Clinical Radiology xxx (2017) e1–e9

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort

C.S. Kosack [a,*], S. Spijker [a], J. Halton [a], M. Bonnet [b,c], S. Nicholas [b], K. Chetcuti [d], A. Mesic [e], W.E. Brant [f], E. Joekes [g], S. Andronikou [h]

[a] *Médecins Sans Frontières, Diagnostic Network, Amsterdam, The Netherlands*
[b] *Epicentre, Paris, France*
[c] *Institute of Research for Development, UMR233, University of Montpellier, INSERM U1175, France*
[d] *Department of Radiology, Alder Hey Children's Hospital, Liverpool, UK*
[e] *Médecins Sans Frontières, Public Health Department, Amsterdam, The Netherlands*
[f] *Department of Radiology and Medical Imaging, University of Virginia Health System, Charlottesville, VA, USA*
[g] *Diagnostic Imaging department, Royal Liverpool University Hospital & Liverpool School of Tropical Medicine, Liverpool, UK*
[h] *Department of Paediatric Radiology, University or Bristol and Bristol Royal Hospital for Children, Bristol, UK*

ARTICLE INFORMATION

Article history:
Received 14 July 2016
Received in revised form
28 November 2016
Accepted 15 January 2017

AIM: To assess the impact of introducing a chest radiograph reading and recording system (CRRS) with a short training session, on the accuracy and inter-reader variability of tuberculosis (TB) interpretation of chest radiographs (CXRs) by a group of non-expert readers in a human immunodeficiency virus (HIV)-positive cohort.

MATERIALS AND METHODS: A set of 139 CXRs was reviewed by a group of eight physicians pre- and post-intervention at two clinics in Shan State, Myanmar, providing HIV/TB diagnosis and treatment services. The results were compared against the consensus of expert radiologists for accuracy.

RESULTS: Overall accuracy was similar pre- and post-intervention for most physicians with an average area under the receiver operating characteristic curve difference of 0.02 (95% confidence interval: −0.03, 0.07). The overall agreement among physicians was poor pre- and post-intervention (Fleiss κ=0.35 and κ=0.29 respectively). The assessment of agreement for specific disease patterns associated with active TB in HIV-infected patients showed that for intrinsically subtle findings, the agreement was generally poor but better for the more intrinsically obvious disease patterns: pleural effusion (Cohen's kappa range = 0.37–0.67) and milliary nodular pattern (Cohen's kappa range = 0.25–0.52).

CONCLUSION: This study demonstrated limited impact of the introduction of a CRRS on CXR accuracy and agreement amongst non-expert readers. The role in which CXRs are used for TB diagnosis in a HIV-positive cohort in similar clinical contexts should be reviewed.

* Guarantor and correspondent: Cara Kosack, Médecins sans Frontières, Plantage Middenlaan 14, 1018 DD Amsterdam, The Netherlands. Tel.: +31 20 520 8700.
E-mail address: cara.kosack@oca.msf.org (C.S. Kosack).

http://dx.doi.org/10.1016/j.crad.2017.01.008

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

Introduction

Tuberculosis (TB) is a major global health hazard. The majority of reported cases are from low and middle income countries. Immunosuppressed patients, such as those with human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS), are particularly at risk with the majority of these patients living in Asia and sub-Saharan Africa.[1]

Confirming TB in HIV-positive patients with microbiological tests in resource-constrained settings is particularly challenging due to higher rates of smear negative results, difficulties in obtaining sputum specimens, limited access to molecular testing, such as GeneXpert and mycobacterial culture, and the time taken for culture results.[2,3] In many settings, a chest radiograph (CXR) remains the primary diagnostic tool and provides an important contribution to the combined criteria for diagnosing TB in this patient group, and is therefore endorsed by the World Health Organization (WHO).[4]

The usefulness of CXRs in the diagnosis of TB, however, is complicated by the non-specific presentation of the disease in HIV-positive patients, and more so if the technical quality of the CXR is limited.[5–7] CXR interpretation is prone to high subjectivity, inter- and intra-reader variability and over-reading, especially when read by less experienced, non-expert readers, who are largely responsible for reporting in resource-constrained settings.[2,3,8–11]

Aside from the provision of training, the use of a CXR reading and recording system (CRRS), in prevalence studies by non-expert readers, has demonstrated improvements and satisfactory inter- and intra-reader agreement.[12] CRRS encourages a systematic approach to reporting using identical descriptive terms by all readers. The resultant data can be used for follow-up and comparative studies. Thus, CRRS could potentially improve the diagnostic clinical validity of CXRs.[10–12]

The aim of the present study was to determine if the introduction of a CRRS improves the interpretation accuracy of CXRs for TB, by non-radiologist physicians, against interpretation by expert radiologists, in an HIV-positive cohort in a resource-constrained setting. Furthermore, a further aim was to assess whether the application of a CRSS reduces the inter-reader variability of CXR interpretation, in the group of non-radiologist physicians by comparing the inter-reader agreement, before and after the intervention.

Materials and methods

Study population

Myanmar has one of the highest TB burdens in the world with an estimated 525 cases per 100 000 reported in 2010.[13] In addition, the HIV/AIDS burden is among the most serious in Asia, with an estimated 216,000 adults and children living with HIV in 2011.[14] A medical humanitarian non-governmental organisation, supports the Lashio and Muse clinics, in Shan state, Myanmar, where this study was conducted.

Study procedures

A sample of 139 CXRs was calculated from the receiver operating characteristic (ROC) diagnostic accuracy table proposed by Obuchowsky for area under the curve (AUC), assuming an expected pre-intervention AUC for CXR reading of 0.75, a 10% difference of AUC between post- and pre-intervention, a 25% frequency of TB suggestive CXR, a correlation of 0.47 between the pre- and post-intervention measure and a 5% alpha risk and 80% power.[15]

The 139 CXRs were randomly selected from a total of 618 conventional screen-film CXRs of HIV-infected adults who came to either clinic for TB screening within the last 12 months. All CXRs were performed at external facilities, not at the clinics. The result of smear microscopy was unknown. TB culture or molecular testing for TB infection was not available.

Eight non-radiologist physicians from Myanmar without a specialisation, whose work currently includes interpreting CXRs at either clinic were recruited. The physicians were asked to record in writing any radiological features detected and whether the CXR was “normal”, “abnormal but not suggestive of TB”, or “abnormal and suggestive of TB”. They subsequently received a four-hour group training session on CXR interpretation of TB and the application of the CRRS (Fig 1). After a period of at least 1 month to avoid recall, the CXRs were reported again in a random order utilising the CRRS.

Reference standard

One hundred and thirty-nine CXRs were digitised following standardised instructions for the digitisation of film images and sent via the internet to three consultant radiologists, all with extensive experience of reading TB films. The CXRs were read digitally permitting the readers any digital enhancement. The reference standard was the consensus opinion of two independent, expert radiologists. The third radiologist was consulted in the case of discrepancies. CXRs were excluded from analysis if the consensus was overall “poor quality” (Fig 1).[16] All readers were blinded to clinical findings, laboratory results, and any previous diagnosis; however, all were aware that CXRs were from patients with presumptive TB from a high prevalence area.

Data collection and statistical analysis

Data were collected from the CRRS forms and double-entered using EpiData 3.1 software (EpiData, Odense, Denmark) and analysis carried out using STATA version 13 (StataCorp, College Station, TX, USA).

The reference standard was classified as radiologically “TB positive” if the expert consultant radiologists interpreted the CXR as “abnormal and TB suggestive” and “TB negative” if interpreted as “normal” or “abnormal but not TB suggestive”.

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

Patient study ID: _____ Reviewer code: _____ Date of review: __/__/_____

Overall quality of radiograph		yes	no
Good quality	No radiographic faults		
Not good but readable	Some faults but readable		
Poor quality: unreadable	Not readable and must be repeated		

Specific quality of radiograph		yes	no
Too dark	Cannot see peripheral vessels		
Too light	Cannot see vertebral ends behind the heart		
Rotated	Medial clavicles different distances from spinous process		
Under-inspired	Less than 9-10 posterior ribs at diaphragm		
Parts cut off	E.g. costophrenic angles		
Artefacts seen	Any dots, lines, metal, clothing that is not anatomy or disease		
Motion artefact	Like a moving photo e.g. loss of sharp edges		

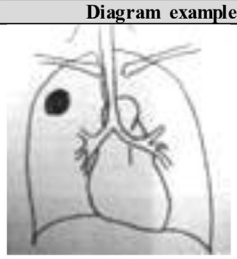
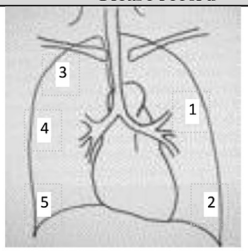
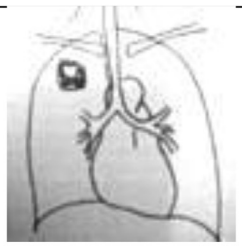
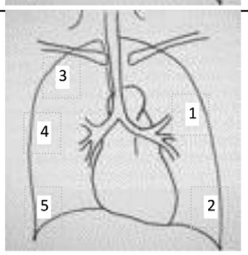
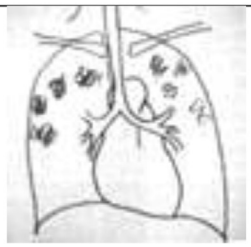
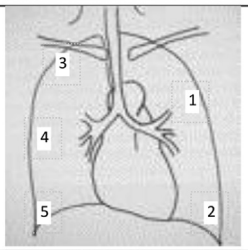
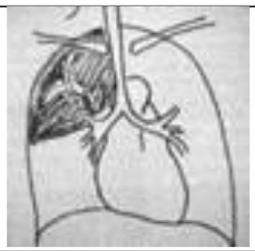
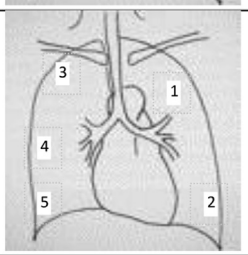
	Diagram example	Please record	Distribution Zones	
Granuloma oval lesion			**1**	
			2	
			3	
			4	
			5	
Cavitation (within a granuloma, consolidation or infiltrate)			**1**	
			2	
			3	
			4	
			5	
Broncho-pneumonic pattern, infiltration			**1**	
			2	
			3	
			4	
			5	
Consolidation			**1**	
			2	
			3	
			4	
			5	

Figure 1 CXR reading and recording system.

For the main accuracy analysis, physicians' pre- and post-intervention CXR readings were compared against the reference standard by plotting ROC curves and calculating the AUCs using the non-parametric method. The equality of ROC curves pre- and post-intervention per physician was assessed using the ROCCOMP command in STATA and overall equality assessed using the DBMMRMC 2.2 software developed by Dorfman-Berbaum-Metz for multireader-multicase ROC analysis of variance (ANOVA).[17]

A secondary accuracy analysis was performed after collapsing non-radiologist physicians CXR readings to two categories (i.e., "abnormal TB suggestive" versus "abnormal not TB suggestive of TB" or "normal"). Sensitivity and specificity of the readings were obtained pre- and post-intervention. Further, a matched paired analysis using McNemar's test was performed to assess if each physicians' sensitivity/specificity improved post-intervention. The paired analysis included only CXRs with pre- and post-intervention readings. To assess for heterogeneity of the sensitivities and specificities of the pre- and post-intervention readings, the R-package mada for meta-analysis of diagnostic accuracy was used to calculate a

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

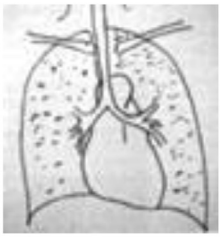
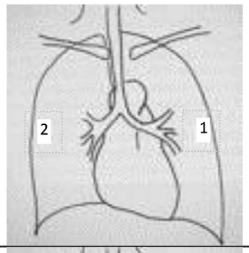
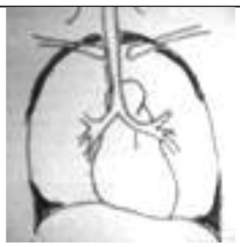
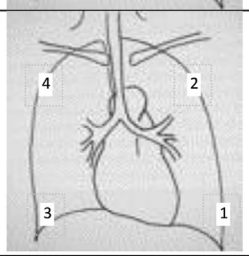
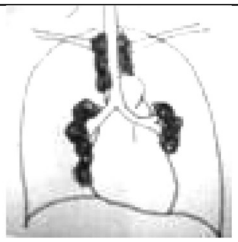
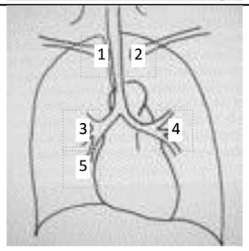
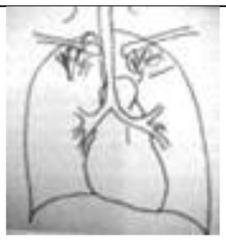
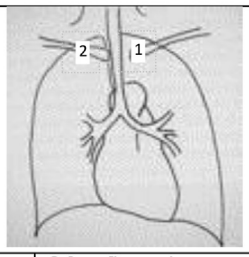
Milliary nodular pattern			**1** **2**	
Pleural effusion			**1** **2** **3** **4**	
Lymph-adenopathy			**1** **2** **3** **4** **5**	
Apical fibrocystic change: scarring and cysts			**1** **2**	
CXR normal			**Other diagnosis**	
CXR abnormal – NOT TB suggestive				
CXR abnormal – TB suggestive				

Figure 1 (*continued*).

summary value and test for equality of sensitivities and specificities.

The inter-reader agreement amongst radiologist and non-radiologist physicians before and after the intervention was assessed using the kappa statistic. Agreement was defined as two readings being either radiologically TB positive or TB negative. To obtain an overall kappa value pre- and post-intervention, Fleiss kappa was calculated. The following interpretation was used to define the strength of agreement for the kappa coefficient[18]: <0.21 = poor, 0.21–0.4 = fair, 0.41–0.6 = moderate, 0.61–0.8 = good and 0.81–1.0 = very good agreement.

The consensus opinion of the radiologists was also compared to each physician for individual disease patterns: broncho-pneumonic pattern/infiltration, lymphadenopathy, consolidation, cavitation, pleural effusion, granuloma oval lesion, miliary disease, and apical fibrocystic change.

Ethics statement

The study was formally approved by the Ethical Review Board of the non-governmental organisation and the Department of Medical Research at the Ministry of Health in Myanmar. Participation was voluntary for physicians and signed informed consent was obtained. CXRs were derived from routine practice of patients with a diagnosis and were subsequently anonymised. Patient informed consent was not obtained but an information letter was provided at each clinic.

Results

Of the 139 CXRs, 13 had to be excluded from further analysis either due to poor image quality (n=4) or no

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

consensus reached (n=9) by the radiologists. Of the 126 CXRs, 39 (31%) were classified as "abnormal TB suggestive" and 87 (69%) as "normal" or "abnormal but not TB suggestive". The non-radiologists rated between 120 and 126 CXRs. Physician 3 did not complete the study and was excluded from the analysis.

Accuracy

Individual ROC curves for each non-radiologist physician pre- and post-intervention demonstrated similar and overlapping curves (Fig 2). Each point corresponds to the accuracy of each reading category against the reference standard. ROC curves were compared by calculating the AUC using the non-parametric method and the results are displayed in Table 1. AUCs were similar pre- and post-intervention, except for physician 4 where the AUC increased from 0.67 to 0.75 (p=0.07). Of the seven, physicians 1 and 6 showed the greatest accuracy with AUCs ≥0.82 pre- and post-intervention. The average AUC difference was 0.02 (95% confidence interval [CI]: −0.03, 0.07) and this was statistically not significant (p=0.41).

The results of the diagnostic accuracy in the form of sensitivity and specificity accounting for paired matching are displayed in Table 2. Differences in sensitivity and specificity between pre- and post-intervention were assessed using McNemar's exact test. Heterogeneity of sensitivities/specificities for pre- and post-intervention were significant (p<0.01), meaning that there was variation amongst the physicians. Therefore, no pooled result is represented.

Table 1
Area under non-parametric (empiric) receiver operating characteristic (ROC) curve.

NRP	n	Pre-Intervention		Post-Intervention		p-Value[a]
		AUC	[95% CI]	AUC	[95% CI]	
1	124	0.86	[0.79,0.92]	0.82	[0.75,0.89]	0.29
2	122	0.78	[0.70,0.85]	0.74	[0.66,0.81]	0.33
4	119	0.67	[0.59,0.75]	0.75	[0.67,0.82]	0.07
5	121	0.74	[0.65,0.82]	0.71	[0.63,0.80]	0.46
6	121	0.89	[0.83,0.94]	0.83	[0.76,0.90]	0.11
7	122	0.74	[0.65,0.82]	0.78	[0.70,0.86]	0.39
8	122	0.76	[0.68,0.84]	0.70	[0.61,0.79]	0.22

NRP, non-radiologist physician; n, number of chest radiographs; AUC, area under curve; CI, confidence interval

[a] p-Value from statistical test comparing equivalence of ROC curves using STATA ROCCOMP command

Inter-reader agreement

Cohen's kappa was calculated after aggregating the "normal" and "abnormal but not TB" categories (Table 3). Physicians 1 and 6, and 6 and 8 had the highest level of

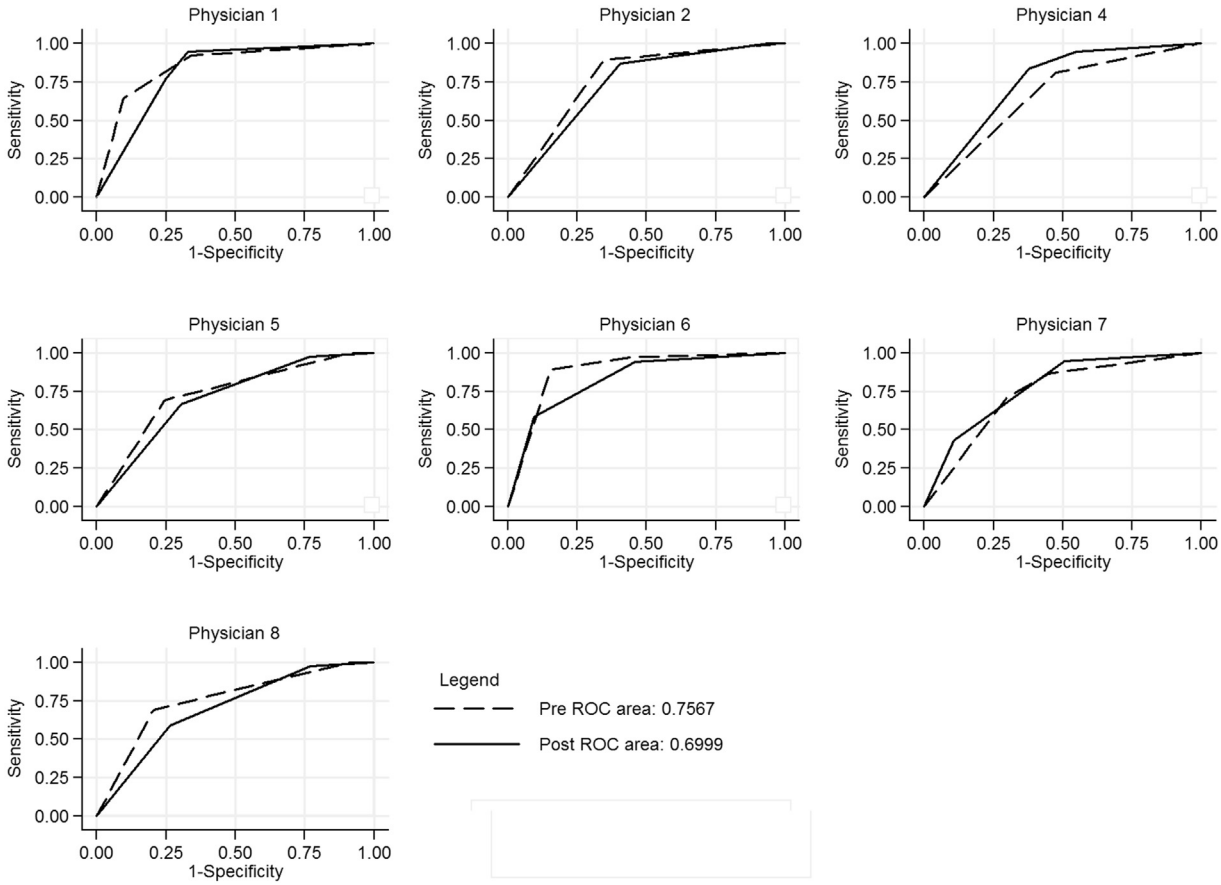


Figure 2 Non-parametric (empiric) ROC curves of non-radiologist physician CXR readings pre- and post-intervention.

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

Table 2
Sensitivity and specificity pre- and post- intervention by physician: matched data.

		Pre-intervention	Post-intervention	Exact McNemar *p*-value
NRP 1	Sensitivity	64.1%	76.9%	0.23
	Specificity	90.6%	75.3%	<0.01
NRP 2	Sensitivity	89.5%	86.8%	1.00
	Specificity	65.5%	59.5%	0.49
NRP 4	Sensitivity	81.1%	83.8%	1.00
	Specificity	52.4%	62.2%	0.19
NRP 5	Sensitivity	69.2%	66.7%	1.00
	Specificity	75.6%	69.5%	0.30
NRP 6	Sensitivity	88.9%	58.3%	0.01
	Specificity	84.7%	90.6%	0.30
NRP 7	Sensitivity	70.3%	43.2%	<0.01
	Specificity	70.6%	89.4%	<0.01
NRP 8	Sensitivity	69.2%	59.0%	0.34
	Specificity	79.5%	73.5%	0.33

NRP, non-radiologist physician.

Table 3
Pair-wise unweighted Cohen's kappa statistic comparing chest radiographs rated as "abnormal TB" or "normal/abnormal not TB".

Comparison		Pre-intervention		Post-intervention	
		Agreement %	Cohen's kappa	Agreement %	Cohen's kappa
Radiologist 2	vs radiologist 1	-	-	61.6	0.18
NRP 2	vs NRP 1	67.5	0.36	62.8	0.27
NRP 4	vs NRP 1	58.1	0.22	70.8	0.42
NRP 4	vs NRP 2	64.5	0.29	63.3	0.26
NRP 5	vs NRP 1	67.5	0.27	66.7	0.32
NRP 5	vs NRP 2	68.3	0.37	63.6	0.28
NRP 5	vs NRP 4	58.9	0.20	67.2	0.35
NRP 6	vs NRP 1	79.4	0.53	71.4	0.37
NRP 6	vs NRP 2	69.1	0.39	64.1	0.31
NRP 6	vs NRP 4	66.1	0.35	66.7	0.35
NRP 6	vs NRP 5	69.1	0.34	65.8	0.24
NRP 7	vs NRP 1	71.2	0.38	65.6	0.22
NRP 7	vs NRP 2	68.8	0.38	56.7	0.17
NRP 7	vs NRP 4	70.7	0.43	67.0	0.35
NRP 7	vs NRP 5	65.6	0.29	65.6	0.23
NRP 7	vs NRP 6	71.2	0.40	78.0	0.36
NRP 8	vs NRP 1	75.4	0.44	66.4	0.30
NRP 8	vs NRP 2	60.3	0.22	66.7	0.35
NRP 8	vs NRP 4	66.9	0.36	64.4	0.29
NRP 8	vs NRP 5	73.0	0.42	65.3	0.28
NRP 8	vs NRP 6	78.6	0.54	75.2	0.43
NRP 8	vs NRP 7	70.4	0.38	66.7	0.20
All physicians			0.35[a]		0.29[a]

TB, tuberculosis; NRP, non-radiologist physician; vs, versus.
[a] Fleiss kappa

agreement (kappa >0.5) and physicians 5 and 4, the lowest (kappa 0.2). The agreement between the first two radiologists on result/diagnosis was low (61.6%). The overall kappa statistic (Fleiss kappa) statistic showed poor agreement pre-intervention (κ=0.35) and even worse agreement post-intervention (κ=0.29; Table 3).

Agreement for specific disease pattern

Agreement for each specific disease pattern was assessed separately. A specific disease pattern was defined as present if one or more physicians or radiologists identified the disease pattern in at least one zone in the CXR. Agreement was defined if at least one radiologist and at least one physician identified a specific disease pattern as present. The agreement between each physician and the radiologist consensus, for specific disease patterns for subtle findings was generally poor but better for the more obvious disease patterns: pleural effusion (88.1–92.9%; κ range = 0.37–0.67) and miliary pattern (91.3–96.8%; κ range = 0.25–0.52; Table 4).

Discussion

CRRS was originally designed for prevalence studies and, by its design, does not lead the reader directly to a radiological diagnosis in order not to bias the final outcome.[11,12] When boxes are ticked for each particular disease presentation (e.g., infiltration, cavitation, and distribution zones), there is no intrinsic algorithm of disease patterns leading the user to a final radiological diagnosis. The user still has to subjectively decide the final radiological diagnosis based on his/her findings, experience, and knowledge. In the absence of a specific agreement between all readers, prior to using CRRS, on which findings constitute a positive radiological finding of TB, individual readers may agree on the disease presentation and distributable zone, but end up with a different final diagnosis. Alternatively, they could disagree on the individual disease findings yet still end up with the same final radiological diagnosis.

In the present study, most physicians had a reasonable level of sensitivity and specificity pre-intervention. No increase in accuracy was detected for the physician group as a whole. Other trials have produced mixed results: from limited improvement,[19] to improvement in sensitivity, but not specificity, for non-expert readers using a CXR image reference set rather than a CRRS tick sheet.[20]

The inter-reader agreement of the physicians decreased slightly post-intervention. Thus, in this setting, the introduction of the CRRS and a short training module actually had a negative impact on the inter-reader agreement. Previous studies that have examined agreement with the use of a CRRS showed varying results but none as poor as in the present study.[12,21]

The agreement between each physician and the radiologist consensus for specific disease patterns for subtle findings (i.e., broncho-pneumonic pattern, consolidation, and granuloma oval lesion) was generally poor, but the agreement was higher for the more obvious disease patterns (i.e., pleural effusion, miliary pattern). These results highlight the problem that physicians show little agreement on the specific radiological signs of TB and tend to over-diagnose subtle CXR findings compared to expert readers. The fact that CRRS has no intrinsic algorithm to guide the reader from individual disease patterns to the final radiological diagnosis likely adds to this issue and may in part explain why the intervention did not lead to an increase in accuracy, despite supporting a more systematic

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

Table 4
Reading agreement between each non-radiologist physician and radiologist per pathology.

Comparison	Lymph-adenopathy		Pleural effusion		Consolidation		Broncho-pneumonic pattern, infiltration		Cavitation		Granuloma oval lesion		Milliary nodular pattern		Apical fibrocystic change	
	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa
NRP 1 vs radiologist	82.5	0.12	92.9	0.67	69.8	0.10	76.2	0.22	84.1	0.28	91.3	−0.03	96.8	0.48	94.4	0.34
NRP 2 vs radiologist	25.4	0.01	89.7	0.48	60.3	0.22	39.7	0.10	92.1	0.34	87.3	0.04	96.8	0.32	93.7	0.00
NRP 4 vs radiologist	89.7	0.00	89.7	0.48	69.8	0.18	64.3	0.13	88.1	0.30	86.5	0.12	91.3	0.44	93.7	0.00
NRP 5 vs radiologist	74.6	0.18	88.9	0.58	68.3	0.09	44.4	0.12	95.2	0.38	91.3	0.11	92.1	0.25	92.1	−0.03
NRP 6 vs radiologist	58.7	0.08	89.7	0.53	71.4	0.16	73.8	0.25	96.0	0.43	85.7	0.02	96.0	0.52	93.7	0.30
NRP 7 vs radiologist	69.1	0.13	88.1	0.39	73.0	0.30	72.2	0.12	78.6	−0.01	92.9	0.16	96.0	0.52	93.7	0.00
NRP 8 vs radiologist	41.3	0.06	87.3	0.37	71.4	0.32	77.0	0.05	92.9	−0.03	92.9	0.16	96.8	0.48	95.2	0.48

NRP, non-radiologist physician; vs, versus.

approach to film interpretation. The high prevalence of TB in the study cohort seems to have led physicians to assume any abnormalities, real or perceived, being suggestive of TB. This could also explain the high sensitivity of the physicians.

Expert radiologists were used as the reference standard. More accurate TB diagnostic tools, such as TB culture or GeneXpert (MTB/Rif) tests, were not available at the study site, a common scenario in resource-constrained settings. The radiologists themselves had poor agreement and the poor quality of the CXRs no doubt contributed to the lack of agreement amongst all readers. CXR interpretation in TB diagnosis has well known limitations, even among experts with better-quality images, the results vary in the literature. Rarely do studies demonstrate an agreement more than "good" for overall diagnosis or specific disease patterns. Most studies, actually find only a "fair" to "moderate", sometimes "good" inter-reader agreements.[6,8,9,11,12,19–25]

A potential limitation of the present study was that the majority of the CXRs were marked as "not good but readable" by the expert radiologists, suggesting that outcomes may have been different if film quality had been better. Many of the images showed significant radiographic errors from poor acquisition and processing techniques, which is not uncommon in resource-constrained settings (Fig 3). The overall poor quality of CXRs compounds the difficulty of interpreting subtle findings in particular, which no doubt contributed to the lack of agreement between all readers including the experts; however, this unfortunately reflects the reality in many similar settings, and it is neither practical nor realistic in a study such as this to simply rate all images with errors as "poor and unreadable". It could therefore also be viewed as a strength of this study. It also reinforces the well-known limitations of poor-quality CXRs.

A further limitation was that the physicians viewed the original CXR on a light box, while the expert group reviewed digitised copies. This is both an advantage and disadvantage for readers; there is a reduction in resolution on the digitised copies, but digitised copies allow for computer manipulation to improve visualisation. Two previous studies comparing traditional film CXRs reporting to digitised copies suggest largely comparable results.[24,25]

In conclusion, accurate reporting of CXRs for TB is a complex and challenging task. This is especially true, even amongst expert readers, in HIV-positive patients with presumptive TB in resource-constrained settings where suboptimal CXR quality is common. In the present study, the impact of a short training session and the introduction of a CRRS did not show an increase in average overall interpretation accuracy among a group of non-radiologist physicians. It actually decreased the level of agreement post-intervention. The use of poor quality CXRs in TB diagnostic algorithms in such circumstances should be questioned: it could be adapted to focus only on identifying a normal film and easily recognisable disease patterns, such as pleural effusions, or else be discontinued.

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

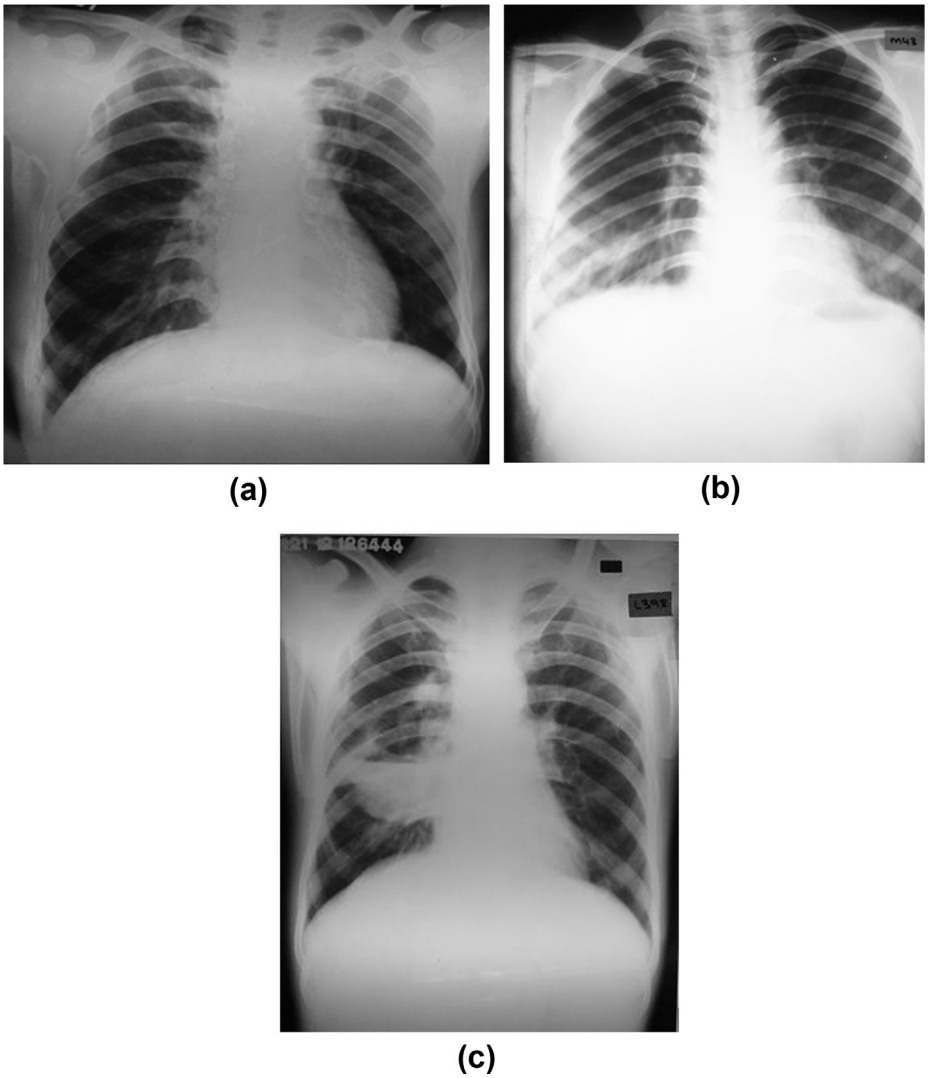


Figure 3 Suboptimal CXR image examples. (a) CXR L226 marked as "normal" by one radiologist and as "TB suggestive" by another radiologist. It was also rated in terms of image quality: "good", "not good but readable", and "poor" by three different radiologists. (b) CXR M43 marked "normal" and "TB suggestive" by two different radiologists. The third radiologists marked it too poor to evaluate. (c) CXR L398 was marked as a cavitation in distribution zone 4, and "TB suggestive" by the first two radiologists, both who also rated it as "not good but readable" in terms of image quality.

Acknowledgements

The authors would like to thank all participating non-radiologist physicians for undertaking the training and reading the images before and after the training. The authors thank also Dr Linn Htet and Dr Khin Nyien Chan for support during the study in Myanmar. The study was funded by MSF's Innovation Fund (Grant number: OR 09 07). The findings and conclusions of this report are those of the authors and do not necessarily represent the views of the finding agencies.

References

1. World Health Organization (WHO). *Global Tuberculosis Report 2014*. Geneva, Switzerland: World Health Organisation; 2014. Available at: http://apps.who.int/iris/bitstream/10665/137094/1/9789241564809_eng.pdf. Accessed 30 November 2016.
2. Lucian Davis J, Worodria W, Kisembo H, *et al*. Clinical and radiographic factors do not accurately diagnose smear-negative tuberculosis in HIV-infected inpatients in Uganda: a cross-sectional study. *PLoS One* 2010;**5**(3):1–8.
3. World Health Organization (WHO). *Automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF assay for the diagnosis of pulmonary and extrapulmonary TB in adults and children. Policy Update*. Geneva: World Health Organization; 2013. Available at: http://apps.who.int/iris/bitstream/10665/112472/1/9789241506335_eng.pdf. Accessed 30 November 2016.
4. World Health Organisation (WHO). *TB/HIV- A clinical manual 2nd Edition*. Geneva, Switzerland: World Health Organization; 2004. Available at: http://apps.who.int/iris/bitstream/10665/42830/1/9241546344.pdf. Accessed 30 November 2016.
5. Kivihya-Ndugga LEA, van Cleeff MRA, Ng'ang'a LW, *et al*. Sex-specific performance of routine TB diagnostic tests. *Int J Tuberc Lung Dis* 2005;**9**(3):294–300.
6. Van Cleeff MRA, Kivihya-Ndugga LE, Meme H, *et al*. The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. *BMC Infect Dis* 2005 Jan;**5**:111.
7. Toman K. *Tuberculosis case-finding and chemotherapy: questions and answers*. Geneva, Switzerland: World Health Organization; 1980.
8. Agizew TB, Arwady MA, Yoon JC, *et al*. Tuberculosis in asymptomatic HIV-infected adults with abnormal chest radiographs screened for tuberculosis prevention. *Int J Tuberc Lung Dis* 2010;**14**(1):45–51.

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008

9. Lam PK, LoBue PA, Catanzaro A. Clinical diagnosis of tuberculosis by specialists and non-specialists. *Int J Tuberc Lung Dis* 2009;**13**(5):659–61.
10. Pinto LM, Pai M, Dheda K, *et al*. Scoring systems using chest radiographic features for the diagnosis of pulmonary tuberculosis in adults: a systematic review. *Eur Respir J* 2013;**42**(2):480–94.
11. Den Boon S, Bateman ED, Enarson D, *et al*. Development and evaluation of a new chest radiograph reading and recording system for epidemiological surveys of tuberculosis and lung disease. *Int J Tuberc Lung Dis* 2005;**9**(10):1088–96.
12. Dawson R, Masuka P, Edwards DJ, *et al*. Chest radiograph reading and recording system: evaluation for tuberculosis screening in patients with advanced HIV. *Int J Tuberc Lung Dis* 2010;**14**(1):52–8.
13. World Health Organization (WHO). *Myanmar Tuberculosis profile*. 2012. Available at: https://extranet.who.int/sree/Reports?op=Replet&name=/WHO_HQ_Reports/G2/PROD/EXT/TBCountryProfile&ISO2=MM&outtype=html. Accessed 30 August 2012.
14. UNAIDS. *Global AIDS Response Progress Report Myanmar*. 2012. Available at: http://www.unaids.org/sites/default/files/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_MM_Narrative_Report.pdf. Accessed 30 August 2012.
15. Obuchowski NA. Computing sample size for receiver operating characteristic studies. *Invest Radiol* 1994;**29**(2):238–43.
16. Spijker S, Andronikou S, Kosack C, *et al*. Quality assessment of X-rays interpreted via teleradiology for Médecins Sans Frontières. *J Telemed Telecare* 2014;**20**:82–8.
17. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 1992 Sep;**27**(9):723–31.
18. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;**85**:257–68.
19. Seddon JA, Padayachee T, Du Plessis M, *et al*. Teaching chest X-ray reading for child tuberculosis suspects. *Int J Tuberc Lung Dis* 2014;**18**(7):763–9.
20. Waitt CJ, Joekes EC, Jesudason N, *et al*. The effect of a tuberculosis chest X-ray image reference set on non-expert reader performance. *Eur Radiol* 2013;**23**(9):2459–68.
21. Balabanova Y, Coker R, Fedorin I, *et al*. Variability in interpretation of chest radiographs among Russian clinicians and implications for screening programmes: observational study. *BMJ* 2005;**331**(7513):379–82.
22. Du Toit G, Swingler G, Iloni K. Observer variation in detecting lymphadenopathy on chest radiography. *Int J Tuberc Lung Dis* 2002;**6**(9):814–7.
23. van't Hoog AH, Meme HK, Deutekom H Van, *et al*. High sensitivity of chest radiograph reading by clinical officers. *Int J Tuberc Lung Dis* 2011;**15**(10):1308–14.
24. Corr P, Couper I, Beningfield SJ, Mars M. A simple telemedicine system using a digital camera. *J Telemed Telecare* 2000 Jan;**6**(4):233–6.
25. Szot A, Jacobson FL, Munn S, *et al*. Diagnostic accuracy of chest X-rays acquired using a digital camera for low-cost teleradiology. *Int J Med Inform* 2004;**73**(1):65–73.

Please cite this article in press as: Kosack CS, et al., Evaluation of a chest radiograph reading and recording system for tuberculosis in a HIV-positive cohort, Clinical Radiology (2017), http://dx.doi.org/10.1016/j.crad.2017.01.008