Observer variation in detecting lymphadenopathy on chest radiography

G. Du Toit, G. Swingler, K. Iloni
School of Child and Adolescent Health, Red Cross Children’s Hospital and University of Cape Town, Cape Town, South Africa

SUMMARY

OBJECTIVE: To assess inter- and intra-observer agreement in the detection of lymphadenopathy on chest radiography in children at risk for tuberculosis.

PATIENTS AND METHODS: Retrospective examination of the antero-posterior and lateral chest radiographs of children aged 1 month to 11 years discharged from the short-stay ward of the Red Cross Children’s Hospital, Cape Town, with a diagnosis of tuberculosis or pneumonia. Four paediatric pulmonologists viewed the radiographs independently. The main outcome measures were inter- and intra-observer agreement on the presence or absence of lymphadenopathy, reported as present, absent or equivocal, and expressed as weighted kappa statistics.

RESULTS: Weighted kappa for the six pairs of observers ranged from 0.14 (95%CI 0.02–0.30) to 0.52 (95%CI 0.35–0.69). After a 3-month interval, intra-observer agreement ranged from 0.44 (95%CI 0.25–0.62) to 0.71 (95%CI 0.56–0.87). The average weighted kappa for inter-observer agreement was 0.33, and the average intra-observer kappa was 0.55.

CONCLUSIONS: There was ‘fair’ inter- and ‘moderate’ intra-observer agreement among paediatric pulmonologists in detecting lymphadenopathy on chest radiography in children. Caution is necessary when basing clinical decisions on the presence of lymphadenopathy on chest radiography.

KEY WORDS: observer variation; radiography; lymphatic diseases; lymphadenopathy; tuberculosis

TUBERCULOSIS is difficult to diagnose in children because of the non-specific symptoms and infrequent isolation of organisms. Strong reliance is thus often placed on chest radiography,1 with mediastinal lymphadenopathy regarded as the radiological hallmark of primary tuberculosis.2,3 The World Health Organization’s proposed diagnostic criteria for pulmonary TB include pulmonary lymph nodes as one of five criteria for probable tuberculosis.4 Other diagnostic scoring systems for childhood tuberculosis also include radiographic lymphadenopathy.5-7 The validity and reliability of the detection of radiographic lymphadenopathy is thus of central importance to the diagnosis of childhood tuberculosis. A search of the Medline, HealthStar, LILACS and AIDSLINE databases and the African Health Anthology (October 2000) failed to identify any studies of either validity or observer variation in the detection of lymphadenopathy on chest radiography in children aged less than 18 years.

This study was performed to measure inter- and intra-observer agreement in the detection of lymphadenopathy on chest radiography in children at risk for tuberculosis.

METHODS

The study was performed at the Red Cross Children’s Hospital (RCCH) in the Western Cape Province of South Africa. This province had a tuberculosis incidence of 468 cases per 100 000 population in 1998, which is one of the highest in the world.8

Radiographs of children with pulmonary tuberculosis have many features in common with those of children with non-tuberculous pneumonia, with few recognised distinguishing features of tuberculosis other than lymphadenopathy. Therefore radiographs of children with a diagnosis of either pulmonary tuberculosis or pneumonia were studied. The aim was not to recruit children with known tuberculosis and controls without tuberculosis, but rather to assemble a representative spectrum of radiographs from a clinical setting in which tuberculosis was part of the differential diagnosis.

Children discharged or transferred from the short-stay ward at RCCH from 1 January 1996 to 31 December 1998 with a diagnosis of pulmonary tuberculosis or pneumonia were identified retrospectively from the hospital’s electronic database. In order to assemble a

Correspondence to: Dr George Du Toit, School of Child & Adolescent Health, Institute of Child Health Building, Red Cross Hospital, Rondebosch 7700, South Africa. Tel: (+27) 21-6585318. Fax: (+27) 21-6891287. e-mail: gdutoit@ich.uct.ac.za

sample with a statistically efficient prevalence of lymphadenopathy, we selected 60 patients with a diagnosis of tuberculosis and 40 with a diagnosis of pneumonia. Selection was random, using a computer generated list of random numbers and sampling frames of all patients discharged during the study period with a diagnosis of tuberculosis or pneumonia. Radiographs were included if both a lateral and antero-posterior radiograph were available from that admission.

Four paediatric pulmonologists attached to RCCH interpreted the radiographs. Each observer independently viewed all of the 100 films in random order. Radiographs were labelled from one to 100, and the viewing sequence was assigned using a computer generated list of random numbers. No clinical information was provided other than that the radiographs were from a random sample of patients who had been admitted to the overnight ward with chest disease. The observers categorised lymphadenopathy as present, absent or equivocal. The same observers viewed the films again in different random order after a 3-month interval. Viewing occurred in batches of 50 to avoid viewer fatigue. No time constraints were applied, and no criteria for the presence of lymphadenopathy were prescribed.

Agreement between individual observers and within each observer were expressed as weighted kappa statistics.\(^9\) Kappa is a measure of the degree of inter-rater agreement, over and above that expected by chance. It has a maximum value of 1, indicating complete agreement. A kappa of zero means that agreement is no better than if it were due solely to chance; a positive value means that the observed agreement is greater than would be expected by chance alone. The minimum kappa value is $-1$, indicating complete disagreement.

Weighted kappa takes into account the degree of disagreement. Because the categories ‘yes’, ‘equivocal’ and ‘no’ are ordered, all disagreements are not equally severe; for example, a ‘yes-no’ disagreement is more severe than an ‘equivocal-yes’ disagreement, and has different clinical implications.

### RESULTS

Of the 100 children, 56 were male and 44 female. Their median age was 22 months with a range of 1 month to 11 years. Forty-nine were discharged home from the overnight ward and the other 51 were referred for further hospital care.

At the first viewing lymphadenopathy was reported as present in 18% of cases, absent in 62% and equivocal in 20%.

Inter- and intra-observer agreement is shown in the Table. Weighted kappa for the six pairs of observers ranged from 0.14 (95% CI 0.02–0.30) to 0.52 (95% CI 0.35–0.69) and for intra-observer agreement from 0.44 (95% CI 0.25–0.62) to 0.71 (95% CI 0.56–0.87). The average weighted kappa for inter-observer agreement was 0.33, and the average intra-observer kappa was 0.55.

Kappa scores can also be computed from the perspective of the response.\(^10\) These are measures for the particular response categories, and are not weighted.

Agreement between individual observers and within each observer were expressed as weighted kappa statistics.\(^9\) Kappa is a measure of the degree of inter-rater agreement, over and above that expected by chance. It has a maximum value of 1, indicating complete agreement. A kappa of zero means that agreement is no better than if it were due solely to chance; a positive value means that the observed agreement is greater than would be expected by chance alone. The minimum kappa value is $-1$, indicating complete disagreement.

**Table** Inter- and intra-observer agreement. Weighted kappas

<table>
<thead>
<tr>
<th>Inter-observer agreement</th>
<th>Intra-observer agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer pairs</td>
<td>Kappa</td>
</tr>
<tr>
<td>A–B</td>
<td>0.19</td>
</tr>
<tr>
<td>A–C</td>
<td>0.52</td>
</tr>
<tr>
<td>A–D</td>
<td>0.37</td>
</tr>
<tr>
<td>B–C</td>
<td>0.14</td>
</tr>
<tr>
<td>B–D</td>
<td>0.39</td>
</tr>
<tr>
<td>C–D</td>
<td>0.36</td>
</tr>
</tbody>
</table>

DISCUSSION

Observer agreement has been categorised as ‘slight’ if kappa is $<0.20$, ‘fair’ if kappa is $0.21–0.40$, ‘moderate’ from $0.41$ to $0.60$, ‘substantial’ from $0.61$ to $0.80$ and ‘almost perfect’ from $0.81$ to $1.0$.\(^11\) Inter-observer agreement in this study was thus ‘fair’ (on average 0.33). The use of weighted kappa resulted in higher measures of agreement than the use of unweighted kappa. This level of inter-observer agreement is nevertheless lower than that reported for other chest radiographic features of respiratory infection in children: for example, unweighted kappas of 0.43 to 0.55 for bronchial wall thickening, 0.46 to 0.79 for consolidation or pneumonia and 0.78 to 0.83 for hyperinflation.\(^12\) Agreement in our study may also be higher than in the circumstances in which tuberculous lymphadenopathy is usually assessed, because the observers were all very experienced at assessing chest radiographs, had similar backgrounds and training, and worked in the same institution in a province with one of the highest incidences of childhood TB in the world. Our sample included a spectrum of radiographs from a setting in which tuberculosis was part of the differential diagnosis. This is likely to have provided a more meaningful estimate of agreement than if we had used only films with lymphadenopathy known to be present or absent, because the latter sample would have excluded many equivocal films.

There was greater agreement on the presence of lymphadenopathy (0.40) than its absence (0.28). As expected, intra-observer agreement was higher than inter-observer agreement, but was mostly ‘moderate’, ranging from 0.44 to 0.71.

Our findings suggest that caution is necessary when
basing clinical decisions on the presence or absence of lymphadenopathy on chest radiography. Although many factors in addition to lymphadenopathy are involved in the diagnosis of childhood TB, these findings have direct relevance to the diagnosis of TB in children. The findings may also be applicable to the radiological detection of lymphadenopathy in children with other conditions.

This investigation addressed the repeatability of the observations and not the validity, for which comparison with a credible reference standard would be necessary. The low inter-observer agreement, however, suggests that the validity of detection by some or all of the observers was low. No criteria for the diagnosis of lymphadenopathy were prescribed, and the wide variation in intra-observer agreement suggests that different observers used different (formal or informal) criteria in their assessment of lymphadenopathy. When questioned after the completion of the study, the observers varied in the criteria they had used in detecting lymphadenopathy. This suggests that the use of explicit criteria for lymphadenopathy could increase the repeatability of detection.

Computerised tomography (CT) with contrast injection is regarded as more accurate than chest radiography in detecting lymphadenopathy.3,13 In settings where CT is readily available, attempts should nevertheless be made to define and improve the accuracy of chest radiography, given the higher radiation dose, the higher cost and the need for contrast injection associated with CT. In settings where CT is not available, and where the vast majority of children with tuberculosis live, the need for further research into the accuracy of chest radiography is clear.

We conclude that lymphadenopathy in children is detected on chest radiography with fair inter- and moderate intra-observer agreement. Further research is needed using a credible reference standard to test the validity of detection. Such research could also enable the development and evaluation of criteria to maximise the validity and repeatability of such detection.

References


RÉSUMÉ

OBJECTIF: Déterminer les concordances entre observateurs et chez le même observateur dans la détection des adénopathies sur les clichés thoraciques d’enfants à risque de tuberculose.

PATIENTS ET MÉTHODES: Dans une étude rétrospective, les clichés thoraciques comportant un film antéro-postérieur et de profil des enfants âgés un mois à 11 ans et sortant d’une salle « bref séjour » de l’Hôpital Red Cross Children, Cape Town et avec un diagnostic de tuberculose ou de pneumonie, ont été examinés indépendamment par quatre pneumologues pédiatiques. Les mesures principales de résultats étaient la concordance entre divers observateurs et chez le même observateur sur la présence ou l’absence d’adénopathies, décrites comme présentes, absentes ou douteuses et exprimées sous forme de statistiques kappa pondérées.

RÉSULTATS: Le kappa pondéré pour les six paires d’observateurs se situe entre 0,14 (IC95% 0,02–0,30) et 0,52 (IC95% 0,35–0,69). Après un intervalle de 3 mois, la concordance par le même observateur se situe entre 0,44 (IC95% 0,25–0,62) et 0,71 (IC95% 0,56–0,87). Le kappa moyen pondéré pour une concordance entre observateurs est de 0,33. Le kappa pour le même observateur est de 0,55.

CONCLUSIONS: Parmi les pneumologues pédiatiques, on observe une concordance « satisfaisante » entre observateurs et « modérée » pour le même observateur, pour ce qui concerne la détection d’adénopathies dans les clichés thoraciques d’enfants. Il est dès lors nécessaire d’être prudent lorsque l’on base des décisions cliniques sur la présence d’adénopathies au cliché thoracique.
OBJETIVO: Evaluar la concordancia inter e intra-observador en la detección de las adenopatías en las radiografías de tórax en niños con riesgo de tuberculosis.

MARCO DE REFERENCIA Y PACIENTES: Estudio retrospectivo. La radiografía de tórax ántero-posterior y lateral de niños de un mes a 11 años de edad, dados de alta de una sala de corta estadía del Hospital de Niños de la Cruz Roja, Ciudad del Cabo, con diagnóstico de tuberculosis o neumonía, fueron examinados independientemente por cuatro neumólogos pediátricos. Las principales mediciones de los resultados fueron la concordancia inter e intra-observador sobre la presencia o ausencia de linfadenopatías, descritas como presentes, ausentes o dudosas y expresadas en estadísticas kappa ponderadas.

RESULTADOS: El kappa ponderado para los seis pares de observadores se sitúa entre 0,14 (IC95% 0,02–0,30) y 0,52 (IC95% 0,35–0,69). Después de un intervalo de 3 meses, la concordancia intra-observador se situaba entre 0,44 (IC95% 0,25–0,62) y 0,71 (IC95% 0,56–0,87). El kappa ponderado promedio de la concordancia inter-observador fue de 0,33. El kappa promedio intra-observador fue de 0,55.

CONCLUSIÓN: Los neumólogos pediátricos se observó una concordancia inter-observador «satisfactoria» e intra-observador «moderada», con respecto a la detección de linfadenopatías en las radiografías de tórax de niños. Es necesario ser prudente cuando las decisiones clínicas se basan en la presencia de linfadenopatías en la radiografía de tórax.