

Research article

Observer variation in chest radiography of acute lower respiratory infections in children: a systematic review

George H Swingler

Address: School of Child and Adolescent Health, University of Cape Town and Red Cross Children's Hospital, Klipfontein Road, Cape Town, 7700 South Africa

E-mail: swingler@ich.uct.ac.za

Published: 12 November 2001

BMC Medical Imaging 2001, 1:1

Received: 17 September 2001

Accepted: 12 November 2001

This article is available from: <http://www.biomedcentral.com/1471-2342/1/1>

© 2001 Swingler; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any non-commercial purpose, provided this notice is preserved along with the article's original URL. For commercial use, contact info@biomedcentral.com

Abstract

Background: Knowledge of the accuracy of chest radiograph findings in acute lower respiratory infection in children is important when making clinical decisions.

Methods: I conducted a systematic review of agreement between and within observers in the detection of radiographic features of acute lower respiratory infections in children, and described the quality of the design and reporting of studies, whether included or excluded from the review.

Included studies were those of observer variation in the interpretation of radiographic features of lower respiratory infection in children (neonatal nurseries excluded) in which radiographs were read independently and a clinical population was studied. I searched MEDLINE, HealthSTAR and HSRPROJ databases (1966 to 1999), handsearched the reference lists of identified papers and contacted authors of identified studies. I performed the data extraction alone.

Results: Ten studies of observer interpretation of radiographic features of lower respiratory infection in children were identified. Seven of the studies satisfied four or more of the seven design and reporting criteria. Six studies met the inclusion criteria for the review. Inter-observer agreement varied with the radiographic feature examined. Kappa statistics ranged from around 0.80 for individual radiographic features to 0.27–0.38 for bacterial vs viral etiology.

Conclusions: Little information was identified on observer agreement on radiographic features of lower respiratory tract infections in children. Agreement varied with the features assessed from "fair" to "very good". Aspects of the quality of the methods and reporting need attention in future studies, particularly the description of criteria for radiographic features.

Background

Chest radiography is a very common investigation in children with lower respiratory infection, and knowledge of the diagnostic accuracy of radiograph interpretation is consequently important when basing clinical decisions on the findings. Inter- and intra-observer agreement in the interpretation of the radiographs are necessary com-

ponents of diagnostic accuracy. Observer variation is however not sufficient for diagnostic accuracy. The key element of such accuracy is the concordance of the radiological interpretation with the presence or absence of pneumonia. Unfortunately there is seldom a suitable available reference standard for pneumonia (such as histological or gross anatomical findings) against which to

compare radiographic findings. Diagnostic accuracy thus needs to be examined indirectly, including assessing observer agreement.

Observer variation in chest radiograph interpretation in acute lower respiratory infections in children has not been systematically reviewed.

The purpose of this study was to quantify the agreement between and within observers in the detection of radiographic features associated with acute lower respiratory infections in children. A secondary objective was to assess the quality of the design and reporting of studies of this topic, whether or not the studies met the quality inclusion criteria for the review.

Methods

Inclusion criteria

Studies meeting the following criteria were included in the systematic review:

1. An assessment of observer variation in interpretation of radiographic features of lower respiratory infection, or of the radiographic diagnosis of pneumonia.
2. Studies of children aged 15 years or younger or studies from which data on children 15 years or younger could be extracted. Studies of infants in neonatal nurseries were excluded.
3. Data presented that enabled the assessment of agreement between observers.
4. Independent reading of radiographs by two or more observers.

5. Studies of a clinical population with a spectrum of disease in which radiographic assessment is likely to be used (as opposed to separate groups of normal children and those known to have the condition of interest).

Literature search

Studies were identified by a computerized search of MEDLINE from 1966 to 1999 using the following search terms: observer variation, or intraobserver (text word), or interobserver (text word); and radiography, thoracic, or radiography or bronchiolitis/ra, or pneumonia, viral/ra, or pneumonia, bacterial/ra, or respiratory tract infections/ra. The search was limited to human studies of children up to the age of 18 years. The author reviewed the titles and abstracts of the identified articles in English or with English abstracts (and the full text of those judged to be potentially eligible). A similar search was performed of HealthSTAR, a former on-line database of published health service research, and the HSRPROJ (Health Services Research Projects in Progress) database. Reference lists of articles retrieved from the above searches were examined. Authors of studies of agreement between independent observers on chest radiograph findings in acute lower respiratory infections in children were contacted with an inquiry about the existence of additional studies, published or unpublished.

Data collection and analysis

The author evaluated for inclusion potentially relevant studies identified in the above search. Characteristics of study design and reporting listed in Table 1 were recorded in all studies of observer variation in the interpretation of radiographic features of lower respiratory infection in children aged 15 years or younger (except infants in neonatal nurseries). The criteria for validity were

Table 1: Characteristics of study design and reporting

	Present ^a	Absent ^b	Unclear ^c
Validity eligibility criteria			
Independent assessment of radiographs	9	1	0
Relevant clinical population (not case-control design)	7	3	0
Other validity characteristics			
Description of study population (3 of age, M:F ratio, clinical features and eligibility criteria)	6	4	0
Description of criteria for radiological signs	4	6	0
Presentation of indeterminate results	7	2	1
Applicability			
Meaningful measures of agreement (kappa or equivalent)	8	2	0
Confidence intervals for measures of agreement	1	9	0
Assessment of intra-observer variability	3	7	0

^a Study characteristic present, according to research report ^b Study characteristic absent, according to research report ^c Insufficient information to determine whether the characteristic was present.

Table 2: Characteristics of included studies

Author	Subjects	Observers
Simpson et al 1974 [14] ^a	330 children under 14 years hospitalized with acute lower respiratory infection	2 radiologists
McCarthy et al 1981 [15]	128 of 1566 children seen in a pediatric emergency room with a pulmonary infiltrate in chest radiography (as judged by the duty radiologist)	2 radiologists
Crain et al 1991 [9]	230 of 242 febrile infants under 8 weeks evaluated in an emergency room and who received a chest radiograph	2 radiologists
Kramer et al 1992 [12]	287 unreferred febrile children, aged 3–24 months, in an emergency unit	1 pediatrician, 1 duty radiologist, 1 "blind" pediatric radiologist
Davies et al 1996 [10] ^b	40 children under 6 months, 25 with pneumonia and 15 with bronchiolitis, admitted to a tertiary care pediatric hospital	3 pediatric radiologists
Coakley et al 1996 [8]	113 previously well children under 3 years hospitalized with acute respiratory infections and no focal abnormality on radiography	2 radiologists

^aKappa calculated from data extracted from the report ^bAverage weighted kappa

Table 3: Observer agreement: kappa statistics (95% confidence intervals)

Radiographic features	Davies 1996 [10]	Simpson 1974 [14]	Coakley 1996 [8]	Kramer 1992 [12]	Crain 1991 [9]	McCarthy 1981 [15]
Inter-observer variation						
Consolidation	0.79					
Pneumonia				0.46 (0.34–0.58) 0.47 (0.35–0.60)		
Collapse/consolidation		0.83 (0.72–0.94)				
Collapse/atelectasis	0.78					
Hyperinflation/air trapping	0.83	0.78 (0.67–0.89)				
Peribronchial/ bronchial wall thickening	0.55	0.55 (0.44–0.66)	0.43 (0.25–0.61)			
Perihilar linear opacities	0.82					
Abnormal					0.61 (0.48–0.74)	
Bacterial vs. viral etiology						0.27–0.38
Intra-observer variation						
Consolidation	0.91					
Collapse/atelectasis	0.86					
Hyperinflation/air trapping	0.85					
Peribronchial /bronchial wall thickening	0.76					
Perihilar linear opacities	0.87					

^aTwo observer pairs

those for which empirical evidence exists for their importance in the avoidance of bias in comparisons of diagnostic tests with reference standards, and which were relevant to tests of observer agreement. The selected criteria for applicability were those featured by at least two of five sources of such recommendations. No weighting was applied to the criteria, except the use of the two most frequently recommended validity criteria (recommend-

ed by at least four out of five sources) as the methodological inclusion criteria [1–5]. No attempt was made to derive a quality score.

In studies meeting all the inclusion criteria for the review, the author extracted the following additional information: number and characteristics of the observers and children studied, and measures of agreement. When no

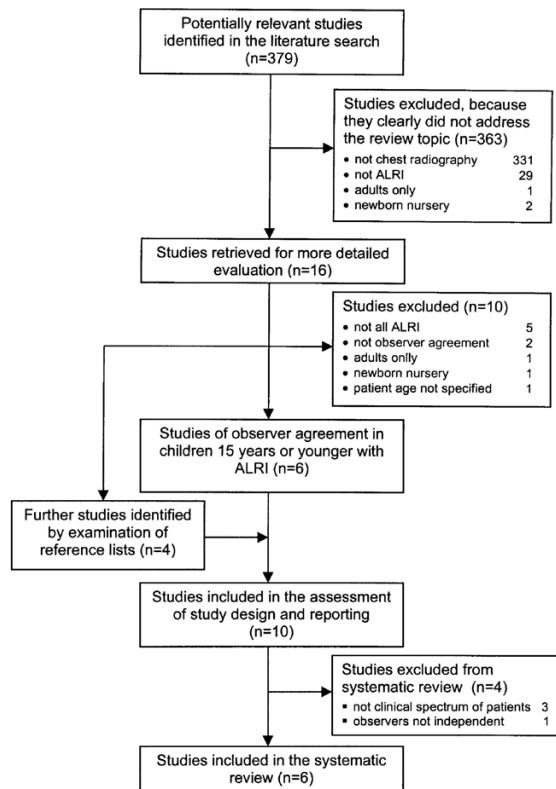


Figure 1
Review profile

measures of agreement were reported, data were extracted from the reports and kappa statistics were calculated using the method described by Fleiss [6]. Kappa is a measure of the degree of agreement between observations, over and above that expected by chance. If agreement is complete, kappa = 1; if there is only chance concordance, kappa = 0.

Results

A review profile is shown in Figure 1. For a list of rejected studies, with reasons for rejection, see Additional file 1: Rejected studies. Ten studies of observer variation in the interpretation of radiographic features of lower respiratory infection in children aged 15 years or younger were identified [7–16]. Contact was established with five of nine authors in whom it was attempted. No additional studies were included in the systematic review as a result of this contact.

The characteristics of the study design and reporting of the 10 studies of observer interpretation of radiographic features of lower respiratory infection in children are summarized in Table 1. Seven of the studies satisfied four

or more of the seven design and reporting criteria. Four studies described criteria for the radiological signs. Six of the studies satisfied the inclusion criteria for the systematic review [8–10,12,14,15]. Of the remaining four studies, three were excluded because a clinical spectrum of patients had not been used [7,13,16] and one because observers were not independent [11]. The characteristics of included studies are shown in Table 2.

A kappa statistic was calculated from data extracted from one report [14], and confidence intervals in three studies in which they were not reported but for which sufficient data were available in the report [8,9,14]. A summary of kappa statistics is shown in Table 3. Inter-observer agreement varied with the radiographic feature examined. Kappas for individual radiographic features were around 0.80, and lower for composite assessments such as the presence of pneumonia (0.47), radiographic normality (0.61) and bacterial *vs.* viral etiology (0.27–0.38). Findings were similar in the two instances in which more than one study examined the same radiographic feature (hyperinflation/air trapping and peribronchial/bronchial wall thickening). When reported, kappa statistics for intra-observer agreement were 0.10–0.20 higher than for inter-observer agreement.

Discussion

The quality of the methods and reporting of studies was not consistently high. Only six of 10 studies satisfied the inclusion criteria for the review. The absence of any of the validity criteria used in this study (independent reading of radiographs, the use of a clinical population with an appropriate spectrum of disease, description of the study population and of criteria for a test result) has been found empirically to overestimate test accuracy, on average, when a test is compared with a reference standard [1]. A similar effect may apply to the estimation of inter-observer agreement, in that two observers may agree with each other more often when aware of each other's assessment, and radiographs drawn from separate populations of normal and known affected children will exclude many of the equivocal radiographs in a usual clinical population, thereby possibly falsely increasing agreement. Only four of ten studies described criteria for the radiological signs, with potential negative implications for both the validity and the applicability of the remaining studies.

The data from the included studies suggest a pattern of kappas in the region of 0.80 for individual radiographic features and 0.30–0.60 for composite assessments of features. Kappa of 0.80 (i.e. 80% agreement after adjustment for chance) is regarded as "good" or "very good" and 0.30–0.60 as "fair" to "moderate" [17]. The small number of studies in this review however makes the de-

tection and interpretation of patterns merely speculative. Only two radiographic features were examined by more than one study. There is thus insufficient information to comment on heterogeneity of observer variation in different clinical settings.

The range of kappas overall is similar to that found by other authors for a range of radiographic diagnoses⁷. 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 be wrong.

Conclusions

Little information was identified on inter-observer agreement in the assessment of radiographic features of lower respiratory tract infections in children. When available, it varied from "fair" to "very good" according to the features assessed. Insufficient information was identified to assess heterogeneity of agreement in different clinical settings.

Aspects of the quality of methods and reporting that need attention in future studies are independent assessment of radiographs, the study of a usual clinical population of patients and description of that population, description of the criteria for radiographic features, assessment of intra-observer variation and reporting of confidence intervals around estimates of agreement. Specific description of criteria for radiographic features is particularly important, not only because of its association with study validity but also to enable comparison between studies and application in clinical practice.

Competing interests

None declared

Additional material

Additional file 1

Studies excluded during the literature search and study selection, listed according to reason for exclusion.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2342-1-1-S1.doc>]

Acknowledgements

Financial support from the University of Cape Town and the Medical Research Council of South Africa is acknowledged.

References

1. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM: **Empirical evidence of design-related bias in studies of diagnostic tests.** *JAMA* 1999, **282**:1061-1066

2. Jaeschke R, Guyatt G, Sackett DL, for the Evidence-Based Medicine Working Group: **Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid?** *JAMA* 1994, **271**:389-391
3. Greenhalgh T: **How to read a paper. Papers that report diagnostic or screening tests.** *BMJ* 1997, **315**:540-543
4. Reid MC, Lachs MS, Feinstein AR: **Use of methodological standards in diagnostic test research. Getting better but still not good.** *JAMA* 1995, **274**:645-651
5. **Cochrane Methods Working Group on Systematic Reviews of Screening and Diagnostic Tests: Recommended methods,** 6 June 1996 [<http://wwwsom.fmc.flinders.edu.au/FUSA/COCHRANE/cochrane/sadtdoc1.htm>]
6. Fleiss JL: **Statistical methods for rates and proportions, 2nd edn.** New York, John Wiley & Sons 1981:212-225
7. Coblenz CL, Babcock CJ, Alton D, Riley BJ, Norman G: **Observer variation in detecting the radiographic features associated with bronchiolitis.** *Invest Radiol* 1991, **26**:115-118
8. Coakley FV, Green J, Lamont AC, Rickett AB: **An investigation into perihilar inflammatory change on the chest radiographs of children admitted with acute respiratory symptoms.** *Clin Radiol* 1996, **51**:614-617
9. Crain EF, Bulas D, Bijur PE, Goldman HS: **Is a chest radiograph necessary in the evaluation of every febrile infant less than 8 weeks of age?** *Pediatrics* 1991, **88**:821-824
10. Davies HD, Wang EE, Manson D, Babyn P, Shuckett B: **Reliability of the chest radiograph in the diagnosis of lower respiratory infections in young children.** *Pediatr Infect Dis J* 1996, **15**:600-604
11. Kiekara O, Korppi M, Tanska S, Soimakallio S: **Radiographic diagnosis of pneumonia in children.** *Ann Med* 1996, **28**:69-72
12. Kramer MM, Roberts-Brauer R, Williams RL: **Bias and "overcall" in interpreting chest radiographs in young febrile children.** *Pediatrics* 1992, **90**:11-13
13. Norman GR, Brooks LR, Coblenz CL, Babcock CJ: **The correlation of feature identification and category judgments in diagnostic radiology.** *Mem Cognit* 1992, **20**:344-355
14. Simpson W, Hacking PM, Court SDM, Gardner PS: **The radiographic findings in respiratory syncytial virus infection in children. Part I. Definitions and interobserver variation in assessment of abnormalities on the chest x-ray.** *Pediatr Radiol* 1974, **2**:155-160
15. McCarthy PL, Spiesel SZ, Stashwick CA, Ablow RC, Masters SJ, Dolan TF: **Radiographic findings and etiologic diagnosis in ambulatory childhood pneumonias.** *Clin Pediatr (Phila)* 1981, **20**:686-691
16. Stickler GB, Hoffman AD, Taylor WF: **Problems in the clinical and roentgenographic diagnosis of pneumonia in young children.** *Clin Pediatr (Phila)* 1984, **23**:398-399
17. Altman DG: **Practical statistics for medical research.** London, Chapman & Hall 1991:404

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedCentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com