

# Chest Roentgenogram in the Evaluation of Heart Defects in Asymptomatic Infants and Children With a Cardiac Murmur: Reproducibility and Accuracy

Niels H. Birkebæk, MD\*; Lars K. Hansen, MD\*; Bo Elle, MD†; Poul E. Andersen, MD, PhD‡; Merete Friis, MD§; Mogens Egeblad, MD§; Karen Damgaard-Pedersen, MD, DMSc||; Nina T. Nielsen, MD¶; and Henrik Oxhøj, DMSc\*

**ABSTRACT.** *Objectives.* To evaluate the reproducibility and the accuracy of pediatric radiologists' assessments of chest radiographs with respect to the presence or absence of heart defects in children with an asymptomatic heart murmur.

*Design.* Ninety-eight children, ages 1 month to 15 years (median, 30.1 months), referred for evaluation of a heart murmur were consecutively included. They all had a standard chest radiograph and a color Doppler echocardiograph (CDE) performed. Six specialists in pediatric radiology evaluated the chest radiographs independently on two occasions 6 months apart. The radiologists were asked to classify each set of films into one of two categories: heart disease or no heart disease. The outcome of the CDE was considered the definite diagnosis.  $\kappa$  statistics were used to analyze the reproducibility of the radiologic assessments. Sensitivity, specificity, and the predictive value of a positive and a negative test were used for evaluation of the accuracy of the radiologic assessments.

*Results.* Mean intra- and interobserver  $\kappa$  values were all  $<0.6$ , and the majority were  $<0.4$ . Mean sensitivity was 0.3 (range: 0.17–0.52), mean predictive value of a positive test was 0.4, implying that 60% of the positive assessments were falsely positive. Mean specificity was 0.86 (range: 0.75–0.93) and the mean predictive value of a negative test was 0.80 implying that 20% of the negative assessments were falsely negative.

*Conclusion* We found a low reproducibility, as well as a low accuracy, of the radiologic assessments of the chest radiographs of children with an asymptomatic heart murmur with respect to the presence or absence of heart disease. A false-positive radiologic assessment of the chest radiograph with respect to heart defects causes unnecessary anxiety and further examinations, whereas a false-negative assessment might result in omission of relevant investigations and proper identification of the heart defect. We cannot recommend the use of chest radiographs in the initial evaluation of the asymptomatic child with a heart murmur. If a heart defect cannot be excluded by clinical examination a CDE must be performed. *Pediatrics* 1999;103(2). URL: <http://www.pediatrics.org/cgi/content/full/103/2/e15>;

*child, cardiac murmur, chest radiograph, reproducibility, accuracy.*

---

ABBREVIATION. CDE, color Doppler echocardiography.

---

We recently studied the value of various methods in the evaluation of otherwise asymptomatic children with a heart murmur, and we found that chest radiography did not contribute usefully to the evaluation of these children.<sup>1</sup> The chest radiographs were evaluated blindly by various radiologists during the daily routine. Our observations made us wonder whether the failing diagnostic contribution of the chest radiograph in the primary evaluation of asymptomatic children with a heart murmur could be ascribed to the method or whether it should be ascribed to variations in the diagnostic standards of the radiologists working in the daily routine.

In the present study we examined the reproducibility and accuracy of specialists in pediatric radiology in their evaluation of chest radiographs with respect to the presence or absence of heart defects in asymptomatic children with a heart murmur.

## PATIENTS AND METHODS

One hundred children, ages 1 month to 15 years (median, 30.1 months), referred for the first time to the Department of Pediatrics at Odense University Hospital for further evaluation because of a heart murmur were consecutively included in the study. The children had no symptoms or other abnormal findings.

A standard chest radiograph (anteroposterior and lateral projections) and a color Doppler echocardiograph (CDE) were routinely obtained in all children. The echocardiographer did not know the result of the clinical examination when making the CDE.

Six specialists in pediatric radiology, with experience in pediatric radiologic radiology, evaluated all the chest radiographs independently. The radiologists were informed that the children had a heart murmur, but did not have access to any other data (including the results of the cardiologic investigation). Each radiologist classified each set of films into one of two categories: heart disease or no heart disease.

Six months after the initial evaluation, the radiographs were circulated to the same radiologists for a repeated evaluation. In this way, a total of 12 evaluations were obtained for each set of films. An abnormality score, giving 1 point for heart disease and 0 for no heart disease, was calculated for each set of films, which in this way could obtain a maximum score of 12 points.

Two sets of films could not be retrieved, so the analysis is comprised of 98 sets.

From the \*Department of Pediatrics, Odense University Hospital, Denmark; the †Department of Radiology, Odense University Hospital, Denmark; the ‡Department of Radiology, Glostrup Hospital, Denmark; the §Department of Radiology, Rigshospitalet, Denmark; and the ||Department of Radiology, Køge Hospital, Denmark.

Received for publication Feb 24, 1998; accepted Sep 11, 1998.

Address correspondence to Niels H. Birkebæk, MD, Skovgårdsparken 44, 7080 Børkop Denmark.

PEDIATRICS (ISSN 0031 4005). Copyright © 1999 by the American Academy of Pediatrics.

## Statistics

The reproducibility of the assessments of the chest radiographs (intra- and interobserver variation) was evaluated by  $\kappa$  statistics (see Appendix 1).<sup>2</sup>

The accuracy of the assessments of the chest radiographs was evaluated by comparison with the outcome of the CDE, which was considered the definitive diagnosis. The sensitivity, specificity, and the predictive value of a positive and a negative test were calculated according to standard methods.<sup>3</sup>

## RESULTS

Table 1 shows the results of the CDE. Twenty-three of the 98 children had a heart abnormality.

The  $\kappa$  values comparing each of the six radiologists' duplicate evaluations are shown in Table 2. The values have been calculated for all the 98 sets of films on the one hand, and for the sets from children with normal and abnormal CDE, respectively, on the other. When all the films were considered, the  $\kappa$  values ranged from 0.33 to 0.64, indicating a low level of agreement between the first and second evaluation of the individual radiologists. When only films from children with abnormal CDE were considered, the  $\kappa$  values ranged from 0.49 to 0.78, indicating a slightly better agreement.

In Table 3 the mean intraobserver (within observer)  $\kappa$  values are shown together with the mean interobserver (between observer)  $\kappa$  values obtained by comparing each of the radiologists' evaluations with those of the others. Again, the calculations have been performed for the whole group of children as well as separately for those with normal and abnormal CDE, respectively. The overall agreement between the radiologists (interkappa) was lower than the average agreement found for the individual radiologist (intrakappa). However, when only films from children with abnormal CDE were considered, the average interobserver agreement was at the same level as the average intraobserver agreement.

Table 4 shows the distribution of the abnormality scores. Eight points or more were obtained in 5 of 23 chest radiographs from children with an abnormal CDE, but only in 1 of 75 of those with a normal CDE. On the other hand, 7 of 23 chest radiographs from children with an abnormal CDE and only 26 of 75 with a normal CDE scored 0 points.

Table 5 shows the radiologists' assessments com-

**TABLE 1.** CDE Diagnoses of the 23 Children With Cardial Structural Abnormalities\*

Heart Defect	<i>n</i>
VSD	9
ASD	4
PS	3
AS	2
AVSD	1
PDA	1
TR	1
MR	1
PPAS	1
Total	23

Abbreviations: CDE, color Doppler echocardiography; VSD, ventricular septal defect; ASD, atrial septal defect; PS, pulmonary stenosis; AS, aortic stenosis; AVSD, atrioventricular septal defect; PDA, patent ductus arteriosus; TR, tricuspid regurgitation; MR, mitral regurgitation; PPAS, peripheral pulmonary artery stenosis.

**TABLE 2.** Intraobserver  $\kappa$  Values for Six Radiologists When Considering All Chest Radiographs, and the Chest Radiographs From Children With Normal and Abnormal CDE, Respectively\*

	Observer					
	1	2	3	4	5	6
All	0,367	0,635	0,333	0,416	0,470	0,477
Normal CDE	0,286	0,318	0,164	0,324	0,370	0,460
Abnormal CDE	0,495	0,775	0,620	0,485	0,704	0,495

Abbreviation: CDE, color Doppler echocardiography.

\* For further explanation, see text.

**TABLE 3.** Mean Values of the Intraobserver and Interobserver  $\kappa$  Values When Considering all Chest Radiographs, and the Chest Radiographs From Children With Normal and Abnormal CDE, Respectively\*

	Mean Intra $\kappa$	Mean Inter $\kappa$
All	0,450	0,282
Normal CDE	0,320	0,106
Abnormal CDE	0,595	0,531

Abbreviation: CDE, color Doppler echocardiography.

\* For further explanation, see text.

**TABLE 4.** Distribution of Abnormality Scores in Children With Normal CDE and Children With Abnormal CDE

Abnormality Score	Normal CDE ( <i>n</i> )	Abnormal CDE ( <i>n</i> )
0	26	7
1	17	4
2	10	2
3	11	1
4	6	3
6	4	1
9	1	1
10	0	1
11	0	1
12	0	2
Total	75	23

Abbreviation: CDE, color Doppler echocardiography.

**TABLE 5.** Sum, Mean, and Range of True-positive, False-positive, False-negative, and True-negative Radiological Assessments of the 12  $\times$  98 Evaluations as Compared to CDE (23 With Heart Disease and 75 With No Heart Disease)\*

	Sum	Mean	Range
True-positive	83	7	4-12
False-positive	127	11	1-19
False-negative	193	16	11-19
True-negative	773	64	56-74
Total 12 $\times$ 98	1176		
Sensitivity		0.30	0.17-0.52
Specificity		0.86	0.75-0.93
Predicted value of positive test		0.40	0.27-0.86
Predicted value of negative test		0.80	0.78-0.85

Abbreviation: CDE, color Doppler echocardiography.

\* Mean and range of sensitivity, specificity, predictive value of a positive and a negative test of the 12 evaluations are shown.

pared with the CDE diagnosis. Mean number and range of true-positive, false-positive, false-negative, and true-negative answers, respectively, of the 12 evaluations are shown. It has to be remembered, that 23 of the 98 children had structural heart disease diagnosed by CDE. Taking the outcome of the CDE as the definitive diagnosis, the sensitivity, the specificity, and the predictive value of a positive and a negative evaluation of the chest radiograph were

calculated. The mean predictive value of a positive test was 0.4 (range: 0.27–0.86) implying that 14% to 73% of the radiographs classified as abnormal, in fact, represented normal children, whereas the mean predictive value of a negative test was 0.8 (range: 0.78–0.85), which means that 15% to 22% of the radiographs classified as normal, in fact, belonged to children with a heart defect.

## DISCUSSION

When asymptomatic children with a heart murmur are referred for evaluation, a chest radiograph is routinely obtained in many centers.<sup>4–6</sup> Some investigators have questioned the value of chest radiographs in the initial evaluations of these children,<sup>7,8</sup> whereas others find the chest radiograph to be a valuable tool for the pediatric cardiologist in evaluation of patients with heart murmur.<sup>5,6</sup> However, the design of these studies made it difficult to unequivocally answer the question if the chest radiograph is useful in distinguishing between an innocent and a noninnocent heart murmur in asymptomatic children. For this purpose, we tested the reproducibility and accuracy of six pediatric radiologists' assessments of the chest radiographs, with respect to the presence or absence of heart defect, in children with a heart murmur. CDE was considered to be the gold standard.<sup>9</sup>

The reproducibility was evaluated with  $\kappa$  statistics, and we found a mean intraobserver  $\kappa$  of 0.45 and a mean interobserver  $\kappa$  of 0.28.  $\kappa$  values in this range are considered to reflect poor to moderate agreement between observers.<sup>10</sup> The intraobserver  $\kappa$  values were generally higher than the interobserver  $\kappa$  values, indicating a closer agreement between two observations if they had been made by the same radiologist than if they had been made by two different radiologists. Other studies of observer variability have given similar results.<sup>11</sup> In our study, the difference between intra- and interobserver variability is probably explained by individual diagnostic habits of the radiologists, as the radiologist hardly remembers a particular chest radiograph for half a year. The low consistency of the assessments of the chest radiograph might not be surprising. Despite existing guidelines for the assessments of chest radiographs, there remains a substantial amount of subjectivity in the evaluations. Reproducible assessment requires, among other things, standardized techniques. Chest radiographs are assumed to have been obtained in maximal inspiration. In children this assumption rarely holds true, and the various levels of inspiration leads to variable image quality. It should also be noted, that a high level of agreement (high reproducibility) between observers does not necessarily imply high accuracy.

With respect to accuracy, we found a very low sensitivity of the chest radiograph to heart defects as diagnosed by CDE. This might be explained by the absence of significant hemodynamic disturbances in some of the children with heart abnormalities at the time of investigation (eg, small atrial septal defect and small ventricular septal defect). A false-negative chest radiograph might result in the omission of

relevant investigations. However, proper identification of the heart defect is important, as many of the children will require precautions, such as endocarditis prophylactics<sup>12</sup> and/or follow-up examinations. The mean predictive value of a positive test was 0.4. Consequently, 60% of the children radiologically diagnosed as having a heart defect were in fact normal. These results were obtained in children who were referred from general practice for further evaluation. This means that the children were already selected. If chest radiographs were applied in a nonselected population with a lower prevalence of heart defects, the predictive value of a positive test would become even lower, and the number of false-positive outcomes very high.<sup>13</sup> Swenson found chest radiographs a valuable tool in the evaluation of new pediatric patients with heart murmur.<sup>4</sup> However, only 38% of the children received a CDE to confirm the clinical and radiologic evaluation in this study. In addition, the physicians might have been biased with regard to the evaluation of the chest radiograph, because they knew the result of the clinical examination when evaluating the chest radiograph.<sup>4,14</sup>

The mean values of sensitivity and specificity and the predictive value of a positive and a negative test in this study corresponded closely to the values found when a number of radiologists assessed the chest radiographs in the daily routine.<sup>1</sup> This indicates that the assessments of the chest radiographs, with respect to the presence or absence of heart defect, were not more correct when a single specialist in pediatric radiology assessed all of the chest radiographs, than when many radiologists assessed them in the daily routine. None of the six specialists in radiology was more correct in their assessments than the others. These results probably reflect poor accuracy of the test (chest radiograph for detection of heart defects) rather than poor accuracy of the radiologists.

In most clinical tests intra- and interobserver variability can be found.<sup>15,16</sup> The usefulness of a test in clinical practice depends on the reproducibility and the accuracy of the test. The present results do not support the use of a routine chest radiograph in the initial evaluation of children with asymptomatic heart murmurs. It frequently leads to unnecessary anxiety, and it increases the burden of ionizing radiation to no use. If any doubt remains after a thorough clinical evaluation of a child with an asymptomatic heart murmur, a competent CDE should be conducted.<sup>17</sup>

## ACKNOWLEDGMENTS

We thank the National Research Council of Health Sciences (M Frydenberg) for statistical advice.

## Appendix I

The results of a study comprising two observers independently recording the same dichotomous diagnosis in  $n$  patients can be presented in a  $2 \times 2$  contingency table, where  $a$ ,  $b$ ,  $c$ , and  $d$  indicate the number of observations and  $n$  is the number of patients:

Observer 1	Observer 2		
	Yes	No	
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	n

The agreement between observer 1 and 2 is  $P_o = (a + d) \div n$ . The agreement by chance between observer 1 and 2 is

$$P_c = (a + b) \div n \times (a + c) \div n + (b + d) \div n \times (c + d) \div n.$$

$\kappa$  is defined as  $(P_o - P_c) \div (1 - P_c)$ .  $\kappa$  can vary between  $-1$  and  $1$ . If the agreement between two observers is 100%,  $\kappa$  becomes  $1$ . With high agreement between the two observers,  $\kappa$  approaches  $1$ , although if agreement is poor,  $\kappa$  approaches  $0$ . A  $\kappa$  value of  $0$  indicates chance agreement, whereas negative  $\kappa$  values indicate that the observed agreement is less than chance.<sup>6,15</sup>

In the present study where 12 sets of assessments are available, it is possible to make  $66 = 12 \times 11 \div 2$  pairs of comparisons of assessments. Six of these will be comparisons between the first and second assessment by an individual radiologist, resulting in intraobserver  $\kappa$  values. The remaining 60  $\kappa$  values are comparisons of different radiologists' assessments, resulting in interobserver kappa values. One must have in mind, that the interobserver  $\kappa$  values are calculated from 12 sets of assessments from 6 radiologists. The interobserver  $\kappa$  values will therefore be correlated to some degree. This is not expected to invalidate our results because only the mean values are considered.

## REFERENCES

1. Birkebæk NH, LK Hansen, Oxhøj H. Diagnostic value of chest roentgenogram in the evaluation of asymptomatic children with a cardiac murmur. *Acta Paediatr Scand.* 1995;84:1379–1381
2. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *Br Med. J.* 1992;304:1491–1494
3. Foldspang A, Juul S, Olsen J, Sabroe S. Vurdering af screening programmer. In: Foldspang A, Juul S, Olsen J, Sabroe S, eds. *Epidemiologi.* Munksgård, Denmark: 1986;171–190
4. Geva T, Hegesh J, Frand M. Reappraisal of the approach to the child with heart murmurs: is echocardiography mandatory?. *Intern J Cardiol.* 1988;19:107–113
5. Danford DA, Nasir A, Gumbiner C. Cost assessment of the evaluation heart murmurs in children. *Pediatrics.* 1993;91:365–368
6. Swenson JM, Fischer DR, Miller SA, Boyle GJ, Ettetdgui JA, Beerman LB. Are chest radiographs and electrocardiograms still valuable in evaluating new pediatric patients with heart murmurs or chest pain? *Pediatrics.* 1997;99:1–3
7. Newburger JW, Rosenthal A, Williams RG, Fellows K, Miettinen OS. Non-invasive tests in the initial evaluation of heart murmurs in children. *N Engl J Med.* 1983;308:61–64
8. Temmerman AM, Mooyaart EL, Taverne PP. The value of the routine chest roentgenogram in the cardiological evaluation of infants and children. A prospective study. *Eur J Pediatr.* 1991;150:623–626
9. Sahn DJ. Applications of colour flow mapping in pediatric cardiology. *Cardiol Clin.* 1989;7:255–264
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–174
11. Koran LM. The reliability of clinical methods, data and judgments. I-II. *N Engl J Med.* 1975;293:642–646
12. Stanford T, Shulman MD, Amren PD, et al. Prevention of bacterial endocarditis. *Circulation.* 1984;70:1123A–1127A
13. Ades AE. Evaluating screening tests and screening programs. *Arch Dis Child.* 1990;65:792–795
14. Kramer MS, Roberts-Brauer R, Williams RL. Bias and "overcall" in interpreting chest radiographs in young febrile children. *Pediatrics.* 1992;90:11–13
15. Gjørup T, Brahm M, Fogh J, Munck O, Jensen AM. Interobserver variation in the detection of metastases on liver scans. *Gastroenterology.* 1986;90:166–172
16. Hansen M, Sindrup SH, Christensen PB, Olsen NK, Kristensen O, Friis ML. Interobserver variation in the evaluation of neurological signs: observer dependent factors. *J Neurol.* 1994;241:492–496
17. Hansen LK, Birkebæk NH, Oxhøj H. Initial evaluation of children with heart murmurs by the non-specialised paediatrician. *Eur J Pediatr.* 1995; 154:15–17