OBSERVER VARIATION IN DETECTING LYMPHADENOPATHY ON CHEST RADIOGRAPHY

DISSERTATION SUBMITTED IN PART FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF MEDICINE IN THE SCHOOL OF CHILD AND ADOLESCENT HEALTH, UNIVERSITY OF CAPE TOWN, JANUARY 2003

Masters submission by:

Dr Lionel George Du Toit

MBBCh (Wits) DCH (S.A) FC Paed (S.A.)

Department of Paediatrics and Child Health,

Red Cross Children's Hospital

Mowbray, Cape Town

DECLARATION:

I, Lionel George Du Toit, hereby declare that the work on which this thesis is based is original (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for any other degree at this or any other university.

I empower the University of Cape Town to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: _____ Date: _____

ABSTRACT:

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

Methods

Diagnostic test. Chest radiography using antero-posterior and lateral films.

Observers. Four paediatric pulmonologists independently viewing the radiographs.

Main outcome measures. 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). Average weighted Kappa for inter-observer agreement was 0.33. 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.

ACKNOWLEDGEMENTS:

Professor George Swingler (study supervisor and mentor) was responsible for the study concept and design. He initiated the study, supervised the collating of the data, and contributed towards the statistical processing and analysis. He also enthusiastically supervised and contributed to the writing of the original report.

Dr Karin Iloni assisted with the study design, and performed the majority of the statistical analysis.

The four Paediatric Pulmonologists who voluntarily participated as study observers: **Professors Max Klein, Louis Reynolds, Andrew Argent and Dr Heather Zar**, all affiliated to the Red Cross Children's Hospital, School of Child and Adolescent Health, University of Cape Town

Professor's Eugene Weinberg, Cassiem Motala and Paul Potter for encouragement and support.

Ms Vivienne Flynn for processing study data, and managing the storing, labelling and reproduction of the chest radiographs.

Radiographers and filing clerks of the RCCH Radiology Department, for radiograph search and retrieval.

Ms Iliam Farqhur, Ward S11 clerk, for assistance in tracing ward register details

The study was supported by a research grant from the School of Child and Adolescent Health, University of Cape Town.

CONTENTS: **PAGE No:**

LIST OF ABBREVIATIONS:

AFB	Acid Fast Bacilli
AIDS	Acquired Immune Deficiency Syndrome
BCG	Bacillus Calmette-Guerin Vaccine
CI	Confidence Interval
EQUIV	Equivalent
HIV	Human Immuno-deficiency Virus
MTB	Mycobacterium Tuberculosis
PCR	Polymerase Chain reaction
PPD	Purified Protein Derivative
PPV	Positive Predictive Value
RCCH	Red Cross War Memorial Children's Hospital
SCAH	School of Child and Adolescent Health, University of Cape Town
TB	Tuberculosis
WHO	World Health Organisation

CHAPTER 1: INTRODUCTION

The Mycobacterium tuberculosis (MTB) organism is a slow growing aerobic pleomorphic rod characterised by acid fast staining, an inability to produce pigment and a complex lipid and protein cell wall. The organism can be identified by Ziehl-Neelsen staining, Lowenstein Jensen medium culture and by Polymerase Chain Reaction techniques. Despite the availability of the Bacillus Calmette-Guerin Vaccine, effective antibiotics and the recent mapping of the tuberculous genome, tuberculosis remains responsible for substantial morbidity and mortality, particularly in the developing world. (Starke et al 1992, Cole 1998)

Tuberculous (TB) infection in humans occurs predominantly by inhalation of air borne bacilli, which lodge in the pulmonary acinus to set up a chain of potentially devastating reactions in the host. (Riley1959).

1. The initial host response may be to completely kill all *M tuberculosis* bacilli, eliminating the infection;
2. The organism can begin to multiply soon after infection, causing clinical disease known as *primary tuberculosis*. A hallmark of this form of disease is enlargement of regional lymph nodes;
3. Primary tuberculosis does not always follow a self-terminating course, but can develop into progressive primary TB characterised by rapid enlargement of the primary lung focus, erosion into the bronchial tree and consequent satellite lung lesions. These highly active lesions may lead to miliary dissemination.

4. Bacilli may thereafter become dormant, a state known as *latent infection.* Approximately one third of the world's population has *latent tuberculosis.* (Raviglione et al 1995)
5. The latent organisms can again become active and cause disease, known as *reactivation tuberculosis.*

All these potential disease outcomes may be associated with chest radiograph findings, so strong reliance is often placed on chest radiography (Coulter 1992). However, between 10-43% of children with confirmed tuberculosis will have a normal chest radiograph. (Merino 2001, Zarabi et al 1971). Mediastinal lymphadenopathy is regarded as the radiological hallmark of primary tuberculosis (Lamont et al 1992, Cremin 1995) Children are at increased risk of disseminated tuberculosis with potentially devastating clinical sequele. As radiographic features suggestive of primary tuberculosis may precede extra-pulmonary dissemination, the accurate interpretation of radiographic features such as mediastinal adenopathy is important to ensure early therapy and prevention of dissemination. Enlarged mediastinal lymphnodes may however be detected when using contrasted CT in 60% of children with TB infection who have a 'normal' chest radiograph. (Delacort 1993)

Most children undergoing primary infection have few, if any, disease specific symptoms. (Starke 1993) They may complain of non-specific symptoms such as lassitude and anorexia in combination with poor weight gain. These clinical features are non-specific among children, especially among children in developing countries who are frequently exposed to recurrent nutritional and infectious insults. (De Onis et al 1997)

1.1 Incidence of TB:

Tuberculosis currently affects more than 20% of the world's population. Every year there are 8-10 million new cases of and 3-5 million deaths attributable to tuberculosis. (Centre for Disease Control 1989) Tuberculosis is particularly a problem in developing countries, where 95% of cases occur. (Murray et al 1996). The WHO is especially concerned about the high incidence of tuberculosis among children. (Bloom et al 1992) There are 1.3 million diseased children under the age of 15 years and 450 000 die annually. These children, infected by adults, represent a reservoir from which future generations will be afflicted (Snider Jr et al, Roper 1992, Ehlers 1993)

The burden of tuberculosis is being exacerbated by the HIV epidemic, as HIV infection has been shown to be the biggest risk factor for tuberculosis. In the absence of HIV infection, the risk of someone developing TB is relatively low (10% lifetime risk), but for individuals co-infected with HIV the reactivation risk increases to a 10% annual risk. (Ellner et al.1993). In the context of an established tuberculosis epidemic and the associated risk posed by the HIV/AIDS epidemic, health care workers can anticipate being more frequently faced with the sometimes-difficult diagnosis of tuberculosis.

1.2 Diagnostic challenges in children with pulmonary tuberculosis:

Tuberculosis is difficult to diagnose in children because of the non-specific symptoms and infrequent isolation of organisms. Although tuberculosis has a propensity for extrapulmonary disease in children, pulmonary tuberculosis is the most common manifestation. (Snider DE. 1988) The finding of acid fast bacilli (AFB) in a sputum smear is strong presumptive evidence of TB, but a definitive diagnosis is made only

on results of polymerase chain reaction (PCR) identification of *M. tuberculosis* or a sputum culture. Both these investigations (PCR and AFB culture) are infrequently available in developing countries, and the time delay required for sputum culture (approximately 3 weeks) is frequently a logistical challenge. Younger children are also unable to spontaneously produce adequate sputum samples. Induced sputum may also be obtained by nasopharyngeal aspiration after hypertonic saline nebulisation and cough induction. These techniques of post induction-sputum collection can considerably improve the diagnostic accuracy of pulmonary tuberculosis, but do require clinical and laboratory expertise. (Zahrani et al 2001). Despite the advent of sophisticated invasive diagnostic tools, such as the flexible fibreoptic bronchoscope, the yield of AFB's remains low and bacteriological confirmation is not significantly aided. (Bibi et al. 2002) Sputum gathering, after nebulisation and cough induction, poses a potentially serious infective risk to the attending health care worker collecting the specimen particularly in the era of drug resistant tuberculosis. In contrast with adults who characteristically present with "open" cavitatory lung lesions associated with tuberculosis reactivation, children infrequently produce AFB in a sputum smear. AFB staining of sputum from adults with pulmonary TB may be positive in up to 75%, but even within secondary and tertiary care institutions; the diagnosis of tuberculosis can be confirmed in only 30-40% of children with tuberculosis. (Strumpf et al. 1979, Lipsky et al. 1984, Starke et al. 1988, Schaaf 1995)

The tuberculin skin test using purified protein derivative (PPD), although far from definitive, is an essential adjunct to the diagnosis of tuberculosis. The correct intradermal administration and interpretation of the reaction elicited by the PPD on the volar surface of the forearm, is not without problems. Reactions can be difficult to

interpret in children who have received BCG vaccination. There are also numerous potential causes of false negative PPD tests, particularly as a result of a break in the cold chain, incorrect technique or immuno-suppression. Children in developing countries are commonly undernourished, and the incidence of HIV/AIDS and/or overwhelming M tuberculosis infection is rising. These insults are all known to be acquired causes of impaired cell mediated immunity, increasing the potential of subsequent false negative PPD results.

1.3 The importance of the presence of mediastinal lymphadenopathy on chest radiography as a diagnostic feature for the diagnosis of tuberculosis:

Considering the constraints of the readily available tuberculosis diagnostic investigations, it is understandable that strong reliance is often placed on chest radiograph signs. Radiographic features associated with pulmonary TB on chest radiography vary, but radiographs characteristic of primary tuberculosis usually show enlargement of hilar, mediastinal, or subcarinal lymph nodes and lung parenchymal changes. These radiographic changes are caused by a combination of lung disease and the mechanical changes induced by partial or complete airway obstruction resulting from enlarging intra-thoracic nodes. (Schaaf 1995, Parisi 1994)

The appearance of chest radiographs in children differs from those of adult chest radiographs. The differences may be due to both technical and anatomical factors. Most children less than five years of age will have a chest radiograph taken in the suspended Anterior-Posterior view, as opposed to that of older children and adults who co-operatively stand with splayed shoulders, for a Posterior-Anterior view. More important is the frequent anterior mediastinal opacification created by the thymus,

which involutes with age. If present, the thymus may obscure the presence of mediastinal lymphnodes. The superior mediastinum is also wider relative to the thoracic cage in infants, as opposed to that of older children and adults, thereby allowing a cardiothoracic ratio of >50% as normal. In addition, normal anatomical structures may also complicate the interpretation of mediastinal adenopathy; for example, deviation of the trachea to the right is considered abnormal in adults but is frequently seen in normal infants, , especially in expiratory films. The pulmonary vasculature may also create superimposed mediastinal findings, in particular the descending branch of the right pulmonary artery. (Grzybowski 1954)

There are numerous diagnostic scoring systems for childhood tuberculosis, all of which include chest radiograph findings among the criteria needed to suspect or confirm a diagnosis of pulmonary tuberculosis. (Migliori et al. 1992, Osborn 1995, Donald 1991, World Health Organisation. 1983. Table 1.1) A broadly used diagnostic scoring system is that of the WHO, which includes a "suggestive appearance on chest radiograph" as one of four criteria for probable tuberculosis (World Health Organisation. 1983)

TABLE 1.1: DIAGNOSTIC SCORING SYSTEMS FOR CHILDHOOD TUBERCULOSIS

1.1.1 World Health Organisation provisional guidelines for the diagnosis of pulmonary tuberculosis in children. (World Health Organisation. 1983)

Suspected tuberculosis

An ill child

- with a history of contact with a confirmed case of pulmonary tuberculosis

Any Child

- Not regaining normal health after measles or whooping cough
- With loss of weight, cough and wheeze not responding to antibiotic therapy for respiratory disease
- With painless swelling in a group of superficial nodes

Probable tuberculosis

A suspected case and any of the following:

- Positive (>10mm) induration on tuberculin testing
- Suggestive appearance on chest radiograph
- Suggestive histological appearance of biopsy material
- Favourable response to specific anti-tuberculous therapy

Confirmed tuberculosis

- Detection by microscopy or culture of tubercle bacilli from secretions or tissues.
- Identification of tubercle bacilli as *M tuberculosis* by culture characteristics

1.1.2 A set of criteria for the diagnosis of pulmonary tuberculosis(TB) in children when culture is not available. (Migliori et al 1992)

Positive acid-fast stain of sputum or gastric aspirate

Or

Two or more of the following:

- History of contact with a tuberculous positive adult
- Cough lasting longer than 2 weeks
- A reactive tuberculin test

  ≥ 10mm in children without prior BCG vaccination

  ≥ 15mm in children with prior BCG vaccination
- Radiographic findings compatible with TB
- Response to anti TB therapy (increased body weight by 10% after 2 months, decrease in symptoms)

1.1.3 Diagnostic pathways for children with suspected tuberculosis: (Osborn 1995)

Suspected tuberculosis:

Ill child + suggestive history

+/- suggestive physical examination

+/- contact with a case of pulmonary tuberculosis

Confirmed tuberculosis:

Mtuberculosis smear/culture positive disease

Probable Tuberculosis:

Suggestive chest radiograph

Positive tuberculin test

Suggestive histology

None of the above mentioned diagnostic scoring systems have been validated. A prospective evaluation of the WHO criteria, as compared with expert opinion, revealed a 63% positive predictive value (PPV) of three of five probable WHO criteria. (Houwert et al 1998)

The radiographic features required by these diagnostic scoring systems are generally non-specific e.g. "suggestive appearance on chest radiography" "radiographic findings compatible with tuberculosis" and "suggestive chest radiograph" respectively. The validity and reliability of the detection of radiographic lymphadenopathy, as a key feature of childhood tuberculosis, is thus of central importance.

1.4 Specific literature search (Detailed in Appendix 1)

A search of the MEDLINE, LILACS databases and the African Health Anthology was performed in March 2002, using the search strategy described in Appendix 1. This strategy failed to to identify any studies of either validity or observer variation in the detection of lymphadenopathy on chest radiography in adults or children. Expert opinion and a further search making use of the OVID database (strategy as in Appendix 1) identified two studies which introduce the concept of inter and intra-observer variation when viewing chest radiographs. However, neither study is performed among children or in populations at increased risk of tuberculosis, nor is use made of both an AP and lateral chest radiograph. In addition, neither study reports specifically on the presence or absence of mediastinal adenopathy. (Garland 1949, Yerushalmy J et al 1950)

CHAPTER 2: METHODS

2.1 Objectives:

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.

2.2 Study population and setting:

This 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 per annum in 1998, which is one of the highest in the world (Epidemiological Comments 1999). 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. This group represents the population in whom the detection of lymphadenopathy usually needs to be performed in actual practice. 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 from the hospital's routine electronic database. In order to assemble a sample with a statistically efficient prevalence of lymphadenopathy, random samples were selected of 60 patients with a diagnosis of

tuberculosis and 40 with a diagnosis of pneumonia. Selection was random, using computer generated lists of random numbers and separate sampling frames of all patients discharged during the study period with a diagnosis of tuberculosis or pneumonia. These diagnoses were made by the attending physicians in the overnight ward of the RCCH, and were assigned ICD-9 codes at the time of discharge. Codes used; pneumonia 486, tuberculosis 011,9 (WHO 1983). Radiographs were included if both a lateral and antero-posterior radiograph were available from that admission, and if the radiograph quality was deemed technically suitable to allow for fair interpretation. The study co-ordinator was the sole judge of technical quality. These patient details on the electronic database were confirmed from the ward discharge register, in which the ward clerk had recorded the date, diagnosis at time of discharge from the short stay ward and destination after discharge. The search commenced in 1998 (as study initially commenced in 1999) and continued backwards in time by completed year until the required number of films were retrieved. Care was taken to locate all films, regardless of the health facility to which they had been discharged. This was done to reduce selection bias, as the films stored at RCCH would more likely represent those children requiring tertiary level medical facilities with clinically, and by implication radiologically, more severe disease. Had only these films been included, the sample may have been biased towards chest radiographs with more obvious features of pulmonary tuberculosis.

2.3 Observers:

Four paediatric pulmonologists attached to RCCH interpreted the radiographs. These observers were experienced qualified paediatric pulmonologists with extensive clinical experience gained from daily clinical practice among children with chest disease who reside in an area with a high incidence of tuberculosis.

2.4 Viewing of 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. (Appendix 2) 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. A 3-month interval was allowed for between viewing in order to minimise recall bias. Viewing occurred in batches of 50 to avoid observer fatigue. No time constraints were applied when viewing radiographs and, in order to optimise observer independence, no criteria for the presence of lymphadenopathy were prescribed.

In an attempt to identify the criteria (or lack thereof) used by the observers, they were each informally requested to complete a questionnaire at the end of the study. (Appendix 4) The following two questions were posed:

Q1) What criteria (if any) do you use when coming to a decision concerning the presence or absence of mediastinal lymphnodes on chest radiographs?

Q2) What is your opinion concerning the value of a lateral chest radiograph in addition to the AP radiograph when viewing for mediastinal lymphadenopathy

2.5 Analysis

Agreement between individual observers and within each observer were expressed as weighted kappa statistics (Fleiss 1981). 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. A weight of 1 was given to complete agreement, 0.5 to discrepancies of one category, and 0 for discrepancies of two categories.

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. The use of the weighted kappa statistic allowed observers to equivocate on the presence or absence of mediastinal adenopathy, thereby more closely simulating the normal clinical setting, particularly in this sample of films which included those of children diagnosed with pneumonia, from which tuberculosis must be differentiated. Kappa statistics for agreement between individual observers and within each observer were computed. Calculations were performed with Microsoft Excel 5.0™ and multidimensional scaling was performed using Statistica 5.5™. The estimate of target sample size was based on the assumption of two observers, Kappa of 0.60, a 50% prevalence of nodes and the absence of an equivocal category. One hundred cases would have given a 95% CI of 0.4-0.80, fifty cases a 95% CI of 0,32-0.88 (too wide to be meaningful) and two hundred cases a 95% CI of 0.46-0.74. We did not think that the relatively small increase in precision justified the extra effort, and a sufficient number of radiographs may not have been accessible.

CHAPTER 3: RESULTS

3.1 Total number of patients screened

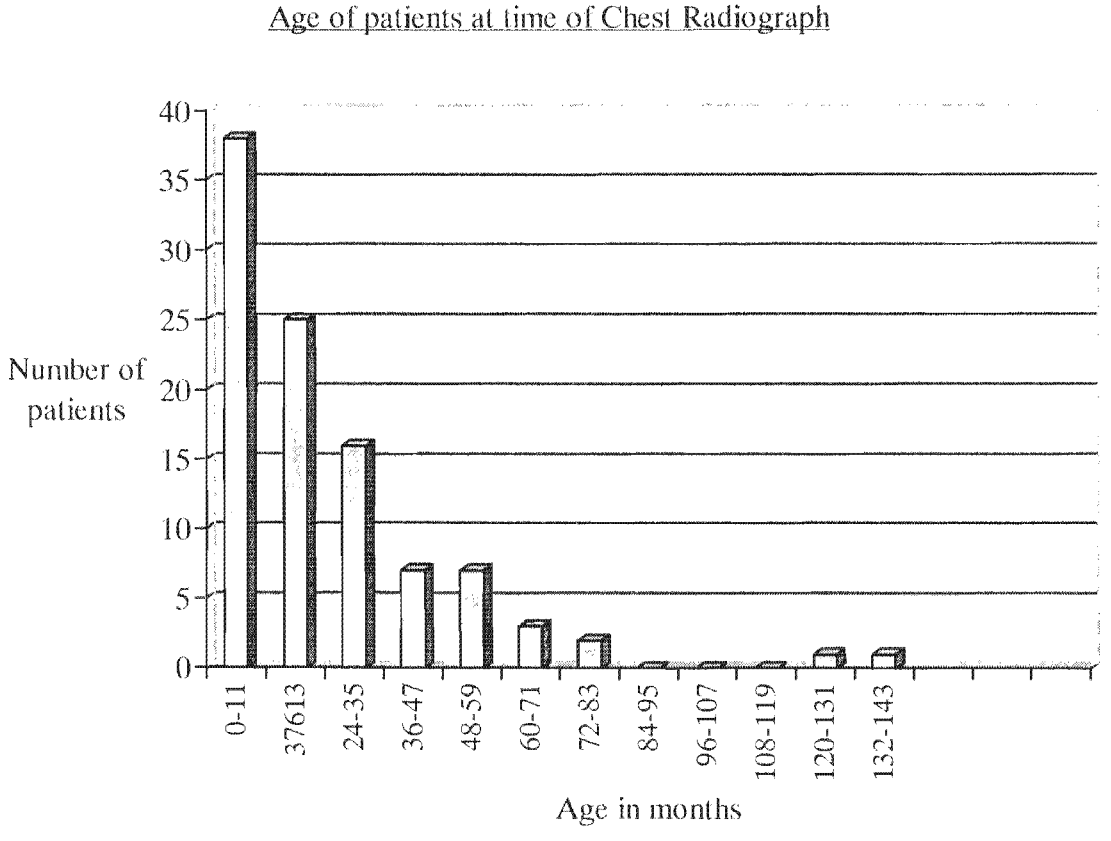
Reasons for the exclusion of radiographs were; those radiographs lost or destroyed by the hospital or clinic, unavailability of both an AP and lateral radiograph in the retrieved radiograph packet, and inadequate technical radiograph quality. In addition, many patients identified through the medical informatics database could not be reconciled with the ward admission and discharge register and were therefore not included.

Table 3.1 Analysis of radiograph inclusion.

ICD 9 Coding (1996-1998)	**Pulmonary Tuberculosis (011,9)**	**Pneumonia (486)**
Number on database	116	2977
Number randomly selected	All selected	171
Number found	90	171
Number of acceptable quality	60	150
Number included	60	40

Of the 100 children whose radiographs were studied, 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 referred for further hospital care.

Figure 3.1 Age of patients at time of Chest Radiograph



3.2 Summary of observer responses

At the first viewing lymphadenopathy was reported as present in 18.5% of cases, absent in 61.5% and equivocal in 20%. Proportions at second viewing were 17.25%, 63.75% and 19% respectively.

Table 3.2 Proportions of responses from individual observers.

3.2.1 Observer A.

Viewing		First	Second	Mean.
Decision	Yes	14%	14%	14%
Decision	Equiv	16%	16%	16%
Decision	No	70%	70%	70%

3.2.2 Observer B.

Viewing		First	Second	Mean.
Decision	Yes	17%	19%	18%
Decision	Equiv	28%	35%	31.5%
Decision	No	55%	46%	50.5%

3.2.3 Observer C

Viewing		First	Second	Mean.
Decision	Yes	21%	24%	22.5%
Decision	Equiv	8%	5%	6.5%
Decision	No	71%	71%	71%

3.2.4 Observer D

Viewing		First	Second	Mean.
Decision	Yes	22%	12%	17%
Decision	Equiv	28%	20%	24%
Decision	No	50%	68%	59%

Table 3.3 Summary of distribution of responses (%) from individual observer at first and second viewings.

Observer		A		B		C		D		
Viewing		First	Second	First	Second	First	Second	First	Second	Overall %
Decision	Yes	14	14	17	19	21	24	22	12	17.8
Decision	Equiv	16	16	28	35	8	5	28	20	19.5
Decision	No	70	70	55	46	71	71	50	68	62.62

Table 3.4 Pooled observer results (Observer A+B+C+D)

	Combined observer results First viewing		Combined observer results Second viewing	
Yes	74	18.5%	69	17.25%
Equivocal	80	20%	76	19%
No	246	61.5%	255	63.75%
Total	400	100%	400	100%

Table 3.5: Inter- and intra-observer agreement weighted kappa's

Inter-observer agreement			Intra-observer agreement		
	Kappa	**95% CI**		**Kappa**	**95% CI**
Average Kappa	0.21	0.15-0.27			
Observer pairs A-B	0.19	0.04-0.35	Observer A	0.58	0.41-0.74
A-C	0.52	0.35-0.69	B	0.71	0.56-0.87
A-D	0.37	0.22-0.52	C	0.44	0.25-0.62
B-C	0.14	-0.02-0.30	D	0.47	0.32-0.62
B-D	0.39	0.23-0.54			
C-D	0.36	0.2-0.51			

Table 3.6 Observers risk profiles.

Observer	Proportion of 'equivocal' responses in rank order	Intra-observer variation Kappa value
C	6.5% (95% CI 3.5 to 10.9%)	0.44
A	16% (95% CI 10.9 to 21.1%)	0.58
D	24% (95% CI 18.1 to 29.9%)	0.47
B	31.5% (95%CI 25.1 to 37.9%)	0.71

The Spearman rank order correlation co-efficient is 0.80 (p=0.20). Although this correlation co-efficient is high, the p value indicates that it could quite easily be due to chance.

Figure 3.1 represents a multidimensional scaling of the Kappa values, which form a similarity matrix. The aim of multidimensional scaling is to find a configuration of the points in 2 dimensions that best approximates the observed similarities. This facilitates visual interpretation of the data. The distances between the plotted points represents the observed distances. Therefore points close together represent close agreement between raters. The points are well represented in 2 dimensions, with a stress of 0.04.

For each of the 4 viewers, the subscript represents the first and second viewing sessions. It is striking that there is not much difference between the intra and inter observer Kappa's. The plot also demonstrates that there is no clear trend between the 2 sessions for greater consensus amongst the viewers.

Table 3.7. Tables demonstrating agreement between the six observer pairs

Pair A-B (Kappa 0.19 CI 0.04-0.35)

		Observer A		
		Yes	Equiv	No
Observer B	Yes	5	2	10
	Equiv	5	6	17
	No	4	8	43

Pair A-C (Kappa 0.52 CI 0.35-0.69)

		Observer A		
		Yes	Equiv	No
Observer C	Yes	12	3	6
	Equiv	1	11	5
	No	1	2	59

Pair A-D (Kappa 0.37 CI 0.22-0.52)

		Observer A		
		Yes	Equiv	No
Observer D	Yes	10	5	6
	Equiv	10	5	6
	No	1	6	43

Pair B-C (Kappa 0.37 CI 0.22-0.52)

		Observer B		
		Yes	Equiv	No
Observer C	Yes	7	5	9
	Equiv	2	1	5
	No	8	22	41

Pair B-D (Kappa 0.39 CI 0.23-0.54)

		Observer B		
		Yes	Equiv	No
Observer D	Yes	9	8	5
	Equiv	6	10	12
	No	2	10	38

Pair C-D (Kappa 0.36 CI 0.2-0.51)

		Observer C		
		Yes	Equiv	No
Observer D	Yes	12	2	8
	Equiv	5	2	20
	No	4	3	43

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). Average weighted Kappa for inter-observer agreement was 0.33. Average intra-observer Kappa was 0.55.

Kappa scores can also be computed from the perspective of the response (Altman DG 1991). These measures are not weighted. Kappa was 0.40 (95% CI 0.32 to 0.48) for the presence of mediastinal lymphadenopathy, 0.28 (95% CI 0.20 to 0.36) for its absence and 0.03 (95% CI –0.05 to 0.11) for equivocal viewings.

Table 3.7 Intra-observer agreement

	Kappa	95% CI
Observer A	0.58	0.41-0.74
B	0.71	0.56-0.87
C	0.44	0.25-0.62
D	0.47	0.32-0.62

Table 3.8 Number of cases with unanimous decisions

Answer Category	Yes	Equivocal	No
First Viewing	5	0	29
Second Viewing	2	0	29
Combined Viewing	2	0	19

CHAPTER 4: DISCUSSION

4.1 Validity of the findings

A strength of this study is that the sample was of a clinically meaningful spectrum of patients. This is important because a selection of patients known either to have or not have the condition of interest has been shown empirically to be the most important source of bias in studies of diagnostic validity, leading to an overestimate of agreement because the more difficult borderline cases have been screened out. (Lijmer 1999) The lack of such a clinically meaningful sample has been described as the key limitation of most diagnostic studies. (Hernandez-Aguado 2002) Although no empirical data are available for studies of observer agreement, it appears likely that issues that apply to studies of validity in this respect will also apply to those of agreement.

A limitation of the study is the retrospective x-ray collection. This required reliance on diagnostic coding rather than on a defined clinical presentation to identify x-rays. Random selection of patients from sampling frames including all children with a diagnosis of either tuberculosis or pneumonia should have produced a meaningful spectrum of patients with the two diagnoses. Prospective identification of children with a defined clinical presentation was not feasible within the time and resource constraints of this project. Although the collection of the films was retrospective (and could have lost some representativeness because of missing films), the measurement

of agreement was prospective and under controlled conditions. We could thus ensure independent viewing by the observers and blinding to clinical details.

The prevalence of lymphadenopathy in the sample appears representative of that in childhood tuberculosis in general. Given an approximate 50% prevalence of radiographic lymphadenopathy in childhood tuberculosis (Merino 2001, Zarabi et al 1971), and a 60% prevalence of tuberculosis in our sample, 30% of films in our sample would be expected to show lymphadenopathy. The observed 18% with lymphadenopathy plus 20% equivocal films is compatible with the expected prevalence.

4.2 Meaning of the findings

Average Kappa for inter-observer agreement was 0.33, with that for individual viewer pairs ranging from 0.14 to 0.52. Observer agreement has been categorised as "slight" if Kappa<0.20, "fair" if Kappa is 0.21-0.40, "moderate" from 0.41-0.60, "substantial " from 0.61-0.80 and "almost perfect " from 0.81-1.0 (Sackett 1991). 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 if unweighted Kappa had been used. 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. (Swingler 2001).

As expected intra-observer agreement was higher than inter-observer agreement but was mostly "moderate", ranging from 0.44 to 0.71.

There was greater agreement on the presence of lymphadenopathy (0.40) than its absence (0.28). This finding is difficult to interpret with confidence but the greater uncertainty in excluding lymphadenopathy could be due to difficulties in distinguishing normal radiological structures from abnormal lymph nodes.
When accessing overall agreement by all four viewers on both occasions nineteen radiographs were reported as 'No', two as 'Yes' and none as 'Equivocal'. However even if consensus were to be reached for a decision, in the absence of a reference standard, it is impossible to know if this was indeed the correct decision.

This study addressed only the repeatability of the observations and not the validity (sensitivity and specificity) of the detection of lymphadenopathy. No reference standard for the presence of mediastinal lymphadenopthy was used, so it was not possible to assess whether some observers' assessments were correct (valid) or whether all observers were incorrect. The poor inter-observer and only moderate intra-observer agreement however suggests low validity, on average. Assessment of validity would have required prospective patient enrolment and the use of a credible reference standard such as contrast-enhanced computerised tomography (Cremin 1995, Kuhn 1993). This was not feasible within the time and resource constraints of this project.

This study does not directly address the diagnosis of tuberculosis itself. The primary focus of the study was to measure inter- and intra-observer agreement in the detection of mediastinal lymphadenopathy on chest radiography in children at risk for tuberculosis. Although lymphadenopathy is an important element in the diagnosis of

tuberculosis in children, it is not sufficient in itself to diagnose or exclude tuberculosis. This limitation reflects an important difficulty in the assessment of all diagnostic tests for tuberculosis, that is, the difficulty in establishing a reference standard for tuberculosis. Bacterial culture is the most credible reference standard, but even within secondary and tertiary care institutions; the diagnosis of tuberculosis can be confirmed in only 30-40% of children with tuberculosis. (Strumpf et al. 1979, Lipsky et al. 1984, Starke et al. 1988, Schaaf 1995). Many other children with tuberculosis have a diagnosis made presumptively, usually including the use of chest radiography. Use of such reference standards (without microbiological diagnosis) would result in an overestimate of the accuracy of lymphadenopathy for diagnosing tuberculosis, because the reference standard is partially determined by the assessed presence of lymphadenopathy on chest radiography.

This study does also not address the additional diagnostic value provided by adding a lateral chest radiograph to the standard anterior-posterior chest radiograph, when viewing for mediastinal lymphadenopathy. Some paediatric radiology texts highlight the value of the lateral chest radiograph when assessing the mediastinal anatomical regions. (Burton E 1999, Cremin 1995, Caffey 1993) In this study the four examiners participating in this study had varying opinions concerning the value the lateral chest radiograph when viewing for mediastinal lymphadenopathy. (Appendix 5) Lateral chest radiographs have been found to increase the detection of lymphadenopathy in children with tuberculosis, but the accuracy of detection was not compared with a reference standard. (Smuts 1994)

4.3 Applicability of the findings to other settings

Observer agreement in this study is likely to be higher than in most settings. The observers were all experts in this field with similar background and training who were working in the same institution. The x-rays were probably of better than average technical quality because they were taken at a tertiary level paediatric hospital with an excellent radiology service, and films of sub-optimal quality were excluded. Knowledge that their assessments were being closely examined could also have led to the observers taking extra care in assessment.

4.4 Implications for clinical practice

In the absence of evidence of the validity of the detection of lymphadenopathy, the findings of poor observer agreement in this study suggest that caution is necessary when basing clinical decisions in children on the presence or absence of lymphadenopathy on chest radiography.

4.5 Implications for further research

In settings in which CT is readily available as an alternative investigation, 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 (where the vast majority of children with tuberculosis live), the need for further research into the accuracy of chest radiography is clear.

Direct research on the validity of the detection of lymphadenopathy is needed. Such studies will require the use of a credible reference standard, such as contrast-enhanced computerised tomography, in a clinically meaningful population.

No criteria for the diagnosis of lymphadenopathy were prescribed in this study, 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 widely 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, and possibly also the validity. Any assessment of the validity of the detection of lymphadenopathy (or of the diagnosis of tuberculosis) should thus include an assessment of specific radiological criteria, and combinations thereof.

CHAPTER 5: CONCLUSIONS

1. Lymphadenopathy in children was detected on chest radiography with low inter- and moderate intra-observer agreement.
2. The low inter-observer agreement provides no direct information about the validity of detection of lymphadenopathy, but suggests that validity was low for some or all of the observers.
3. Agreement in this study is likely to be higher than in most settings where lymphadenopthy is used to diagnose tuberculosis in children.
4. The findings may also be applicable to the radiological detection of lymphadenopathy in children with conditions other than tuberculosis.

CHAPTER 6: RECOMMENDATIONS

1. Caution is necessary when basing clinical decisions in children on the presence or absence of lymphadenopathy on chest radiography alone.
2. Research on the validity of the detection of lymphadenopathy is needed, using a credible reference standard such as contrast-enhanced computerised tomography.
3. Such assessment should include an evaluation of specific criteria for the presence of lymphadenopathy, to maximise the validity and repeatability of such detection.
4. In addition, research is also needed to compare the validity and reliability of the detection of chest lymphadenopathy in children when assessed by a) antero-posterior plus lateral views and b) by antero-posterior view alone.

CHAPTER 7: REFERENCES

Al-Zahrani K, AL Jahdali H, Poirier L, Rene P, Menzies D. Yeild of smear, culture and amplification test from repeated sputum induction for the diagnosis of pulmonary tuberculosis. Int J Tuberc Lung Dis 2001 Sep; 5 (9):855-60.

Altman DG. Practical statistics for medical research. 2nd ed. London: Chapman & Hall; 1991. p.403–409.

Bamford L. Health Systems Trust. South African Health Review 1999 Available from: URL: http://www.hst.org.za/sahr/99/chap23.htm

Bibi H, Feigenbaum D. Should Bronchoscopy be performed in the evaluation of Suspected Paediatric Tuberculosis? Chest 2002 Nov; 122 (5):1604-8.

Bloom BR, Murray CJL. Tuberculosis: commentary on a re-emergent killer. Science 1992; 257:1055-1064.

Burton EM, Brody SA. Approach to image acquisition and interpretation. In: Burton EM, Brody SA, editors. Essentials of Paediatric Radiology. 1st ed. New York: Thieme; 1999 .p.1

Kuhn JP, Slovis TL, Silverman FN, Kuns LR. The Neck and Respiratory System. In: Silverman FN, Kuhn JP, Caffey J, editors. Paediatric X-ray diagnosis. 9th edition. St Louis: Mosby; 1993. p.511

Centre for Disease Control (1989) A strategic plan for the elimination of tuberculosis in the United States. MMWR 38 (suppl. S-3):1-25. Available from: URL: www.cdc.gov/mmwr/

Colbentz CL, Babcook CJ, Alton D, Riley BJ, Norman G. Observer variation in detecting the radiologic features associated with bronchiolitis.
Investi Radiol 1991 Feb;26(2):115-118

Coulter JBS. Tuberculosis. In: Campbell AGM, McIntosh N, editors. Forfar and Arneil's Textbook of Paediatrics. 4th ed. Edinburgh: Churchill Livingstone; 1992. p.1390-1403

Cremin BJ, Jamieson DH. Imaging of pulmonary tuberculosis. In: Cremin BJ, Jamieson DH, editors. Childhood tuberculosis: modern imaging and clinical concepts. 1st ed. London: Springer; 1995. p.19-31

De Onis M, Blossner M. WHO global database on child growth and malnutrition. Geneva: World Health Organisation; 1997. Unpublished document WHO/NUT/97.4. Available from: URL: http://www.who.int/whosis/cgrowth/bulletin.htm

Department of Health, Republic of South Africa. Epidemiological comments Vol.1 No.2. Notifiable Medical Conditions. June 1999

Delacourt C, Mani TM, Bonneort U, de Blic J, Sayeg N, Lalleman D, et al. Computed tomography with normal chest radiograph in tuberculous infection. Arch Dis Child 1993 Oct; 69(4): 430-2

Donald PR. Diagnostic considerations in management and epidemiology. In: Coovadia HM, Benatar SR, editors. A Century of Tuberculosis: South African Perspectives. 1st ed. Cape Town: Oxford University Press; 1991.

Beyers JA. The radiological features of pulmonary tuberculosis in childhood. In Donald PR, Fourie PB, Grange JM, editors. Tuberculosis in Childhood. 1st ed. Pretoria: JL van Schaik Publishers; 1999 p.103-117

Ehlers MRW. The wolf at the door. Some thoughts on the biochemistry of the tubercle bacillus. S Afr Med J 1993; 82:900-903.

Ellner JJ, Hinmann AR, Dooley SW, Fischl MA, Sepkowitz KA, Goldberger MJ, et al. Tuberculosis Symposium: emerging problems and promise. J Infect Dis 1993 Sep; 168:537-551

Fleiss J.L. Statistical methods for rates and proportions, 2nd ed. New York: Wiley; 1981.p.212-236.

Garland LH. On the scientific evaluation of diagnostic procedures. Radiology 1949; 3:309-327

Grzybowski, S. Radiology in childhood tuberculosis. Tubercle 1954; 35:113-117

Hernandez-Aguado I. The winding road towards evidence based diagnoses. J Epidemiol Community Health. 2002 May; 56(5): 323-5

Houwert KAF, Borggreven PA, Schaaf HS, Nel E, Donald PR, Stolk J. Prospective evaluation of World Health Organisation criteria to assist diagnosis of tuberculosis in children. Eur Respir J 1998 May; 11(5):1116-1120

ICD 9 = International Classification of Disease, 9th revision
Published by the World Health Organisation (WHO).
Available from: URL: http://www.cdc.gov/nchs/about/otheract/icd9/

Kuhn JP, Slovis TL, Silverman FN, Kuns LR. Mediastinum. In: Silverman FN, Kuhn JP, Caffey J, editors. Paediatric X-ray diagnosis. 9th edition. St Louis: Mosby; 1993.

Lamont AC, Cremin BJ, Pelteret RM: Radiological patterns of Pulmonary Tuberculosis in the Paediatric Age Group. In: **, editors. Paediatric Radiology. 1st ed. Great Britain: Springer; 1986. p.2-7

Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282:1061-1066

Lipsky BA, Bates J, Tenover FC, Plorde JJ. Factors affecting the clinical value of microscopy for acid fast bacilli. Rev Infect Dis 1984; 6: 214-22

Migliori GB, Borghesi A, Rossanigo P, Adriko C, Neri M, Santini S, et al. Proposal of an improved score method for the diagnosis of pulmonary tuberculosis in childhood in developing countries. Tuber Lung Dis 1992 Jun; 73 (3):145-9.

Murray CJL, Lopez AD. The Global Burden of Disease, A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Harvard School of Public Health, World Health Organization and The World Bank, 1996.

Osborn CM. The challenge of diagnosing childhood tuberculosis in a developing country. Arch Dis Child 1995 Apr; 72 (4):369-74.

Parisis MT, Jensen MC, Wood BP. Pictoral review of the usual and unusual roentrogram manifestations of childhood tuberculosis. Clin Imag 1994; 18:149-54

Raviglioni MC, Snider DE, Kochi A. Global epidemiology of tuberculosis. Morbidity and Mortality of a worldwide epidemic. JAMA 1995; 273:220-6

Riley RL, Mills W, Nyka N, Weinstock PB, Storey LU, Sultan MC. Aerial dissemination of pulmonary tuberculosis: a two year study in a tuberculosis ward. Am J Epidemiolog 1959; 142:3-14

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology. A basic science for clinical medicine. 2nd ed. Boston; Little Brown and Company, 1991 p. 30.

Schaaf HS, Beyers N, Gie RP, et al. Respiratory tuberculosis in childhood: the diagnostic value of clinical features and special investigations. Pediatr Infect Dis J 1995; 14:189-94

Smuts NA, Beyers N, Gie RP, Schaaf HS, Talent JM, Nel E, et al. Value of the lateral chest radiograph in tuberculosis in children. Pediatr Radiol 1994; 24 (7):478-480

Snider DE Jr, Rieder HL, Gombs D, Bloch AB, Hayden CH, Smith M. Tuberculosis in children. Pediatr Infect Dis J 1988; 7:271-8

Snider Jr DE, Roper WL. The new tuberculosis, N Engl J Med 1992; 326:703-705

Starke JR. Childhood tuberculosis: a diagnosis dilemma. Chest 1993; 104:329-330

Starke JR, Jacobs RF, Jereb J. Resurgence of tuberculosis in children. J Pediatr 1992; 120:889-852

Starke JR. Modern approach to the diagnosis and treatment of tuberculosis in children. Pediatr Clin N Am 1988; 35:441-466

Strumpf IJ, Tsang AY, Syre JW. Re-evaluation of sputum staining for the diagnosis of pulmonary tuberculosis. Am Rev Respir Dis. 1979; 119:599-602

Swingler GH. Observer variation in chest radiography of acute lower respiratory infections in children: a systematic review. BMC Medical Imaging 2001; 1:1 Available from: URL: http://www.biomedcentral.com/1471-2342/1/1/

Shaw D. The Chest in Children. In: Sutton D, editor. Textbook of Radiology and Imaging. 6th Edition. London: Churchill Livingston; 1998. p.529-531

The worldwide magnitude of protein-energy malnutrition: an overview from the World Health Organisation. Provisional guidelines for the diagnosis and classification of the EPI target diseases for primary health care, surveillance and special studies. EPI/GEN/83/4. Geneva. World Health Organisation. 1983

Raviglione MC, Snider DEJ, Kochi A. Global epidemiology of tuberculosis: morbidity and mortality of a worldwide epidemic. JAMA 1995; 273: 220-226

Merino JM, Alvarez T, Marrero M et al. Microbiology of Primary Pulmonary Tuberculosis. Chest 2001; 119:1434-1438

Cole ST, Brosch R, Parkhill J et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998 Jun 11; 393 (6685):537-44.

Yerushalmy J, Harkness JT, Cope JH, Kennedy BR. The role of dual reading in mass radiography. Am Rev Tuberc 1950; 4:443-465

Zarabi M, Sane S, Girdany BR. The chest roentrogram in the early diagnosis of tuberculis meningitis in children. AM J Dis Child 1971; 121:389

CHAPTER 10: APPENDICES

APPENDIX 1

Literature search strategy:

PubMed Search	Most recent Query	Result
#13	#4 AND #7 AND #10 Field: All Fields, Limits: All Adult: 19+ years	25
#12	#4 AND #7 AND #1010 Field: All Fields, Limits: All Child: 0-18 years	16
#11	#4 AND #7 AND #10	33
#10	#8 OR #9	226615
#9	Lymphadenopathy [Text Word]	9211
#8	Lymphatic diseases	222756
#7	#5 OR #6	529672
#6	Radiography, thoracic	16958
#5	Radiography	529672
#4	#1 OR #2 OR #3	13981
#3	Intraobserver [Text Word]	1683
#2	Interobserver Variation [Text Word]	3991
#1	Observer Variation	11949

APPENDIX 2

Randomisation of the radiograph viewing order. (Generated using Epi Info™)

Original radiograph numbering	Randomised order First viewing	Randomised order Second viewing
1	77	41
2	25	82
3	91	32
4	60	78
5	5	88
6	27	92
7	71	24
8	41	71
9	42	42
10	30	22
11	56	73
12	57	29
13	47	90
14	1	87
15	31	95
16	72	4
17	78	97
18	12	2
19	99	6

20	90	37
21	88	83
22	15	77
23	2	3
24	51	23
25	100	53
26	63	34
27	29	65
28	75	52
29	59	28
30	76	36
31	65	81
32	16	38
33	13	54
34	37	80
35	44	51
36	86	67
37	17	11
38	19	10
39	80	61
40	52	72
41	55	44
42	95	20
43	87	15
44	79	25

45	10	7
46	85	75
47	11	46
48	18	48
49	70	96
50	35	58
51	83	13
52	98	85
53	82	27
54	20	49
55	45	47
56	66	55
57	73	98
58	92	31
59	22	57
60	67	19
61	23	89
62	58	93
63	94	30
64	53	18
65	24	84
66	69	66
67	50	99
68	7	33
69	62	43

70	40	74
71	8	35
72	96	45
73	38	91
74	89	26
75	64	1
76	68	70
77	3	76
78	39	100
79	84	17
80	28	86
81	6	12
82	21	14
83	46	40
84	36	8
85	49	21
86	9	63
87	61	94
88	93	5
89	97	9
90	54	64
91	33	62
92	32	68
93	26	50
94	43	16

95	4	59
96	14	60
97	34	79
98	74	56
99	48	69
100	81	39

APPENDIX 3

Summary of observer responses*: Composite table.

Xray	Observer A		Observer B		Observer C		Observer D	
	First viewing	Second Viewing	First viewing	Second viewing	First viewing	Second viewing	First viewing	Second viewing
1	YES	NO	YES	YES	YES	YES	YES	EQUIV
2	NO	NO	NO	NO	NO	NO	EQUIV	NO
3	YES	YES	EQUIV	EQUIV	NO	YES	YES	EQUIV
4	NO	NO	EQUIV	EQUIV	NO	NO	EQUIV	NO
5	NO	EQUIV	NO	NO	EQUIV	NO	NO	NO
6	NO	NO	YES	YES	NO	NO	NO	NO
7	NO	NO	NO	NO	NO	NO	NO	NO
8	EQUIV	NO	NO	NO	EQUIV	NO	EQUIV	NO
9	EQUIV	NO	EQUIV	EQUIV	NO	YES	EQUIV	EQUIV
10	YES	YES	EQUIV	EQUIV	EQUIV	YES	YES	NO
11	NO	NO	YES	YES	NO	NO	EQUIV	YES
12	NO	NO	NO	NO	NO	NO	NO	NO
13	NO	NO	EQUIV	EQUIV	NO	NO	NO	NO
14	NO	EQUIV	NO	NO	NO	NO	NO	NO
15	NO	NO	NO	NO	YES	NO	EQUIV	EQUIV
16	EQUIV	NO	EQUIV	NO	NO	NO	NO	NO
17	NO	NO	NO	NO	NO	NO	EQUIV	EQUIV
18	YES	YES	YES	YES	YES	YES	YES	YES
19	NO	NO	NO	EQUIV	NO	NO	EQUIV	NO
20	YES	YES	NO	NO	YES	YES	EQUIV	NO
21	NO	EQUIV	EQUIV	EQUIV	NO	NO	NO	NO

22	EQUIV	YES	EQUIV	EQUIV	NO	YES	YES	YES
23	NO	EQUIV	YES	YES	NO	NO	EQUIV	NO
24	YES	NO	EQUIV	EQUIV	YES	NO	NO	NO
25	YES	YES	YES	YES	YES	YES	YES	YES
26	NO	NO	NO	NO	NO	NO	NO	NO
27	NO	NO	NO	EQUIV	NO	NO	NO	EQUIV
28	NO	NO	NO	EQUIV	NO	NO	EQUIV	NO
29	NO	EQUIV	YES	YES	EQUIV	YES	YES	EQUIV
30	NO	NO	EQUIV	YES	YES	YES	YES	EQUIV
31	YES	YES	NO	NO	YES	YES	YES	YES
32	NO	E	NO	EQUIV	NO	NO	NO	NO
33	NO	NO	NO	NO	NO	YES	NO	EQUIV
34	EQUIV	EQUIV	NO	EQUIV	NO	NO	NO	NO
35	NO	EQUIV	EQUIV	EQUIV	NO	NO	NO	EQUIV
36	NO	NO	NO	NO	NO	NO	NO	NO
37	EQUIV	EQUIV	NO	EQUIV	YES	NO	YES	NO
38	NO	EQUIV	NO	EQUIV	NO	NO	NO	NO
39	NO	NO	NO	NO	NO	NO	NO	NO
40	NO	NO	NO	NO	YES	NO	NO	NO
41	NO	NO	EQUIV	EQUIV	NO	EQUIV	EQUIV	NO
42	NO	NO	NO	NO	NO	NO	EQUIV	NO
43	NO	NO	NO	EQUIV	NO	NO	NO	NO
44	NO	NO	NO	NO	NO	NO	NO	NO
45	NO	NO	NO	NO	NO	NO	NO	NO
46	NO	NO	EQUIV	NO	NO	NO	NO	NO
47	NO	NO	NO	EQUIV	NO	NO	NO	NO
48	NO	NO	YES	YES	EQUIV	YES	NO	NO
49	NO	EQUIV	EQUIV	NO	NO	NO	EQUIV	NO
50	NO	NO	NO	NO	NO	NO	NO	NO

51	YES	NO	EQUIV	EQUIV	YES	EQUIV	EQUIV	EQUIV
52	NO	NO	NO	NO	NO	NO	NO	NO
53	NO	NO	NO	NO	NO	NO	NO	NO
54	NO	NO	EQUIV	EQUIV	NO	YES	EQUIV	NO
55	EQUIV	NO	NO	NO	YES	YES	NO	NO
56	NO	NO	NO	NO	E	NO	EQUIV	NO
57	NO	NO	NO	EQUIV	NO	NO	NO	NO
58	YES	YES	NO	EQUIV	YES	YES	YES	EQUIV
59	NO	NO	NO	NO	NO	NO	NO	NO
60	NO	NO	NO	NO	NO	NO	NO	NO
61	NO	NO	NO	NO	NO	NO	NO	NO
62	NO	NO	EQUIV	EQUIV	YES	NO	YES	EQUIV
63	YES	YES	EQUIV	EQUIV	YES	YES	YES	YES
64	YES	YES	NO	NO	YES	EQUIV	EQUIV	YES
65	NO	NO	NO	NO	NO	NO	NO	NO
66	YES	YES	YES	YES	YES	YES	YES	YES
67	NO	NO	EQUIV	NO	NO	NO	NO	EQUIV
68	EQUIV	NO	EQUIV	EQUIV	NO	NO	EQUIV	EQUIV
69	NO	EQUIV	NO	NO	EQUIV	NO	EQUIV	NO
70	EQUIV	NO	NO	EQUIV	NO	NO	YES	NO
71	NO	NO	NO	NO	NO	NO	EQUIV	NO
72	NO	NO	YES	YES	NO	NO	EQUIV	NO
73	EQUIV	EQUIV	NO	EQUIV	EQUIV	YES	NO	NO
74	NO	NO	EQUIV	EQUIV	NO	NO	NO	NO
75	NO	NO	NO	NO	NO	NO	NO	NO
76	NO	NO	EQUIV	NO	NO	NO	NO	NO
77	NO	NO	EQUIV	EQUIV	NO	EQUIV	YES	YES
78	NO	NO	YES	YES	NO	YES	EQUIV	NO
79	EQUIV	EQUIV	YES	YES	NO	NO	YES	YES

80	E	NO	EQUIV	EQUIV	NO	EQUIV	YES	EQUIV
81	NO	NO	NO	NO	NO	NO	NO	NO
82	NO	NO	NO	NO	NO	NO	NO	NO
83	NO	NO	NO	NO	NO	NO	NO	NO
84	NO	NO	NO	NO	NO	NO	NO	NO
85	NO	NO	NO	NO	NO	NO	NO	NO
86	NO	NO	NO	NO	NO	NO	NO	NO
87	NO	NO	YES	YES	YES	NO	YES	YES
88	NO	NO	YES	YES	NO	NO	YES	EQUIV
89	NO	YES	NO	NO	NO	NO	NO	NO
90	NO	NO	EQUIV	EQUIV	NO	NO	EQUIV	NO
91	NO	NO	YES	YES	NO	YES	EQUIV	EQUIV
92	YES	YES	YES	YES	YES	YES	YES	YES
93	NO	NO	NO	EQUIV	NO	NO	NO	NO
94	NO	NO	EQUIV	YES	NO	NO	EQUIV	NO
95	NO	NO	NO	NO	YES	YES	NO	NO
96	NO	NO	EQUIV	NO	NO	NO	EQUIV	NO
97	EQUIV	EQUIV	EQUIV	EQUIV	NO	NO	NO	EQUIV
98	EQUIV	NO	YES	YES	YES	NO	EQUIV	NO
99	EQUIV	YES	NO	EQUIV	NO	YES	YES	EQUIV
100	EQUIV	EQUIV	NO	NO	NO	NO	NO	NO

* Explanation to the terms 'Yes', 'No' and 'Equivocal'.

The question asked was "Is mediastinal lymphadenopathy present?" Answers provided included the following categories; 'Yes' for unequivocally present, 'No' for unequivocally absent, and an 'equivocal' answer category was included to allow for an uncertain decision. Observer codes (known only to the study co-ordinator) and

individual performance remained anonymous, unless the particular individuals involved requested feedback on their own performance.

APPENDIX 4

Examiner Questionnaire:

Inter and Intra Observer agreement between Paediatric Pulmonologists in the detection of mediastinal lymphadenopathy when viewing chest radiographs of children.

Dear study examiners,

With reference to the above study in which you kindly took part, please answer the following questions in an attempt to further understand the difficulty in commenting on mediastinal lymphadenopathy when viewing radiographs. Your answers will also assist us in the drafting of a further study, which includes the use of contrasted CT as a reference standard.

Q1. What criteria (if any) do you use when coming to a decision concerning the presence or absence of mediastinal lymphnodes on chest radiographs?

Q2. What is your opinion concerning the value of a lateral chest radiograph in addition to the AP radiograph when viewing for mediastinal lymphadenopathy

Kind regards

G Du Toit, G Swingler, K Iloni

APPENDIX 5

List of criteria cited by each observer when answering the questionnaire:

Q1) What criteria (if any) do you use when coming to a decision ceoncerning the presence or absence of mediastinal lymphnodes on chest radiographs?
Observer A : Intra Observer Agreement: Kappa 0.58 (95% CI 0.41-0.74)
In deciding about the presence of mediastinal lymphadenopathy I always consider the clinical history and Mantoux reaction into account, in the absence of which I have difficulty and in practice will generally not commit myself. I make a certain radiological diagnosis of lymphadenopathy only in the presence of : 1. Tracheo-bronchial compression compatible with lymphadenopathy 2. Splayed carina with retro-cardiac density suggestive of lymphnodes 3. Rounded paratracheal shadows with calcification 4. Multiple rounded paratracheal masses
Observer B: Intra Observer Agreement: Kappa 0.71 (95% CI 0.56-0.87)
1. Impression of widened mediastinum 2. Indentation of hilar structures e.g. bronchi on lateral Chest Radiograph
Observer C: Intra Observer Agreement: Kappa 0.44. (95% CI 25-0.62)

Presence of mediastinal lymphadenopathy if: 1. Obvious widening of the mediastinum or hilar shadows by what looks like nodes 2. Abnormal shape or outline or course of the trachea and main bronchi. Usually localised narrowing, with or without above findings. 3. Difficulty in seeing the outlines of the trsachea or main bronchi on repeated Chest Radiographs 4. Localised air trapping with suggestion of 1 and/or 2 and/or 3 Absence of mediastinal lymphadenopathy if none of the above 1-5 features are present
Observer D: Intra Observer Agreement: Kappa 0.47 (95% CI 0.32-0.62)
1. Check if adenopathy/swelling is present in the neck and axillae 2. Look at the mediastinum on AP: if widened and age not appropriate for the presence of a thymus, I consider this suggestive of mediastinal lymphadenopathy, particularly if the lateral chest radiograph doesn't demonstrate anterior mediastinal opacification. 3. Look at the trachea and bronchi-this is suggestive of lymphadenopathy if narrowed. 4. Look at the hilum on AP, if opacified consider lymphadenopathy. 5. Look at the lateral chest radiograph for well defined lobular structures.

Q2) What is your opinion concerning the value of a lateral chest radiograph in addition to the AP radiograph when viewing for mediastinal lymphadenopathy.
Observer A: Intra Observer Agreement: Kappa 0.58 (95% CI 0.41-0.74)
I find the lateral chest radiograph useful only for positive identification of the thymus, and rarely in cases of a splayed carina to exclude left atrial enlargement.
Observer B: Intra Observer Agreement: Kappa 0.71 (95% CI 0.56-0.87)
Very valuable
Observer C: Intra Observer Agreement: Kappa 0.440.(95% CI 25-0.62)
In most cases not much use. The lateral chest radiograph does help in localising mediastinal masses into anterior, middle and posterior mediastinum. The lateral chest radiograph is also useful in identifying thymic shadows that look unusual on the AP chest radiograph.
Observer D: Intra Observer Agreement: Kappa 0.47 (95% CI 0.32-0.62)
Essential in most cases. Difficult to interpret in the presence of significant lung pathology where nodes may have been seen.

MMED Thesis Submission: Observer variation in detecting Lymphadenopathy on chest radiography.

Du Toit LG DTTLIO002, Submission: May 2003

University of Cape Town
Faculty of Health Sciences

A schedule of the completed corrections is attached indicating where all proposed changes have been made:

Reviewer 1:

- Issue 1: The following description of the Kappa weighting has been added to Page 13, Para 1: "A weight of 1 was given to complete agreement, 0.5 to discrepancies of one category, and 0 for discrepancies of two categories."
- Issue 2: The 'unclear' multidimensional graph has been omitted.
- Issue 3 is resolved by correcting the data entries in the table 3.7. Tables 3.7 should have included an "Equivocal" column, which has now been corrected. (Page 21)
- Minor issues:
- Issue 4: The point is taken. The purpose of this study was to assess agreement in actual practice. Criteria will be specified to maximise agreement in a follow-up study.
- Issue 5: The selection of pneumonia radiographs was random. This has been clarified on Page 11, para 1.
- Issue 6: The specific age distributions for the 2 groups are not available and therefore are omitted.
- Issue 7: The value 13.25% was incorrect in Table 3.3, and has been corrected.
- Issue 8: The interpretation of the Spearman rank correlation has been changed to read, " Although this correlation co-efficient is high, the p value indicates that it could quite easily be due to chance." Page 19, Final para
- Ideas for future: The ideas for future projects are appreciated

Reviewer 2:

- Terminology changes include:
  a) M Tuberculosis is now presented as *M tuberculosis*
  b) All Kappa references are now spelt with a capital 'K'
- (Page 3, first para) The sentence " 1.3 million infected children" has been altered to read "1.3 million diseased children"
- (Page 1, last sentence) The last sentence has been altered from "highly active lesions that seed the blood and lymph" to "highly active lesions may lead to miliary dissemination"
- (Page 16, first para) The previously transposed numbers 19% and 63.75% have been corrected.

- (Page 26, last para) The sentence has been altered to more precisely identify the focus of the study and now reads, " The primary focus of the study was to measure inter- and intra-observer agreement in the detection of mediastinal lymphadenopathy on chest radiography in children at risk for tuberculosis"
- (Page 9, final para) The 3 additional references proposed have been incorporated into the literature review. They were not identified in the original literature because they pre-date MEDLINE

Additional change:

(Page 55) The last page no longer contains the final Appendix 6 (Page 55), which was of 4 scanned images (Quality deemed too poor to add to thesis)