Research Article

Inter-observer Variability for Radiography in Pediatric Acute Respiratory Distress Syndrome and Improvement with a Computer-Aided Diagnosis System

Zaglam N[1,2], Essouri S[2,4], Fléchelles O[2], Emeriaud G[2,3], Cheriet F[1,2] and Jouvet P[2,3]*

[1]Department of computer engineering, École Polytechnique, Canada

[2]Research Center of Ste Justine Hospital, Canada

[3]Pediatric ICU, Sainte-Justine Hospital, Canada

[4]Pediatric ICU, CHU Kremlin Bicêtre, France

*Corresponding author: Philippe Jouvet, Soins Intensifs Pédiatriques, Hôpital Sainte Justine, 3175 chemin Côte Sainte Catherine, Montréal (Québec) H3T 1C5, Canada, Tel: + 1 514 345 4927; Fax: + 1 514 342 7731; Email: philippe.jouvet@umontreal.ca

Abstract

Acute respiratory distress syndrome (ARDS) is the most severe form of acute respiratory failure both in adult and children. The Consensus Conference on ARDS definition requires the presence of bilateral pulmonary infiltrates on chest X-ray (CXR). To be consistently useful, interpretation of the CXR must be reliable. Adult studies on radiographic interpretation in ARDS have shown limited inter-observer agreement and concluded that intensivists without formal consensus training can only achieve a moderate level of agreement. In order to improve this agreement level, a computer-aided diagnosis (CAD) system was developed.

Objective: To compare the reliability of radiological diagnosis of P-ARDS between clinical assessment and a CAD system.

Design: Retrospective radiological study.

Patients: Chest X ray of children admitted in a pediatric intensive care unit between April 1, 2009 and April 30, 2010.

Measurements: a CXR database was developed using 90 CXR selected among children included in a previous study. We developed a methodology to create a gold standard for the radiological diagnosis of ARDS. We compared the inter-observer variability for radiological ARDS diagnosis between two intensivists and the CAD.

Results: Inter-observer variability was moderate between two intensivists (kappa: 0.55). The CAD system was able to significantly improve the kappa score either alone or as second reader (0.77 and 0.79-0.86 respectively) and reach a good agreement level.

Conclusion: Our study confirms the inter-observer variability with clinical assessment alone. The use of a CAD system for CXR interpretation in pediatric ARDS is able to reduce variability.

Keywords: ARDS; Children; Chest X Ray; Intensive care; Critical care; Computer aided diagnosis systems

Introduction

Acute respiratory distress syndrome (ARDS), the most severe form of acute respiratory failure both in adult and children, is characterized by increased capillary permeability, inflammation and alveolar damage. The incidence of Pediatric ARDS (P-ARDS) is lower than in adults and ranges from 2 to 12.8 cases per 100,000 per year [1-4]. ARDS mortality in children appears to be lower than in adults but is still high (18 – 27% versus 27-45%) [4-10]. Variability in defining and identifying ARDS has led to difficulties in comparing clinical trials. Both the American-European Consensus Conference (AECC) and the recent Berlin definitions of ARDS require the presence of bilateral pulmonary infiltrates on chest radiography [11-13]. To be consistently useful, interpretation of the chest radiography must be reliable. Adult studies on radiographic interpretation in ARDS have shown limited inter-observer agreement and concluded that intensivists without formal consensus training can only achieve a

moderate level of agreements [14,15]. Angoulvant et al., reported similar results in a pediatric ARDS population [16].

The lack of strong agreement for the radiographic interpretation of P-ARDS can impact the delivery of clinical care (delayed recognition of ARDS condition) and becomes crucial in clinical studies. This last point has been stated by the Pediatric Acute Lung Injury Consensus Conference (PALICC) in 2014 with the following recommendation: "Future clinical trials for P-ARDS should stratify patients by the presence or absence of bilateral infiltrates on chest imaging. In order to minimize variability in these studies, investigators should standardize interpretation of all chest imaging" [17].

Radiological evaluations are affected by subjective interpretation and affect the reproducibility of this diagnostic test. Computer-aided diagnosis (CAD) is currently a leading topic of research in medical imaging that can help to the standardization of interpretation. The consensus conference PALICC also stated that "Future studies

are needed to determine the optimal common training or effect of automated methodologies to reduce inter-observer variability in the interpretation of chest imaging for PARDS" [17]. Several CAD have been developed for detection of nodules or texture analysis in adult [18]. A specific CAD was developed to help intensivists with the early recognition of P-ARDS. This CAD is based on texture analysis of semi-automatic selection of region of interest (ROI). The selection of ROI is made by the initial segmentation of ribs which are then removed from the Chest X-ray (CXR) to obtain the inter-costal areas where patches are automatically extracted and analysed. This CAD has been developed and previously validated by our team [19].

The main objective of the present study was to assess the inter-observer variability and agreement with a gold standard of experienced intensivists and a CAD system for the radiological diagnosis of P-ARDS.

Materials and Methods

Patients

Chest X-ray selection was done within a database of children previously included in the TGRPP study (Transfusion de Globules Rouges Plaquettes et al. Plasma) [20]. This database contains 916 patients aged between 7 days and 18 years and admitted in the unit between April 1, 2009 and April 30, 2010. General characteristics, primary diagnosis and clinical conditions of all patients enrolled in the TGRPP study were recorded prospectively into the database, was approved by the Institutional Review Board of Sainte-Justine Hospital n°2870 and parental consent was waived. This current study was approved by the Institutional Review board of Sainte-Justine Hospital n°3424.

Chest X ray gold standard database development

The first 120 CXR performed upon admission of patients to the PICU were consecutively selected among the TGRPP study database and included in the study. Three experienced (≥ 10 years in PICU i.e. Reader 1: 20 years, Reader 2: 11 years, Reader 3: 10 years) pediatric intensivists assessed these chest X-Rays for the study. All 3 intensivists did a pediatric residency, an intensive care fellowship with training on chest X-ray interpretation and were working more than 20 weeks a year in PICU during the last 10 years, with a daily interpretation of chest X-Ray of all the patients they had in charge including patients with or without ARDS. The readers were not aware of the clinical diagnosis of the patient. All protected health information, including name, age, date of examination were masked before evaluation. Two pediatric intensivists read the CXR in the same order without additional formal consensus training. Interpretation was done independently from one to another. Each reader was asked to evaluate the four quadrants. The horizontal plane of the ipsilateral pulmonary artery defined the limit between upper and lower quadrant of the lung field. If this landmark was obscured, the midpoint of the height of the lung fields was used. To create a gold standard database of ARDS/non ARDS chest X-Rays (classification as CXR ARDS/non ARDS and location of affected quadrants), the following steps were performed (Figure 1):

Step 1: For each CXR, both intensivists determined the number and the position of affected quadrant and then classified the CXR as ARDS positive or negative. In case of agreement on all quadrants, the chest X-Ray was included in the gold standard database.
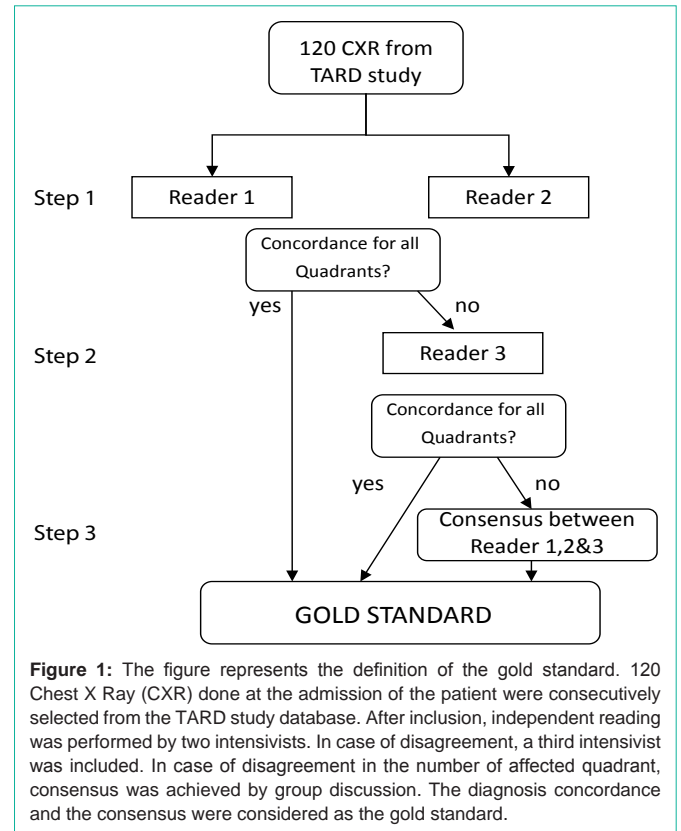

Figure 1: The figure represents the definition of the gold standard. 120 Chest X Ray (CXR) done at the admission of the patient were consecutively selected from the TARD study database. After inclusion, independent reading was performed by two intensivists. In case of disagreement, a third intensivist was included. In case of disagreement in the number of affected quadrant, consensus was achieved by group discussion. The diagnosis concordance and the consensus were considered as the gold standard.

Step 2: In case of disagreement on the number and/or location of affected quadrants, the CXR was analysed by the third intensivist. If there was an agreement on all quadrants with one of the two previous intensivists, the chest X-Ray was included in the gold standard database.

Step 3: In case of persistent disagreement on the number and/or location of affected quadrant, the final diagnosis was achieved by the three intensivists using a consensus oriented decision making. This whole process of chest X-Ray classification resulted in the classification of the 120 chest X-Ray.

Among the 120 CXR selected, 30 of them were used for the development of the CAD system [19]. Thus, 90 CXR were used to validate the CAD (comparison of the CAD diagnosis to the gold standard).

CAD system characteristics

The CAD system characteristics have been published previously [19]. In summary, the CAD system performed a texture analysis of extracted inter-costal patches of lung tissues. Inter-costal lung tissues patch are extracted from the CXR using an algorithm for semi-automatic segmentation of ribs developed by Plourde et al. [21]. The extremities of the inter-costal space are removed from the CAD analysis to avoid confounders due to the scapulae or pleural effusion. The lungs are then subdivided in four quadrants: upper right lung quadrant=1, lower right lung quadrant=2, upper left lung quadrant=3, lower left lung quadrant=4 and patches (32x32 mm or 16x16 mm according to patient size) are automatically selected. The CAD system calculates for each patch of each quadrant the histogram characteristics, co-occurrence matrix characteristics and the spectral

Table 1: General characteristics of the studied population. ARDS and no ARDS diagnosis correspond to the final chest X-Ray classification in the Gold Standard database (see Figure 1).

Characteristics	ARDS (n=53)	No ARDS (n=37)	p value
Male gender, n (%)	33 (62.3)	22 (59.5)	0.79
Age (months)	47.2 ± 63.2	86.9 ± 76	0.009
Weight (kg)	17.9 ± 19.5	25.1 ± 18.3	0.091
PRISM score	7.57 ± 7.3	6.7 ± 5.5	0.54
PELOD score	5.23 ± 6.9	6.1 ± 8.3	0.63
Pre-existing cardiopathy n (%)	26 (49)	9 (24.3)	0.15
Primary diagnosis at PICU admission			
respiratory disease, n	28 (34)	9 (24.3)	0.33
sepsis, n	17 (32.1)	12 (32.4)	0.79
post cardiac surgery	11 (20.8)	4 (10.8)	0.24
MOF, n	24 (45.2)	14 (37.8)	0.19
Organ failure other than respiratory, n			
1	16	11	
2	6	2	
3	2	1	
PICU length of stay (days)	6.9 ± 6.4	7.6 ± 16.9	0.82
Mortality in PICU n (%)	2 (3.8)	1 (2.7)	0.72

PRISM: Pediatric RISk of Mortality [31], PELOD: Pediatric Logistic Organ Dysfunction [32], PICU: Pediatric Intensive Care Unit, MOF: Multiple Organ Failure

characteristics. According to the authors' previous work [19], the threshold which gives the best performance (sensitivity=90.9% at specificity=94.7%) is 34% of affected patch per quadrant. The final diagnosis of radiological ARDS is retained if the opacities affected at least 1 quadrant, on both side.

Inter-observer variability study

To assess the clinical impact of the CAD use for CXR diagnosis of ARDS, we studied the inter-observer variability of reader 1 versus reader 2, reader 1 & 2 versus gold standard, reader 1&2 with a second reading by the CAD versus gold standard (in this scheme the radiological ARDS diagnosis was considered as positive if either the intensivist or the CAD interpreted the CXR as positive) and reader 1 with a second reading by the CAD versus reader 2 with a second reading by the CAD.

We also compared quadrant by quadrant the classification performed by the two first readers versus the gold standard, in order to determine the quadrants that are the most challenging to diagnose.

Statistics

For comparisons of rating of the presence or absence of bilateral alveolar infiltrates we calculated raw agreement and chance corrected agreement to measure inter-observer variability using the K_{appa}-statistic [22].

Raw agreement can be misleading if the two observers both make a high or low proportion of positive rating, raw agreement will be high even if they are just guessing. High agreement by chance tends to occur when the observers believe the prevalence of the clinical entity of interest is high or low in the studied population. To avoid this

Table 2: Different schemes for using the CAD to improve radiologic diagnosis of P-ARDS.

	Observed agreement	Expected agreement	kappa
Reader 1/Reader 2	0.74	0.49	0.49
Reader 1/gold standard	0.94	0.51	0.89
Reader 2/gold standard	0.76	0.49	0.52
CAD/gold standard	0.88	0.51	0.77
Reader 1+ CAD/gold standard	0.93	0.53	0.86
Reader 2 + CAD/gold standard	0.90	0.53	0.79
Reader 1+CAD/ Reader 2 +CAD	0.94	0.55	0.88

CAD: Computer-Aided Diagnosis system. To study the optimal use of the CAD we analysed different schemes, CAD as exclusive reader or CAD as second reader (Reader + CAD). For each scheme observed, expected and chance-corrected agreements (kappa) were calculated.

problem we used the calculated chance-corrected agreement using k statistics [22]. Variability between the consensus and the CAD system was calculated using the same statistical method. Kappa values for the level of agreement were interpreted using the following categories: 0-0.20 (slight agreement), 0.21-0.40 (fair agreement), 0.41-0.60 (moderate agreement), 0.61-0.80 (good agreement), 0.81-1 (almost perfect agreement) [23].

Results

Subject characteristics

The main clinical features are reported in Table 1. The studied population represented the heterogeneous population of PICU with mainly primary respiratory diseases and sepsis.

Inter-observer variability

Inter-observer variability between reader 1 and 2 was high for P-ARDS. Among the 90 CXR, the 2 experienced readers achieved a moderate level of agreement (k= 0.49) (Table 2). The variability in quadrant classification between reader 1 and 2 was the highest for both lower quadrants (Q2 and Q4) with only a fair agreement, kappa of 0.38 and 0.32 respectively (Table 3).

Compared to the gold standard, the CAD system had a good level of agreement (k=0.77). The use of the CAD system as second reader was able to decrease the variability and improve agreement of reader 2. Furthermore, the use of the CAD as second reader decreased significantly the inter-observer variability between reader 1 and 2 and achieved the highest level of agreement «almost perfect agreement» (Table 2).

Discussion

The chest X-Ray diagnosis of ARDS is still challenging in both adult and pediatric intensive care units. In this study, we demonstrated that the use of a computer-aided diagnosis system dedicated to the radiological diagnosis of ARDS can significantly improved chest X-Ray inter-observer variability from a moderate level of agreement to an almost perfect agreement.

In this study, the performance of experienced readers in the interpretation of diffuse bilateral infiltrate achieved a moderate agreement. This level of agreement corresponds to the level of agreement observed in previous similar studies on radiographic

Table 3: Variability for interpretation for each quadrant.

	Observed agreement	Expected agreement	k
Q1			
Reader 1/Reader 2	0.73	0.48	0.49
Reader 1/gold standard	0.92	0.53	0.83
Reader 2/gold standard	0.74	0.47	0.51
Q2			
Reader 1/Reader 2	0.69	0.5	0.38
Reader 1/gold standard	0.96	0.54	0.9
Reader 2/gold standard	0.71	0.5	0.43
Q3			
Reader 1/Reader 2	0.73	0.48	0.48
Reader 1/gold standard	0.94	0.52	0.88
Reader 2/gold standard	0.77	0.48	0.55
Q4			
Reader 1/Reader 2	0.69	0.54	0.32
Reader 1/gold standard	0.94	0.51	0.89
Reader 2/gold standard	0.71	0.54	0.37

Q1=upper right lung quadrant, Q2=lower right lung quadrant, Q3=upper left lung quadrant, Q4=lower left lung quadrant. For each quadrant, observed, expected and chance-corrected agreements (kappa) were calculated.

diagnosis of ARDS, using the AECC criteria. Angoulvant et al. [16] found high inter-observer variability among experienced pediatricians, intensivists and radiologists, regarding the AECC radiographic criterion for ARDS. Meade et al. [14] described a similar moderate level of agreement on the presence of diffuse bilateral infiltrates suggestive of ARDS on 778 films in adults. However, they demonstrated that consensus training can increase the level of agreement up to 0.88 when CXR are reviewed by intensivists and radiologists. Rubenfeld et al. [15] found only a moderate level of agreement (k=0.55) when experts in the field applied the consensus radiographic definition for ARDS. In the clinical setting, poor or moderate agreement may compromise precision of measurement and can result in misleading findings. This has led to considerable difficulties in comparing epidemiologic data relating to ARDS incidence [14-15]. The recent Berlin definition on ARDS addressed some of the limitations of the AECC definition [12] including the radiologic criteria of ARDS that were modified as follows: "Bilateral opacities-not fully explained by effusions, lobar/lung collapse, or nodules". Bilateral infiltrates on CXR remains part of the definition of ARDS in order to exclude localized pathology (lobar pneumonia) and to select the diffuse inflammatory disease that characterizes ARDS. In order to reduce inter-observer variability, the expert group involved in the Berlin definition developed a reference set of chest radiographs [13]. However, with such an approach, subjectivity remains in the CXR diagnosis of ARDS. This is the reason why we developed a computer aided diagnosis system.

The CAD systems are developed to allow the identification of early-stage disease and to reduce the high level of inter-observer variability. Inter-observer variability is mostly due to human interpretation rather than the technical aspect of imaging patients [24]. CAD was developed for the detection of various pathologies in different organs. However, CAD performances are variable depending on the type of imaging finding. Studies comparing the diagnostic performance of the CAD to clinical diagnosis are critical in order to verify the clinical impact of such CAD [25,26]. Three study design methods can be used: historical control, crossover control and sequential control. In a historical control design, the number of patients to include is very large in order to detect change in performance of CAD. The crossover and sequential designs use a common sample of patients. In the crossover design, images of the first half of patients are interpreted by the observers and images of the second half of patients by the CAD. Then, after a washout period the first half cohort is read by the CAD and the second half cohort by the observers. In a sequential design, such as our study, the observers interpret the images before the CAD interpretation. Obuchowski et al. [25] investigated the difference between a crossover and sequential design when used to estimate the effect of CAD and concluded that the effect of CAD on observer performance can be evaluated without bias using the sequential design.

The CAD systems are mainly labelled for use as second readers but other reading modes, such as concurrent read mode, similar to the mode we used, are more widely used in clinical practice because they may reduce clinical workload. In our study, we evaluated different modes of use of the CAD and we observed that both use of the CAD system as second reader or as concurrent reader achieve a high level of agreement with a Kappa of 0.79 to 0.86 and 0.77 respectively (Table 3). Paquerault et al. [27] compared decision making by the CAD alone or by a reader with or without the CAD diagnosis and concluded that performance increased when a CAD system participated to the diagnosis, with a decrease in variability and subjectivity.

Although the utility of CXR has been demonstrated for verifying the position of endotracheal tubes, catheters and the detection of abnormalities which require an immediate intervention [27], the poor sensitivity of CXR to detect changes in edema, consolidation or atelectasis has been shown by multiple investigators [28-30]. Given the current absence of alternatives to chest X-ray in the pediatric population for the diagnosis for bilateral infiltrates, we believe that a CAD system is an objective means to improve the radiologic diagnosis of P-ARDS.

Our study has several limitations. First, the study did not use the CAD system prospectively to diagnose radiological ARDS. We used a database from a previous study that collected patient data prospectively. The next step will be to assess the improvement of inter-observer variability while using the CAD system prospectively, using a similar design for the gold standard achievement. Secondly, the CAD system did not assess new infiltrates as we analysed the chest-X Ray at intensive care admission without referring to a previous chest X-Ray. This does not influence the interpretation of the chest X-Ray in this study as readers were in the same situation. However, according to the ARDS definition, the chest X–Ray must be interpreted as showing new infiltrates and a refinement of the CAD will be required to take into account previous imaging. Third, we need to confirm these results in various intensive care settings. A study in one center does not reflect the whole spectrum of clinical environments and several validated databases will decrease any bias due to reader's chest X-Ray interpretation. Fourth, the CAD system is semi-automatic and needs clinician intervention to identify the ribs [19]. To improve implementation in clinical practice, a fully automatic

chest X-Ray evaluation by the CAD system is under development and will need further validation. Despite all these limitations, the strength of our study is the high sensitivity and sensibility of the CAD system we have developed and the demonstrated improvement that such a CAD system can offer, which is consistent with findings in the literature on this topic.

Conclusion

Our study revealed that there is still large variability in the radiologic diagnosis of P-ARDS. The use of a validated computer-aided system, either on its own or as second reader, for the radiologic diagnosis of ARDS should be helpful and it warrants further investment.

References

1. Erickson S, Schibler A, Numa A, Nuthall G, Yung M, Pascoe E, et al. Acute lung injury in pediatric intensive care in Australia and New Zealand: a prospective, multicenter, observational study. Pediatr Crit Care Med. 2007; 8: 317-323.

2. Kneyber MC, Brouwers AG, Caris JA, Chedamni S, Plötz FB. Acute respiratory distress syndrome: is it underrecognized in the pediatric intensive care unit? Intensive Care Med. 2008; 34: 751-754.

3. Bindl L, Dresbach K, Lentze MJ. Incidence of acute respiratory distress syndrome in German children and adolescents: a population-based study. Crit Care Med. 2005; 33: 209-312.

4. Rubenfeld GD, Caldwell E, Peabody E, Weaver J, Martin DP, Neff M, et al. Incidence and outcomes of acute lung injury. N Engl J Med. 2005; 353: 1685-1693.

5. Bersten AD, Edibam C, Hunt T, Moran J; Australian and New Zealand Intensive Care Society Clinical Trials Group. Incidence and mortality of acute lung injury and the acute respiratory distress syndrome in three Australian States. Am J Respir Crit Care Med. 2002; 165: 443-448.

6. Flori HR, Glidden DV, Rutherford GW, Matthay MA. Pediatric acute lung injury: prospective evaluation of risk factors associated with mortality. Am J Respir Crit Care Med. 2005; 171: 995-1001.

7. López-Fernández Y, Azagra AM, de la Oliva P, Modesto V, Sánchez JI, Parrilla J, et al. Pediatric Acute Lung Injury Epidemiology and Natural History study: Incidence and outcome of the acute respiratory distress syndrome in children. Crit Care Med. 2012; 40: 3238-3245.

8. Zimmerman JJ, Akhtar SR, Caldwell E, Rubenfeld GD. Incidence and outcomes of pediatric acute lung injury. Pediatrics. 2009; 124: 87-95.

9. Li Y, Wang Q, Chen H, Gao HM, Zhou T, Qian SY. Epidemiological features and risk factor analysis of children with acute lung injury. World J Pediatr. 2012; 8: 43-46.

10. Dahlem P, van Aalderen WM, Hamaker ME, Dijkgraaf MG, Bos AP. Incidence and short-term outcome of acute lung injury in mechanically ventilated children. Eur Respir J. 2003; 22: 980-985.

11. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al: The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. Am J Respir Crit Care Med. 1994; 149: 818-824.

12. ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E. Acute respiratory distress syndrome: the Berlin Definition. JAMA. 2012; 307: 2526-2533.

13. Ferguson ND, Fan E, Camporota L, Antonelli M, Anzueto A, Beale R. The Berlin definition of ARDS: an expanded rationale, justification, and supplementary material. Intensive Care Med. 2012; 38: 1573-1582.

14. Meade MO, Cook RJ, Guyatt GH, Groll R, Kachura JR, Bedard M, et al. Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. Am J Respir Crit Care Med. 2000; 161: 85-90.

15. Rubenfeld GD, Caldwell E, Granton J, Hudson LD, Matthay MA. Interobserver variability in applying a radiographic definition for ARDS. Chest. 1999; 116: 1347-1353.

16. Angoulvant F, Llor J, Alberti C, Kheniche A, Zaccaria I, Garel C, et al. Inter-observer variability in chest radiograph reading for diagnosing acute lung injury in children. Pediatr Pulmonol. 2008; 43: 987-991.

17. The Pediatric Acute Lung Injury Consensus Conference Group: Pediatric Acute Respiratory Distress Syndrome: Consensus Recommendations from the Pediatric Acute Lung Injury Consensus Conference. Pediatr Crit Care Med. 2015 [Epub ahead of print].

18. van Ginneken B, ter Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging. 2001; 20: 1228-1241.

19. Zaglam N, Jouvet P, Flechelles O, Emeriaud G, Cheriet F. Computer-aided diagnosis system for the Acute Respiratory Distress Syndrome from chest radiographs. Comput Biol Med. 2014; 52: 41-48.

20. Kleiber N, Lefebvre E, Emeriaud G, Jouvet P. Incidence des complications respiratoires post-transfusion de concentrés érythrocytaires aux soins intensifs pédiatriques. Réanimation. 2012; 21: S131.

21. Plourde F, Cheriet F, Dansereau J. Semiautomatic detection of scoliotic rib borders from posteroanterior chest radiographs. IEEE Trans Biomed Eng. 2012; 59: 909-919.

22. Fleiss J. Statistical methods for rates and proportions. 2nd edition, New-York Wiley. 1981.

23. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971; 76: 378-382.

24. Robinson PJ. Radiology's Achilles' heel: error and variation in the interpretation of the Röntgen image. Br J Radiol. 1997; 70: 1085-1098.

25. Obuchowski NA, Meziane M, Dachman AH, Lieber ML, Mazzone PJ. What's the control in studies measuring the effect of computer-aided detection (CAD) on observer performance? Acad Radiol. 2010; 17: 761-767.

26. Wagner RF, Metz CE, Campbell G. Assessment of medical imaging systems and computer aids: a tutorial review. Acad Radiol. 2007; 14: 723-748.

27. Paquerault S, Hardy PT, Wersto N, Chen J, Smith RC. Investigation of optimal use of computer-aided detection systems: the role of the "machine" in decision making process. Acad Radiol. 2010; 17: 1112-1121.

28. Sheard S, Rao P, Devaraj A. Imaging of acute respiratory distress syndrome. Respir Care. 2012; 57: 607-612.

29. Desai SR. Acute respiratory distress syndrome: imaging of the injured lung. Clin Radiol. 2002; 57: 8-17.

30. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. Anesthesiology. 2004; 100: 9-15.

31. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. Crit Care Med. 1988; 16: 1110-1116.

32. Leteurtre S, Martinot A, Duhamel A, Proulx F, Grandbastien B, Cotting J, et al. Validation of the paediatric logistic organ dysfunction (PELOD) score: prospective, observational, multicentre study. Lancet. 2003; 362: 192-197.