA, gestational age; wk, week; RDOR, relative diagnostic odds ratios; CI, confidence interval.
[a] P < 0.05 indicates that the number of cases among data affects the results of meta-regression.

even so-called "black slips" (deterioration of the LUS grade) occur in the first hours of life. Complete clearance of airway fluid cannot be achieved in the first four hours. Premature neonates are more affected by this "black slip" and even seem to have fluid refill in the airways without end-expiratory pressure (35). Also, the changes are most pronounced 1 - 3 hours compared to 5 - 10 minutes after birth, probably due to ventilator fatigue (30). These findings help understand why most studies have chosen to perform the LUS two hours after birth around the LUS time to predict surfactant needs better.

Although the meta-analysis found good pooled sensitivity and specificity of 86% and 79%, respectively, with a diagnostic odds ratio of 45.99, reflecting the high diagnostic correctness of LUS, some parameters showed a high degree of heterogeneity, such as I^2
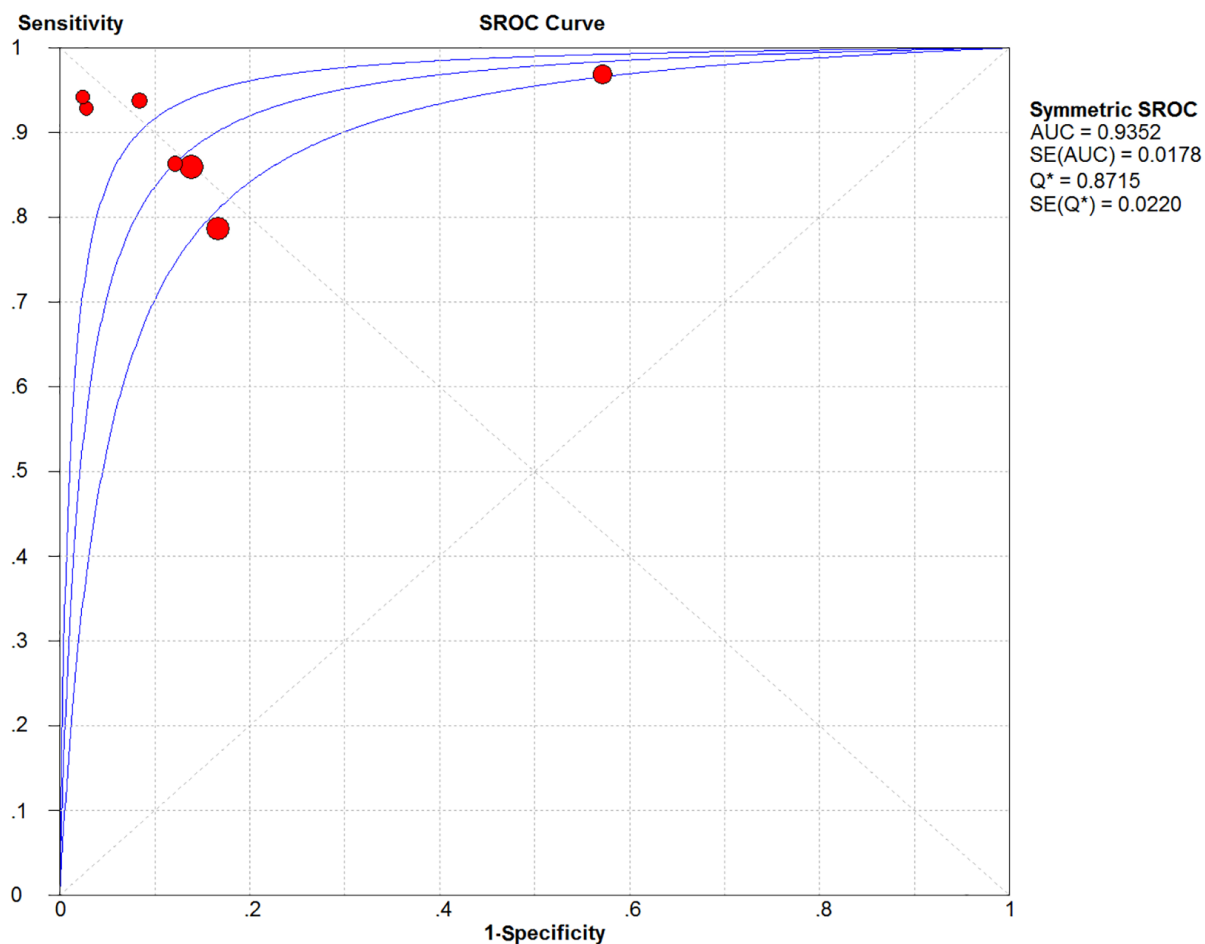
Figure 5. A summary receiver operating characteristic (SROC) curve. The pooled receiver operating characteristics curve summarized seven independent diagnostic trials of the same metric, plotted using the random-effects method (DerSimonian-Laird), yielding an area under a curve of 0.94 (95% confidence interval (CI): 0.92 - 0.95).
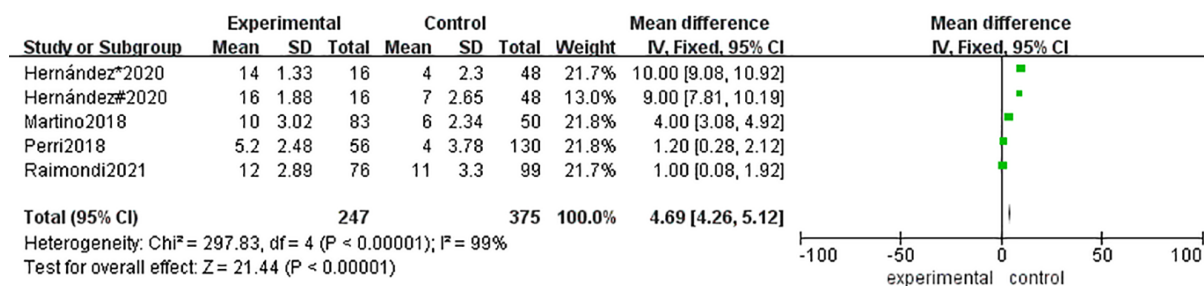

Figure 6. Experimental control forest plot. A comparison of the cut-offs of 4-6 score vs. 8 - 12 score yielded a Z value of 21.44, P < 0.01, with heterogeneity I^2 = 99%.

(specificity) of 93.6% and I^2 (positive likelihood ratio) of 95.5% (Figure 3). These heterogeneities may be due to cut-off or non-cut-off effects. We analyzed the non-cut-off effect using meta-regression and Deeks' funnel plots. We categorized the data into subgroups according to "year of publication," "cases," "study design," and "gestational age" and compared within-group differences to identify sources of heterogeneity (Table 3). We found no significant
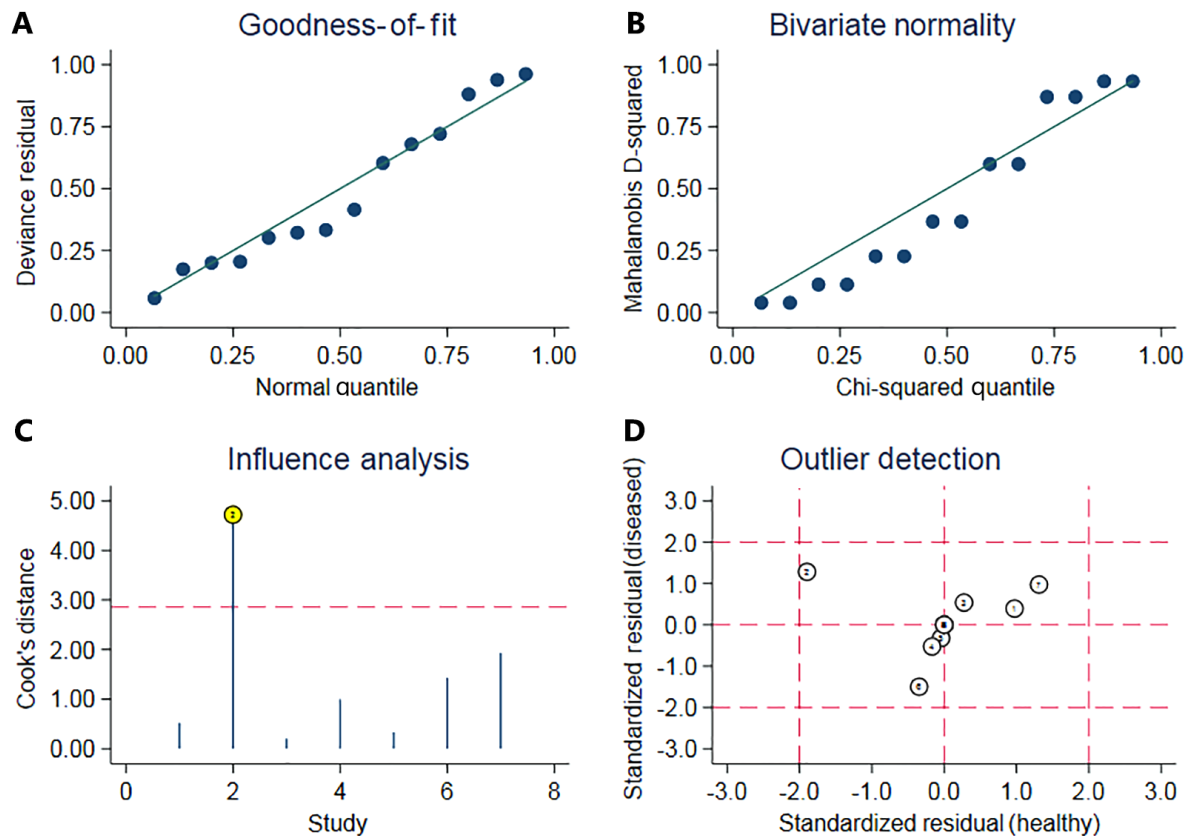
Figure 7. Influence analysis plot. The sensitivity of the included study groups was analyzed by A, goodness-of-fit; B, bivariate normality; C, influence analysis; and D, outlier detection, respectively. The larger outlier detected in (D) is article number 2, and the analysis from (C) shows that removing article number 2 has a greater effect on the results.

differences in the meta-regressions for the subgroup analyses, except for a slight deviation in the number of cases (P = 0.04). Deeks' funnel plot (Figure 4) shows the publication bias of the articles, which may be due to the lack of negative reviews, as the selected articles all had positive feedback on LUS (36, 37). The sources of heterogeneity in the non-cut-off effect are complex and include study populations, indicator tests, reference standards (38, 39), and so on. As shown in Figure 7, the articles by Brat11 on the LUS prediction of 100% sensitivity to surfactant may introduce heterogeneity.

To explore the accuracy of the LUS and sources of heterogeneity, we plotted summary receiver characteristic curves (Figure 5) with an AUC of 94% and no "shoulder arm" feature (40). However, it is clear from the articles that the cut-offs are inconsistent, and a possible cut-off effect may exist. First, the effect of gestational age on LUS was considered, and although previous meta-regression showed no significant difference in LUS between < 34 and > 34 weeks, it cannot be ruled out that there is no

difference in cut-offs of 28 - 34, 26 - 28, or 34 - 37 weeks. De Martino et al. (26) argued that the different cut-offs of < 28 weeks vs. > 28 weeks would impact the indications for substance use. Secondly, the disease severity may impact the cut-offs, as Vardar et al. (27) found that a score of 4 - 6 was suitable for predicting surfactant use in mild patients, with a score of 10 or higher indicating severe lung changes and emphasized that the LUS was suitable for predicting the early detection of RDS, but not for predicting late complications of chronic lung disease. Apparently, two ranges of cut-offs can be seen, with three papers (11, 21, 26) choosing 4 - 6 as the predictive value and four papers (8, 27-29) choosing 8 - 12 as the predictive value, respectively. Nevertheless, a score of 8 - 12 has not been shown to be superior to a score of 4 - 6, although some studies have argued that choosing cut-offs greater than 6 reaches the optimal range in terms of specificity and sensitivity (31). In any case, their study was statistically based on their own data, and the optimal cut-offs can be reliably obtained from the summary receiver operating

characteristic curve with no standardized range of cut-offs. In practice, though, mechanically ventilated patients have a higher mean airway pressure and may require a higher cut-off. In addition, the lungs of premature neonates are inherently immature, and the cut-offs for very premature neonates may be different from those for late premature neonates, making it inappropriate to use the same cut-offs for prediction. We compared two different cut-offs using forest plots, which showed a difference between scores of 4 - 6 and 8 - 12 (Figure 6), and this inconsistency in cut-offs led to a cut-off effect. Uncertainty in the range of cut-offs can greatly affect the objectivity of LUS, and therefore, reasonable cut-offs should be established for different cases. For example, different cut-offs are used to predict the likelihood of RDS according to different gestational ages and lesion degrees. Considering all data, it is concluded that cut-offs of 4 - 6 scores apply to premature neonates at 34 weeks, whereas cut-offs of 8 - 12 scores apply to premature neonates around 28 weeks or those with disease progression, suggesting the need for mechanical ventilation.

Since the LUS is a quantitative indicator of the diagnostic description of lung ultrasound, many authors have questioned its applicability and accuracy (41). Nevertheless, the sensitivity and specificity of the LUS for predicting the onset and progression of RDS remain high when analyzed from the study data. The LUS can help neonatologists predict disease early and guide the use of surfactants and mechanical ventilation. The LUS is currently used in RDS, transient tachypnea of the newborn, pneumonia and chronic lung disease, etc. However, the disadvantage of the LUS is that it only provides a numerical quantitative assessment of the severity of lung lesions. It is difficult to accurately assess patients with severe and complex lesions, such as pulmonary hemorrhage, pneumothorax, and hernias; therefore, the original images should be kept for analysis and combined with CXR or even computed tomography (CT) if necessary (42, 43). Additionally, evaluating lung diseases by ultrasound is closely related to the ultrasound equipment and the operator's experience. For example, when using ultrasound to probe lung regions, the B-line may show hypertrophy inhomogeneity at different frequencies or always prefer open harmonics, affecting the imaging of both the B-line and the A-line (44, 45). Improper selection of ultrasound probes can also affect the scoring accuracy seriously, and it is important to understand how to select convex, micro-convex, and sector probes.

6. Conclusions

In conclusion, the LUS was highly reliable and sensitive in predicting surfactant use in the meta-analysis. The heterogeneity was mainly attributable to the cut-off effect, but there were also inconsistent case selection criteria and publication bias. The LUS is suitable for early prediction of RDS, and inconsistent cut-offs may be related to gestational age or disease severity. However, we can develop a well-established set of scoring diagnostic criteria, with cut-offs ranging from low to high, suggesting progressively increased risk factors. As summarized by the meta-analysis, a score of 4 - 6 is appropriate for predicting neonates with RDS at around 34 weeks gestation, and a score of 8 - 12 is appropriate for predicting neonates at around 28 weeks gestation or those with severe disease requiring intervention, such as the possible need for ventilator support in addition to surfactant. The current European gold standard for the diagnosis of RDS does not include the LUS, and it is hoped that a consensus on the LUS will be reached in the future so that the LUS can be better used in clinical practice.

Footnotes

References

1. Berman S. Otitis media, shared decision making, and enhancing value in pediatric practice. *Arch Pediatr Adolesc Med*. 2008;**162**(2):186–8. [PubMed ID: 18250248]. https://doi.org/10.1001/archpediatrics.2007.19.

2. Dilli D, Cakmakci E, Akduman H, Oktem A, Aydogan S, Citli R, et al. Comparison of three natural surfactants according to lung ultrasonography scores in newborns with respiratory distress syndrome. *J Matern Fetal Neonatal Med*. 2021;**34**(10):1634–40. [PubMed ID: 31296073]. https://doi.org/10.1080/14767058.2019.1643313.

3. Kutanzi KR, Lumen A, Koturbash I, Miousse IR. Pediatric Exposures to Ionizing Radiation: Carcinogenic Considerations. *Int J Environ Res Public Health*. 2016;**13**(11):1057. [PubMed ID: 27801855]. [PubMed Central ID: PMC5129267]. https://doi.org/10.3390/ijerph13111057.

4. Neerhof MG, Dohnal JC, Ashwood ER, Lee IS, Anceschi MM. Lamellar body counts: a consensus on protocol. *Obstet Gynecol*. 2001;**97**(2):318–20. [PubMed ID: 11165603]. https://doi.org/10.1016/s0029-7844(00)01134-0.

5. 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):341–7. [PubMed ID: 20851923]. https://doi.org/10.1164/rccm.201003-0369OC.

6. 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):1282–96. [PubMed ID: 22858877].

7. Wang XT, Ding X, Zhang HM, Chen H, Su LX, Liu DW, et al. Lung ultrasound can be used to predict the potential of prone positioning and assess prognosis in patients with acute respiratory distress syndrome. *Crit Care*. 2016;**20**(1):385. [PubMed ID: 27899151]. [PubMed Central ID: PMC5129200]. https://doi.org/10.1186/s13054-016-1558-0.

8. Raimondi F, Migliaro F, Corsini I, Meneghin F, Pierri L, Salome S, et al. Neonatal Lung Ultrasound and Surfactant Administration: A Pragmatic, Multicenter Study. *Chest*. 2021;**160**(6):2178–86. [PubMed ID: 34293317]. https://doi.org/10.1016/j.chest.2021.06.076.

9. Rodriguez-Fanjul J, Jordan I, Balaguer M, Batista-Munoz A, Ramon M, Bobillo-Perez S. Early surfactant replacement guided by lung ultrasound in preterm newborns with RDS: the ULTRASURF randomised controlled trial. *Eur J Pediatr*. 2020;**179**(12):1913–20. [PubMed ID: 32710304]. [PubMed Central ID: PMC7378405]. https://doi.org/10.1007/s00431-020-03744-y.

10. Gortner L, Schuller SS, Herting E. Review demonstrates that less invasive surfactant administration in preterm neonates leads to fewer complications. *Acta Paediatr*. 2018;**107**(5):736–43. [PubMed ID: 29172232]. https://doi.org/10.1111/apa.14161.

11. Brat R, Yousef N, Klifa R, Reynaud S, Shankar Aguilera S, De Luca D. Lung Ultrasonography Score to Evaluate Oxygenation and Surfactant Need in Neonates Treated With Continuous Positive Airway Pressure. *JAMA Pediatr*. 2015;**169**(8):e151797. [PubMed ID: 26237465]. https://doi.org/10.1001/jamapediatrics.2015.1797.

12. Sweet DG, Carnielli V, Greisen G, Hallman M, Ozek E, Te Pas A, et al. European Consensus Guidelines on the Management of Respiratory Distress Syndrome - 2019 Update. *Neonatology*. 2019;**115**(4):432–50. [PubMed ID: 30974433]. [PubMed Central ID: PMC6604659]. https://doi.org/10.1159/000499361.

13. Ma H, Yan W, Liu J. Diagnostic value of lung ultrasound for neonatal respiratory distress syndrome: a meta-analysis and systematic review. *Med Ultrason*. 2020;**22**(3):325–33. [PubMed ID: 32399541]. https://doi.org/10.11152/mu-2485.

14. Bahadue FL, Soll R. Early versus delayed selective surfactant treatment for neonatal respiratory distress syndrome. *Cochrane Database Syst Rev*. 2012;**11**(11):CD001456. [PubMed ID: 23152207]. [PubMed Central ID: PMC7057030]. https://doi.org/10.1002/14651858.CD001456.pub2.

15. Challis P, Nydert P, Hakansson S, Norman M. Association of Adherence to Surfactant Best Practice Uses With Clinical Outcomes Among Neonates in Sweden. *JAMA Netw Open*. 2021;**4**(5):e217269. [PubMed ID: 33950208]. [PubMed Central ID: PMC8100866]. https://doi.org/10.1001/jamanetworkopen.2021.7269.

16. Rodriguez-Fanjul J, Balcells C, Aldecoa-Bilbao V, Moreno J, Iriondo M. Lung Ultrasound as a Predictor of Mechanical Ventilation in Neonates Older than 32 Weeks. *Neonatology*. 2016;**110**(3):198–203. [PubMed ID: 27220313]. https://doi.org/10.1159/000445932.

17. Raimondi F, Migliaro F, Sodano A, Ferrara T, Lama S, Vallone G, et al. Use of neonatal chest ultrasound to predict noninvasive ventilation failure. *Pediatrics*. 2014;**134**(4):e1089–94. [PubMed ID: 25180278]. https://doi.org/10.1542/peds.2013-3924.

18. Raimondi F, Yousef N, Rodriguez Fanjul J, De Luca D, Corsini I, Shankar-Aguilera S, et al. A Multicenter Lung Ultrasound Study on Transient Tachypnea of the Neonate. *Neonatology*. 2019;**115**(3):263–8. [PubMed ID: 30731475]. https://doi.org/10.1159/000495911.

19. Chiumello D. Bedside ultrasound assessment of positive end expiratory pressure-induced lung recruitment. *Am J Respir Crit Care Med*. 2012;**185**(4):457. [PubMed ID: 22336681]. https://doi.org/10.1164/ajrccm.185.4.457.

20. Jones BP, Tay ET, Elikashvili I, Sanders JE, Paul AZ, Nelson BP, et al. Feasibility and Safety of Substituting Lung Ultrasonography for Chest Radiography When Diagnosing Pneumonia in Children: A Randomized Controlled Trial. *Chest*. 2016;**150**(1):131–8. [PubMed ID: 26923626]. https://doi.org/10.1016/j.chest.2016.02.643.

21. Perri A, Riccardi R, Iannotta R, Di Molfetta DV, Arena R, Vento G, et al. Lung ultrasonography score versus chest X-ray score to predict surfactant administration in newborns with respiratory distress syndrome. *Pediatr Pulmonol*. 2018;**53**(9):1231–6. [PubMed ID: 29870158]. https://doi.org/10.1002/ppul.24076.

22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;**372**:n71. [PubMed ID: 33782057]. [PubMed Central ID: PMC8005924]. https://doi.org/10.1136/bmj.n71.

23. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;**155**(8):529–36. [PubMed ID: 22007046]. https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

24. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med*. 2002;**21**(11):1525–37. [PubMed ID: 12111918]. https://doi.org/10.1002/sim.1185.

25. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. 2010;**1**(2):112–25. [PubMed ID: 26061377]. https://doi.org/10.1002/jrsm.11.

26. De Martino L, Yousef N, Ben-Ammar R, Raimondi F, Shankar-Aguilera S, De Luca D. Lung Ultrasound Score Predicts Surfactant Need in Extremely Preterm Neonates. *Pediatrics*. 2018;**142**(3). e20180463. [PubMed ID: 30108142]. https://doi.org/10.1542/peds.2018-0463.

27. Vardar G, Karadag N, Karatekin G. The Role of Lung Ultrasound as an Early Diagnostic Tool for Need of Surfactant Therapy in Preterm Infants with Respiratory Distress Syndrome. *Am J Perinatol*. 2021;**38**(14):1547–56. [PubMed ID: 32674204]. https://doi.org/10.1055/s-0040-1714207.

28. Aldecoa-Bilbao V, Balcells-Esponera C, Herranz Barbero A, Borras-Novell C, Izquierdo Renau M, Iriondo Sanz M, et al. Lung ultrasound for early surfactant treatment: Development and validation of a predictive model. *Pediatr Pulmonol*. 2021;**56**(2):433–41. [PubMed ID: 33369257]. https://doi.org/10.1002/ppul.25216.

29. Gregorio-Hernandez R, Arriaga-Redondo M, Perez-Perez A, Ramos-Navarro C, Sanchez-Luna M. Lung ultrasound in preterm infants with respiratory distress: experience in a neonatal intensive care unit. *Eur J Pediatr*. 2020;**179**(1):81–9. [PubMed ID: 31655870]. https://doi.org/10.1007/s00431-019-03470-0.

30. Badurdeen S, Kamlin COF, Rogerson SR, Kane SC, Polglase GR, Hooper SB, et al. Lung ultrasound during newborn resuscitation predicts the need for surfactant therapy in very- and extremely preterm infants. *Resuscitation*. 2021;**162**:227–35. [PubMed ID: 33548362]. https://doi.org/10.1016/j.resuscitation.2021.01.025.

31. Razak A, Faden M. Neonatal lung ultrasonography to evaluate need for surfactant or mechanical ventilation: a systematic review and meta-analysis. *Arch Dis Child Fetal Neonatal Ed*. 2020;**105**(2):164–71. [PubMed ID: 31248960]. https://doi.org/10.1136/archdischild-2019-316832.

32. Singh Y, Tissot C, Fraga MV, Yousef N, Cortes RG, Lopez J, et al. International evidence-based guidelines on Point of Care Ultrasound (POCUS) for critically ill neonates and children issued by the

POCUS Working Group of the European Society of Paediatric and Neonatal Intensive Care (ESPNIC). *Crit Care*. 2020;**24**(1):65. [PubMed ID: 32093763]. [PubMed Central ID: PMC7041196]. https://doi.org/10.1186/s13054-020-2787-9.

33. Raschetti R, Yousef N, Vigo G, Marseglia G, Centorrino R, Ben-Ammar R, et al. Echography-Guided Surfactant Therapy to Improve Timeliness of Surfactant Replacement: A Quality Improvement Project. *J Pediatr*. 2019;**212**:137–43.e1. [PubMed ID: 31079857]. https://doi.org/10.1016/j.jpeds.2019.04.020.

34. Roberts CL, Badgery-Parker T, Algert CS, Bowen JR, Nassar N. Trends in use of neonatal CPAP: a population-based study. *BMC Pediatr*. 2011;**11**:89. [PubMed ID: 21999325]. [PubMed Central ID: PMC3206424]. https://doi.org/10.1186/1471-2431-11-89.

35. Blank DA, Kamlin COF, Rogerson SR, Fox LM, Lorenz L, Kane SC, et al. Lung ultrasound immediately after birth to describe normal neonatal transition: an observational study. *Arch Dis Child Fetal Neonatal Ed*. 2018;**103**(2):F157–62. [PubMed ID: 28659360]. https://doi.org/10.1136/archdischild-2017-312818.

36. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials*. 1993;**Doc No 50**:4967 words. 53 paragraphs. [PubMed ID: 8306005].

37. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ*. 2005;**331**(7507):19. [PubMed ID: 15967761]. [PubMed Central ID: PMC558532]. https://doi.org/10.1136/bmj.38488.385995.8F.

38. Song F, Parekh-Bhurke S, Hooper L, Loke YK, Ryder JJ, Sutton AJ, et al. Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med Res Methodol*. 2009;**9**:79. [PubMed ID:19941636]. [PubMed Central ID: PMC2789098]. https://doi.org/10.1186/1471-2288-9-79.

39. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess*. 2005;**9**(12). https://doi.org/10.3310/hta9120.

40. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;**6**:31. [PubMed ID: 16836745]. [PubMed Central ID: PMC1552081]. https://doi.org/10.1186/1471-2288-6-31.

41. Liu X, Si S, Guo Y, Wu H. Limitations of Bedside Lung Ultrasound in Neonatal Lung Diseases. *Front Pediatr*. 2022;**10**:855958. [PubMed ID: 35558371]. [PubMed Central ID: PMC9086677]. https://doi.org/10.3389/fped.2022.855958.

42. Bozkaya D, Dizdar EA, Korkut S, Ceran B, Alkan M, Oguz SS. Evaluation of Different Types of Natural Surfactants by Lung Ultrasound in Respiratory Distress Syndrome. *Am J Perinatol*. 2021;**38**(6):590–6. [PubMed ID: 31770784]. https://doi.org/10.1055/s-0039-1700856.

43. Liu J. The Lung Ultrasound Score Cannot Accurately Evaluate the Severity of Neonatal Lung Disease. *J Ultrasound Med*. 2020;**39**(5):1015–20. [PubMed ID: 31737918]. https://doi.org/10.1002/jum.15176.

44. Kameda T, Kamiyama N, Kobayashi H, Kanayama Y, Taniguchi N. Ultrasonic B-Line-Like Artifacts Generated with Simple Experimental Models Provide Clues to Solve Key Issues in B-Lines. *Ultrasound Med Biol*. 2019;**45**(7):1617–26. [PubMed ID: 31031034]. https://doi.org/10.1016/j.ultrasmedbio.2019.03.003.

45. Matthias I, Panebianco NL, Maltenfort MG, Dean AJ, Baston C. Effect of Machine Settings on Ultrasound Assessment of B-lines. *J Ultrasound Med*. 2020;**40**(10):2039–46. [PubMed ID: 33289208]. [PubMed Central ID: PMC8518047]. https://doi.org/10.1002/jum.15581.