INT J TUBERC LUNG DIS 21(8):880–886

http://dx.doi.org/10.5588/ijtld.16.0851

Automatic versus human reading of chest X-rays in the Zambia National Tuberculosis Prevalence Survey

J. Melendez,*† R. H. H. M. Philipsen,*† P. Chanda-Kapata,‡§ V. Sunkutu,¶ N. Kapata,‡§ B. van Ginneken*†

*Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, †Thirona, Nijmegen, The Netherlands; ‡Ministry of Health Headquarters, Lusaka, Zambia; §Centre for Tropical Medicine and Travel Medicine, Department of Infectious Diseases, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; ¶University Teaching Hospital, Lusaka, Zambia

SUMMARY

SETTING: Tuberculosis (TB) prevalence survey in Zambia between 2013 and 2014.

OBJECTIVE: To compare the performance of automatic software (CAD4TB 5) in chest X-ray (CXR) reading with that of field (general practitioners) and central (radiologists) readers.

DESIGN: A retrospective study comparing the performance of human and automatic reading was conducted. Two scenarios for central reading were evaluated: abnormalities not consistent with TB were considered to be 'normal' or 'abnormal'. Sputum culture was defined as the reference standard. Measures derived from receiver operating characteristic analysis were used to assess readers' performances.

RESULTS: Of 46 099 participants, 23 838 cases included all survey information; of these, 106 cases were culture-confirmed TB-positive. The performance of CAD4TB 5 was similar to that of field and central readers. Although there were significant differences in specificity when compared with field readings ($P = 0.002$) and central readings considering the first scenario ($P < 0.001$), these differences were not substantial (93.2% vs. 92.6% and 98.4% vs. 99.6%, respectively).

CONCLUSION: The performance of automatic CXR readings is comparable with that of human experts in a TB prevalence survey setting using culture as reference.

KEY WORDS: TB; prevalence survey; automatic chest radiograph reading

TUBERCULOSIS (TB) is a major cause of mortality and morbidity worldwide, and the epidemic is larger than previously estimated. In 2015, 10.4 million people fell ill with TB, and 1.8 million people died, with most of the disease burden in middle- and low-income countries.[1] The best methods to measure TB burden are national TB prevalence surveys. Between 2009 and August 2016, 22 surveys had been completed, and five were planned for 2017.[1]

TB prevalence surveys are considerable undertakings, with samples of up to 100 000 screened persons. Clusters of up to 1000 adults must be screened for TB within short time frames of 1–2 weeks. Furthermore, multiple clusters need to be screened in parallel. To be successful, the logistics of such programmes must be optimised. Most prevalence survey teams consist of a field team and a central team. The former performs, among other tasks, symptom questionnaires, chest X-ray (CXR) acquisition and reading; the latter often includes CXR reading by radiologists and culture testing.[2]

Field CXR reading is usually done by clinical/medical officers. However, several countries lack well-trained personnel;[3] if available, these personnel may be costly, particularly in Africa, and screening high numbers of subjects may lead to reader fatigue.[4,5] In the present study, we validated the use of automatic CXR reading software, CAD4TB 5, in the context of the Zambia National TB Prevalence Survey. This software can automatically process CXRs and thus reduce the workload and costs of human readers, both in the field and at the central level. We retrospectively compared automatic CXR reading with field and central human readings using culture as the reference standard.

METHODS

Ethics statement

The Zambia Biomedical Research Ethics Committee (UNZABREC), Lusaka, Zambia, gave ethical ap-

JM and RHHMP contributed equally to this work.

Correspondence to: Jaime Melendez, Diagnostic Image Analysis Group, Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Internal Postal Code 766, Radiology, Postbus 9101, 6500 HB Nijmegen, The Netherlands. e-mail: Jaime.MelendezRodriguez@radboudumc.nl

Article submitted 22 November 2016. Final version accepted 6 April 2017.

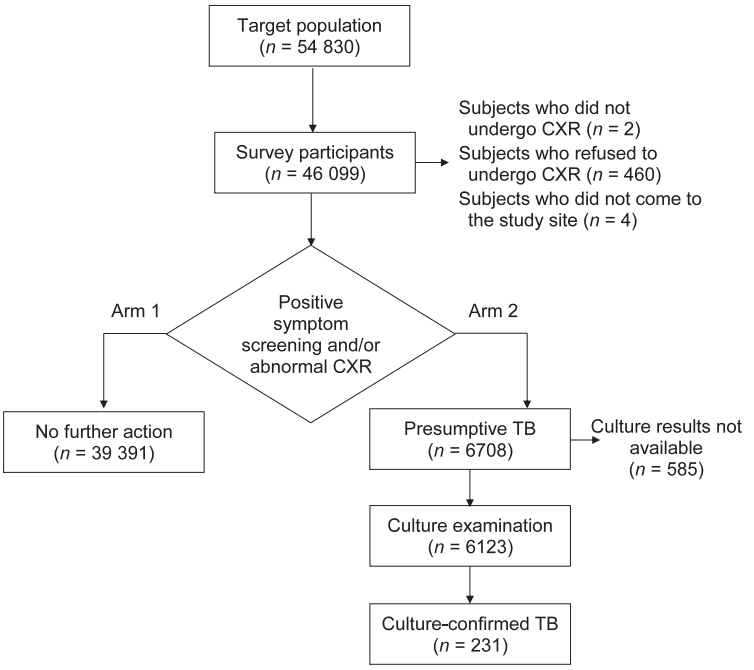


Figure 1 Simplified flowchart of the prevalence survey (adapted from Kapata et al.[6]). CXR = chest X-ray.

proval for the study. All participants provided written informed consent.

Survey and study design

The Zambia National TB Prevalence Survey, a countrywide survey including rural and urban areas, was carried out between September 2013 and July 2014. Although the target sample size was 54 830 adults aged ⩾15 years, the actual number of participants was 46 099. Data were collected across 66 clusters in all provinces. Participants were screened for TB symptoms using a structured questionnaire. Symptom screening was positive if the participant had at least one of cough, fever or chest pain for ⩾2 weeks. Subjects also underwent posterior-anterior (PA) CXR using direct digital radiography (EasyDR X-ray system; Delft Imaging Systems, Veenendaal, The Netherlands). The survey screening protocol followed World Health Organization (WHO) guidelines,[2] and has been described by Kapata et al.,[6] who reported that 265 TB cases had been identified and it was estimated that the TB prevalence for all forms and age groups was 455 per 100 000.

A simplified flowchart of the prevalence survey is shown in Figure 1. If a participant had a negative symptom screen and a normal CXR, no further action was taken (Arm 1). Those who screened positive for symptoms and/or had abnormalities suggestive of TB on their CXR were asked to submit two sputum specimens (one on the spot and one the next morning) for bacteriological examinations (Arm 2). Sputum samples were processed at any of three central reference laboratories. Positive *Mycobacterium tuberculosis* growth was confirmed using acid-fast Ziehl-Neelsen staining and Capillia Test (Tauns Laboratories, Shizuoka, Japan).[6]

In the present study, a TB case was defined as a patient with positive culture results for *M. tuberculosis*. A non-TB case was defined as a patient sent to Arm 1 (no symptoms and normal CXR on field reading) or a case sent to Arm 2 (with symptoms and/or abnormal CXR on field reading), but no *M. tuberculosis* detected on culture.

Chest radiograph reading

Field readings were carried out by general practitioners with at least 2 years of post-qualification experience. They classified radiological findings as 'normal', 'abnormal (suggestive of TB)', 'other abnormalities (not suggestive of TB)' and 'not interpretable', and were encouraged to over-read the CXRs to increase sensitivity.

Central readings were performed by one of four specialised radiologists at the central radiology unit of the University Teaching Hospital, Lusaka, Zambia. All the radiologists possessed postgraduate-level qualifications and had more than 10 years of experience in reading X-rays. The central readers categorised all the CXRs as 'normal', 'abnormal consistent for TB' or 'abnormal inconsistent for TB'. They were blinded to laboratory results but not to field readings.

Automatic CXR readings were performed using CAD4TB 5 (Thirona, Nijmegen, The Netherlands), which analyses the unobscured lung fields of a PA

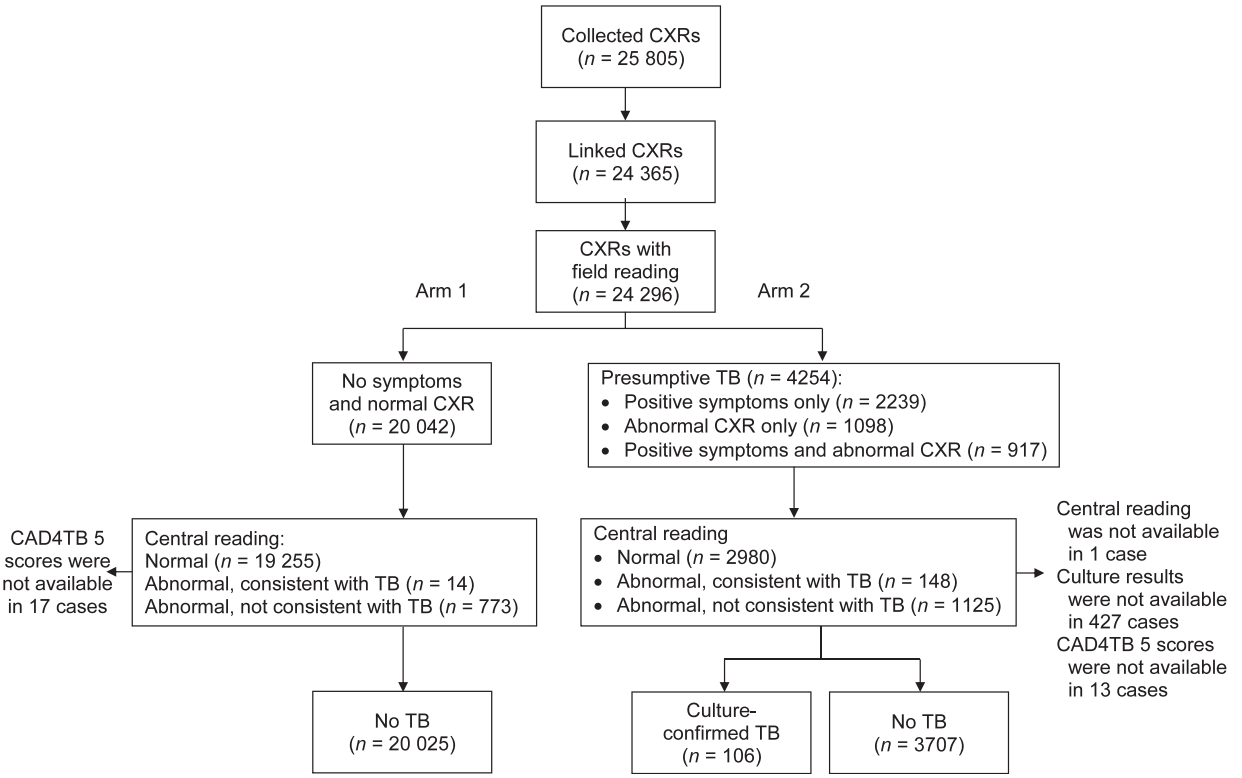


Figure 2 Stratification of the cases included in this study. CXR = chest X-ray; TB = tuberculosis.

CXR for the presence of abnormalities and outputs a TB score of between 0 and 100. Computation time is less than 1 min on a standard personal computer. The system's main method is 'supervised classification', in which a classifier is trained with labelled examples of normal and abnormal regions. In the test phase, the trained classifier processes previously unseen images and assigns them a likelihood of belonging to one of the learned classes. CAD4TB 5 uses a more powerful classifier (a support vector machine) than other CAD4TB versions reported in the literature.[7–9]

CAD4TB 5 consists of two main components: quality check and TB analysis. The quality check component assesses whether the input is an appropriate PA CXR. Prior to quality assessment, energy-based normalisation is applied to reduce scanner-related differences among images.[10] Valid PA CXRs continue to the next component. The TB analysis component starts by automatically segmenting the unobscured lung fields,[11] with the aim of restricting further analysis to this area and providing an anatomical context for abnormality detection. 'Abnormalities' are broadly defined as textural changes in the appearance of the lung parenchyma because of disease. As mentioned above, a classifier is applied to highlight these changes. Its output is a heat map indicating the likelihood that a pixel belongs to an abnormal region. The pixel likelihoods are then summarised into a single score by applying a quantile rule.[12] Apart from textural analysis, assessment of the shape of the segmented lung fields is carried out. The rationale is that large abnormalities may affect this feature. The final TB score is a result of fusing the outputs of the texture and shape analysis subcomponents.

Data analysis and evaluation

Receiver operating characteristic (ROC) analysis was performed on the TB scores generated by automatic reading. The area under the ROC curve (AUC) was computed from the raw curve (i.e., without fitting). AUC 95% confidence intervals (CIs) were calculated using the DeLong method.[13] Detection performance of field and central readings was evaluated by calculating sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) with 95%CIs. Field readings were categorised as 'abnormal' if abnormalities suggestive of TB were indicated, and as 'normal' otherwise. Central readings were categorised as normal or abnormal using the following criteria: 1) considering 'abnormal consistent for TB' readings as abnormal and the rest as normal; and 2) considering 'abnormal consistent for TB' and 'abnormal inconsistent for TB' readings as abnormal, and the rest as normal. With these criteria, two sets of performance measures were obtained for the central readings. When evaluating the performance of automatic readings, specificity, PPV and NPV were

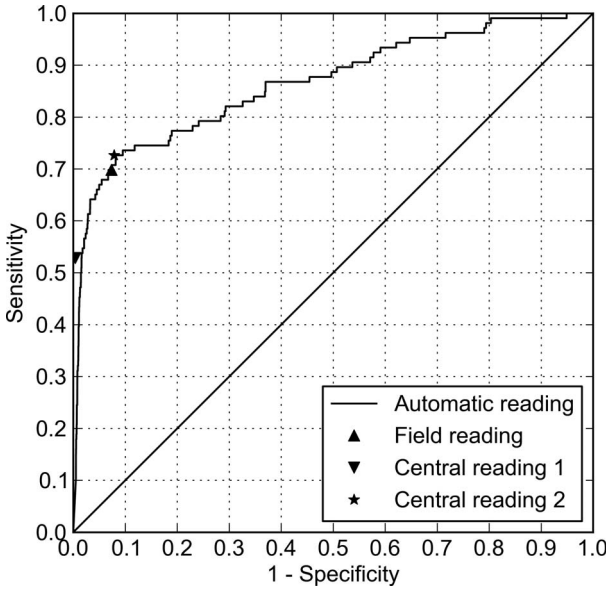


Figure 3 ROC curve yielded by automatic reading and operating points corresponding to field and central reading, the latter using two definitions of an abnormal CXR. Central reading 1 refers to abnormalities not consistent with TB being considered normal, whereas Central reading 2 refers to abnormalities not consistent with TB being considered abnormal. ROC = receiver operating characteristic; CXR = chest X-ray.

computed at cut-off points corresponding to the sensitivity of the field and central readings. Significant differences in specificity were determined using the McNemar test ($P < 0.05$). Data were processed using R v3.3.1 (R Foundation for Statistical Computing, Vienna, Austria) with extension packages 'pROC',[14] 'epiR'[15] and 'stats'.

RESULTS

Of the 46 099 participants in the prevalence survey, 45 633 (99.0%) underwent a CXR (Figure 1). Of the 46 099 participants, 6708 (14.6%) were identified as presumptive TB cases and were eligible for culture examination. Culture results were available for 6123 (91.3%) of the 6708 presumptive TB cases. In total, 231 instances were considered positive in accordance with the definition given above.

Figure 2 shows stratification of the CXRs gathered from the survey. From the 45 633 CXRs taken, 25 805 CXRs could be collected (the remainder were lost because of data-extraction issues) and 24 365 CXRs could be linked with survey information. Moreover, 69 CXRs had no field readings available. Of the remaining 24 296 CXRs, 4254 (17.5%) corresponded to subjects sent to Arm 2 and 20 042 (82.5%) to subjects sent to Arm 1. Among the 4254 subjects sent to Arm 2, 2239 (52.6%) had positive symptoms but a normal CXR, 1098 (25.8%) had no symptoms but an abnormal CXR and 917 (21.6%) had both positive symptoms and an abnormal CXR.

Table 1 Demographics of the survey data included in this study

Characteristic	*n* (%)
Total	23 838
Sex	
Male	10 440 (43.8)
Female	13 398 (56.2)
Age, years, mean ± SD	36.0 ± 17.0
Cough	
No	17 904 (75.1)
1 week	4 495 (18.9)
2 weeks	1 439 (6.0)
Fever	
No	19 750 (82.8)
1 week	3 402 (14.3)
2 weeks	686 (2.9)
Chest pain	
No	16 360 (68.6)
1 week	5 285 (22.2)
2 weeks	2 193 (9.2)
Bacteriological findings (culture)	3 813 (16.0)
No TB	3 707 (97.1)
TB	106 (2.9)
Central reading	
Normal	21 888 (91.8)
Abnormal, consistent with TB	149 (0.6)
Abnormal, not consistent with TB	1 801 (7.6)
Field reading	
Normal	21 952 (92.1)
Abnormal	1 814 (7.6)
Other abnormalities	63 (0.3)
Not interpretable	9 (0.0)
Symptom screening	
Negative	21 006 (88.1)
Positive	2 832 (11.9)

SD = standard deviation; TB = tuberculosis.

Only 148 CXRs in Arm 2 (3.5%) were categorised as 'abnormal consistent for TB' by central reading, which suggested over-reading by the field readers. Culture results were not available for 427 cases in Arm 2. In addition, central reading was missing for one CXR and the quality check component of CAD4TB 5 rejected 30 CXRs. These cases were excluded from our analysis. Evaluation was thus carried out with 23 838 CXRs: 23 732 were catego-

Table 2 Performance results of automatic reading at different thresholds on the TB score

Threshold	Sensitivity %	Specificity %	PPV %	NPV %
99	10.4	99.5	8.5	99.6
90	23.6	99.3	13.4	99.7
80	34.9	99.0	13.8	99.7
70	48.1	98.5	12.8	99.8
60	57.5	97.5	9.4	99.8
50	67.9	94.0	4.8	99.8
40	73.5	88.4	2.7	99.9
30	78.3	76.2	1.4	99.9
20	96.2	26.3	0.6	99.9
10	100.0	0.0	0.4	—
0	100.0	0.0	0.4	—

TB = tuberculosis; PPV = positive predictive value; NPV = negative predictive value.

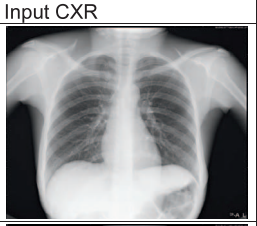
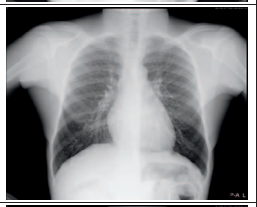
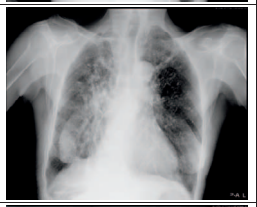
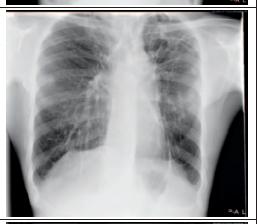
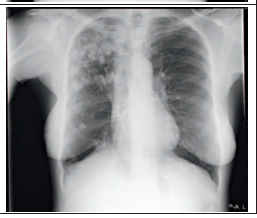
	Input CXR	Reference standard and readings
A		Non-TB case Field reading: normal Central reading: normal Automatic reading (TB score): 13.7/100
B		Culture-confirmed TB case Field reading: normal Central reading: normal Automatic reading (TB score): 22.8/100
C		Culture-confirmed TB case Field reading: abnormal Central reading: abnormal inconsistent for TB Automatic reading (TB score): 86.6/100
D		Non-TB case Field reading: abnormal Central reading: abnormal inconsistent for TB Automatic reading (TB score): 89.5/100
E		Culture-confirmed TB case Field reading: abnormal Central reading: abnormal consistent for TB Automatic reading (TB score): 95.5/100

Figure 4 Examples of input CXRs and their reference standard, human readings and automatic TB score. CXR = chest X-ray; TB = tuberculosis.

rised as normal (associated with non-TB cases) and 106 categorised as abnormal (associated with culture-confirmed TB cases). Table 1 shows the demographics of the corresponding participants.

Automatic reading

The ROC curve computed from the scores assigned by automatic reading of CXRs is shown in Figure 3. The AUC was 0.87 (95%CI 0.82–0.91). Table 2 lists the performance of the software at various thresholds: a low threshold led to high sensitivity at the expense of low specificity and vice versa. Examples of CXRs and TB scores assigned by automatic reading are shown in Figure 4.

Automatic reading vs. field and central readings

Figure 3 shows that the operating points of central and field readings lay very close to the ROC curve yielded by automatic reading, which indicates that either alternative performed comparably. Table 3 lists the sensitivity, specificity, PPV and NPV with 95%CIs for automatic, field and central readings. Sensitivity and specificity for central reading were substantially affected by whether readings categorised as 'abnormal inconsistent for TB' were considered to be normal or abnormal. Automatic reading achieved a specificity of 93.2% (95%CI 92.9–93.6) at a field reading sensitivity of 69.8%, with a slight but significant difference with respect to a field reading specificity of 92.6% ($P = 0.002$). Similar specificities were observed on comparing automatic reading with central reading at the two sensitivity points of 52.8% and 72.6%. At the first point, where 'abnormal inconsistent for TB' readings were considered to be normal, the specificity of central readings was 99.6%, whereas that of automatic readings was 98.4% ($P < 0.001$). At the second point, where 'abnormal inconsistent for TB' readings were considered to be

Table 3 Sensitivity, specificity, PPV and NPV by field, central and automatic readings*

	Sensitivity % (95%CI)	Specificity % (95%CI)	PPV % (95%CI)	NPV % (95%CI)
Field reading	69.8 (60.1–78.3)	92.6 (92.3–93.0)	4.1 (3.2–5.1)	99.9 (99.8–99.9)
Central reading				
'Abnormal, not consistent with TB' as normal	52.8 (42.9–62.6)	99.6 (99.5–99.7)	3.8 (3.0–4.6)	99.8 (99.7–99.8)
'Abnormal, not consistent with TB' as abnormal	72.6 (63.1–80.9)	92.1 (91.8–92.4)	3.9 (3.1–4.9)	99.9 (99.8–99.9)
Automated reading				
At field reading sensitivity (69.8%)	69.8 (60.1–78.3)	93.2 (92.9–93.6)	4.4 (3.5–5.5)	99.9 (99.8–99.9)
At central reading sensitivity (52.8%; abnormal, not consistent with TB' as normal)	52.8 (42.9–62.6)	98.4 (98.3–98.6)	13.0 (10.0–16.5)	99.8 (99.7–99.8)
At central reading sensitivity (72.6%; 'abnormal, not consistent with TB' as abnormal)	72.6 (63.1–80.9)	91.8 (91.5–92.2)	3.8 (3.0–4.8)	99.9 (99.8–99.9)

* Scores obtained by automatic reading were thresholded to obtain the same sensitivities as those of field and central readings to allow direct comparison of their specificities. Central reading included two operating points: 'abnormal, not consistent with TB' readings considered to be normal, and 'abnormal, not consistent with TB' readings considered to be abnormal.
PPV = positive predictive value; NPV = negative predictive value; CI = confidence interval; TB = tuberculosis.

abnormal, the specificity of central readings was 92.1%, whereas that of automatic readings was 91.8% ($P = 0.1492$).

DISCUSSION

The ROC analysis showed that automatic, field and central readings scored CXRs in a similar way. The operating points of human readers lay very close to the ROC curve of automatic reading, which suggested that the software could achieve comparable sensitivities and specificities after application of suitable thresholds to its output. The selection of a threshold should be based on the operating conditions. It is expected that different thresholds would be applied, for example, in a passive case finding scenario and in a prevalence survey. This selection may also depend on the availability of a diagnostic test. If the availability of the diagnostic test (e.g., sputum culture or molecular diagnostics such as Xpert® MTB/RIF [Cepheid, Sunnyvale, CA, USA]) is limited, a threshold leading to high specificity could be appropriate. On the contrary, in a prevalence survey, where the aim is to obtain high sensitivity, a low threshold could be set.

Given that our results indicate that the performance of automatic reading is comparable to that of human reading, it is envisioned that the former could play a major part, for example, in large-scale TB detection endeavours, such as prevalence surveys. Similar conclusions were reached in a related study.[16] In addition, as stated previously,[7,17] the consistent scores provided by automatic reading can contribute to the elimination of well-known issues in human CXR interpretation, such as inter- and intra-observer variability.

Although automatic reading performed well in this and other studies,[7–9,16] some aspects could be improved. Inspection of culture-negative cases with high scores (potential false-positive readings) and culture-positive cases with low scores (potential false-negative readings) provide insights into this matter. A problem in the first situation is the system's inability to differentiate between TB-related abnormalities and abnormalities related to other diseases (Figure 4D). As this image was correctly categorised by central reading as a non-TB instance despite the lesions depicted, it is reasonable to expect that such knowledge could be incorporated into the automatic system. However, the extent to which this would be realisable seems limited, as such differentiation might not always be possible, as shown in Figure 4C, where central reading erroneously categorised the CXR of a culture-confirmed case as 'abnormal inconsistent for TB'. It is possible that the limit would be determined by the radiographic test itself and its restricted specificity for TB detection. In the case of false-negative readings, the issue is the unusual radiological manifestations of TB or their complete absence. This feature was a particularity of the current study, leading to relatively low sensitivity also for human readings. An example is shown in Figure 4B, in which both human readings categorised this CXR as 'normal' and the score assigned by automatic reading was low. For this, and perhaps the problem highlighted above, the addition of clinical information to automatic CXR reading may be helpful.[18]

Our study had four main imitations. First, culture examination was not performed for all participants. Thus, the CXRs of Arm 1 participants were considered normal based on symptoms and CXR screening without a confirmatory test. This may have introduced a bias in favour of field reading, as their readings consequently became correct for 84% of the CXRs. Unfortunately, this was a limitation of the survey protocol itself and could not be corrected retrospectively. Second, central readers were not blinded to field readings, and this may have biased their results, although, given the large difference in the cases read as abnormal (consistent for TB) by both groups, such an outcome seems unlikely. Third, detection performance was evaluated on a subset of the CXRs acquired during the survey, as not all of them were available. However, although this may have had an effect on our conclusions from a prevalence survey evaluation perspective, the results obtained with the collected sample still supported the main idea developed throughout the study, i.e., that automatic reading can perform comparably to human experts. Fourth, our study provided evidence for a single population only. To reliably assess the capabilities of automatic reading, further studies involving populations with different TB incidences should be conducted. This has also been recognised by other researchers[19] and the WHO,[20] which encourages research on automatic reading following a protocol that contributes to gathering the required evidence.

In conclusion, the performances of automatic and human CXR reading when screening for TB in a prevalence survey setting are comparable. This suggests that it would be suitable to use automatic CXR reading and symptom screening in TB prevalence surveys.

Acknowledgements

The authors thank B C M'ule for undertaking the extraction and archiving of the images used for our study.

Field-level data collection was funded by the Government of The Republic of Zambia (Lusaka, Zambia) through the Ministry of Health and The United States Government through the US Agency for International Development/Centers for Disease Control and Prevention/TBCARE I. No funding was obtained for the post-field automated analysis for this work. The contents of this manuscript are the sole responsibility of the authors, and the funding body had no role in the decision to publish our study.

Conflicts of interest: JM and RP are employees of Thirona. BvG is co-founder and shareholder of Thirona, which developed the CAD4TB software. The other authors report no conflicts.

References

1 World Health Organization. Global tuberculosis report, 2016. WHO/HTM/TB/2016.13. Geneva, Switzerland: WHO, 2016.
2 World Health Organization. Tuberculosis prevalence surveys: a handbook. Geneva, Switzerland: WHO, 2011.
3 Kawooya M G. Training for rural radiology and imaging in Sub-Saharan Africa: addressing the mismatch between services and population. J Clin Imaging Sci 2012; 2: 37.
4 Lee C S, Nagy P G, Weaver S J, Newman-Toker D E. Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol 2013; 201: 611–617.
5 Krupinski E A, Berbaum K S, Caldwell R T, Schartz K M, Kim J. Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol 2010; 7: 698–704.
6 Kapata N, Chanda-Kapata P, Ngosa W, et al. The prevalence of tuberculosis in Zambia: results from the first national TB prevalence survey, 2013–2014. PLOS ONE 2016; 11: e0146392.
7 Muyoyeta M, Maduskar P, Moyo M, et al. The sensitivity and specificity of using a computer aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka Zambia. PLOS ONE 2014; 9: e93757.
8 Maduskar P, Muyoyeta M, Ayles H, et al. Detection of tuberculosis with digital chest radiography: automatic reading versus interpretation by clinical officers. Int J Tuberc Lung Dis 2013; 17: 1613–1620.
9 Breuninger M, van Ginneken B, Philipsen R H H M, et al. Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. PLOS ONE 2014; 9:e106381.
10 Philipsen R H H M, Maduskar P, Hogeweg L, Melendez J, Sánchez C I, van Ginneken B. Localized energy-based normalization of medical images: application to chest radiography. IEEE Trans Med Imaging 2015; 34: 1965–1975.
11 van Ginneken B, Stegmann M B, Loog M. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Med Image Anal 2006; 10: 19–40.
12 Loog M, van Ginneken B. Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest radiographs. In: Proceedings of the 17th International Conference on Pattern Recognition, 23–26 August 2004, Cambridge, UK: pp 644–647.
13 DeLong E R, DeLong D M, Clarke-Pearson D L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44: 837–845.
14 Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12: 77.
15 Stevenson M, Nunes T, Heuer C, et al. epiR: an R package for the analysis of epidemiological data. Vienna, Austria: R Computing, 2013.
16 Maduskar P. Automated analysis of tuberculosis in chest radiographs. PhD Thesis. Nijmegen, The Netherlands: Radboud University, 2015.
17 van't Hoog A H, Langendam M W, Mitchell E, et al. A systematic review of the sensitivity and specificity of symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative persons and persons with unknown HIV status. Geneva, Switzerland: World Health Organization, 2013.
18 Melendez J, Sánchez C I, Philipsen R H H M, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. Sci Rep 2016; 6: 25265.
19 Pande T, Cohen C, Pai M, Ahmad Khan F. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. Int J Tuberc Lung Dis 2016; 20: 1226–1230.
20 World Health Organization. Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. WHO/HTM/TB/2016. 20. Geneva, Switzerland: WHO, 2016.

RESUME

CONTEXTE : Enquête de prévalence de la tuberculose (TB) en Zambie entre 2013 et 2014.

OBJECTIF : Comparer la performance de la lecture de la radiographie pulmonaire (CXR) par un logiciel automatique (CAD4TB 5) avec celle du terrain (médecins généralistes) et celle réalisée au niveau central (radiologues).

SCHÉMA : Une étude rétrospective comparant la performance de la lecture humaine et automatique a été réalisée. En ce qui concerne la lecture centrale, deux scénarios ont été évalués : les anomalies non-cohérentes avec la TB ont été considérées soit comme normales soit comme anormales. La culture de crachats a été définie comme le standard de référence. Les mesures dérivées de l'analyse de la courbe de la fonction d'efficacité du récepteur ont été utilisées pour évaluer la performance des lecteurs.

RÉSULTATS : Sur 46 099 participants, 23 838 cas ont inclus toutes les informations relatives à l'enquête et parmi eux, 106 cas ont été des TB confirmées par culture. La performance du CAD4TB 5 a été similaire à celle des lecteurs du terrain et du centre. Bien qu'il y ait eu des différences significatives en termes de spécificité en ce qui concerne la comparaison de la lecture de terrain ($P = 0{,}002$) et de la lecture centrale considérant le premier scénario ($P < 0{,}001$), ces différences n'ont pas été substantielles (93,2% contre 92,6% et 98,4% contre 99,6%, respectivement).

CONCLUSION : La lecture automatique des CXR a une performance comparable à celle des experts humains dans le cadre d'une enquête de prévalence de la TB utilisant la culture comme référence.

RESUMEN

MARCO DE REFERENCIA: La encuesta de prevalencia de tuberculosis (TB) realizada en Zambia del 2013 al 2014.

OBJETIVO: Comparar la eficacia de la lectura automática de las radiografías de tórax (CXR) mediante una programa informático (CAD4TB 5), con la lectura en el terreno (practicada por médicos generales) y la lectura central (practicada por radiólogos).

MÉTODO: Fue este un estudio retrospectivo de comparación de la lectura CXR automática y humana. En la lectura central se evaluaron dos hipótesis, a saber: las radiografías con anomalías no indicativas de TB se consideraban normales o se consideraban anormales. Se adoptó el cultivo de esputo como patrón de referencia. Se utilizaron las medidas derivadas del análisis de las características funcionales con el fin de analizar el desempeño de los lectores.

RESULTADOS: De los 46 099 participantes, 23 838 contaban con toda la información de la encuesta y de estos, 106 casos presentaron una TB confirmada mediante cultivo. El desempeño del programa CAD4TB 5 fue equivalente al de la lectura en el terreno y la lectura central. Aunque se observaron diferencias significativas en materia de especificidad, cuando se comparó la lectura automática con la lectura en el terreno ($P = 0{,}002$) y con la lectura central ($P < 0{,}001$) con respecto a la primera hipótesis, estas diferencias no fueron importantes (93,2% contra 92,6% y 98,4% contra 99,6%, respectivamente).

CONCLUSIÓN: La lectura de CXR automática ofreció un rendimiento comparable al de la lectura humana por expertos, en el contexto de una encuesta de prevalencia de TB, al adoptar el cultivo como norma de referencia.