Eur J Pediatr (2012) 171:369–374
DOI 10.1007/s00431-011-1543-1

ORIGINAL PAPER

Evaluation of the World Health Organization criteria for chest radiographs for pneumonia diagnosis in children

Shalom Ben Shimol · Ron Dagan · Noga Givon-Lavi · Asher Tal · Micha Aviram · Jacob Bar-Ziv · Vadim Zodicov · David Greenberg

Received: 16 May 2011 / Accepted: 26 July 2011 / Published online: 26 August 2011

Abstract Our objective was to compare the inter-observer level of agreement in diagnosing pneumonia using the World Health Organization (WHO) guidelines for the interpretation of radiographs. We conducted a prospective study in a pediatric emergency room. Fifteen observers (13 pediatricians, 2 radiologists) interpreted 200 pediatric (<5 years old) chest radiographs using the WHO guidelines. Observers were blinded to the clinical presentation. Results were analyzed for kappa values. Individual readings were compared to two "gold standard" teams: (1) radiologist and pediatrician and (2) two radiologists. *Results*: Alveolar pneumonia, non-alveolar pneumonia, and no pneumonia were found (by radiologists) in 12.8%, 2.7%, and 78.6% of readings, respectively. The mean kappa values for alveolar pneumonia, non-alveolar pneumonia, and no pneumonia of observers versus the team consisting of a radiologist and a pediatrician were 0.73, 0.23, and 0.61, respectively. For non-alveolar pneumonia, the mean kappa value was higher for the gold standard consisting of a radiologist and a pediatrician when compared to the two-radiologist team. Pediatricians overdiagnosed "non-alveolar pneumonia" compared with radiologists. In contrast, for the alveolar pneumonia and no-pneumonia diagnoses, no significant differences were found. *Conclusions*: The WHO guidelines for interpretation of chest radiographs result in high level of agreement between readers for the definition of "alveolar pneumonia" and "no pneumonia" but poor agreement for non-alveolar pneumonia. The disagreement with regard to the latter was associated with overdiagnosis by pediatricians, which may lead to overtreatment. We believe that radiographic non-alveolar pneumonia should not be an endpoint for clinical trials and research, nor should it be implemented in clinical setting.

Keywords Community-acquired pneumonia · Chest radiographs · Pneumonia diagnosis · World Health Organization (WHO)

S. Ben Shimol · R. Dagan · N. Givon-Lavi · D. Greenberg (✉)
Pediatric Infectious Disease Unit, Soroka University Medical Center and the Faculty of Health Sciences,
Ben-Gurion University of the Negev,
Beer-Sheva, Israel
e-mail: dudi@bgu.ac.il

A. Tal · M. Aviram
Department of Pediatrics, Soroka University Medical Center and the Faculty of Health Sciences,
Ben-Gurion University of the Negev,
Beer-Sheva, Israel

V. Zodicov
Department of Radiology, Soroka University Medical Center and the Faculty of Health Sciences,
Ben-Gurion University of the Negev,
Beer-Sheva, Israel

J. Bar-Ziv
Department of Radiology, Hadassah University Medical Center,
Jerusalem, Israel

Introduction

The diagnosis of pneumonia usually derives from the clinical presentation: cough, fever, increased respiratory rate, crackles, and decreased respiration sounds. In young children, some of these clinical signs and symptoms can be absent [21]. Radiography still remains the best available tool to diagnose pneumonia, although observer variation is high [3, 4, 6, 13, 20].

It is believed, but not proven, that alveolar pneumonia represents the radiographic pattern of pneumonia most frequently associated with bacterial infections [7, 11, 15,

23, 26]. The World Health Organization (WHO) Department of Immunization, Vaccines and Biologicals established in 2001 a working group to standardize the categorization of radiological pneumonia for the purpose of estimating burden of likely bacterial pneumonia and vaccine impact [6, 24]. This working group stated that by using standardized definitions and training, it is possible to achieve agreement in identifying radiological alveolar pneumonia. These definitions may enable comparison of results between different epidemiological studies that use radiological pneumonia as an outcome. A previous study demonstrated that training of doctors, using standardized features with the help of software, substantially improved agreement in identifying radiological pneumonia, with kappa value increasing from a range of 0.1 to 0.2 before training to a range of 0.37 to 0.52 after training, indicating moderate to good post-training agreement [18].

The WHO working group recommended two readers (one radiologist and one pediatrician) as the "gold standard" observers, but in cases of disagreement between those two readers, the consensus reading was made by two radiologists [6, 24]. However, several previous studies required two radiologists as observers for the gold standard for pneumonia diagnosis [1, 9, 27, 28]. Other studies used several different combinations of readers: pediatric residents, experienced pediatricians, or pediatric pulmonologists [4, 8, 19, 28].

Studies evaluating the efficacy of various pneumococcal and *Haemophilus influenzae* type b (Hib) vaccines, using pneumonia as an endpoint, require gold standard definition of pneumonia [5, 16, 17, 25]. These vaccine trials used the WHO criteria for diagnosis of radiographic pneumonia in children as an endpoint, demonstrating efficacy between 20% and 37% in preventing alveolar pneumonia [8, 24].

The objectives of our study were (1) to compare the inter-observer variations for three WHO definitions—alveolar pneumonia, non-alveolar pneumonia, and no pneumonia—and (2) to compare the level of agreement for the diagnosis of pneumonia according to the WHO guidelines (a radiologist and a pediatrician) versus two radiologists for each category of diagnosis.

Materials and methods

Study design

Setting The study was conducted at the Soroka University Medical Center in Beer Sheva, Israel. Chest radiographs of children <5 years old obtained at the pediatric emergency room because of clinical suspicion of pneumonia were included. We analyzed chest radiographs from 200 children: 100 chest radiographs (anteroposterior and lateral views) each month for two consecutive months (January and February 2001), regardless of their diagnosis.

All X-rays were processed and visualized in Hipax 3.27.1 X-ray Image processing software, Ateinhart Medizinsysteme, Germany and were stored for future analyses. All images were visualized by readers as Jpeg file format at a resolution of 1,024×768 pixels.

Readers Fifteen readers (observers) took part in this study: 13 pediatricians and 2 radiologists specializing in pediatric radiology. Of the 13 pediatricians, two were certified pediatric pulmonologists, two were emergency medicine pediatricians, two were certified infectious disease specialists interested in pneumonia vaccine studies, and seven were senior residents in their last year of pediatric training.

We chose the group of observers according to their specialty, with emphasis given on those who interpret chest radiograph on a daily basis, as part of their clinical work. All observers were introduced to the WHO criteria for radiographic pneumonia as part of their training for the study.

Observers were asked to review individually all chest radiographs and categorize them according to the WHO criteria [6, 24]. All observers were blinded to the clinical presentation of the patients and to other readers' interpretation. We assigned one pediatrician (reader 13, chosen randomly) and two radiologists (readers 14 and 15) as the gold standard readers, in accordance with the WHO regulation for gold standard readers. In cases of disagreement between the two readers of the gold standard readers, those radiographs were excluded from the analysis.

Study definitions

The WHO radiographic definitions were as follows:

"Category 1" Alveolar pneumonia (endpoint consolidation): dense fluffy consolidation (alveolar infiltrate) of a portion of a lobe or entire lung. This often contains air bronchogram and may be associated with a pleural effusion.

"Category 2" Non-alveolar pneumonia (non-endpoint): linear and patchy densities (interstitial infiltrates) in a lacy pattern involving both lungs, featuring peribronchial thickening and multiple areas of atelectasis. Lung inflation is normal to increased.

"Category 3" No pneumonia: no consolidation/infiltrate, absence of endpoint consolidation, other infiltrate, or pleural effusion.

Technically unreadable chest radiographs were excluded. This study was approved by the Human Ethics Committee of the Soroka University Medical Center.

Statistical analysis

All results were analyzed by SPSS software for calculating kappa values. Agreement rates for radiographic findings were calculated by summing the proportion of all radiographs that read as category 1, 2, or 3 ("alveolar pneumonia", "non-alveolar pneumonia," and "no pneumonia") by the reviewers. The kappa coefficient (k) is an appropriate index of agreement between two (or more) readers who are regarded as more or less interchangeable and between repeat readings by the same reader. Kappa values and their 95% confidence intervals, a measure of inter-reader reliability that adjusts for random agreement, were used to assess the statistical significance of agreement between reviewers. Several different analyses were conducted using different readers (two radiologists; a radiologist and a pediatrician) for each analysis group.

Data from previous studies suggested a pattern of kappa values (inter-observer agreement) in the region of 0.80 for individual radiographic features and 0.30–0.60 for composite assessments of features. Kappa value of 0.75 (i.e., 75% agreement after adjustment for chance) is regarded as "good" or "very good" and 0.30–0.60 as "fair" to "moderate" [2, 22].

Positive and negative predictive values were calculated to compare inter-observer agreement between senior pediatricians, pediatric residents, and the two radiologists (gold standard readers). A two-tailed p value <0.05 was considered statistically significant.

Results

Two hundred chest radiographs were reviewed. Thirteen radiographs were defined as "unreadable chest X-ray" by most readers, thus 187 chest radiographs constituted the final studied sample. Alveolar pneumonia was diagnosed by both radiologists in 12.8% of chest radiographs (within observer variation: 10.2% to 17.1%), non-alveolar pneumonia was found in 2.7% (2.7% to 29.9%), and no pneumonia in 78.6% (62% to 83.9%).

In our data analysis (including mean and median number of readings, kappa values, positive and negative predictive values, and standard deviation), we excluded reader number 2, as this reader's diagnoses (34.2%, 30.5%, and 25.7% for alveolar pneumonia, non-alveolar pneumonia, and no pneumonia, respectively) were markedly different from all other readers' diagnoses. However, we do present all data regarding reader 2 in the graphs.

Correlation between readers

The mean overall kappa value of readers (readers 1–12) versus a gold standard consisting of a radiologist and a pediatrician (readers 13 and 14; WHO definition as gold standard readers) was 0.59±0.14, and the mean overall kappa value of readers (1–13) versus two radiologists (readers 14 and 15 as gold standard) was 0.54±0.09. The p value between the two mean kappa values was not significant.

The best correlation between readers was found for the presence of alveolar pneumonia followed by no pneumonia (Figs. 1 and 2). For these two categories, no significant difference was found when the comparison was made with the gold standard of a radiologist and a pediatrician or two radiologists. For non-alveolar pneumonia, the correlation was very poor but was the worst when the individual readings were compared to the gold standard of two radiologists.

The pediatricians overdiagnosed non-alveolar pneumonia. While only 2.7% (5/187) of all radiographs were interpreted by both radiologists as non-alveolar pneumonia, 14.4% (27/187) and 12.3% (23/187) of all radiographs were interpreted as non-alveolar pneumonia by senior pediatricians and residents, respectively. Generally, more radiographs were read as "negative" (no pneumonia) by the two radiologists (78.6% and 83.9%) versus senior pediatricians and residents (69.5% to 73.8%, respectively).

Deriving from the low kappa values of pediatricians in regard to non-alveolar pneumonia and its overdiagnosis, the positive predictive value of this diagnosis was low among pediatricians (Table 1). In contrast, since most radiographs were read by the radiologists as no pneumonia, the negative predictive value was relatively higher among pediatricians.

Discussion

Currently, the most commonly used and most pragmatic method for diagnosing pneumonia is radiography [10, 12, 14, 25]. However, inter-observer variation is high for interpretation of chest radiographs, and in clinical setting, it is important to realize that chest radiography should not be considered as a single, "objective" tool for pneumonia diagnosis.

Prior to the publication of the "Standardized Radiographic Definitions for Diagnosing Pneumonia in Developed Population" by the WHO Working Group [24], it was universally agreed that there was no strict radiological definition of pneumonia in children. Instead, the diagnosis of pneumonia was based upon a spectrum of clinical signs and symptoms that are consistent with the clinical and pathological diagnosis of pneumonia. At one end of the

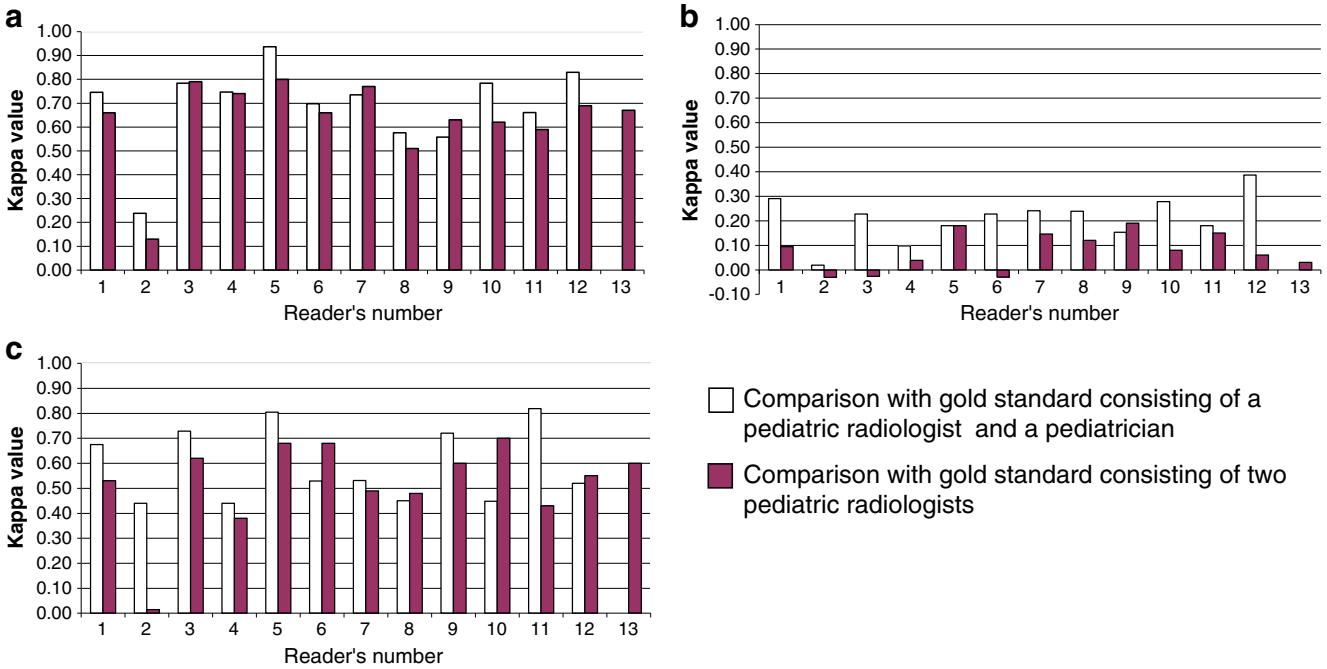

Fig. 1 Kappa values for all categories—two radiologists versus radiologist and pediatrician as gold standard methods. **a** Kappa values for alveolar pneumonia. **b** Kappa values for non-alveolar pneumonia. **c** Kappa values for no pneumonia

chest radiograph, spectrum is the typical appearance of alveolar/lobar consolidation, which is considered to be strongly associated with bacterial pneumonia [10, 12, 14, 25]. At the other end are the mild interstitial and perihilar changes that are often considered to be associated with viral infections or asthma [10, 12, 14, 25].

In our study, we found a mean kappa value of 0.73 between readers and a gold standard consisting of a radiologist and a pediatrician for diagnosing alveolar pneumonia. This suggests a very good general agreement for this entity, when using the WHO criteria. The overall range of the mean kappa values in our study is similar to that found by other authors for a variety of radiographic diagnoses. However, good and very good agreement does not necessarily imply high validity (closeness to the truth). Observer agreement is necessary for validity, but observers may agree and nevertheless both are wrong.

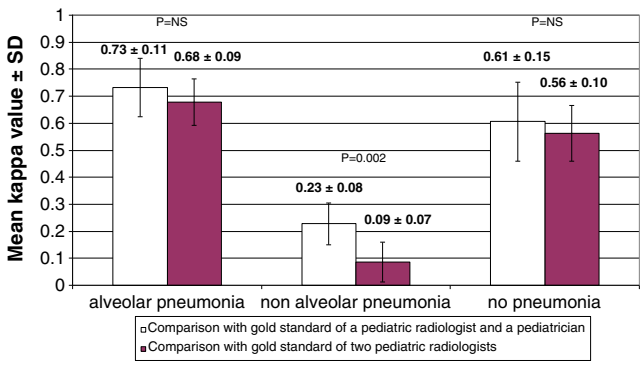

Fig. 2 Mean kappa values comparison of three categories using gold standard consisting of two radiologists versus a radiologist and a pediatrician as gold standard

The WHO definitions for diagnosing radiographic pneumonia are generally acceptable, regardless of the reader's identity (in terms of specialty and experience). These criteria can be implemented by a simple teaching program, and its assimilation may lead to a good level of agreement for diagnosing radiographic pneumonia in children [6].

We found major differences in agreement between readers, when comparing the different categories of criteria. First, with regard to alveolar pneumonia, the reading was characterized by very good level of agreement between readers, as indicated by mean kappa values in the range of 0.66 to 0.79. This entity was also characterized by high level of positive (~80%) and negative (~97%) predictive values when pediatricians were compared to radiologists. However, we also found that experienced pediatricians had a better predictive values and correlation with radiologists compared with the pediatric resident with regard to alveolar pneumonia. Nevertheless, the good level of agreement, with regard to both positive and negative predictive values, suggests that using alveolar pneumonia for the comparison of study outcomes is appropriate.

Second, with regard to non-alveolar pneumonia, the reading was characterized by low level of agreement between all readers, as indicated by mean kappa values in the range of 0.09 to 0.23 between readers and a gold

Table 1 Median number of diagnoses, sensitivity, specificity, and kappa values comparison between senior pediatricians, pediatric residents, and radiologists

	Median number of diagnoses			Median positive predictive value		Median negative predictive value		Median kappa value[a]	
	Radiologists	Senior pediatricians	Pediatric residents	Senior pediatricians (%)	Pediatric residents (%)	Senior pediatricians (%)	Pediatric residents (%)	Senior pediatricians	Pediatric residents
Alveolar pneumonia	24	22	26	89	78	98	97	0.83	0.76
Non-alveolar pneumonia	10	22	21	25	13	97	96	0.08	0.13
No pneumonia	152	132	141	98	98	47	70	0.54	0.72

[a] Comparison to gold standard consisting of two radiologists

standard consisting of a radiologist and a pediatrician or of the two radiologists, respectively, and by very low positive predictive values (~20%) when pediatricians were compared to radiologists. Several previous studies (including a meta-analysis of those studies) also found a poor kappa for non-alveolar pneumonia [10, 12, 14, 25].

These data suggest that the WHO criteria for diagnosis of non-alveolar pneumonia are vague and need better specification, as the current definitions are problematic. Our data demonstrate higher rate of reading interpretation as non-alveolar pneumonia versus no pneumonia by senior pediatricians and pediatric residents compared with radiologists. It seems that pediatricians tend to overdiagnose non-alveolar pneumonia, possibly influenced by their wish not to miss "early" pneumonia (not yet fully demonstrated on chest radiograph) and leave the child untreated. Thus, the WHO criteria could lead to overdiagnosis of pneumonia, especially in the non-alveolar category. We believe that diagnosis of radiographic non-alveolar pneumonia should not be an endpoint for clinical trials and research, nor should it be implemented in clinical settings.

Third, with regard to no pneumonia, the reading was characterized by good level of agreement between readers, as indicated by kappa values in the range of 0.56 to 0.65, high level of positive predictive values (~98%), and moderate levels of negative predictive values (~55%) when pediatricians were compared to radiologists. However, we did find higher rate of no-pneumonia reading by radiologists than by pediatricians (both specialists and residents). This probably reflects again the pediatrician's tendency to overdiagnose pneumonia.

In our study, we did not find any differences in kappa values between the WHO recommendations for gold standard readers (a radiologist and a pediatrician) and two radiologist readers for diagnosing alveolar pneumonia and no-pneumonia endpoints, but we did find a significant difference in kappa values between the two gold standard groups for diagnosing non-alveolar pneumonia endpoint. We therefore believe that using the WHO modality of two readers (a radiologist and a pediatrician) is an acceptable method for diagnosing pneumonia. Moreover, for practical reasons, in the future, we believe that using the WHO modality is the best option, since most studies uses only one or two readers to set their gold standard diagnosis rather than using a consensus of multiple (15 in our study) readers/observers.

Conclusions The WHO guidelines for interpretation of chest radiographs result in high level of agreement between readers for the definition of alveolar pneumonia and no pneumonia but poor agreement for non-alveolar pneumonia. The disagreement with regard to the latter was associated with overdiagnosis by pediatricians, which may lead to overtreat-

ment. We believe that radiographic non-alveolar pneumonia should not be an endpoint for clinical trials and research, nor should it be implemented in clinical setting.

Conflict of interest The authors have no conflict of interest or financial relationships relevant to this article to disclose.

References

1. Albaum MN, Hill LC, Murphy M, Li YH, Fuhrman CR, Britton CA, Kapoor WN, Fine MJ (1996) Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT Investigators. Chest 110:343–350
2. Altman D (1991) Practical statistics for medical research. Chapman & Hall, London, p 404
3. Ayieko P, English M (2007) Case management of childhood pneumonia in developing countries. Pediatr Infect Dis J 26:432–440
4. Bada C, Carreazo NY, Chalco JP, Huicho L (2007) Inter-observer agreement in interpreting chest X-rays on children with acute lower respiratory tract infections and concurrent wheezing. Sao Paulo Med J 125:150–154
5. Black S, Shinefield H, Fireman B, Lewis E, Ray P, Hansen JR, Elvin L, Ensor KM, Hackell J, Siber G, Malinoski F, Madore D, Chang I, Kohberger R, Watson W, Austrian R, Edwards K (2000) Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. Pediatr Infect Dis J 19:187–195
6. Cherian T, Mulholland EK, Carlin JB, Ostensen H, Amin R, de Campo M, Greenberg D, Lagos R, Lucero M, Madhi SA, O'Brien KL, Obaro S, Steinhoff MC (2005) Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. Bull World Health Organ 83:353–359
7. Esposito S, Bosis S, Cavagna R, Faelli N, Begliatti E, Marchisio P, Blasi F, Bianchi C, Principi N (2002) Characteristics of *Streptococcus pneumoniae* and atypical bacterial infections in children 2–5 years of age with community-acquired pneumonia. Clin Infect Dis 35:1345–1352
8. Hansen J, Black S, Shinefield H, Cherian T, Benson J, Fireman B, Lewis E, Ray P, Lee J (2006) Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than 5 years of age for prevention of pneumonia: updated analysis using World Health Organization standardized interpretation of chest radiographs. Pediatr Infect Dis J 25:779–781
9. Hazir T, Nisar YB, Qazi SA, Khan SF, Raza M, Zameer S, Masood SA (2006) Chest radiography in children aged 2–59 months diagnosed with non-severe pneumonia as defined by World Health Organization: descriptive multicentre study in Pakistan. BMJ 333:629
10. Jadavji T, Law B, Lebel MH, Kennedy WA, Gold R, Wang EE (1997) A practical guide for the diagnosis and treatment of pediatric pneumonia. CMAJ 156:S703–S711
11. Korppi M, Heiskanen-Kosma T, Leinonen M (1997) White blood cells, C-reactive protein and erythrocyte sedimentation rate in pneumococcal pneumonia in children. Eur Respir J 10:1125–1129
12. Leventhal JM (1982) Clinical predictors of pneumonia as a guide to ordering chest roentgenograms. Clin Pediatr (Phila) 21:730–734
13. Mandell LA, Marrie TJ, Grossman RF, Chow AW, Hyland RH (2000) Summary of Canadian guidelines for the initial management of community-acquired pneumonia: an evidence-based update by the Canadian Infectious Disease Society and the Canadian Thoracic Society. Can J Infect Dis 11:237–248
14. Margolis P, Gadomski A (1998) The rational clinical examination. Does this infant have pneumonia? JAMA 279:308–313
15. Mulholland K (1999) Magnitude of the problem of childhood pneumonia. Lancet 354:590–592
16. Mulholland K, Hilton S, Adegbola R, Usen S, Oparaugo A, Omosigho C, Weber M, Palmer A, Schneider G, Jobe K, Lahai G, Jaffar S, Secka O, Lin K, Ethevenaux C, Greenwood B (1997) Randomised trial of *Haemophilus influenzae* type-b tetanus protein conjugate vaccine [corrected] for prevention of pneumonia and meningitis in Gambian infants. Lancet 349:1191–1197
17. Mulholland K, Levine O, Nohynek H, Greenwood BM (1999) Evaluation of vaccines for the prevention of pneumonia in children in developing countries. Epidemiol Rev 21:43–55
18. Patel AB, Amin A, Sortey SZ, Athawale A, Kulkarni H (2007) Impact of training on observer variation in chest radiographs of children with severe pneumonia. Indian Pediatr 44:675–681
19. Sarria E, Fischer GB, Lima JA, Menna Barreto SS, Flores JA, Sukiennik R (2003) Interobserver agreement in the radiological diagnosis of lower respiratory tract infections in children. J Pediatr (Rio J) 79:497–503
20. Schuchat A, Dowell SF (2004) Pneumonia in children in the developing world: new challenges, new solutions. Semin Pediatr Infect Dis 15:181–189
21. Stein RT, Marostica PJ (2007) Community-acquired pneumonia: a review and recent advances. Pediatr Pulmonol 42:1095–1103
22. Swingler GH (2001) Observer variation in chest radiography of acute lower respiratory infections in children: a systematic review. BMC Med Imaging 1:1
23. Virkki R, Juven T, Rikalainen H, Svedstrom E, Mertsola J, Ruuskanen O (2002) Differentiation of bacterial and viral pneumonia in children. Thorax 57:438–441
24. World Health Organization Pneumonia Vaccine Trial Investigator's Group (2001) Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children. WHO/V&B/01.35. World Health Organization, Geneva
25. Wilkins TR, Wilkins RL (2005) Clinical and radiographic evidence of pneumonia. Radiol Technol 77:106–110
26. Wubbel L, Muniz L, Ahmed A, Trujillo M, Carubelli C, McCoig C, Abramo T, Leinonen M, McCracken GH Jr (1999) Etiology and treatment of community-acquired pneumonia in ambulatory children. Pediatr Infect Dis J 18:98–104
27. Young M, Marrie TJ (1994) Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia. Arch Intern Med 154:2729–2732
28. Zenteno AD (2008) Utility of chest x-ray like approach to etiology of community acquired pneumonia in children. Rev Chilena Infectol 25:17–21 [Article in Spanish]