

Chest radiographs for paediatric TB diagnosis: Inter-rater agreement and utility

Grace Kaguthi, Videlis Nduba, Jocelyn Nyokabi

Franklin Onchiri, Robert Gie, Martien W. Borgdorff

Interdisciplinary Perspectives on Infectious Diseases, Volume 2014, Article ID 291841

<http://dx.doi.org/10.1155/2014/291841>

Introduction: The chest radiograph (CXR) is considered a key diagnostic tool for paediatric Tuberculosis (TB) in clinical management and endpoint determination in TB vaccine trials.

Methods: We set out to compare inter-rater agreement for TB diagnosis in western Kenya. A paediatric pulmonologist and radiologist (experts), a medical officer (M.O) and four clinical officers (C.Os) with basic training in paediatric CXR reading, blindly assessed CXRs of infants who were TB suspects in a cohort study. C.Os had access to clinical findings for patient management. Weighted kappa scores summarized inter-rater agreement on lymphadenopathy and abnormalities consistent with TB. Sensitivity and specificity of raters was determined using microbiologically confirmed TB as the gold standard (n=8).

Results: A total of 691 radiographs were reviewed. Agreement on abnormalities consistent with TB was poor; $k=0.14$; (95%CI: 0.10 - 0.18), and on lymphadenopathy moderate $k=0.26$ (95%CI: 0.18 - 0.36), M.O [75% (95 %CI: 34.9%-96.8%)] and C.Os [63% (95%CI: 24.5%-91.5%)] had high sensitivity for culture confirmed TB.

Conclusion: TB vaccine trials utilizing expert agreement on CXR as a non-microbiologically confirmed endpoint will have reduced specificity and underestimate vaccine efficacy. C.Os detected many of the bacteriologically confirmed cases, however this must be interpreted cautiously as they were unblinded to clinical features.

INTRODUCTION

Paediatric TB remains a challenging disease to diagnose despite advances in molecular techniques in mycobacterial identification and antigen based tests for latent TB infection(1). Classical TB symptoms are nonspecific [2] and more so in settings with high HIV prevalence and malnutrition. Atypical presentation with acute severe pneumonia in young children has been observed [3]. Childhood TB is characterized by paucibacillary disease and microbiological confirmation is only possible in <50% of paediatric cases [1]. Chest imaging is therefore of great importance in identifying smear negative, culture negative TB. Among adults with suspected TB, several clearly defined chest radiograph features have been identified as having high inter rater reliability and correlation with culture positive TB [4] Unfortunately similar data among infants has been limited. Lymphadenopathy is the hallmark of primary TB [5]. However, it is frequently missed due to inadequate sensitivity of the chest radiograph [6]. Chest CT Scan (CT) has been considered the gold standard for detecting mediastinal lymphadenopathy, detecting up to 60% more lymphadenopathy in children with normal chest radiographs [7]. Despite this, use of CT has been limited in infant TB vaccine trials, which are set up to detect every TB endpoint so as to demonstrate efficacy. Reasons include; the modest agreement on lymphadenopathy on CT [7], cost limitations and the reluctance to use high dose ionizing radiation in young children. Thus, the chest radiograph is the mainstay of radiological diagnosis, and is frequently the only tool available.

There are a limited number of studies that have described inter rater agreement on chest radiograph for TB diagnosis [8, 9]. Existing studies had small sample sizes, were drawn from hospitalized children, compared agreement only among experienced and highly trained raters, and most importantly, they focused entirely on presence of lymphadenopathy as a marker of TB. While lymphadenopathy is a key feature to diagnosing childhood TB, other radiological features also contribute to the diagnosis [5, 6]

We conducted the study in Siaya County, Western Kenya, which has a high burden of both tuberculosis and HIV [10, 11]. The objective was to determine inter rater agreement on any abnormality on chest radiograph, and agreement on abnormalities consistent with TB among experienced and inexperienced raters. We also aimed to compare the raters' sensitivity and specificity against microbiologically confirmed TB in young children.

STUDY POPULATION AND METHODS

Study setting

The study was conducted in Siaya County, Western Kenya from June 2009 to December 2011. TB diagnosis is by the Keith Edward Score Chart [12] which assigns a score to suggestive pulmonary and extra-pulmonary signs/symptoms of TB. Children who score ≥ 7 as well as those who score < 7 but have an abnormal chest radiograph are treated for TB. A total of 2900 BCG vaccinated infants, aged zero to six weeks and weighing at least 1700g were enrolled and followed up for 12-24 months to determine TB Incidence. TB suspects were identified through four monthly scheduled visits, sick visits as well as

by review of TB case records for contact tracing. Suspect criteria included, a history of contact and/or suggestive signs and symptoms of TB, and/or protocol defined hospitalization history for example for severe pneumonia.

Clinical and Laboratory Investigations

TB suspects were admitted into a case verification ward (CVW) for collection of two serial induced sputa specimens, two serial early morning gastric aspirates, DNA PCR HIV (HIV QUAL test 48 Roche Molecular Systems Inc, Switzerland) testing and Rapid HIV tests for those aged less than and greater than 18 months respectively. Tuberculin Skin Testing was also done. Digital antero-posterior (AP) and lateral chest radiographs were taken at admission and images written into CD-ROMs. The CD-ROMs had Digital Imaging and Communication in Medicine (DICOM) software (Phillips) that was used to view images. Readers could optimize the light/greyscale settings, as well as the magnification.

Definitions and Classifications

Microbiologically confirmed TB (Definite TB) was *M. tuberculosis* identified by Xpert MTB/RIF or speciated with either Capilia (FIND and Tauns co. Ltd) or GenoType assay (Hain Diagnostika, Nehren, Germany) after positive sputum culture. Probable TB was a case started on anti-TB treatment based on Keith Edward Score Chart and/or a CXR consistent with TB.

Raters and Training

There were four sets of raters: A Radiologist and a Paediatric Pulmonologist (expert readers), a Medical Officer (M.O) and four Clinical Officers (C.Os). The C.Os reviewed the images while the suspects were admitted to the CVW, all other raters read the images after close of the study and were blinded to the clinical history. All raters were trained on using the electronic interface to enter readings. The Clinical and Medical Officer(s) were trained on reading and identifying TB on paediatric radiographs prior to the start of the study. A chest radiograph reading form developed during a consensus meeting of TB vaccine sites (Cape Town December 2008) was used. An electronic version was developed to standardize reporting and minimize ambivalence in the diagnosis. The technical quality of images was assessed prior to reading. Indicators included: adequacy of collimation (visibility of the lung apices, costophrenic angles with nothing obscuring the lung fields), adequate exposure; the number of visible intervertebral spaces; adequate inspiration, by counting six anterior ribs, absence of rotation by comparing the length of the same rib on the left and right hemidiaphragms. Thereafter, the reader classified the chest radiograph quality as optimal or suboptimal or unreadable. 'Unreadable' radiographs were counted as suboptimal but were still read. Raters then systematically reviewed the images for airway narrowing, left tracheal deviation, lymphadenopathy, airway opacities, calcification and pleural effusion. The pathology items were scored individually as present, absent, and equivocal. Final assessment of the image was; normal radiograph; abnormal TB unlikely; abnormal TB likely. Radiological signs suggestive of TB were: miliary picture, airway narrowing or tracheal deviation to the left, presence of hilar, paratracheal or subcarinal or other lymphadenopathy, evidence of calcification, cavitation, pleural effusion or thickening.

Data collection and analysis

Data quality was assured by edit, logic and validation checks built onto the data entry interface. Data cleaning was also conducted, and an audit trail of changes maintained. Data was saved onto an SQL database and analysed using SAS 9.0 (SAS Institute Inc, Cary NC, USA). To quantify degree of agreement in TB diagnosis among the raters, we estimated the individual kappas for each rating category as well as a generalized (multiple rater) chance-corrected kappa statistic (a multi-rater measure of agreement), which is an extension of Cohen's kappa for assessing reliability or proportion of agreement for multiple raters. The overall/generalized kappa measures agreement across all categories.

Kappa scores were interpreted as follows: Poor 0.01-0.20, moderate 0.21-0.40, fair 0.41-0.60, good 0.61-0.80 or excellent 0.81-1.0 (13). We compared agreement on any abnormality on chest radiograph, as well as agreement on abnormalities consistent with TB. For comparability with previously conducted paediatric chest radiographs studies we examined the agreement on lymphadenopathy on AP and lateral images. Sensitivity and specificity of readers' diagnosis to definite TB was also calculated. To examine each rater's propensity to place a radiograph in a certain category, and thus elucidate patterns of agreement, McNemar Test for 2X2 tables, and Bowker's Test of symmetry were applied for tables with more than two categories. These methods compare marginal frequencies of each rater and test for statistically significant differences (Tests of marginal homogeneity). Where $p < 0.05$, marginal heterogeneity exists that is, differing propensity to rate a category.

The study was reviewed and approved by the Kenya Medical Research Institute Ethics Review Committee, (KEMRI-ERC) SSC 1465. Written Informed consent was obtained from parents and Guardians before study entry and for CVW investigations.

RESULTS

TB suspects comprised 33% of those enrolled (959/2900). Of these, 767 (80%) consented to CVW admission for additional investigations and 837 chest radiographs were taken (Figure 1).

There were similar numbers of male (51.5%) and female (49.5%) suspects. Some of the prevalent co-morbid conditions among suspects included HIV exposure- nearly one-fifth (17.3%), HIV infection (4.4%) and malnutrition (46.0%). Close to half of investigated infants 355/767 (46.3%) had been hospitalized with severe lower respiratory tract infections such as bronchiolitis and pneumonia. Prevalence of acute and chronic malnutrition was as follows; wasting (WHZ < -2) $n=167/767$ (21.7%) stunting (HAZ < -2) $82/767$ (10.2%) and underweight (WAZ < -2) $104/767$ (13.6%).

Agreement on CXR reading was low overall. For abnormalities consistent with TB (Table 1a and 1b), poor to moderate agreement was observed across all rater pairs: pulmonologist-radiologist ($k=0.24$ (95% CI: 0.15-0.34)); pulmonologist-medical officer ($k=0.21$ (95% CI: 0.13-0.29)) and radiologist-medical officer $k=0.18$ (95% CI: 0.10-0.26).

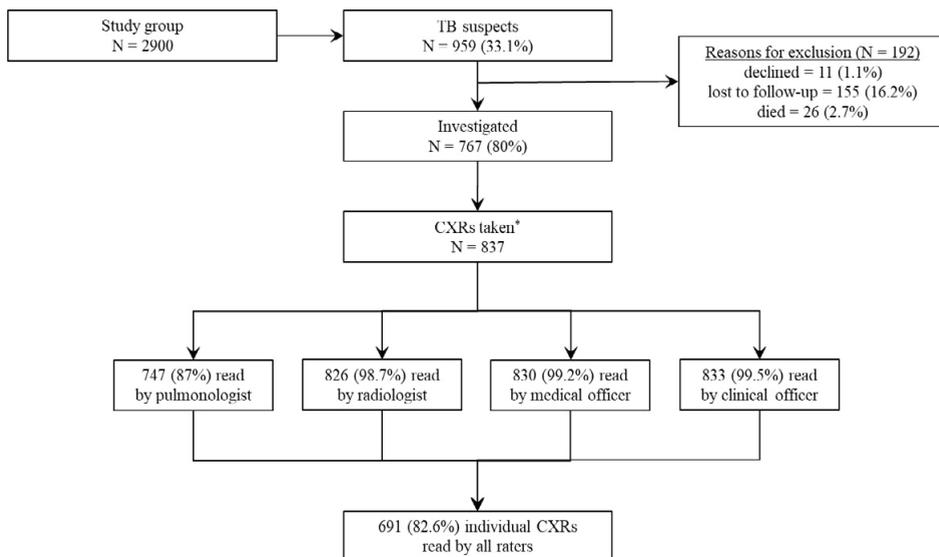


Figure 1 | Profile of presumptive TB cases and chest radiographs read per rater

* per patient more than one CXR could have been taken

Table 1a | Inter-rater agreement on abnormalities consistent with TB

Outcome	Kappa (95% CI)	P value
TB Classification Category		
Abnormal	0.177 (0.124 - 0.237)	<0.001
Abnormal likely TB	0.193 (0.095 – 0.330)	<0.001
Abnormal unlikely TB	0.065 (0.026 – 0.125)	<0.001
Generalized/Weighted Kappa	0.136 (0.100 - 0.176)	<0.001

On any abnormality on the radiograph, agreement was moderate between expert readers [k=0.28 (95% CI: 0.19-0.37)] and between expert readers and M.O. [Pulmonologist vs medical officer (k=0.23 (95% CI: 0.15-0.31)] and [Radiologist vs Medical Officer (k= 0.22 (95% CI: 0.14-0.30))] respectively. (Table 3). Poor- agreement was registered across all other rater pairs. There was similar propensity to rate categories between the expert pair (p=0.14) and between the radiologist and the clinical officer (p=0.24). Regarding quality of radiographs, little or no agreement was registered across all rater pairs. The best kappa score was 0.07 (0.03-0.10), observed between the expert pair. (Table 4). There were eight definite and 40 probable TB cases in the study. One radiograph of a definite TB case was not reviewed by all four raters. The sensitivity and specificity of raters (Table 5) was determined using definite TB cases as the gold standard (n=8).

Table 1b | Inter-rater agreement on abnormalities consistent with TB

Clinical Officer (C.O)	Medical Officer (M.O)	RADIOLOGIST									Total
		Normal			Abnormal likely TB			Abnormal unlikely TB			
		Normal	Abnormal likely TB	Abnormal unlikely TB	Normal	Abnormal likely TB	Abnormal unlikely TB	Normal	Abnormal likely TB	Abnormal unlikely TB	
Normal	Normal	345	5	30	12	0	3	18	0	4	417
	Abnormal likely TB	5	1	3	1	0	1	0	0	2	13
	Abnormal unlikely TB	83	1	20	3	1	5	10	1	6	130
Abnormal likely TB	Normal	12	1	1	0	0	1	0	0	0	15
	Abnormal likely TB	1	0	0	1	3	0	0	1	0	6
	Abnormal unlikely TB	4	1	0	1	1	1	2	0	0	10
Abnormal unlikely TB	Normal	51	1	4	3	0	2	4	2	1	68
	Abnormal likely TB	4	1	0	1	1	2	0	0	0	9
	Abnormal unlikely TB	10	0	2	2	1	1	4	2	1	23
Total		515	11	60	24	7	16	38	6	14	691

Agreement among all raters on the diagnosis of lymphadenopathy was moderate, with a multi-rater weighted $k=0.26$ (Table 2).

Table 2 | Overall agreement amongst all raters on lymphadenopathy and quality of CXR

Rating	Kappa (95% CI)	P value
Lymphadenopathy		
Present	0.27 (0.129–0.448)	<0.001
Absent	0.29 (0.163–0.430)	<0.001
Equivocal	0.17 (0.044–0.423)	<0.001
Overall weighted kappa	0.26 (0.182–0.355)	<0.001
Quality of radiographs		
Optimal	-0.1324	0.999
Suboptimal	-0.1674	0.999
Unreadable	-0.0184	0.8822
Overall weighted kappa	-0.1452	0.998

Table 3 | Agreement on any abnormality on chest radiograph (+) abnormal/ (-) normal

N = 691						
Reader	-/-	-/+	+/+	+/-	Kappa	McNemar's Test
radiologist vs pulmonologist	515	62	43	71	0.28 (0.19-0.37)	0.44
pulmonologist vs M.O	445	132	59	55	0.23 (0.15-0.31)	<0.0001
radiologist vs M.O	450	136	55	50	0.22 (0.14-0.30)	<0.0001
pulmonologist vs CO	477	100	31	83	0.10 (0.01-0.18)	0.21
radiologist vs C.O	493	93	38	67	0.18 (0.10-0.27)	0.04
C.O vs M.O	417	143	48	83	0.10 (0.02-0.17)	<0.0001

Table 4 | Inter-rater agreement on quality of chest radiograph

N = 633						
Reader	Optimal-Optimal	Optimal-Suboptimal	Suboptimal-Suboptimal	Suboptimal-Optimal	Kappa	McNemar's Test
radiologist vs pulmonologist	29	295	301	7	0.07 (0.03-0.10)	<0.0001
pulmonologist vs M.O	10	37	271	314	-0.09 (-0.13- -0.05)	<0.0001
radiologist vs M.O	2	45	551	34	-0.02 (-0.09- 0.05)	0.22
pulmonologist vs CO	7	5	303	317	0.005 (-0.02-0.03)	<0.0001
radiologist vs C.O	2	12	586	34	0.06 (-0.05-0.16)	0.0003
C.O vs M.O	0	12	573	47	-0.03 (-0.05- -0.02)	<0.0001

Table 5 | Accuracy of TB diagnosis by the various care providers – based on culture confirmed TB

CXr reader	Culture results		Sensitivity (95%CI)	Specificity (95%CI)	PPV*	NPV*
	Positive N = 8	Negative N = 683				
Clinical officer						
Positive	5	126	62.5%	81.6%	3.8%	99.5%
Negative	3	557	(24.5% - 91.5%)	(78.4% - 84.4%)	(1.41%-9.13%)	(98.3%-99.9%)
Medical officer						
Positive	6	185	75.0%	72.9%	3.14%	99.6%
Negative	2	498	(34.9% - 96.8%)	(69.4%-76.2%)	(1.28%-7.03%)	(98.4%-99.9%)
Radiologist						
Positive	4	101	50.0%	85.2%	3.96%	99.3%
Negative	4	582	(15.7%-84.3%)	(82.3% -87.8%)	(1.23%-10.0%)	(98.1-99.8%)
Pulmonologist						
Positive	4	110	50.0%	83.9%	3.5%	99.5%
Negative	4	573	(15.7%-84.3%)	(80.9%- 86.6%)	(1.13%-9.27%)	(98.1%-99.8%)

*PPV = positive predictive value, NPV = negative predictive value - disease incidence 1.12% (0.54-2.36)₂

The clinical and medical officer(s) detected the largest proportion of definite TB, while specificity was highest for expert raters. Due to the small number of definite cases, sensitivity was imprecisely measured. The same table shows low positive predictive values (PPV) (<4.0% for all raters) and high negative predictive values (>99.0% for all raters) for the chest radiograph of 28/40 (70%) probable TB cases whose chest radiographs were read by all raters, experts agreed on only two as being consistent with TB. Such would be the stringent case definition applied in infant TB vaccine trials [14] where only radiographs in which experts agreed (Figure 2) would be admissible as non-microbiologically confirmed TB cases.

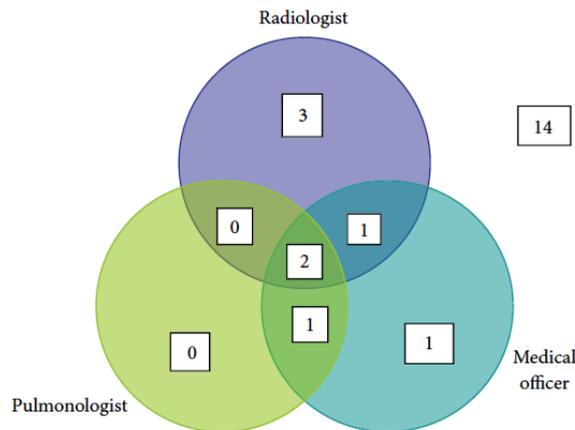


Figure 2 | Venn diagram showing radiographs of probable TB cases reviewed by all raters (n=28) and classified as 'abnormal likely TB'.

DISCUSSION

Overall, we observed poor to moderate agreement between experts and between expert, non-expert pairs. This was consistent across the rater's summary opinion of radiograph:-as well as on specific pathology such as lymphadenopathy. Our findings fit in with other studies comparing inter-rater agreement on lymphadenopathy for TB diagnosis [9]. Additionally, we demonstrate that agreement is also poor for the composite assessment of the radiograph, beyond individual radiographic abnormalities. This undermines the reliability of the chest radiograph for TB diagnosis in infant vaccine trials, as well as possibly for clinical diagnosis and patient care. We suggest several contributing factors. In an infant cohort study with active case finding, suspects are likely to be picked up and investigated before advanced disease with well-defined radiographic features sets in. Early tuberculosis is likely associated with greater diagnostic uncertainty and risk of misclassification [15]. Inter rater agreement studies in a clinical setting where TB suspects may present with more advanced disease are needed to confirm this.

A study of two large infant cohorts in South Africa showed frequent discordance between radiological and microbiological features of TB [2, 16]. The absence of clear, defined radiological abnormalities

that correlate well with microbiologically confirmed disease contributes to lack of a reproducible, standardized criteria that raters can use with certainty to evaluate radiographs.

In radiographs of young children, mediastinal abnormalities are difficult to assess and interpret [17] particularly in inexperienced readers; in this case the C.O and M.O. However, even experienced readers have been found to have poor agreement on lymphadenopathy on chest radiographs [9].

Previous studies had smaller sample sizes for assessment of inter rater variability[8, 9] One of the strengths of this study is the inclusion of a large sample of radiographs for evaluation, from young children with a broad range of respiratory illness and co morbidities. The varying levels of raters' expertise represents the general clinical care structure from primary level to specialized referral hospitals. The findings are therefore applicable to a broad range of settings.

The study had limitations. It was not possible to obtain consensus from expert readers on radiographs on which they differed. This would have increased agreement scores and validity of the chest radiograph as a diagnostic tool. It would also have elucidated patterns of disagreement, in order to refine criteria for identifying pathologies consistent with TB on chest radiographs as has been previously recommended [9]. Obtaining high inter rater agreement for paediatric chest radiographs in acute paediatric respiratory illness is difficult. Pneumococcal vaccine trials have succeeded in this for opacities consistent with pneumonia on paediatric chest radiographs[18] Unfortunately, it has been difficult to replicate this success in TB vaccine trials.

Conventionally, TB vaccine trial efficacy sample sizes are calculated based on composite endpoints. These include bacteriological and non-bacteriological criteria as it is expected microbiologically confirmed cases will contribute a limited number of endpoints. To increase the number of endpoints, and thus reduce sample size requirements, non-bacteriologically confirmed endpoints that rely on chest radiograph findings are included. The latter is defined as radiographic findings compatible with tuberculosis identified independently by two experts [15, 19]. This approach has some limitations. We found that experts agreed only on five of 35 radiographs, as being consistent with TB. Of these, three were bacteriologically confirmed, therefore the chest radiograph contributed only two additional endpoints. Poor agreement and high variability in interpreting paediatric CXRs for TB diagnosis among experts increases the probability of misclassifying true disease status and thus underestimating vaccine efficacy [20]. C.Os and M.Os, picked up majority of the cases, that were later bacteriologically confirmed. This would seem like a positive outcome, given that they work as primary health care providers and would therefore accelerate diagnosis and treatment of infants with TB. However, the limited number of definite cases results in imprecise estimates of sensitivity and should be cautiously interpreted. The high sensitivity also trades off on specificity and could result in unnecessary TB treatment.

Prevailing disease rates influence predictive values. Among infants, the TB incidence of 1.12% (95% CI 0.54%-2.36%) would be considered high, however low PPV relative to sensitivity is attributed to low disease rates. While the NPV was high, where prevalence is not much above 1%, even a non-informative test may have a NPV close to 100%.

CONCLUSION

Poor agreement and high variability in classifying paediatric radiographs underscores need for caution in diagnosing TB in clinical settings where bacteriological confirmation is unavailable; as in most resource limited settings. It further demonstrates that addition of radiographic, non-bacteriologically confirmed endpoints will be of low benefit in decreasing sample sizes for TB vaccine trials.

REFERENCES

1. Marais BJ, Pai M. Recent advances in the diagnosis of childhood tuberculosis. *Arch Dis Child*. 2007 May;92(5):446-52.
2. 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 Mar;14(3):189-94.
3. Jeena PM, Pillay P, Pillay T, et al. Impact of HIV-1 co-infection on presentation and hospital-related mortality in children with culture proven pulmonary tuberculosis in Durban, South Africa. *Int J Tuberc Lung Dis*. 2002 Aug;6(8):672-8.
4. Pinto LM, Dheda K, Theron G, et al. Development of a simple reliable radiographic scoring system to aid the diagnosis of pulmonary tuberculosis. *PLoS One*. 2013;8(1):e54235.
5. Leung AN, Muller NL, Pineda PR, et al. Primary tuberculosis in childhood: radiographic manifestations. *Radiology*. 1992 Jan;182(1):87-91.
6. Kim WS, Choi JI, Cheon JE, et al. Pulmonary tuberculosis in infants: radiographic and CT findings. *AJR Am J Roentgenol*. 2006 Oct;187(4):1024-33.
7. S.Andronikou BB, J.Galpin, S.Brachmeyer, S.Lucas, E.Joseph, G.Du Toit, G.Swinger. Pathological correlation of CT-detected mediastinal lymphadenopathy in children: the lack of size threshold criteria for abnormality. *Pediatr Radiol*. 2005.
8. Swinger GH, du Toit G, Andronikou S, et al. Diagnostic accuracy of chest radiography in detecting mediastinal lymphadenopathy in suspected pulmonary tuberculosis. *Arch Dis Child*. 2005 Nov;90(11):1153-6.
9. Du Toit G, Swinger G, Iloni K. Observer variation in detecting lymphadenopathy on chest radiography. *Int J Tuberc Lung Dis*. 2002 Sep;6(9):814-7.
10. Division of Leprosy and Tuberculosis and Lung Disease (DLTLD) Annual Report. 2011.
11. Kenya National Bureau of Statistics (KNBS). Kenya Demographic and Health Survey 2008-2009.
12. World Health Organization (WHO). TB/HIV: a Clinical Manual. 1996: Chapter 4.3.
13. Altman DG. Practical Statistics for Medical Research: In: Inter-Rater Agreement 1991:403-4.
14. Tameris MD, Hatherill M, Landry BS, et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet*. Mar 23;381(9871):1021-8.
15. Hatherill M, Verver S, Mahomed H. Consensus statement on diagnostic end points for infant tuberculosis vaccine trials. *Clin Infect Dis*. Feb 15;54(4):493-501.
16. Mulenga H, Moyo S, Workman L, et al. Phenotypic variability in childhood TB: implications for diagnostic endpoints in tuberculosis vaccine trials. *Vaccine*. Jun 2010;29(26):4316-21.
17. Williams H. J HMA. Imaging of paediatric Mediastinal Abnormalities. *Paediatric Respiratory Reviews*. 2003(4):55-66.
18. World Health Organization (WHO) Pneumonia Vaccine Trial Investigators' Group. Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children. October 2001.
19. Graham SM, Ahmed T, Amanullah F, et al. Evaluation of Tuberculosis diagnostics in children: 1. Proposed clinical case definitions for classification of intrathoracic tuberculosis disease. Consensus from an expert panel. *J Infect Dis*. May 2015;205 Suppl 2:S199-208.
20. Behr MA, Schwartzman K, Pai M. Tuberculosis vaccine trials. *Lancet Jun2003*. 29;381(9885):2252-3.