Title Computer-aided interpretation of chest radiography to detect TB in rural South Africa

Authors Jana Fehr[1,2], Stefan Konigorski[1], Stephen Olivier[2], Resign Gunda[2,3,4], Ashmika Surujdeen[2], Dickman Gareta[2], Theresa Smit[2], Kathy Baisley[2,5], Sashen Moodley[2], Deenan Pillay[2,4], Alison D. Grant[2,3,5,6], Mark J. Siedner[2,6,7,8], Christoph Lippert[1], Emily B. Wong[2,4,9] and the Vukuzazi Team*.

1 Digital Health & Machine Learning, Hasso Plattner Institute, University of Potsdam, Germany
2 Africa Health Research Institute, KwaZulu-Natal, South Africa
3 School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, South Africa
4 Division of Infection and Immunity, University College London, London, UK
5 Dept. of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
6 School of Clinical Medicine, College of Health Sciences, University of KwaZulu-Natal, South Africa
7 Harvard Medical School, Boston, MA, USA.
8 Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, USA.
9 Division of Infectious Diseases, University of Alabama at Birmingham, AL, USA.
* Supplementary Appendix

Summary

Background Computer-aided digital chest radiograph (CXR) interpretation can facilitate high-throughput screening for tuberculosis (TB), but its use in population-based screening has been limited. We applied an automated image interpretation algorithm, CAD4TBv5, prospectively in an HIV-endemic area.

Methods Participants underwent CXR and, for those with symptoms or lung field abnormality, microbiological assessment of sputum collected at a mobile camp in rural South Africa. CAD4TBv5 scored each CXR on a 0-100 scale in the field. An expert radiologist, blinded to the CAD4TBv5 score and other data, assessed CXRs for 1) lung field abnormality and 2) findings diagnostic of active TB (R+). We estimated the performance of CAD4TBv5 for triaging (identifying lung field abnormality as a criteria for sputum examination) and diagnosis (detection of active TB as defined by microbiologic (M+) or radiologic (R+) gold standards).

Findings For triaging, a CAD4TBv5 threshold of 25 identified abnormal lung fields with a sensitivity of 90·3% and specificity of 48·2%. For diagnosis, CAD4TBv5 had less agreement with the microbiological reference standard (M+) used to define definite TB (AUC 0·78) than with the radiological reference standard (R+) used to define probable TB (AUC 0·96). HIV-serostatus did not impact CAD4TB's performance.

Interpretation A low CAD4TBv5 threshold was required to achieve acceptable triaging sensitivity. Low specificity at this threshold led to high rates of sputum collection despite normal lung fields. CAD4TBv5 had difficulty identifying microbiologically-confirmed TB cases with subtle radiological features but had excellent agreement with the radiologist in identifying radiologically-defined TB cases. We conclude that computer-aided CXR interpretation can be useful in population-based screening in HIV-endemic settings, but threshold selection should be guided by setting-specific piloting and priorities. CXR interpretation algorithms require refinement for the identification of radiologically-subtle early TB.

Funding Funded by the Africa Health Research Institute, Wellcome Trust, Bill and Melinda Gates Foundation and NIAID/NIH.

40

Research in context

Evidence before this study

Community-based screening has the potential to increase active TB case finding and advance the goal of eradicating TB by 2030. The recent development of mobile digital chest radiography hardware allows decentralized high-throughput CXR imaging. Computer-aided interpretation of chest radiography offers the opportunity for cost- and labour-efficient population-based TB-screening by preselecting individuals with abnormal lung fields for microbiological sputum examination. Previous studies evaluated the performance of the computer-aided detection system CAD4TB in clinic and hospital settings and retrospectively in a population-based screening programme in Zambia. Prospective assessment of CAD4TB in a community-based setting is necessary to evaluate its usefulness for TB screening in these settings.

Added value of this study

This is the first study applying CAD4TB prospectively in a community-based screening programme in an HIV-endemic area. In this programme, a pilot phase identified a CAD4TB triaging cut-off that provided a high sensitivity for triaging participants for sputum testing. This threshold is much lower than the previously reported thresholds useful in clinical settings. Because of the population-based design of this study, the microbiologically-proven cases of TB represented a wider spectrum of radiological findings than are typically observed in clinical settings. In this context we found that the ability of CAD4TB to detect active TB was highly congruent with radiological findings but not with microbiological test results.

Implications of all the available evidence

CXR reading algorithms can identify radiologically-evident TB, likely because they have been trained on datasets obtained from clinical settings where patients present with relatively late disease. In population-based screening programmes, early TB disease occurs with radiologically-subtle characteristics. Triaging thresholds and imaging algorithms that have been developed for the clinics need to be adjusted to be most useful for population-based screening.

Introduction

Tuberculosis (TB) continues to cause over 1 million deaths annually, challenging the WHO strategy to eliminate the disease by 2030.[1] In resource-limited settings with high TB burden, community-based screening programmes have been established to increase case-finding.[2–4] However, these programmes are challenged by high costs for medical staff and diagnostics including molecular tests such as Xpert MTB/RIF® and microbiological culture.[5,6,7] To reduce the cost and staffing requirements associated with diagnostic testing, digital chest radiography (CXR) has become an important tool to identify individuals with lung field abnormalities who require sputum testing according to WHO-guidelines for TB screening.[7,8] However, this approach requires a workforce of experienced clinicians or radiologists which keeps the cost of outreach programmes high. Another challenge is that CXRs of HIV-positive TB patients may show atypical radiological signs.[9,10]

Computer-aided detection (CAD) systems to support clinicians in detecting TB-related abnormalities in digital CXRs have the potential to make health screening programmes more efficient.[11] One commercial product, CAD4TB (©Thirona, Nijmegen, the Netherlands) scores CXRs according to abnormal signs, such as cavities and consolidation.[12] One ideal use-case for CAD is to rapidly identify people with lung abnormalities on CXRs, who should be triaged for bacteriological sputum examination according to WHO guidelines.[8] Previous retrospective studies have highlighted that CAD4TB can identify CXRs with abnormal lung fields and TB-related abnormalities[13–15] and reduce the cost for diagnostics.[16,17] Using the tool for triage requires selecting a decision threshold score above which participants are referred for sputum testing. Currently, no official guidelines exist for selecting a decision threshold. Several studies suggested the importance of obtaining a study-specific decision threshold adjusted to the underlying TB-prevalence and demographic parameters.[18,19] So far, CAD4TB has been mainly evaluated in clinical settings in which symptomatic patients seek care. One study at a diabetes-care centre in Indonesia retrospectively evaluated CAD4TB and suggested a decision threshold score of 65-70.[20] Data on its performance in population-based screening is limited. Another study applied CAD4TB retrospectively in a non-clinical population-based setting and suggested that a lower CAD4TB score threshold would be required for triaging.[21] An independent, real-world prospective analysis of CAD4TB scores for triaging study participants is important to establish how computer-automated chest radiography performs in population-based screenings.

Here we report the prospective application of computer-aided CXR reading with CAD4TB version 5 (CAD4TBv5) during the first year of a community-based TB screening programme in rural South Africa. We also report the performance of CAD4TB version 6 (v6), which was performed retrospectively. We evaluated CAD4TB as a tool to 1) triage participants for sputum testing based on lung abnormalities, and 2) diagnose active TB based on microbiological and radiological gold standards.

Materials and Methods

Study design

The community-screening programme 'Vukuzazi' used mobile vans to provide free health assessments in the rural uMkhanyakude district of KwaZulu-Natal in South Africa. The data for this analysis was collected during the first year of the project (between 25 May 2018 and 24 May 2019)[22]. Study field workers visited households to explain and provide the study description and invite eligible residents to participate. Eligibility criteria included a minimum age of 15 years and ongoing residency in the area. At the camp, participants answered questions about smoking, TB symptoms and history, and HIV history and treatment. HIV status was assessed by 4th generation antibody/antigen test (Genscreen Ultra HIV Ag-Ab, Bio-Rad, Marnes-la-Coquette, France) on venous blood.

Posterior-anterior digital CXRs were obtained using a mobile unit (Canon CXDI-NE) and saved in DICOM-format in cloud storage. A score indicating lung abnormalities and likelihood of active pulmonary tuberculosis[23] was calculated on-site using CAD4TBv5. CAD4TB's methodology is based on initial lung field segmentation and subsequent analysis of the lung shape, symmetry, and costophrenic angles, resulting in an abnormality score between 0 and 100 (increasing with abnormality).[24] Within seven days of enrolment, an expert radiologist with more than 35 years of local experience reviewed all CXRs in a central setting blinded to the CAD4TBv5 score and any other patient information. The radiologist categorized each CXR as 1) having either normal or abnormal lung fields and 2) as having radiological signs typical of active TB (R+) or not (R-). CAD4TBv6 an updated version of the image interpretation software that uses deep neural networks,[25] became available after data collection. CAD4TBv6 scores were calculated and analysed retrospectively.

The aim of our study was to maximize TB case-finding and to capture the full spectrum of active TB. Following WHO guidelines for TB-prevalence surveys[8], participants were referred for sputum examination if they endorsed any cardinal TB-symptom (fever, weight loss, cough, or night sweats) or if they had an abnormal CXR (indicated by a CAD4TBv5 score above the triaging threshold in the camp). If the expert radiologist indicated abnormal lung fields despite a CAD4TBv5 score below the threshold, a follow-up team contacted participants for sputum collection at home. Sputum specimens were analysed for *Mycobacterium tuberculosis* using Xpert Ultra MTB/RIF® (XpertUltra) (Cepheid, Sunnyvale, CA, USA) and liquid MGIT culture (MGIT) (Becton Dickinson, UK), held for 42 days).

During a pre-designated pilot phase, based on literature review[20] and consultation with experts in the field, a CAD4TBv5 threshold of 60 was selected to triage participants for sputum examination. For the main phase of the study, the CAD4TBv5 triaging threshold was adjusted based on analysis of the pilot data to obtain 90% sensitivity for detection of lung field abnormalities.

Definitions of TB

Definite TB was defined in participant's whose sputum was microbiologically positive regardless of the radiological status. Sputum was defined as microbiologically positive (M+) if *M. tuberculosis* was detected by either XpertUltra or MGIT. Because of emerging questions about the significance of XpertUltra "trace" positive[26], we performed a sensitivity analysis that excluded participants whose M+ evidence was solely due to a "trace" result. Probable TB was defined if TB was diagnosed radiologically (R+) but sputum was microbiologically negative (M-) or not obtained (M0).

Data analysis

To assess performance of CAD4TBv5 to triage participants for sputum collection, sensitivity, specificity, negative predictive value (NPV), positive predictive values (PPV) were calculated compared a gold standard based on the radiologist's assessment of 'abnormal' or 'normal lung fields'. Additionally we calculated the number of definite TB cases that would have been detected or missed at selected thresholds.

To assess CAD4TBv5's performance for diagnosing definite or probable TB, the area under the receiver operating curve (AUC) with estimations of 95% confidence intervals (CI) was calculated and AUCs were compared using the DeLong method. CAD4TB scores between groups of participants with specified TB and HIV statuses were performed using Mann-Whitney-Wilcoxon tests.

As a secondary analysis, for both the triaging and diagnostic analyses we compared CAD4TBv5 scores to the retrospectively calculated CAD4TBv6 scores.

Data analysis was performed with R (version 3.6.1) using the packages 'epiR' and 'pROC'.

Ethic statement

Ethics approval was obtained from the University of KwaZulu-Natal Biomedical Research Ethics Committee (BE560/17), the London School of Hygiene & Tropical Medicine Ethics Committee (14722), and the Partners Institutional Review Board (2018P001802).

Role of the funding source

The community-screening programme 'Vukuzazi' is funded by the Africa Health Research Institute, the Wellcome Trust and the Bill and Melinda Gates Foundation. Additional funders include NIAID/NIH (K08AI118538 to EBW). The funders had no role in study design, data analysis and interpretation, or writing of the report. The mobile CXR unit including the real-time use of the CAD4TBv5 software was leased from Aurum Innova in partnership with Delft. The retrospective calculation of CAD4TBv6 scores was provided by Delft free of charge. Delft provided no funding and had no other role in this study.

Results

Demographics and TB categorization

We report results from Vukuzazi's first year, during which 10,320 participants were enrolled (Table 1). The first 1,132 participants were enrolled in a pilot phase. Among all participants, 406 were excluded from chest radiography due to pregnancy or physical inability to climb into the mobile CXR van. A total of 9,914 participants completed CXR and survey measurements and are therefore included in this analysis. Of these, 7912 (79·8%) had normal and 2002 (20·2%) had abnormal CXRs. The classification of participants by their microbiologic (M) and radiologic (R) characteristics is detailed in Figure 1. Ninety-nine participants (1·0%) had microbiologically-positive sputum (M+) and met the criteria for definite TB. Of these, 24 had a XpertUltra trace as only microbiological evidence of TB and were therefore excluded from sensitivity analysis of the definite TB (excluding trace) group. 172 (1·7%) had probable TB with radiological evidence of TB (R+) without microbiological evidence (M-/0).

Table 1: Summary of demographics, symptoms, CXR abnormality and TB diagnoses among study participants. *Shown are absolute frequencies (relative frequencies), or *median (min; max).*

Characteristic	**Total (%)**
N	10,320
Female	7,049 (68·3)
Age*	38 (15; 101)
HIV seropositivity	3,105 (30·1)
On suppressive ART	2,522/3,105 (81·2)
History of TB	1,228 (11·9)
Current smoker	737 (7·1)
TB assessment	
Cough currently	696 (6·7)
Fever	306 (3·0)

Weight loss (in prior 6 months)	237 (2·3)
Night sweat	337 (3·3)
Any TB symptom	1,091 (10·6)
Any lung field abnormality on CXR	2002 (20·2)
Definite TB (M+)	99 (1·0)
Definite TB (M+ excluding trace)	75 (75·7)
Probable TB (M- R+)	172 (1·7)

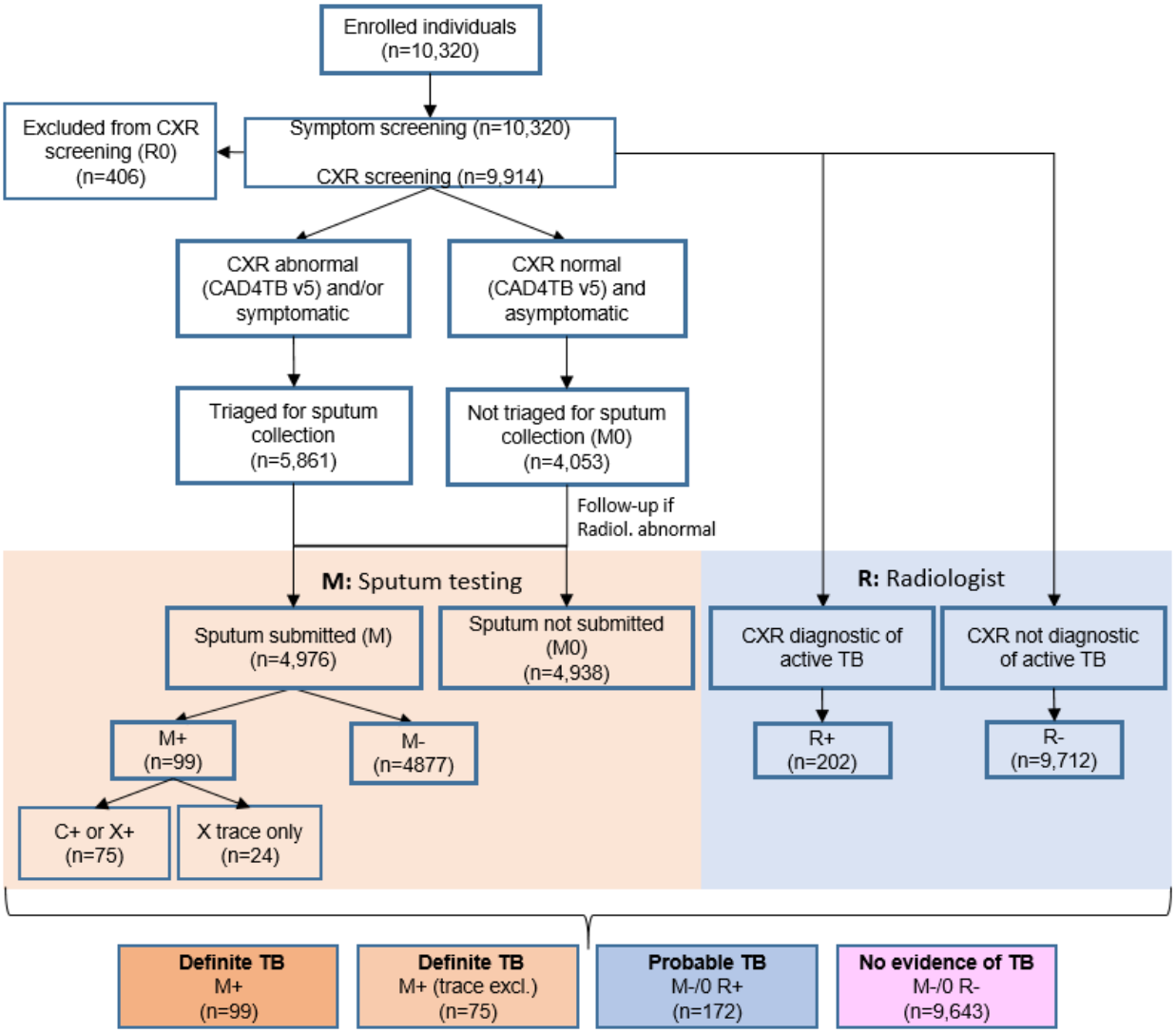

Figure 1: Participation procedure and grouping of participants. *Participants were triaged for sputum collection in the camp if patients exhibited symptoms or if the CAD4TBv5 score was above the triaging threshold. Additional sputums were collected in a follow-up home visit if the independent radiologist indicated lung abnormalities not being*

triaged in the camp. Sputum was tested with Xpert MTB/RIF Ultra (X) and liquid culture (C). M+ indicates at least one positive result among both tests. Missing sputum (M0) was due to not meeting triaging criteria for sputum examination or an inability to produce sputum. The radiologist indicated whether chest x-rays showed signs diagnostic of active TB (R+) or not (R-). Participants were grouped into definite TB (M+) definite TB excluding Xpert Ultra trace only results from M+, probable TB (M-/0 R+) and no evidence of TB (M-/0 R-).

1) Evaluation of CAD4TB as a tool to triage participants for sputum examination

During the pilot phase, 1,090 participants underwent chest radiography and were triaged for sputum examination at a CAD4TBv5 score threshold above 60 and/or the presence of symptoms. At this threshold, CAD4TBv5 correctly identified only 54 of 198 (27·3%) CXRs with abnormal lung fields. This low rate of field triage required a large number of sputum samples to be collected at a separate follow-up visit (144, of which 109 were successfully collected), a significant impediment to study operations. Table 2 compares the radiologist's assessment of lung field abnormality with CAD4TBv5 at a threshold of 60. At this threshold, CAD4TBv5 had a sensitivity=27·3% (CI: 21·2-34·0) and a specificity=99·1% (CI: 98·2-99·6). To achieve the targeted triage sensitivity of 90%, the threshold of 25 was selected for the main phase.

Table 2: Contingency table of chest x-rays during the pilot phase classified into abnormal and normal by the radiologist and CAD4TBv5 triaging threshold 60. *Chest x-rays from the pilot phase with CAD4TBv5 scores higher or below 60 were compared to the radiologist classification of normal and abnormal lung fields. The table contains absolute numbers n of samples and percentages of the total n. In addition, absolute numbers of positive results (M+) are given.*

Pilot phase

n (%)	Abnormal lung fields	Normal lung fields	
CAD≥60	54 (5·0) M+ 7	8 (0·7) M+ 0	62 (5·7) M+ 7
CAD<60	144 (13·2) M+ 1	884 (81·1) M+ 1	1,028 (94·3) M+ 2
	198 (18·2) M+ 8	892 (81·8) M+ 1	1,090 (100) M+ 9

Table 3 compares CAD4TBv5's performance at a triaging threshold of 25 and the radiologist’s assessment of lung abnormality among participants who underwent chest radiography during the entire study (n=9,914). Sensitivity and specificity at this threshold resulted in 90·3% (CI: 88·9-91·6) and 48·2% (CI: 47·1-49·3), respectively. At this threshold, 5,906 (59·6%) CXRs were categorized as abnormal, 3·0 times more than necessary based on the radiologist’s assessment (n=2,002). Among the 99 definite TB (M+) cases, 95 were triaged for sputum collection by CAD4TBv5, whereas only 80 would have been by the radiologist.

Table 3: Contingency table of chest x-rays from the entire study classified into abnormal and normal by the radiologist and CAD4TBv5 triaging threshold 25. *Chest x-rays from the entire study with CAD4TBv5 scores higher or below 25 were compared to the radiologist classification of normal and abnormal lung fields. The table contains absolute numbers n of samples and percentages of the total. In addition, absolute numbers of positive results (M+) are given.*

Pilot and main phase

n (%)	Abnormal lung fields	Normal lung fields	
CAD≥25	1,808 (18·2) M+ 79	4,098 (41·3) M+ 16	5,906 (59·6) M+ 95
CAD<25	194 (2·0) M+ 1	3,814 (38·5) M+ 3	4,008 (40·5) M+ 4
	2,002 (20·2) M+ 80	7,912 (79·8) M+ 19	9,914 (100) M+ 99

If the initial triage threshold of 60 had been used for the entire study, only (751) 7·6% of the 9,914 participants would have been referred for sputum testing by CAD4TBv5, whereas the radiologist would have referred 3·8 times more participants. Among the 99 definite TB cases, 44 had CAD4TBv5 score<60. Of these, 11 met criteria for sputum collection due to symptoms but the remaining 33 individuals were asymptomatic and would have been missed if a triaging threshold of 60 had been used throughout the study.

CAD4TBv6 scores were calculated retrospectively and similarly compared to the radiologist's assessment of normal and abnormal lung fields (Figure 2). CAD4TBv6 performed slightly better than CAD4TBv5 (AUC v6: 0·88 (CI: 0·87-0·89), AUC v5: 0·84 (CI: 0·83-0·85), p-value<0·001). Table S1 lists different triaging thresholds and respective sensitivities and specificities of CAD4TBv5 and v6. With CAD4TBv6, a threshold of 35 achieved an acceptable triaging sensitivity (90·3% (CI: 88·9-91·6)), which provided a slightly improved, but still sub-optimal specificity of 59·1% (CI: 58·0-60·2).

We did not find significant AUC differences between HIV- and HIV+ participants for either version of CAD4TB (p-value v5=0·69 and v6=0·05, figure S1).

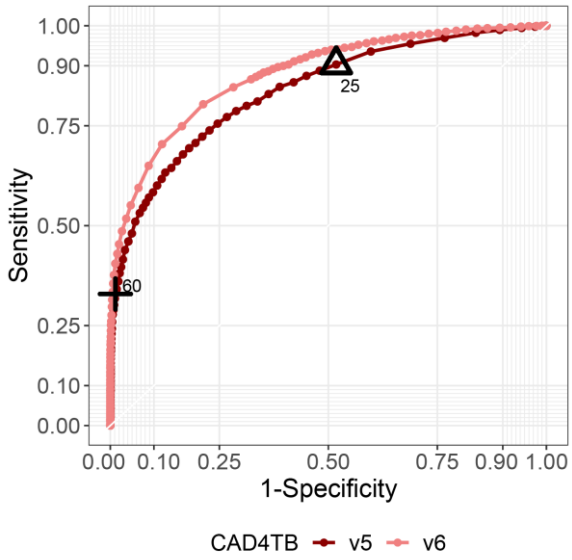

Figure 2: Receiver operating curves of identifying abnormal lung fields with CAD4TB compared to the radiologist. *Annotations depict the operating points with CAD4TBv5 triaging thresholds 25 (Δ) and 60 (+).*

2) CAD4TB as a tool to identify active TB in CXRs

Based on microbiological and radiological findings, participants were defined as having definite TB (microbiologically-proven, regardless of radiological diagnosis (M+ R-/+), probable TB (M-/0 R+) or no evidence of TB (M-/0 R-). The distribution of CAD4TBv5 scores for each group is depicted in figure 3. Scores of the probable TB group were significantly higher than those for the definite TB group (p-value<0·001). The sensitivity analysis in which XpertUltra "trace"-only cases were excluded from the definite TB group did not meaningfully shift the distribution of CAD4TBv5 scores. Scores were lowest in the group with no evidence of TB. We compared CAD4TBv5 score distributions of each group between HIV- and HIV+ individuals and found no significant difference (Figure 3B). CAD4TBv6 scores showed similar trends to CAD4TBv5 (Figure S2).

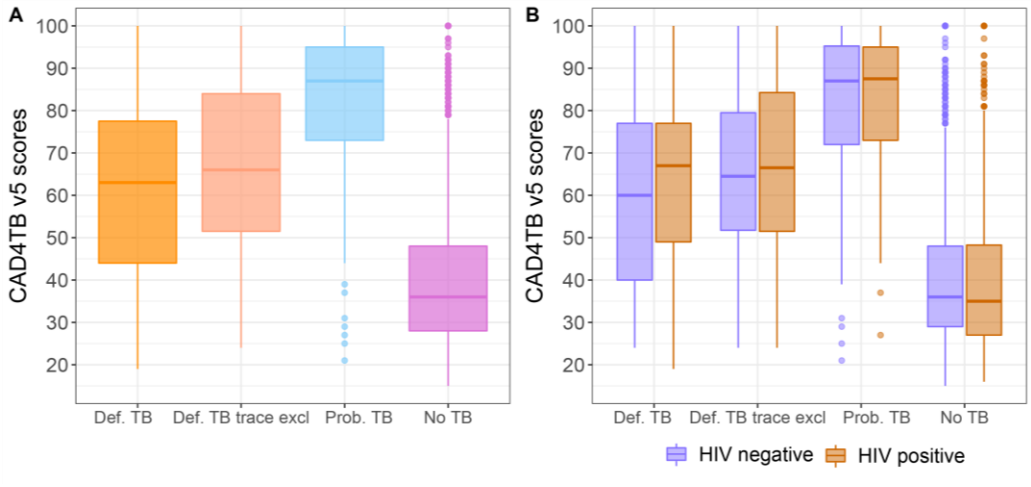

Figure 3: CAD4TB v5 scores between diagnostic groups of microbiological and/or radiological evidence for active TB. *Participants were grouped by into: Definite TB with microbiological evidence (M+), definite TB excluding samples that only had a XpertUltra trace result, probable TB with no microbiological evidence but radiological signs of TB (M-/0 R+) and no evidence of TB (M-/0 R-). Horizontal lines mark the median and the 25-75% quartile. A) shows all participants, B) shows all participants stratified by HIV status.*

We assessed the diagnostic performance of CAD4TBv5 to identify definite and probable TB (Figure 4, Table S2). The AUC for definite TB of 0·78 (0·73-0·83) was significantly lower than that of probable TB (0·95-0·98), p-value<0·001). The AUC for definite TB that excluded XpertUltra trace only results did not significantly differ from that of definite TB (0·82 (0·77-0·87), p-value=0·28). The diagnostic performance of CAD4TBv6 was similar (Figure S3, Table S2).

Pairwise AUC comparison showed no significant differences between CAD4TBv5 and v6 (p-values: Def. TB=0·59, Def. TB trace excl.=0·28, Prob. TB=0·71). Sensitivity, specificity, PPV, and NPV of CAD4TB scores 25, 50, 60, 65, and 70 compared to TB-definitions are listed in Table S3 (CAD4TBv5) and S4 (CAD4TBv6). AUCs from CAD4TBv5 and v6 scores for definite and probable TB were not significantly different between HIV- and HIV+ participants (p-values>0·1, table S5).

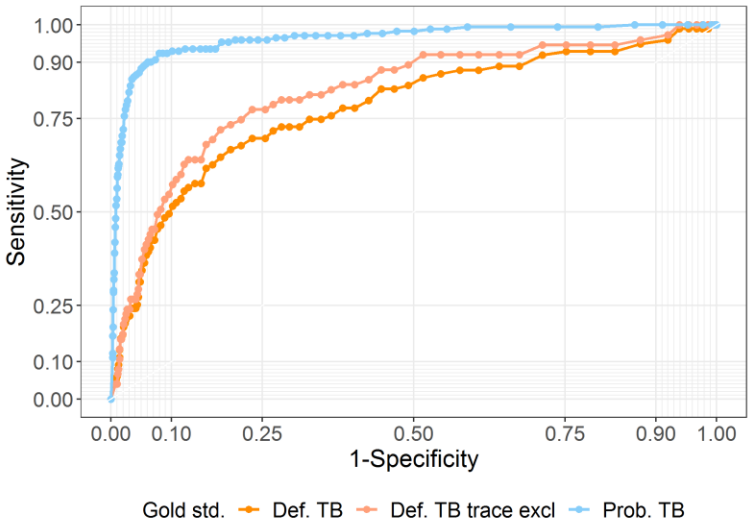

Figure 4: Receiver operating curve of CAD4TBv5 scores compared to diagnostic microbiological and/or radiological evidence. *Positive TB was defined as either definite TB with microbiological evidence (M+), definite TB excluding samples that only had a XpertUltra trace result, or probable TB with no microbiological evidence but radiological signs of TB (M-/0 R+).*

Discussion

In a setting of community-based TB screening where HIV prevalence was high, we prospectively applied CAD4TBv5 and found that using previously recommended CAD4TB thresholds of 60-85 would have missed a high proportion of microbiologically-confirmed TB cases. To achieve 90% sensitivity compared to an expert radiologist's assessment of lung field abnormality, we needed to apply a CAD threshold of 25. CAD4TBv5 had similar performance to an expert radiologist in identifying radiologically-typical TB (AUC 0·96 (CI: 0·95-0·98)) but performed much less well compared to a microbiological gold standard (AUC 0·78 (CI: 0·73-0·83)).

Applying CAD4TB to triage participants for sputum examination requires defining a triage threshold. To date, CAD4TB has mostly been utilized in healthcare centres to triage symptomatic patients.[13,14,16,18,19,21,25]. Previous studies have suggested thresholds between 60 and 85.[13,20,21] Several reports suggested that CAD4TB's threshold needs to be adjusted to study aims, the number of available microbiological tests, and underlying TB prevalence.[16,18,19,21,22] There has been no recommendation of a triage threshold for community-based screening programmes. The goal of our study was to maximize TB case-finding and to capture the full spectrum of active TB. Therefore, we conducted a pilot phase to identify a threshold identifying abnormal lung fields at 90% sensitivity and found that only a low threshold of 25 was able to achieve this aim. The AUC of CAD4TBv5 to identify abnormal lung fields was 0·84 (CI: 0·83-0·85) and only slightly improved with CAD4TBv6 (0·88 (CI: 0·87-0·89)). These results are similar to one reported study from a TB-prevalence survey in Zambia with CAD4TBv5 which found an AUC=0·87.[22] The high HIV-prevalence in our study population (30·1%) did not seem to affect the triaging performance of CAD4TB.

Using CAD4TBv5 at a triage threshold of 25 identified a 1·0% prevalence of previously undiagnosed active TB. Because the specificity for abnormality was low (48·2% (CI: 47·1-49·3)) at this threshold, 3,904 participants were triaged for sputum collection despite CXRs that the radiologist assessed as normal. This

had cost implications but also identified 16 cases of radiologically-subtle active TB that would have been missed if sputum had only been collected from individuals with CXR abnormality. If we had not conducted a pilot and used initial triaging threshold of 60, we would have significantly underestimated the rate of undiagnosed active TB in the community (0·7%).

There are multiple potential explanations for the low CAD4TB threshold that was required to achieve 90% sensitivity for the identification of abnormal CXRs in our setting. It is possible that that the population has an unusually high frequency of subtle lung-field abnormalities due to previous TB or other disease (i.e. costophrenic angle blunting) or that the radiologist over-called lung abnormality. Another reason could be that CAD4TB has been trained on symptomatic clinical patients with radiologically-significant features of TB. A refinement of this algorithm for triaging applications in population-based settings where most active TB cases are asymptomatic and radiologically-subtle may be required. Further prospective analyses in different demographic settings are necessary to determine which factors impact CAD4TB scores and to guide the choice of study-specific thresholds based on study objectives and capacities.

We assessed the ability of CAD4TB to identify definite and probable TB. Like others, we found that CAD4TB performed very well in identifying probable TB based on radiologic diagnosis of TB (AUC v5 0·96 (CI: 0·95-0·98) and v6 0·96 (CI: 0·95-0·98). This is comparable to AUCs reported from a retrospective clinic-based cohort in Pakistan[25] using CAD4TBv5 and v6 compared to radiologic diagnosis (v5: 0·95 (0·93-0·96), v6: 0·99 (CI: 0·98-0·99)). We found that CAD4TB had significantly lower performance for identifying definite microbiologically-confirmed TB, which was similar to the study from Pakistan (CAD4TBv5 0·87 (CI: 0·85-0·88), CAD4TB v6 0·89 (CI: 0·87-0·89)), although the AUCs from our study were generally even lower (CAD4TBv5 0·78 (CI: 0·73-0·83), CAD4TBv6 0·79 (0·73-0·84)). We found no significant effect of HIV-serostatus on CAD4TB's performance.

It is known that identifying active TB from CXRs is a non-trivial task as the characteristics overlap with prior TB and other pathologies.[27] Radiologists are trained in clinical settings where patients present with symptoms and generally advanced disease. In our community-based study, only 20·2% of definite cases endorsed symptoms and only 30·3% of these had CXR features that the radiologist interpreted as significant of active TB. A recent review emphasized that the whole clinical spectrum of TB disease has not been fully characterized and that symptoms may appear intermittently over the course of TB but are consistently present only during the latest stage of disease.[28] We hypothesize that population-based screening identifies cases of subclinical TB which can be asymptomatic and radiologically-subtle and that this may explain the differences between our study and those previously reported.

Limitations of our study are that only one radiologist performed independent CXR reading, that a single spot sputum was the basis of microbiological data, and that we may have misclassified definite TB cases based on a false-positive XpertUltra trace result. XpertUltra is reported to perform at a higher sensitivity than the predecessor Xpert MTB/RIF[29], even among HIV co-infected participants.[26] We attempted to address the uncertainty about trace results in a sensitivity-analysis that excluded these cases and did not find significantly different results. More research is necessary to investigate if XpertUltra trace captures subclinical TB. Another limitation is that microbiological testing was not performed for asymptomatic participants whose CAD4TBv5 score was below 25.

Conclusions

In a prospective population-based study we assessed the performance of computer-aided CXR interpretation with CAD4TBv5 and found that to triage abnormal chest x-rays for sputum collection with 90% sensitivity required using a threshold of 25. Using previously recommended CAD4TB thresholds of 60 and higher would have missed a high proportion of microbiologically-confirmed TB cases.

Diagnostically, CAD4TBv5 had similar performance to an expert radiologist in identifying radiologically-typical TB but performed much less well compared to a microbiological gold standard. Piloting was necessary to obtain a CAD4TB triaging threshold identifying abnormal lung fields at 90% sensitivity. This threshold was much lower than expected and prompted a large number of sputum examinations for individuals with normal lung fields but also identified additional asymptomatic and radiologically-subtle cases of TB. CAD4TB's performance to identify active TB in CXRs had high agreement with radiologically-diagnosed probable TB, but poor recognition of microbiologically-positive definite TB. We hypothesize that computer-aided radiography for TB which has been developed for clinical settings is currently not trained to detect subclinical stages that may comprise the majority of case-findings during population-based screening programmes. To fulfil its full potential, CXR-interpretation software requires fine-tuning to detect subclinical TB. Additional research is needed to clarify radiological and microbiological manifestations of subclinical TB to combat transmission and morbidity and to achieve the WHO goal of eliminating TB by 2030.

Contributions

JF performed data analysis with figures and tables and writing of the report. SO and KB contributed to data analysis. DG contributed to data management. RG, AS, DP, and MJS contributed to study design. TS and SM contributed to the laboratory study setup and provision of test results. ADG contributed to study design and analysis conceptualization. SK and CL contributed to data analysis and supervision. EW, study design, analysis conceptualization, revision and supervision. All authors contributed to the revision of the manuscript.

Declaration of interests

We declare no competing interests. CAD4TBv5 scores were purchased from Delft but CAD4TBv6 scores were provided free of charge. Delft did not contribute any funding and was not involved in the analysis and writing of the report.

Data sharing

The Vukuzazi screening protocol and anonymized data can be requested at https://data.ahri.org/index.php/catalog/990.

Acknowledgments

'Vukuzazi' was the collective effort of a large team (see Supplemental Appendix). We thank the community of the uMkhanyakude district in KwaZulu-Natal for participating in this study. We thank Delft for providing CAD4TBv6 scores free of charge. We dedicate this paper to the memory of Anand Ramnanan who made a major contribution to the Vukuzazi Project.

References

1 World Health Organization. Global Tuberculosis report. 2018.

2 Corbett EL, Bandason T, Duong T, Dauya E, Makamure B, Churchyard GJ, et al. Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): A cluster-randomised trial. Lancet. 2010;**376**:1244–53.

3 Creswell J, Ottmani SE, Wei D, Sahu S, Lönnroth K, Uplekar MW. Programmatic approaches to screening for active tuberculosis. Int J Tuberc Lung Dis. 2013;**17**:1248–56.

4 Creswell J, Sahu S, Blok L, Bakker MI, Stevens R, Ditiu L. A multi-site evaluation of innovative approaches to increase tuberculosis case notification: Summary results. PLoS One. 2014;**9**.

5 Hanna BA, Ebrahimzadeh A, Elliott LB, Morgan MA, Novak SM, Rusch-Gerdes S, et al. Multicenter

evaluation of the BACTEC MGIT 960 system for recovery of mycobacteria. J Clin Microbiol. 1999;**37**:748–52.

6 Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, et al. Rapid Molecular Detection of Tuberculosis and Rifampin Resistance. N Engl J Med. 2010;**363**:1005–15.

7 van't Hoog AH, Meme HK, Laserson KF, Agaya JA, Muchiri BG, Githui WA, et al. Screening strategies for tuberculosis prevalence surveys: The value of chest radiography and symptoms. PLoS One. 2012;**7**:1–9.

8 World Health Organisation. Tuberculosis prevalence surveys: a handbook. Tuberculosis. Geneva: WHO; 2011.

9 Kisembo HN, Den Boon S, Davis JL, Okello R, Worodria W, Cattamanchi A, et al. Chest radiographic findings of pulmonary tuberculosis in severely immunocompromised patients with the human immunodeficiency virus. Br J Radiol. 2012;**85**:130–9.

10 Kistan J, Laher F, Otwombe K, Panchia R, Mawaka N, Lebina L, et al. Pulmonary TB: Varying radiological presentations in individuals with HIV in Soweto, South Africa. Trans R Soc Trop Med Hyg. 2017;**111**:132–6.

11 van't Hoog AH, Meme HK, Van Deutekom H, Mithika AM, Olunga C, Onyino F, et al. High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. Int J Tuberc Lung Dis. 2011;**15**:1308–14.

12 van Ginneken B, Schaefer-Prokop CM, Prokop M. Computer-aided Diagnosis: How to Move from the Laboratory to the Clinic. Radiology. 2011;**261**:719–32.

13 Muyoyeta M, Maduskar P, Moyo M, Kasese N, Milimo D, Spooner R, 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**:16–8.

14 Breuninger M, Van Ginneken B, Philipsen RHHM, Mhimbira F, Hella JJ, Lwilla F, 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**.

15 Philipsen R, Ginneken B Van, Melendez J. Computer Aided Detection of Tuberculosis CAD4TB. 2018;

16 Qin ZZ, Sander MS, Mph BR, Titahong CN, Sudrungrot S, Laah SN, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. Sci Rep. 2019;1–10.

17 Murphy KM, Travers P, Walport M. Janeway Immunologie. 7th ed. Springer Spektrum; 2014.

18 Rahman MT, Codlin AJ, Rahman MM, Nahar A, Reja M, Islam T, et al. An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients. Eur Respir J. 2017;**49**.

19 Zaidi SMA, Habib SS, Van Ginneken B, Ferrand RA, Creswell J, Khowaja S, et al. Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan. Sci Rep. 2018;**8**:1–9.

20 Koesoemadinata RC, Kranzer K, Livia R, Susilawati N, Annisa J, Soetedjo NNM, et al. Computer-assisted chest radiography reading for tuberculosis screening in people living with diabetes mellitus. Int J Tuberc Lung Dis. 2018;**22**:1088–94.

21 Melendez J, Philipsen R, Chanda-Kapata P, Sunkutu V, Kapata N, van Ginneken B. Automatic versus human reading of chest X-rays in the Zambia National Tuberculosis Prevalence Survey. Int J Tuberc Lung Dis. 2017;**21**:880–6.

22 Pillay D, Koole O, Wong EB. Performance of computer-aided chest radiography for detection of TB in a community-based multimorbidity cohort in rural KwaZulu-Natal [Data set]. Africa Health Res. 2020. https://doi.org/10.23664/AHRI.VUKUZAZI.CHESTXRAYPAPERDATA.

23 Delft. CAD4TB white paper. 2016

24 Hogeweg L, Sánchez CI, Maduskar P, Philipsen R, Story A, Dawson R, et al. Automatic Detection of Tuberculosis in Chest Radiographs Using a Combination of Textural, Focal, and Shape Abnormality Analysis. IEEE Trans Med Imaging. 2015;**34**:2429–42.

25 Murphy K, Habib SS, Zaidi SMA, Khowaja S, Khan A, Melendez J, et al. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. arXiv 2019;1–11.

26 Dorman SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, et al. Xpert MTB/RIF Ultra for

detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. Lancet Infect Dis. 2018;**18**:76–84.

27 Philipsen RHHM, Sánchez CI, Maduskar P, Melendez J, Peters-Bax L, Peter JG, 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**:1–8.

28 World Health Organization. Chest Radiography in Tuberculosis. Geneva; 2016.

29 Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG, et al. Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection. Clin Microbiol Rev. 2018;**31**:1–24.

30 Chakravorty S, Simmons AM, Rowneki M, Parmar H, Cao Y, Ryan J, et al. The new Xpert MTB/RIF ultra: Improving detection of Mycobacterium tuberculosis and resistance to Rifampin in an assay suitable for point-of-care testing. MBio. 2017;**8**:1–12.