



An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients

Md Toufiq Rahman¹, Andrew J. Codlin², Md Mahfuzur Rahman¹, Ayenun Nahar¹, Mehdi Reja¹, Tariqul Islam³, Zhi Zhen Qin², Md Abdus Shakur Khan⁴, Sayera Banu¹ and Jacob Creswell²

Affiliations: ¹International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh. ²Stop TB Partnership, Geneva, Switzerland. ³National Institute of Neurosciences and Hospital, Dhaka, Bangladesh. ⁴National Institute of Diseases of the Chest & Hospital, Dhaka, Bangladesh.

Correspondence: Jacob Creswell, Stop TB Partnership, Chemin de Blandonnet 2, Geneva, Switzerland 1214. E-mail: jacobc@stoptb.org



@ERSpublications

Automated CXR reading reduces expensive TB test use; differentiated scoring may be needed and costs must be reduced <http://ow.ly/jPC8309Rx3F>

Cite this article as: Rahman MT, Codlin AJ, Rahman MM, *et al.* An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients. *Eur Respir J* 2017; 49: 1602159 [<https://doi.org/10.1183/13993003.02159-2016>].

ABSTRACT Computer-aided reading (CAR) of medical images is becoming increasingly common, but few studies exist for CAR in tuberculosis (TB). We designed a prospective study evaluating CAR for chest radiography (CXR) as a triage tool before Xpert MTB/RIF (Xpert).

Consecutively enrolled adults in Dhaka, Bangladesh, with TB symptoms received CXR and Xpert. Each image was scored by CAR and graded by a radiologist. We compared CAR with the radiologist for sensitivity and specificity, area under the receiver operating characteristic curve (AUC), and calculated the potential Xpert tests saved.

A total of 18036 individuals were enrolled. TB prevalence by Xpert was 15%. The radiologist graded 49% of CXRs as abnormal, resulting in 91% sensitivity and 58% specificity. At a similar sensitivity, CAR had a lower specificity (41%), saving fewer (36%) Xpert tests. The AUC for CAR was 0.74 (95% CI 0.73–0.75). CAR performance declined with increasing age. The radiologist grading was superior across all sub-analyses.

Using CAR can save Xpert tests, but the radiologist's specificity was superior. Differentiated CAR thresholds may be required for different populations. Access to, and costs of, human readers must be considered when deciding to use CAR software. More studies are needed to evaluate CAR using different screening approaches.

This article has supplementary material available from erj.ersjournals.com

Received: Nov 03 2016 | Accepted after revision: Jan 09 2017

Support statement: This study was supported by the Stop TB Partnership's TB REACH initiative and was funded by the Government of Canada and UNITAID. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

Copyright ©ERS 2017. This ERJ Open article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

Introduction

Modelling and research suggest that chest radiography (CXR) can be a good screening tool to identify people in need of further testing to diagnose tuberculosis (TB), especially when the follow-on diagnostic test is expensive [1, 2]. However, although diagnostic testing for TB in high burden settings is generally free, the cost of CXR is almost always borne by the care seeker and can be quite costly [3, 4]. In many settings there is also a dearth of trained staff who can quickly read CXR images, and there is a high degree of inter-reader variability, resulting in both over- and under-diagnosis of TB [5, 6]. These limitations have generated interest in scoring systems for chest radiography [7] as well as a growing awareness of, interest in, and use of computer-aided reading (CAR) software.

CAR software can quickly evaluate digital radiographs to recognise abnormalities in lungs fields to identify people in need of further diagnostic testing. CAR software is already commonplace in mammography services [8], but CAR for identifying TB abnormalities on chest radiographs is still in the development stage. All of the published literature for CAR for TB has evaluated CAD4TB [9], developed by Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, the Netherlands. CAD4TB is currently the only commercially available CAR software for TB. Evaluations of CAD4TB in high TB burden countries have been conducted in Africa [10–14] in collaboration with the software's developers. Briefly, CAD4TB identifies shape and textural abnormalities in CXR images to produce an abnormality score ranging from 0 (normal) to 100 (highly abnormal). A cut-off score can be selected, above which people should receive a diagnostic TB test. The optimal cut-off score may depend on the prevalence of abnormal lung images in the screened population and may vary by setting. Given the lack of evidence around its use, no formal guidance from the World Health Organization (WHO) is available advising to which abnormality scores to set the cut-off. The literature evaluating CAD4TB demonstrates the software's sensitivity has improved with each version and that its performance is similar to that of human readers [15].

Smear microscopy is overwhelmingly used as the diagnostic test in most high TB burden countries and is inexpensive [16], but it is not nearly as sensitive as the Xpert MTB/RIF (Xpert) assay [17, 18]. WHO currently recommends Xpert as an initial diagnostic test for TB “where resources permit” [19]. This condition on the recommendation highlights the usefulness of a sensitive triage test (*i.e.* CXR) to reduce costs by minimising the number of expensive Xpert tests performed [20, 21]. The Xpert assay has a concessional price of USD9.98 for public-sector providers, with shipping and importation costs increasing the cost [22]. There has been much global interest around Xpert [23], with huge increases in procurement of the assay [22].

In many Asian settings, especially in urban areas, most people choose to first access the private healthcare market because it is more convenient and/or is perceived as having a higher standard of care than public facilities [24, 25]. However, private providers are not eligible for concessional pricing, and higher procurement costs are passed onto patients, making the test unaffordable for many people with TB in high burden countries [26]. As a result, most countries have not engaged the private sector for Xpert use [27]. Referral systems for free Xpert testing may be an attractive way to link public- and private-sector providers.

To address the lack of evidence around CAD4TB in Asian settings, and the high use of private-sector care in Asia, we designed a prospective study to link patients visiting private providers and public facilities to Xpert testing and evaluated the performance of CAD4TB software as a triage tool in Dhaka, Bangladesh.

Methods

Study sites

With support from the Stop TB Partnership's TB REACH initiative, three screening centres were established across Dhaka. The screening centres operate as a social enterprise, where any funds generated through the sale of laboratory tests are then used to fund programmatic TB care. A Delft EZ DR X-ray system linked with CAD4TB software version 3.07 was installed in each screening centre, as well as multiple four-module GeneXpert systems. A custom-built mHealth application connected to a medical record system (OpenMRS) systematically captured demographic and symptom data from all individuals visiting the screening centres. A referral network of more than 2000 private providers was established to refer clients for testing services. Any private provider could refer patients from testing to a screening centre for testing based on clinical suspicion. In addition, 133 facilities linked to the National TB Control Program (NTP) were engaged to encourage referral of smear-negative individuals for Xpert testing. The screening centres were also open to walk-in clients with no referral history.

Screening, diagnostic and treatment algorithm

Adults (>15 years) presenting at any of the screening centres were verbally screened for a cough of more than 2 weeks duration as both public- and private-sector referrals were made independently of the study based on the clinical decisions of the referring physicians. Symptomatic individuals paid approximately

USD6.40 (BDT 500) for a CXR, which was immediately analysed using CAD4TB software. The cost was on the lower end of the digital CXR market in Dhaka, and Xpert testing and TB care were provided without additional charges. Regardless of the CAD4TB abnormality score, all symptomatic individuals were asked to submit a sputum sample and received a free Xpert test. If the Xpert test failed (invalid, error or no result), it was performed again to obtain a valid outcome if enough sample remained. A Bangladeshi, board-certified radiologist with 10 years of experience read all CXR images offsite at a cost of USD0.50 per image. The radiologist was blinded to both the CAD4TB score and the Xpert result. CXRs were graded as highly TB suggestive, possible signs of TB, non-TB abnormality, and normal. In addition, the radiologist provided a standard radiology report to guide clinical care. Those unable or unwilling to pay for the CXR were referred to a nearby public-sector facility.

People with positive Xpert results were actively referred to the NTP-linked health facility nearest to their home, but were also allowed to receive anti-TB treatment from their private referring physician. If a TB patient opted to receive care with his/her referring physician, treatment initiation and outcomes were reported to the NTP by the study. Individuals with rifampicin-resistant TB were referred to the National Institute of Diseases of the Chest and Hospital for additional testing, clinical evaluation and second-line treatment initiation.

All enrolled participants provided informed written consent. The study protocol was reviewed and approved by the research review committee and the ethical review committee at the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b).

Data analysis

De-identified data were abstracted from OpenMRS and analysed using R version 3.2.4 (R Foundation for Statistical Computing, Vienna, Austria). The radiologist's grades were dichotomised with highly suggestive of TB and possible TB grading considered "abnormal". Basic demographic data, the yields from Xpert testing, and the grading of CXRs by the radiologist and CAD4TB software were compared across age groups and referral groups for people with valid Xpert results. We calculated the sensitivity, specificity, and positive and negative predictive values (PPV and NPV) of the radiologist's grading and different CAD4TB abnormality scores using Xpert as the bacteriological standard. Receiver operating characteristic (ROC) graphs were generated using the pROC package [28]. We calculated the theoretical numbers of Xpert tests saved by using CXR as a triage tool using different CAD4TB cut-off values.

Results

Between May 2014 and February 2016, 18746 people visited the screening centres and were verbally screened. Only 53 individuals declined to pay for the CXR. Of the remaining, 18036 symptomatic individuals aged 15 years and over were identified, and 17134 (95%) of them were able to provide a sputum specimen for testing with Xpert. Among the people tested, 17066 (99.6%) had a valid Xpert result and were included in the analysis.

The final sample comprised 11368 (67%) men and 5698 (33%) women. The greatest proportion of test results came from people who were referred by their private provider (n=12340, 72%), while other referral groups accounted for smaller and similar proportions (table 1). Just over half of participants were aged 26–54 years, and 26% (n=4485) of the participants were aged 55 years or older.

Overall, 2623 people (15%) had *Mycobacterium tuberculosis*-positive (MTB+) results. The yield was lowest among walk-ins (8% versus 17% for the other referral groups). Young adults (≤ 25 years) had the highest rates of TB (22%), while middle age and older groups had yields of 14% and 12%, respectively. Among people with MTB+ results, 139 (5%) were resistant to rifampicin. The radiologist graded 8428 (49%) CXRs as abnormal, overall. The radiologist's CXR abnormality rate was lowest among walk-ins compared to the NTP facility and private provider referrals (27% versus 48% and 54%, respectively). The radiologist also graded a greater proportion of CXRs as abnormal among participants 55 years and older (61% versus 45%). For all participants, the median (interquartile range) CAD4TB abnormality score was 79 (48–95). Walk-ins had the lowest median score (55, interquartile range 37–82). Those 55 years and older had a median CAD4TB score of 93, whereas it was 65 among young adults.

If CXR had been used as a triage tool before Xpert testing, the radiologist's binary grading would have a sensitivity and specificity of 91% and 58% overall (table 2). The sensitivity of the radiologist was around 90% for all groups, while the specificity ranged from 53% (private providers) to 78% (walk-ins). Among older participants, specificity was 43% and among those aged less than 25 years it was 70%. The overall PPV of the radiologist was 28%, and the NPV was 97%. Using CAD4TB to approximate the radiologist's sensitivity for the whole sample, a threshold score of 63 would have been chosen (table 2), and it would have varied from 56 to 82 for the different sub-groups. However, CAD4TB specificity was inferior to the

TABLE 1 Demographic and clinical characteristics of people screened with chest radiography (CXR) and Xpert MTB/RIF

	All n=17066	Referral groups			Age groups		
		Private provider n=12340 (72%)	NTP facilities n=2316 (14%)	Walk-ins n=2410 (14%)	Senior (≥55 years) n=4485 (26%)	Middle age (25–55 years) n=8991 (53%)	Young adult (≤25 years) n=3590 (21%)
Male	11368 (67)	8212 (67)	1364 (59)	1792 (74)	3330 (74)	6056 (67)	1982 (55)
Female	5698 (33)	4128 (33)	952 (41)	618 (26)	1155 (26)	2935 (33)	1608 (45)
Age years	39 (27–55)	40 (28–56)	34 (25–49)	35 (26–50)	62 (59–70)	37 (30–45)	21 (18–24)
Age group							
Senior (≥55 years)	NA	3638 (29)	394 (17)	453 (19)	NA	NA	NA
Middle age (25–55 years)	NA	6284 (51)	1277 (55)	1430 (59)	NA	NA	NA
Young adult (≤25 years)	NA	2418 (20)	645 (28)	527 (22)	NA	NA	NA
Xpert MTB/RIF results							
Positive	2623 (15)	2048 (17)	391 (17)	184 (8)	558 (12)	1270 (14)	795 (22)
RIF resistant	139 (5)	101 (5)	24 (6)	14 (8)	14 (0)	74 (1)	51 (1)
RIF sensitive	2438 (93)	1918 (94)	354 (91)	166 (90)	531 (12)	1175 (13)	732 (20)
RIF indeterminate	46 (2)	29 (1)	13 (3)	4 (2)	13 (0)	21 (0)	12 (0)
Negative	14443 (85)	10292 (83)	1925 (83)	2226 (92)	3927 (88)	7721 (86)	2795 (78)
Radiologist CXR grading							
Abnormal	8428 (49)	6662 (54)	1114 (48)	652 (27)	2754 (61)	4090 (45)	1584 (44)
Normal	8638 (51)	5678 (46)	1202 (52)	1758 (73)	1731 (39)	4901 (55)	2006 (56)
CAD4TB CXR grading							
Median	79 (48–95)	84 (53–95)	75 (45–94)	55 (37–82)	93 (77–95)	74 (43–94)	65 (42–92)
CAD4TB ≥90	6579 (39)	5375 (44)	763 (33)	441 (18)	2550 (57)	2815 (31)	961 (27)
CAD4TB 80–89	1909 (11)	1405 (11)	273 (12)	231 (10)	633 (14)	996 (11)	332 (9)
CAD4TB 70–79	1570 (9)	1101 (9)	231 (10)	238 (10)	388 (9)	903 (10)	340 (9)
CAD4TB 60–69	1243 (7)	846 (7)	185 (8)	212 (9)	259 (6)	677 (8)	313 (9)
CAD4TB 50–59	1269 (7)	876 (7)	165 (7)	228 (9)	210 (5)	691 (8)	367 (10)
CAD4TB 40–49	1638 (10)	1027 (8)	243 (10)	368 (15)	184 (4)	962 (11)	474 (13)
CAD4TB 30–39	1442 (8)	898 (7)	212 (9)	332 (14)	149 (3)	966 (11)	373 (10)
CAD4TB <30	1416 (8)	812 (7)	244 (11)	360 (15)	112 (2)	430 (12)	1523 (9)

Date are presented as n (%) or median (interquartile range). NTP: National TB Program; RIF: rifampicin; CAD4TB: computer automated detection of TB score; NA: not applicable.

radiologist at each threshold score (for the whole sample it was 41% versus 58%). Specificity of CAD4TB was below the radiologist for all referral types and age groups when holding sensitivity similar. The trade-offs between sensitivity and specificity across the different referral and age groups can be visualised

TABLE 2 Performance and comparison of radiologist and computer automated reading for tuberculosis diagnosis

Radiologist CXR reading	Abnormal %	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
All	49	91 (89–92)	58 (57–59)	28 (27–29)	97 (97–97)
Private providers	54	90 (89–91)	53 (52–54)	28 (27–29)	96 (96–97)
NTP facilities	48	93 (90–95)	61 (59–63)	33 (30–35)	98 (97–98)
Walk-ins	27	89 (83–93)	78 (76–80)	25 (22–29)	99 (98–99)
Senior (≥55 years)	61	90 (80–92)	43 (41–44)	18 (17–20)	97 (96–97)
Middle age (25–55 years)	45	90 (88–92)	62 (61–63)	28 (27–29)	97 (97–98)
Young adult (≤25 years)	44	92 (90–94)	70 (68–71)	46 (44–49)	97 (96–98)
CAD4TB CXR scoring	Cut-off score [#]	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
All	63	90 (90–92)	41 (40–42)	22 (21–23)	96 (95–96)
Private providers	64	90 (89–92)	36 (36–37)	22 (21–23)	95 (94–96)
NTP facilities	60	90 (87–93)	43 (41–45)	24 (22–27)	96 (94–97)
Walk-ins	56	90 (85–94)	54 (52–56)	14 (12–16)	99 (98–99)
Senior (≥55 years)	82	90 (88–93)	33 (32–35)	16 (15–18)	96 (95–97)
Middle age (25–55 years)	61	90 (89–92)	45 (44–46)	21 (20–22)	97 (96–97)
Young adult (≤25 years)	56	90 (89–92)	45 (43–46)	21 (20–22)	97 (96–97)

[#]: chosen to approximate radiologist sensitivity. PPV: positive predictive value; NPV: negative predictive value; CXR: chest radiography; NTP: National TB Program; CAD4TB: computer automated detection of TB score.

in the ROC graphs in figures 1 and 2. The graphs show that, regardless of the CAD4TB cut-off value, the radiologist performed better across all groups. The area under the curve (AUC) for all referral groups is 0.74 (95% CI 0.73–0.75). The CAD4TB software performed significantly better among walk-ins (AUC 0.84, 95% CI 0.81–0.87) compared to people referred from NTP facilities (AUC 0.77, 95% CI 0.74–0.79) and private providers (AUC 0.72, 95% CI 0.70–0.73). The software also performed significantly better among young adults (AUC 0.81, 95% CI 0.80–0.83) than the older age group (AUC 0.66, 95% CI 0.65–0.69). AUC by gender was not significantly different (data not shown).

Table 3 presents the results of the analysis of potential Xpert tests saved using different algorithms. Compared to testing all symptomatic individuals with Xpert, using CXR with the human reader could have saved 51% of Xpert tests, with a total loss of 9% (n=249) of the MTB+ cases. To approximate the sensitivity of the radiologist, a CAD4TB threshold score of 63 would be used and it would have saved 36% of the Xpert tests. To save 50% of Xpert tests using CAD4TB, a higher threshold score of 80 would be needed, resulting in 18% (n=473) of MTB+ individuals being missed. Tables of performance of CAD4TB at each threshold and for each referral and age group are presented in the supplementary files.

Discussion

To our knowledge, this is the first published study to evaluate CAD4TB software in Asia and also the first using a public–private partnership. Our evaluation’s sample size was also large. We identified more people with confirmed TB than the combined enrolment of all other published CAD4TB studies to date [15]. CAD4TB software can quickly identify people who require additional diagnostic testing for TB, with small losses in TB detection, and has the potential to save using many Xpert tests. However, the usefulness of CAD4TB will be heavily dependent on the setting and will probably need further improvements in performance and cost reductions to become a useful triage tool in urban Asian settings.

Published studies using CAD4TB have been conducted in Africa, where there are relatively fewer trained radiologists than in Asia [29]. Although a number of studies have shown CAD4TB software to be comparable to a trained human reader [10–14], the radiologist in our study performed better across every referral group, and was less expensive. The use of CAD4TB software as a triage tool in passive case finding in Dhaka (to rule out unnecessary tests) may not be as beneficial as using it as a screening tool to identify people in need of TB testing in rural settings and/or active case-finding approaches where high throughput, the need for quick decisions, and shortages of highly skilled readers are more likely. More prospective

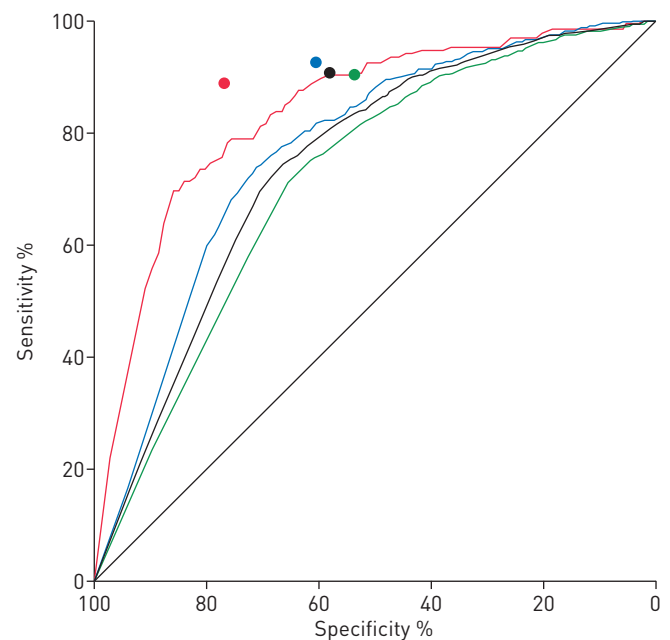


FIGURE 1 Receiver operating characteristic analysis for the detection of *Mycobacterium tuberculosis* stratified by referral type.

Referral groups	AUC [95% CI]	Radiologist's sensitivity and specificity
All groups	0.74 [0.73–0.75]	● 91% and 58%
Private provider	0.72 [0.70–0.73]	● 90% and 53%
NTP facilities	0.77 [0.74–0.79]	● 93% and 61%
Walk-ins	0.84 [0.81–0.87]	● 89% and 78%

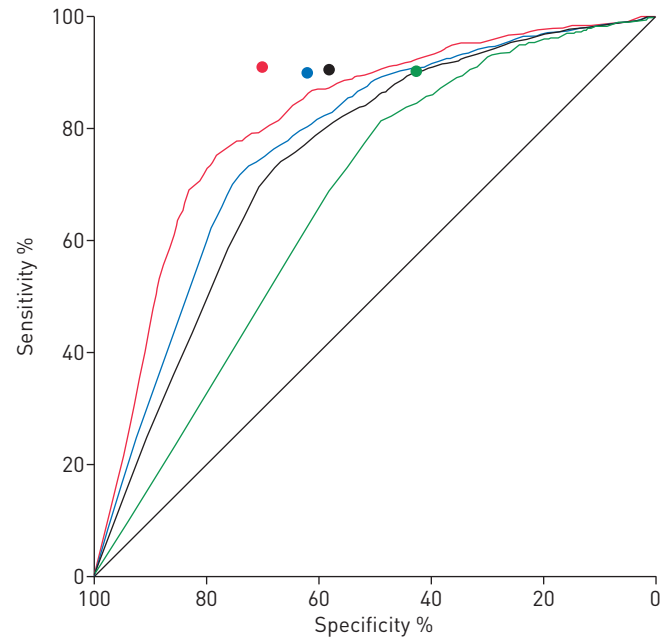


FIGURE 2 Receiver operating characteristic analysis for the detection of *Mycobacterium tuberculosis* stratified by age group.

Age groups	AUC [95% CI]	Radiologist's sensitivity and specificity
All age groups	0.74 [0.73–0.75]	● 91% and 58%
Senior (≥ 55)	0.66 [0.65–0.69]	● 90% and 43%
Middle age	0.77 [0.76–0.78]	● 90% and 62%
Young adult (≤ 25)	0.81 [0.80–0.83]	● 92% and 70%

studies are needed to evaluate CAD4TB in screening and active case-finding interventions, because all of the published studies have been in populations with TB prevalence ranging from 18% to 60% [7].

To our knowledge, this is also the first published evaluation of the CAD4TB software across different sub-groups. The CAD4TB software performed differently depending on the age of the patient and referral group, which should be further evaluated. It is not surprising that CAD4TB performed worse among older individuals, as they have more time to develop non-TB-related abnormalities. We expect the differences in referral type are probably explained by other factors, but could not identify them through the current analysis. Our findings suggest that, to maximise case detection and Xpert test conservation, differentiated CAD4TB threshold scores may be needed, highlighting the need for piloting the software in each setting to identify the ideal cut-off scores. For instance, a study from Zambia achieved 100% sensitivity by including 83% of the individuals for testing [12], while in our evaluation 99.9% of individuals would need testing to achieve 100% sensitivity (supplementary files).

There are several limitations of our study. We used Xpert, not culture, as the end-point of our diagnostic algorithm, meaning that people with Xpert-negative, culture-positive TB have incorrectly been labelled as TB-free. The available culture facilities could not cope with the specimen throughput, and funds were not available to systematically culture even a sample of those tested. Given the evidence for intra-reader variability [5, 13], future work should evaluate CAD4TB against multiple radiologists with varying experience in reading chest radiographs, as our results included only one reader. We did not test for HIV, but Bangladesh's TB-HIV co-infection rate is estimated to be less than 1% [30]. The entry point for our diagnostic algorithm was TB symptoms, meaning that CXR was not used to detect asymptomatic individuals with TB, who make up the majority of TB cases in recent prevalence surveys [31]. Our sample comprises only people willing to pay for CXR. Although this is certainly a different population than recruited for other CAD4TB evaluation studies, it reflects the reality of health services in many Asian settings where people choose to pay for services [32]. Just 53 people came to the screening centres and decided not to pay. Because patients pay for CXR out of their own pocket (whether attending public or private facilities) in almost every high TB burden country [3, 4], our results may be more representative of populations who receive CXR as a triage tool than studies that provide CXR for free. As with other studies evaluating CAD4TB software, we did not measure the possible beneficial effect that having the CXR results immediately available might have had on dropout in a diagnostic algorithm which otherwise would require multiple visits. This is an area, especially for active case finding/screening interventions, that requires more research, as there may be value in having an immediate

TABLE 3 Performance of different algorithms for Xpert tests saved and cases missed in Bangladesh

	CAD score	People tested on Xpert n	Xpert tests saved %	MTB+ cases missed n (%)
Direct Xpert testing				
All referral groups	NA	17 066	NA	NA
Private providers	NA	12 340	NA	NA
NTP facilities	NA	2316	NA	NA
Walk-ins	NA	2410	NA	NA
Senior (≥ 55 years)	NA	4485	NA	NA
Middle age (25–55 years)	NA	8991	NA	NA
Young adult (≤ 25 years)	NA	3590	NA	NA
Radiologist→Xpert				
All	NA	8428	51	249 (9)
Private providers	NA	6662	46	200 (10)
NTP facilities	NA	1114	52	28 (7)
Walk-ins	NA	652	73	21 (11)
Senior (≥ 55 years)	NA	2754	39	58 (10)
Middle age (25–55 years)	NA	4090	55	128 (10)
Young adult (≤ 25 years)	NA	1584	56	63 (8)
CAD4TB→Xpert				
Set CAD to match radiologist sensitivity				
All	63	10 912	36	249 (9)
Private providers	64	8388	32	195 (10)
NTP facilities	60	1452	37	37 (9)
Walk-ins	56	1195	50	18 (10)
Senior (≥ 55 years)	82	3129	30	53 (9)
Middle age (25–55 years)	61	5391	40	123 (10)
Young adult (≤ 25 years)	56	2110	41	79 (10)
Set CAD to reduce Xpert tests by ~50%				
All	80	8488	50	473 (18)
Private providers	84	6226	50	423 (21)
NTP facilities	75	1162	50	68 (17)
Walk-ins	56	1195	50	18 (10)
Senior (≥ 55 years)	94	2220	51	144 (26)
Middle age (25–55 years)	74	4500	50	194 (15)
Young adult (≤ 25 years)	66	1785	50	105 (13)

MTB+: *Mycobacterium tuberculosis*-positive; CAD4TB: computer automated detection of TB software; NTP: National TB Program.

decision to access a diagnostic test rather than waiting for a human reader. Finally, we used version 3.07, and the latest software version released after this evaluation (version 5) may provide better results. Future studies should compare the results in different software versions to demonstrate improvements.

There are a number of lessons for those interested in implementing CAD4TB screening and Xpert testing outside a research setting. Offering the clients of private providers the opportunity for a free Xpert test was an excellent incentive to encourage referrals and can be a useful tool for public–private partnerships. There were 39 GeneXpert systems operating in the public sector in Bangladesh during this intervention, yet none were available at a for-profit price, probably due to lack of market. Educating clinicians about the interpretation of the abnormality score is essential. It was difficult to explain to providers that only one in three people with a score of 95 would test MTB+ (supplementary file) when CAD4TB is marketed for TB. The developers of CAD4TB have produced software updates that are compatible with multiple radiography systems, which will allow for wider adoption of the technology. Developing a version for children may be of interest as well, given the low rates of TB case-finding among children globally [30]. The cost of a CAD4TB reading must be low enough to be an attractive option for screening initiatives, as a radiologist report costs around USD0.5 per image read in Bangladesh. Furthermore, if paying clients demand a radiology report and a printed film, it can add costs for implementers, making the running costs of direct digital systems with CAD4TB higher than what has been reported [14]. Finally, newer versions of CAD4TB require an internet connection, and sending CXR images to telemedicine centres for reading may be an alternative depending on costs, turnaround times and reading quality.

Conclusions

In Bangladesh, the CAD4TB software worked on a high quality, direct digital CXR system, and has the potential to quickly screen out a large number of symptomatic individuals who do not need further diagnostic TB tests. If such an algorithm had been employed, it would have resulted in a significant reduction in the number of Xpert tests used. However, the software's overall performance was inferior and more expensive than a highly trained radiologist in this setting. The software may be more useful in high throughput programmes such as active case-finding interventions where a quick decision on further testing is critical, especially if trained readers are scarce and/or costs are high. The next versions' improvements, costs of CAD4TB, and its compatibility with different radiography systems will help determine its usefulness overall.

References

- 1 van't Hoog AH, Onozaki I, Lonnroth K. Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden. *BMC Infect Dis* 2014; 14: 532.
- 2 Somashekar N, Chadha VK, Praseja P, et al. Role of pre-Xpert screening using chest X-ray in early diagnosis of smear-negative pulmonary tuberculosis. *Int J Tuberc Lung Dis* 2014; 18: 1243-1244.
- 3 Pande T, Pai M, Khan FA, et al. Use of chest radiography in the 22 highest tuberculosis burden countries. *Eur Respir J* 2015; 46: 1816-1819.
- 4 Pedrazzoli D, Lalli M, Boccia D, et al. Can tuberculosis patients in resource-constrained settings afford chest radiography? *Eur Respir J* 2017 47: 1601877.
- 5 WHO Expert Committee on Tuberculosis. Ninth Report. WHO Technical Report Series, No. 552. Geneva, World Health Organization, 1974.
- 6 Waaler HT. Descriptions of the study material and organization of the study. *Bull Int Union Tuberc* 1968; 41: 110-114.
- 7 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: 480-494.
- 8 Karssemeijer N, Bluekens AM, Beijerinck D, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology* 2009; 253: 353-358.
- 9 van Ginneken B, Stegmann MB, 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.
- 10 Breuninger M, van Ginneken B, Philipsen RH, 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.
- 11 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.
- 12 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.
- 13 Steiner A, Mangu C, van den Hombergh J, et al. Screening for pulmonary tuberculosis in a Tanzanian prison and computer-aided interpretation of chest X-rays. *Public Health Action* 2015; 5: 249-254.
- 14 Philipsen RHHM, Sánchez CI, Maduskar P, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. *Sci Rep* 2015; 5: 12215.
- 15 Pande T, Cohen C, Pai M, et al. Computer aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *Int J Tuberc Lung Dis* 2016; 20: 1226-1230.
- 16 Pantoja A, Fitzpatrick C, Vassall A, et al. Xpert MTB/RIF for diagnosis of TB and drug-resistant TB: a cost and affordability analysis. *Eur Respir J* 2013; 42: 708-820.
- 17 Steingart KR, Schiller I, Horne DJ, et al. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2014; 1: CD009593.
- 18 Steingart KR, Henry M, Ng V, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis* 2006; 6: 570-581.
- 19 World Health Organization. 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 extra-pulmonary TB in adults and children. Policy update. WHO/HTM/TB/2013.16. Geneva, WHO, 2013.
- 20 Kik SV, Denkinger CM, Casenghi M, et al. Tuberculosis diagnostics: which target product profiles should be prioritised? *Eur Respir J* 2014; 44: 537-540.
- 21 Theron G, Pooran A, Peter J, et al. Do adjunct tuberculosis tests, when combined with Xpert MTB/RIF, improve accuracy and the cost of diagnosis in a resource-poor setting? *Eur Respir J* 2012; 40: 161-168.
- 22 Albert H, Nathavitharana RR, Isaacs C, et al. Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: what lessons have we learnt and how can we do better? *Eur Respir J* 2016; 48: 516-525.
- 23 Weyer K, Mirzayev F, Migliori GB, et al. Rapid molecular TB diagnosis: evidence, policy making and global implementation of Xpert MTB/RIF. *Eur Respir J* 2013; 42: 252-271.
- 24 Hossain S, Zaman K, Quaiyum A, et al. Care seeking in tuberculosis: results from a countrywide cluster randomised survey in Bangladesh. *BMJ Open* 2014; 4: e004766.
- 25 Shaikh B, Hatcher J. Health seeking behaviour and health services utilization trends in National Health Survey of Pakistan: what needs to be done? *J Pak Med Assoc* 2007; 57: 411-414.
- 26 Puri L, Oghor C, Denkinger CM, et al. Xpert MTB/RIF for tuberculosis testing: access and price in highly privatised health markets. *Lancet Glob Health* 2016; 4: e94-e95.
- 27 Qin ZZ, Pai M, Van Gemert W, et al. How is Xpert MTB/RIF being implemented in 22 high tuberculosis burden countries? *Eur Respir J* 2015; 45: 549-554.
- 28 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.
- 29 WHO. The World Health Report 2006: Working Together for Health. Geneva, WHO, 2006.

- 30 World Health Organization. Global Tuberculosis Report 2015. WHO/HTM/TB/2015.22. Geneva, WHO, 2015. Available from: http://who.int/tb/publications/global_report/en/
- 31 Onozaki I, Law I, Sismanidis C, *et al.* National tuberculosis prevalence surveys in Asia, 1990–2012: an overview of results and lessons learned. *Trop Med Int Health* 2015; 20: 1128–1145.
- 32 Wells WA, Ge CF, Patel N, *et al.* Size and usage patterns of private TB drug markets in the high burden countries. *PLoS One* 2011; 6: e18964.