Diagnostic value of lung ultrasound for neonatal respiratory distress syndrome: a meta-analysis and systematic review

Hai-Ran Ma¹², Jing Liu², Wen-Kang Yan¹

¹Department of Neonatology and Neonatal Intensive Care Unit, Huizhou Municipal Central Hospital, Huizhou Guangdong, ²Department of Neonatology and Neonatal Intensive Care Unit, Beijing Chaoyang District Maternal and Child Healthcare Hospital, Beijing and National Neonatal Lung Ultrasound Training Base, Beijing, China

Abstract

Aim: Neonatal respiratory distress syndrome (NRDS) is one of the most common and severe diseases in neonatal intensive care units worldwide. Increasing evidence suggests that lung ultrasound (LUS) may be a reliable diagnostic tool for neonatal respiratory distress syndrome. The aim of study was to evaluate the diagnostic accuracy of LUS for NRDS with a systematic review and meta-analysis. Material and methods: We searched for articles in EMBASE, PubMed and Cochrane Central from inception until 17 August 2019. The selected studies were diagnostic accuracy studies that reported the utility of LUS in the diagnosis of NRDS. Two researchers independently extracted data and assessed quality using the QUADAS-2 tool. Then, we created a bivariate mixed effects model to calculate the sensitivity and specificity of LUS in diagnosing NRDS. A summary receiver operator characteristic (SROC) curve was constructed to summarize the performance characteristics of LUS. Results: Nine studies involving 703 infants were included in the review. LUS had a pooled sensitivity of 0.99 (CI: 0.92-1.00) and a specificity of 0.95 (CI: 0.87-0.98). The areas under the curve for LUS was 0.99 (0.98-1.0). Meta-regression revealed that LUS had a significant diagnostic accuracy for NRDS. Conclusion: LUS is a promising method that is easily carried out, inexpensive, nonionizing and repeatable and can be performed at the bedside. Current evidence supports LUS as a useful imaging alternative for the diagnosis of NRDS.

Keywords: ultrasonography; respiratory distress syndrome; newborn

Introduction

Neonatal respiratory distress syndrome (NRDS) usually occurs in preterm infants. This disorder is caused primarily by surfactant deficiency that leads to diffuse atelectasis and alveolar collapse shortly after birth. NRDS is one of the most important causes of morbidity and mortality in preterm infants. The clinical features of NRDS are characterized by tachypnoea, nasal flaring, expiratory grunting, chest retractions and cyanosis [1], and the risk for this syndrome increases with decreasing gestational age (GA). The incidence of NRDS was found to be 93% in extremely preterm infants (GA 28 weeks or below) [2] and 10.5%, 6%, 2.8%, 1% and 0.3% in infants born at 34, 35, 36, 37, and ≥38 weeks gestation, respectively [3].

Tremendous efforts have been made to improve the morbidity and mortality of NRDS. There are well-established methods for the diagnosis and treatment of NRDS. Early diagnosis is important to optimize the treatment of infants with NRDS. Chest radiography (CXR) is currently a leading and extensive tool in establishing the diagnosis of NRDS [4,5], although currently, the appearance of “ground glass with air bronchograms” is rarely seen due to early surfactant therapy and early continuous positive airway pressure (CPAP) [5]. However, CXR has the risk of exposing the infant to ionizing radiation. The risk
Diagnostic value of lung ultrasound for neonatal respiratory distress syndrome

Hai-Ran Ma et al

for the potential negative effects of CXR are significantly higher in younger children [6,7]. A 1-year-old child is 10-15 times more at risk of developing carcinoma than an adult who receives the same dose of ionizing radiation [8].

Lung ultrasound (LUS), as a non-invasive routine technology, was used to diagnose neonatal lung diseases at the beginning of the last decade of the 20th century [9,10]. LUS has recently been recommended as a diagnostic procedure in the diagnosis and monitoring of NRDS in clinical practice [5]. The advantages of an ultrasound examination over CXR, such as its non-ionizing, convenient and repeatable nature, in the diagnosis of NRDS in neonates were shown by a number of publications [9,11-15]. Two ultrasonography approaches (trans-thoracic and transabdominal) were used to detect NRDS and both had high sensitivity and relatively high specificity [12,16]. Therefore, LUS appears to be the ideal choice for a screening or diagnostic method in the assessment of NRDS in newborns. Nonetheless, LUS has a tendency to overdiagnose NRDS [17]. A comprehensive systematic review with a structured question is required to identify the diagnostic accuracy, methodological quality of evidence, limitations and clinical application of LUS for detecting NRDS in the contemporary era. The objective of this review is to calculate the pooled sensitivity and specificity of LUS in diagnosing NRDS and facilitate an understanding of the capabilities of LUS regarding the evaluation of NRDS.

Material and method

Data sources and searches

In this systematic review and meta-analysis, two researchers (J. L and HR. M) independently searched the PubMed, EMBASE and Cochrane Library databases for articles published in any language or country [18]. These general bibliographic databases were searched from inception to 17 August 2019. We believe this search strategy is sufficient for this study since searching Medline (part of PUBMED) leads to only minor differences in the estimates of diagnostic test accuracy [19]. Aside from the above search, the reference lists of the included studies were checked manually to identify as many relevant diagnostic test accuracy (DTA) studies as possible [20].

The search used Medical Subject Headings terms for neonates, ultrasound and NRDS. The Boolean operators (AND) and (OR) were used to minimize irrelevant literature and maximize the breadth of the search. The following search terms for ultrasonography were combined using the OR connector: “echotomography” OR “echography” OR “sonography” OR “medical sonography” OR “ultrasonography” OR “ultrasound” OR “ultrasonic tomography” OR “computer echotomography.” Similarly, the search terms for NRDS were defined using the following syntax: “Neonatal Respiratory Distress Syndrome” OR “Respiratory Distress Syndrome” OR “Infantile Respiratory Distress Syndrome” OR “Hyaline Membrane Disease” OR “Hyaline Membrane Diseases” OR “Disease, Hyaline Membrane”. The search results of ultrasonography and neonatal respiratory distress syndrome were combined using the AND connector. Discrepancies between researchers were resolved through a consensus with a third reviewer (WK. Y).

Study selection

All articles identified as potentially eligible by either reviewer were reviewed as abstracts or full-text articles. Eligible studies met the following inclusion criteria: 1) cohort or case-control studies used LUS as a diagnostic tool for neonatal respiratory distress syndrome; 2) NRDS was diagnosed using a combination of clinical signs and symptoms (tachypnoea, nasal flaring, chest retractions and grunting), CXR with/without laboratory blood gas analysis; 3) results were reported in sufficient detail to allow for a reconstruction of contingency tables of the original data (true-positive, true-negative, false-positive and false-negative results); and 4) studies conducted in human newborns. Studies were excluded if they did not meet the inclusion criteria above. Duplicated studies, reviews and conference abstracts were rejected. In cases of data duplication (i.e., the same data was published in two or more reports), only the most recent study with the largest and most recent data set was included.

Data extraction and quality assessment

Two authors (J. L and HR. M) independently carried out the data extraction, including study characteristics, study design, demographic characteristics, blinding parameters, time between CXR and LUS and quantitative data that allowed for the construction of a standard 2x2 table. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was also applied by two authors (J. L and HR. M) independently to evaluate the methodologic quality of the selected studies [21]. This instrument consists of four domains: patient selection, index test, reference standard and flow and timing. Each domain was assessed with respect to the risk of bias. Discrepancies were resolved by discussion and a consensus with a third reviewer (WK. Y). Predefined meta-regression (subgroup analyses) was performed to explore potential sources of heterogeneity.

Statistical analysis

Data analysis was conducted using STATA (Version 14.0, Stata Corp LP) with user-written “midas” programs. The bivariate mixed effects model was used to
calculate pooled sensitivity and specificity with corresponding 95% confidence intervals [22]. The diagnostic accuracy parameters mentioned above are shown graphically. Heterogeneity was analysed among studies using the $I^2$ statistic and the Q test [23,24]. A summary receiver operator characteristic (SROC) curve was constructed to summarize the diagnostic value of LUS for NRDS. Publication bias was assessed using Deek’s tests [25]. All statistical tests were two-sided with the type I error rate set as 0.05. Thus, a $p$-value less than 0.05 was considered statistically significant.

Results

Study flow and characteristics

Our search retrieved 850 articles. After identifying 102 duplicated studies and excluding 738 studies based on predefined criteria, a total of 10 articles were selected for detailed review. In addition, 3 studies were manually added from the reference lists of the known reviews and the included primary studies and 4 studies were excluded due to various reasons summarized in figure 1, leaving a total of 9 studies for the final analysis. A total of 703 patients were included in this review. Six studies were prospective cohort studies [9,12,16,17,26,27] and three studies were prospective case-control studies [11,28,29]. Detailed characteristics of the 9 studies are summarized in Table I and Table II.

Primary outcome

Nine studies assessed the diagnostic accuracy of LUS for NRDS in neonates and the assessment of methodologic quality using QUADAS-2 is illustrated in figure 2.

![Fig 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of the search process](image)

![Fig 2. Methodological quality assessment according to QUADAS-2](image)

### Table I. Study characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Origin</th>
<th>Study type</th>
<th>Sample size</th>
<th>Gestational age (mean±SD), weeks</th>
<th>True positive</th>
<th>False positive</th>
<th>True negative</th>
<th>False negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahuja, 2012 [12]</td>
<td>India</td>
<td>Prospective</td>
<td>88</td>
<td>29.6±1.57</td>
<td>32</td>
<td>6</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>Avni, 1990 [9]</td>
<td>Belgium</td>
<td>Prospective</td>
<td>40</td>
<td>Unknown</td>
<td>22</td>
<td>2</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Bober, 2006 [17]</td>
<td>Poland</td>
<td>Prospective</td>
<td>131</td>
<td>32±4.4</td>
<td>101</td>
<td>8</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Copetti, 2008 [11]</td>
<td>Italy</td>
<td>Case-control</td>
<td>55</td>
<td>27.2±2.7</td>
<td>40</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Liang, 2018 [26]</td>
<td>China</td>
<td>Prospective</td>
<td>120</td>
<td>28.59±1.98</td>
<td>29</td>
<td>4</td>
<td>82</td>
<td>5</td>
</tr>
<tr>
<td>Liu, 2015 [28]</td>
<td>China</td>
<td>Case-control</td>
<td>100</td>
<td>34.9±2.7</td>
<td>50</td>
<td>0</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Lovrenski, 2012 [27]</td>
<td>Serbia</td>
<td>Prospective</td>
<td>47</td>
<td>30.9±3.16</td>
<td>43</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Rachuri, 2017 [29]</td>
<td>India</td>
<td>Case-control</td>
<td>63</td>
<td>34.5 ± 3.2</td>
<td>29</td>
<td>1</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Vergine, 2014 [16]</td>
<td>Italy</td>
<td>Prospective</td>
<td>59</td>
<td>33±4</td>
<td>22</td>
<td>2</td>
<td>34</td>
<td>1</td>
</tr>
</tbody>
</table>

SD – standard deviation
The pooled sensitivity and specificity of LUS for diagnosing NRDS were 0.99 (CI: 0.92-1.00) and 0.95 (CI: 0.87-0.98), respectively. The Q values were 50.30 (p=0.00) and 37.84 (p<0.00), indicating heterogeneity among the studies. The I² statistic values were 84.09% and 78.86%, respectively (fig 3). Since these values were >50%, this was considered to be significant heterogeneity based on the recommendations from the Cochrane handbook [30]. This supports the use of meta-regression (subgroup analyses) to explore meta-analysis results. Additionally, the AUROC was 0.99 (0.98-1.0) (fig 4). Publication bias was evaluated using Deek’s funnel plot asymmetry test and was suggested to be not statistically significant (p=0.80; fig 5). Consequently, we did not apply the trim and fill approach [31].

Fig 3. Forest plots of sensitivity and specificity

The pooled sensitivity and specificity of LUS for diagnosing NRDS were 0.99 (CI: 0.92-1.00) and 0.95 (CI: 0.87-0.98), respectively. The Q values were 50.30 (p=0.00) and 37.84 (p<0.00), indicating heterogeneity among the studies. The I² statistic values were 84.09% and 78.86%, respectively (fig 3). Since these values were >50%, this was considered to be significant heterogeneity based on

Fig 4. Summary receiver operating characteristic curves.
In this study, we found that LUS was valuable for the detection of neonatal respiratory distress syndrome (NRDS). LUS showed high sensitivity (99%) and specificity (95%) in detecting NRDS. CXR is currently a common reference tool in the diagnosis of NRDS [5]. Vergine et al. demonstrated a sensitivity and specificity of 91% and 84% for CXR in diagnosing NRDS, respectively [16]. When using CXR to detect NRDS, Liang et al. [26] showed that this method has a sensitivity of 88.23% and a specificity of 89.53%. Copetti et al [11] described that the simultaneous presence of abnormalities in the pleural line, ultrasound aspect of “white lung” and the absence of “spared areas” has a sensitivity and specificity of 100%.

In comparison, the pooled sensitivity of the three studies that used transabdominal scanning only was 99% (CI: 95-100%) and the specificity was 83% (CI: 76-91%). In four studies with intervals between CXR and LUS less than 6 hours, the pooled sensitivity was 98% (CI: 94-100%) and the specificity was 98% (CI: 95-100%). However, in five studies with intervals less than 24 hours, the pooled sensitivity was 99% (CI: 97-100%) and the specificity was 89% (CI: 82-97%). For sensitivity and specificity, the “type” of study was a statistically significant factor (p<0.05) and the sensitivity of the “technology” in the studies was approximately significant (p=0.05). The detailed data are presented in figure 6.

### Discussion

In this study, we found that LUS was valuable for the detection of neonatal respiratory distress syndrome (NRDS). LUS showed high sensitivity (99%) and specificity (95%) in detecting NRDS. CXR is currently a common reference tool in the diagnosis of NRDS [5]. Vergine et al. demonstrated a sensitivity and specificity of 91% and 84% for CXR in diagnosing NRDS, respectively [16]. When using CXR to detect NRDS, Liang et al. [26] showed that this method has a sensitivity of 88.23% and a specificity of 89.53%. Copetti et al [11] described that the simultaneous presence of abnormalities in the pleural line, ultrasound aspect of “white lung” and the absence of “spared areas” has a sensitivity and specificity of 100%.

Based on these data, the accuracy of LUS was better than that of CXR. The pooled PLR and NLR were 18.1 (7.5, 43.4) and 0.01 (0.00, 0.09), respectively, which also indicated a high level of overall diagnostic accuracy.

Heterogeneity was found in this meta-analysis through the I² statistic and the Q test (I² >50, p<0.01). Subgroup analysis revealed a significant difference between case-control studies and cohort studies. The reason for this is that enrolling participants with known disease and a control group without the condition may similarly exaggerate the diagnostic accuracy [21,32,33]. As such, we believed that a subgroup analysis of the six cohort studies would provide the most accurate reflection of test accuracy, although both subgroup analyses had high sensitivity and specificity.

Apart from the significant differences in study design in the subgroup analysis, there are several possible sources for this heterogeneity. First, studies that adopted the transthoracic technique as the detection method had a non-significantly higher diagnostic accuracy compared with studies that used the transabdominal approach as the detection method. Copetti et al [11] claimed that, in comparison with the transabdominal approach, the transthoracic technique could examine all the lung fields rather than only the base. This finding is in accordance...
with the findings of Hiles et al. [34], suggesting that the transthoracic technique may be the better diagnostic approach for avoiding false-positive diagnoses and has the clinical advantage of reducing unnecessary additional testing and interventions. Second, the subgroup analysis also indicated that studies with a detection interval between CXR and LUS of less than 6 hours had a non-significantly higher specificity than those with a detection interval less than 24 hours. Only one study [28] in this meta-analysis performed LUS and CXR at the same time, which is ideal. The CXR and LUS features of RDS change following progressive pathology exacerbation and/or management of the disease. If a delay occurs or if treatment of NRDS begins between LUS and CXR, the improvement or deterioration of the condition may be misclassified. An inappropriate interval between the index test and the reference standard may be problematic for the diagnosis of acute diseases [21]. Vergine et al. [16] declared that the sonographic appearance of RDS does not change immediately after surfactant administration.

This may be explained by data obtained with an animal model that confirmed that there is no difference in fluid lung content among lungs treated or not treated with surfactants [35]. Oktem et al. [36] demonstrated that LUS did not find any valuable changes or differences 2 hours after surfactant replacement therapy and that changes occurred 4 hours after surfactant therapy. However, the optimal test interval between LUS and CXR is still unknown. Third, three studies [11,12,27] included in this review did not show “blinding” while performing LUS and one study had unknown blinding parameters [26]. Knowledge of the clinic and radiologic data may overestimate the diagnostic accuracy of LUS [37]. Therefore, different LUS methods might contribute to the heterogeneity in the meta-analysis. Fourth, the choice of standard criteria for the diagnosis of NRDS may influence the diagnostic accuracy of LUS. If patients are diagnosed with a different reference standard, verification bias can occur [21]. Studies that adopted only “clinical diagnosis and CXR” as the standard criterion may show a lower diagnostic accuracy than those that used a combination of “clinical diagnosis, CXR and blood results (or gastric aspirate shake test)” as the standard criterion. More studies on the standard criteria would provide more details for the diagnosis of NRDS. Currently, the little information regarding the standard criteria for a reference test may overlook patients with NRDS, resulting in an underestimated diagnostic accuracy of LUS for NRDS.

CXR has potential harmful effects on infants. Infants are more sensitive to radiation than adults [6,8]. Hence, LUS has been recently recommended as a routine procedure for the diagnosis of different neonatal lung pathologies [38-45] in the neonatal intensive care unit (NICU). Oktem et al. [36] and Brat et al. [46] emphasized the importance of LUS in monitoring NRDS treatment. Repeated CXR for infants in a short time is unethical and harmful due to the increased exposure to ionizing radiation. Gargani et al. [47] showed that sonographic abnormalities in the lung precede PaO2/FiO2 changes in an animal model. Therefore, LUS can be easily used for the early detection of patients with NRDS before clinical deterioration and improve the clinical progression and prognosis.

The mechanism of LUS for assessing lung pathologies is based on the artefact phenomenon. Techniques for detecting NRDS have been developed and include the transabdominal approach and the transthoracic approach, and it seems that the diagnostic accuracy of the latter method is better than that of the former. Copetti et al. [48] demonstrated that the transabdominal approach could not evaluate the changes in the pleural line, which are essential for the diagnosis of NRDS. Liu et al. [28] suggested that combined imaging findings of lung consolidation, pleural line abnormalities and white lung by the transthoracic method had a high diagnostic accuracy for NRDS with a sensitivity and specificity of 100%. This might explain why Ahuja et al. [12] found a lower diagnostic accuracy using transabdominal ultrasound for the diagnosis of NRDS. Their studies showed a sensitivity of 84.2% and specificity of 88%.

Despite the favourable performance of LUS in the diagnosis of NRDS, we should be aware of several concerns. For example, subcapular or subclavicular consolidations or perihilar consolidations that do not reach the pleura cannot be evaluated by ultrasound imaging [49]. Copetti et al. [11] maintained that air-leak syndrome (pneumomediastinum, interstitial emphysema, pneumopericardium) and complications of NRDS cannot be detected using LUS. Nonetheless, June et al. [50] claimed that pneumomediastinum could be easily visualized and abnormal air accumulation could be localized by ultrasound.

The severity of NRDS is divided into 4 stages on the basis of CXR. These stages of severity correlate closely with the management of NRDS. Bober et al. [17] indicated that the stages on ultrasound examination were well correlated to the radiographic criteria for stages. Liu et al. [28] argued that LUS may not be proper for the diagnosis of grade 1 NRDS.

The interpretation of the basic LUS patterns is quite important for the diagnosis of NRDS. Tsou et al. [51] indicated that the diagnostic accuracy of lung ultrasound is significantly affected by sonographer experience. This is similar to CXR since there are variations in intra-observer and inter-observer agreement among radiologists.
who use CXR for diagnosis [52]. However, Copetti et al claimed that inter- and intra-observer variability in the interpretation of imaging findings is a theoretical problem, not a real problem [11]. Bedetti et al [53] showed that ultrasound lung comets are equally reliable between experienced operators and new beginners.

This study has several limitations that merit consideration. First, only 9 studies were included in the review, which might not represent all the clinical features of NRDS. This may be because LUS is still a relatively new technique and further studies will be required to guarantee LUS as a helpful diagnostic method for NRDS. Second, there is significant heterogeneity in this meta-analysis. Hence, our findings should be interpreted after taking this factor into consideration and the diagnostic accuracy of LUS for NRDS may not be high enough. In addition to the possible reasons discussed above, heterogeneity may be affected by variations in gestational age and operators [51] and may also depend on the probe that is used. Future studies should explore ways to reduce these variations, ensure repeatability of these results and identify the optimal probe needed for this purpose [54].

**Conclusion**

Our findings show that LUS is an accurate diagnostic tool for NRDS, has the benefits of being a low-cost, non-ionizing and repeatable method that can be performed at the bedside and has a comparable predictive power compared with CXR. Nevertheless, it is important to note that this conclusion is based on a small number of studies and the heterogeneity should be taken into account. Methodological stringency and multicenter studies are required to confirm this conclusion and further assess the performance of LUS in the diagnosis of NRDS.

**Conflict of interest:** none

**Acknowledgment:** This work was supported by the Social Development Projects, Beijing Chaoyang District Bureau of Science, Technology and Information (CYSF1922).

**References**

19. van Enst WA, Scholten RJ, Whiting P, Zwinderman AH, Hoof L. Meta-epidemiologic analysis indicates that MEDLINE searches are sufficient for diagnostic test accu-


