Tuberculosis case finding in a population with high HIV prevalence in western Kenya
van 't Hoog, A.H.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5

High sensitivity of Chest Radiograph Reading by Clinical Officers in a Tuberculosis Prevalence Survey


A.H. van’t Hoog¹, H.K. Meme³, H. van Deutekom⁴, A.M. Mithika¹, C.Olunga¹, F.Onyino¹†, M.W. Borgdorff²,².

¹ Kenya Medical Research Institute, KEMRI/CDC Research and Public Health Collaboration, Kisumu, Kenya
² Academic Medical Centre, University of Amsterdam, The Netherlands
³ Kenya Medical Research Institute, Centre for Respiratory Diseases Research, Nairobi, Kenya
⁴ Municipal Health Service, Department of Tuberculosis Control, Amsterdam, The Netherlands

†Deceased
SUMMARY

Background
Chest radiographs (CXRs) are used in tuberculosis (TB) prevalence surveys to identify participants for bacteriological examination. Expert readers are rare in most African countries. In our survey, clinical officers scored CXRs of 19,216 participants once. We assessed to what extent missed CXR abnormalities affected our TB prevalence estimate.

Methods
Two experts, a radiologist and pulmonologist independently reviewed 1031 randomly selected CXRs, mixed with films of confirmed TB cases. CXRs with disagreement on ‘any abnormality’ or ‘abnormality consistent with TB’ were jointly reviewed during a consensus panel. We compared the final expert and clinical officer classifications with bacteriologically confirmed TB as the gold standard.

Results
After the panel, 199 (19%) randomly selected CXRs were considered abnormal by both expert reviewers and another 82 (8%) by one reviewer. Agreement was good among the experts (κ 0.78, 95%CI 0.73-0.82) and moderate between the clinical officers and experts (κ range 0.50-0.62). The sensitivity of ‘any abnormality’ was 95% for the clinical officers, 83% and 81% for the respective experts. The specificities were respectively 73%, 74% and 80%. TB prevalence was underestimated by 1.5%-5.0%.

Conclusions
Acceptable CXR screening can be achieved with clinical officers. Reviewing a sample of CXRs by two experts allows an assessment of prevalence underestimation.
BACKGROUND

Tuberculosis (TB) prevalence surveys are recommended in selected high TB burden countries to measure progress towards the Millennium Development Goals and Stop TB Partnership targets, which recommend a 50% reduction in TB prevalence by 2015 compared to the 1990 levels.1-2 Prevalence surveys require sample sizes ranging from 20,000 to over 200,000, to estimate the disease burden with sufficient precision and to detect significant reductions 10 years later. Most surveys aim to measure the burden of bacteriologically confirmed pulmonary TB. In settings where performing sputum culture examination on all survey participants is not feasible, a screening strategy including chest radiography (CXR) is recommended to select participants for further bacteriological examination of sputum (usually culture and smear).3,4 The recommended practice is that the films be examined in the field by a medical officer5, as soon as the CXR has been obtained, and at a later time at central level by two independent expert readers, a radiologist and a chest physician, and by a third in case of discrepancies.3,4

In sub-Saharan Africa, although TB rates are high, but professionals qualified to read and interpret CXRs are rare. For logistical and cost purposes it would be useful if other medical personnel could be trained to read the screening CXRs in prevalence surveys with sufficient accuracy so that expert interpretation of the presence of radiographic abnormalities consistent with present or past TB could be limited to a smaller number of CXRs.

We conducted a TB prevalence survey in a rural area in western Kenya.6 Participants were screened by CXR, symptom questionnaire and sputum microscopy. For CXR screening, clinical officers rather than medical officers were trained by a radiologist to read CXRs at the study enrolment sites and classify those as either normal or abnormal. We wanted to compare clinical officer and expert classifications, and assess to what extent CXR abnormalities that were missed by the clinical officers may have contributed to underestimation of the reported prevalence6.

METHODS

Survey

A TB prevalence survey was conducted between August 2006 and December 2007 in the Asembo (Rarieda District) and Gem District areas in western Kenya, where human immunodeficiency virus (HIV) prevalence is high.7 The final analysis included 20,566 residents aged ≥15 years from 40 sampled clusters (Figure 1). A posterior-anterior chest
radiograph (CXR) was taken of 19 216 (93%) participants, in a mobile unit at a nearby location, using a 400 mA / 40 KW X-ray machine, 35 x 43 cm films and automatic film processing. Two sputum samples for microscopy were requested from all participants, and an additional sputum sample for mycobacterial culture was requested from suspects (see definition below). The prevalence of bacteriologically confirmed TB was 6.0 per 1000 (95% confidence interval (CI) 4.6-7.4), and of smear-positive TB 2.5 per 1000 (95%CI 1.6-3.4). HIV testing was requested of prevalent TB cases. Of 101 patients tested, 52 (51%) were HIV-infected.

Definitions and classifications

Suspect: survey participant with either a CXR abnormality and/or symptoms suggestive of TB on interview and/or a positive sputum smear microscopy result.

Confirmed TB case: survey participant with bacteriologically confirmed pulmonary TB: either one culture sample positive for Mycobacterium tuberculosis, or two sputum smears positive for acid-fast bacilli not explained by positive culture for nontuberculous mycobacteria.

CXR quality rating (applied by experts): 1 = optimal; 2 = suboptimal, but no technical defect likely to impair classification; 3 = suboptimal with some technical defect but still adequate; or 4 = unreadable.

Readers and training

Three clinical officers with an education level comparable to paramedical and nursing professions, were trained for 2 weeks by a radiologist in assessing CXRs for quality and presence of common abnormalities, and to identify which abnormal conditions required further care. The CXR of a survey participant was reviewed on site on the same day by one clinical officer blinded to participant information other than age and sex. For CXRs classified as abnormal, the clinical officer indicated whether the abnormality was pulmonary, cardiac, pleural effusion and/or ‘other’. The clinical officers rotated, and alternated CXR reading with clinical tasks on site.

Quality assurance and control

The radiographers applied standard quality procedures. After assessment, the CXRs were transported every day to the Kenya Medical Research Institute/Centers for Disease Control and Prevention offices in Kisumu and stored in an air-conditioned
room. Procedures to maintain CXR and clinical officer reading quality included 2-day supervision visits by the radiologist approximately every 6-8 weeks, and a daily ‘warm-up’ whereby the two or three clinical officers present at the field-site jointly reviewed five CXRs. To assess inter- and intra-reader variation, the clinical officers also re-read a random sample of 780 CXRs during the survey. These were selected separate from the sample for expert reading and not shown in Figure 1.

Sample for expert review
From earlier analysis of a non-random sample read by the radiologist (data not shown), we calculated that to estimate the proportion of abnormalities consistent with TB possibly missed by the clinical officers and with a 95% CI of ±10% of the point estimate, 1000 expert readings would be necessary. To allow for non-retrievable or poor quality films, we randomly selected 1050 (5.4%) films from the database with 19,299 CXRs with reports available and mixed these with the CXRs of the confirmed TB cases (Figure 1). One radiologist (expert1, HM) and a pulmonologist (expert2, HD) reviewed the CXRs in March and April 2009. Both were blinded to any participant information, the reason for selection and the other reader’s results. The expert readers used a standard CXR reading and recording system (CRRS) form11, to note the quality rating and to record details on each abnormality and whether the abnormality was considered consistent with TB. During a joint review and consensus panel both expert readers reviewed the CXRs with discrepant results together, blinded to their earlier report and in presence of a rapporteur (AH), who wrote the new results on pre-printed forms. Disagreements were discussed and recorded, including the possible clinical relevance of the abnormality for TB. Reaching consensus was the aim, but was not mandatory.

Data collection and statistical analysis
Field-readings were recorded on scannable forms (Cardiff Teleforms®, Autonomy Cardiff, San Diego, CA, USA), while the experts recorded their results in electronic forms. Results from the consensus panel were manually entered. Data were analyzed with SAS 9.1 (SAS Institute Inc., Cary, NC, USA). To assess inter-observer agreements, we calculated the percentage agreement, and the kappa (κ) with 95% CIs12: κ 0.21-0.40 represented fair agreement, 0.41-0.60 moderate, 0.61-0.80 substantial and 0.81-1.00 almost perfect agreement.13 We calculated the sensitivity and specificity of the clinical officers and experts in identifying TB cases by CXR, using the CXRs of bacteriologically confirmed TB cases as the gold standard positives, and the randomly selected CXRs of participants without TB as the negatives (derived at as shown in Figure 1).
Estimating possibly missed cases

Among the 12,041 non-suspects in the survey who did not have symptoms suggestive of TB, had negative sputum smears and a normal CXR according to the clinical officers, and who were therefore not considered for sputum culture examination, some cases of prevalent TB may have been missed due to CXR abnormalities that were missed by the clinical officers. Additional missed cases due to missing data (sputum samples, CXRs) were not considered for this calculation.

We compared the clinical officer and expert reports of non-suspects among the randomly selected CXRs, and calculated the proportion of CXRs that were classified as normal by the clinical officers, but as having an abnormality consistent with TB according to one of the experts or both experts. We multiplied the proportions of missed abnormalities consistent with TB by the prevalence of smear-negative, culture-positive TB among participants who did not report symptoms suggestive of TB but whose CXR was classified as ‘abnormal’ by the clinical officers, assuming that participants for whom CXR abnormalities were missed by the clinical officers had a TB prevalence similar to that of participants whose abnormalities were not missed. The 95%CI of the product of the two proportions was calculated according to the Delta method. We then applied these proportions to the 12,041 non-suspects to obtain the number of missed prevalent cases.

Ethical approval

The prevalence survey protocol was approved by the scientific and ethical steering committees of the Kenya Medical Research Institute (protocol number 943) and the institutional review board of the US Centers for Disease Control and Prevention (protocol number 4712).

RESULTS

The analysis included 1143 CXRs, 1031 were randomly selected among participants without TB, while 112 were of persons identified with bacteriologically confirmed prevalent TB (Figure 1). The distribution of CXR abnormalities, symptoms suggestive of TB, suspects and prevalent cases was similar in the random sample and the survey (Table 1). Expert 1 rated the quality of 1105 (97%) as suitable for classification: 393 (34%) as optimal, 712 (62%) as ‘2’, and 38 (3%) as ‘3’. Expert 2 rated 1115 CXRs (98%) as suitable for classification: 1024 (90%) as optimal, 91 (8%) as ‘2’, 27 (2%) as ‘3’ and one as unreadable.
Figure 1. Survey procedures and selection of radiographs for expert reading

Survey procedures

Registered participants (n = 20,710)  
⇒ CXRs available (n = 19,299)

Participants with complete interview data included in analysis: (n = 20,566)
- Spot sputum for FM: 20,409 (99%)
- Overnight sputum for FM: 19,788 (96%)
- Chest radiography: 19,216 (93%)

Participants without PTB ('random negatives') (n = 10,319)
- Bacteriologically confirmed cases (n = 112)
- CXRs of bacteriologically confirmed cases in both sets (n = 3)

Bacteriologically confirmed PTB (120 with CXR) (n = 123)
- CXR+ Symptom+: 65 (36 sm+)
- CXR+ Symptom-: 48 (11 sm+)
- CXR- Symptom+: 7 (1 sm+)
- CXR missing Symptom+: 3 (3 sm+)

No PTB (n = 20,433)

Suspect (n = 7,346 (36%))
- CXR+ Symptom+ (n = 1,492)
- CXR+ Symptom- (n = 3,850)
- CXR- Symptom+ (n = 1,831)
- CXR missing Symptom+ (n = 167)

Cultures result available (n = 6,808 (93%))

Random set
CXRs randomly selected: 1,050 (5.4%)
- Could not be retrieved: 11 (1%)
- Excluded because of missing survey data: 5 (0.5%)
⇒ CXRs included: 1,034 (98%)

CXRs included in analysis (n = 1,143):
- Participants without PTB ('random negatives') (n = 1,031)
- Bacteriologically confirmed cases (n = 112)
- CXRs of bacteriologically confirmed cases in both sets (n = 3)

CXR retrieved and available for analysis: 112 (93%)

*Includes 1 smear negative (MGIT) culture positive case
† Sputum results of persons who were unable to provide a sputum sample for microscopy or culture, and of contaminated cultures, were considered negative. 157 participants did not provide any sputum sample. 162 culture samples were contaminated on both media.

CXR+/− = chest radiograph abnormal/normal (field reading); Symptom+/− = Symptoms suggestive / not suggestive of pulmonary tuberculosis; sm+ = smear positive
FM = fluorescence microscopy; PTB = pulmonary tuberculosis; MGIT = Mycobacterial Growth Indicator Tube
### Table 1. Characteristics of all chest radiographs of the survey analysis and of the random sample selected for expert review

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All CXRs in survey</th>
<th>Random sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Total</td>
<td>19 216</td>
<td></td>
</tr>
<tr>
<td>Abnormal by CO</td>
<td>5 342</td>
<td>27.8</td>
</tr>
<tr>
<td>Type of Abnormality (% of those with any abnormality)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary</td>
<td>4 793</td>
<td>89.7</td>
</tr>
<tr>
<td>Cardiac</td>
<td>883</td>
<td>16.5</td>
</tr>
<tr>
<td>Other</td>
<td>1 169</td>
<td>21.9</td>
</tr>
<tr>
<td>Pleural effusion</td>
<td>400</td>
<td>7.5</td>
</tr>
<tr>
<td>Read by†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Officer 1</td>
<td>7 151</td>
<td>37.2</td>
</tr>
<tr>
<td>Abnormal</td>
<td>2 019</td>
<td>28.2</td>
</tr>
<tr>
<td>Clinical Officer 2</td>
<td>7 199</td>
<td>37.5</td>
</tr>
<tr>
<td>Abnormal</td>
<td>2 207</td>
<td>30.7</td>
</tr>
<tr>
<td>Clinical Officer 3</td>
<td>4 813</td>
<td>25.0</td>
</tr>
<tr>
<td>Abnormal</td>
<td>1 100</td>
<td>22.9</td>
</tr>
<tr>
<td>Symptoms suggestive of TB</td>
<td>3 323</td>
<td>17.3</td>
</tr>
<tr>
<td>Non-suspect (Normal CXR, no symptoms suggestive of TB)</td>
<td>12 041</td>
<td>62.7</td>
</tr>
<tr>
<td>Bacteriologically confirmed PTB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smear Positive</td>
<td>48</td>
<td>0.2</td>
</tr>
<tr>
<td>Smear or Culture Positive</td>
<td>120</td>
<td>0.6</td>
</tr>
</tbody>
</table>

*More than one abnormality is possible
†Missing CO code in 53 (0.3%) of all CXRs and 3 (0.3%) of the random sample.
CXR = Chest X-ray; CO = clinical officer; TB = tuberculosis; PTB = pulmonary TB

### Agreement

Of the 1031 ‘random negatives’, 199 (19%) were classified after the consensus panel as abnormal by both experts, and another 82 (8%) by one of the experts. They agreed on the presence of an abnormality consistent with TB in 77 (8%; Table 2). The κ for inter-expert reader agreement on the presence of any abnormality increased from 0.52 (95%CI 0.46-0.58) before, to 0.78 (95%CI 0.73-0.82) after the consensus panel, and for abnormalities consistent with TB from 0.40 (95%CI 0.32-0.49) to 0.79 (95%CI 0.72-0.85). Of the 112 radiographs of confirmed TB cases, the κ for inter-expert reader agreement
Table 2. Agreement between experts on any abnormality and abnormality consistent with PTB in the randomly selected radiographs* and in the radiographs of persons with PTB

<table>
<thead>
<tr>
<th></th>
<th>Before or after consensus panel</th>
<th>Normal - normal</th>
<th>Normal - abnormal</th>
<th>Abnormal - normal</th>
<th>Abnormal - abnormal</th>
<th>κ (95% CI)</th>
<th>Mean pair agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Randomly selected radiographs (n = 1031)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any abnormality?</td>
<td>Before</td>
<td>690</td>
<td>50</td>
<td>134</td>
<td>157</td>
<td>0.52 (0.46-0.58)</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>750</td>
<td>12</td>
<td>70</td>
<td>199</td>
<td>0.78 (0.73-0.82)</td>
<td>92</td>
</tr>
<tr>
<td>Abnormality consistent with TB?</td>
<td>Before</td>
<td>865</td>
<td>33</td>
<td>83</td>
<td>50</td>
<td>0.40 (0.32-0.49)</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>917</td>
<td>9</td>
<td>28</td>
<td>77</td>
<td>0.79 (0.72-0.85)</td>
<td>96</td>
</tr>
<tr>
<td><strong>Radiographs of persons with PTB (n = 112)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any abnormality?</td>
<td>Before</td>
<td>17</td>
<td>2</td>
<td>4</td>
<td>89</td>
<td>0.82 (0.68-0.96)</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>19</td>
<td>0</td>
<td>2</td>
<td>91</td>
<td>0.94 (0.86-1.00)</td>
<td>98</td>
</tr>
<tr>
<td>Abnormality consistent with TB?</td>
<td>Before</td>
<td>22</td>
<td>4</td>
<td>7</td>
<td>79</td>
<td>0.74 (0.59-0.88)</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>25</td>
<td>0</td>
<td>3</td>
<td>84</td>
<td>0.93 (0.84-1.00)</td>
<td>97</td>
</tr>
</tbody>
</table>

*3 persons with *M. tuberculosis* have been excluded from the random sample and are only included in the radiographs of persons with PTB

PTB = pulmonary TB; CI = Confidence Interval; TB = tuberculosis

on any abnormality was almost perfect before and after the consensus panel, and for abnormalities consistent with TB increased from 0.74 (95% CI 0.59-0.88) to 0.93 (95% CI 0.84-1.00; Table 2).
Among the 1031 'random negatives' agreement between the classifications of each of the three clinical officers and 'any abnormality' classified by at least one of the experts was moderate, with small differences for each individual clinical officer (Table 3).

Table 3. Agreement in the random sample on any abnormality between CXR classifications of each clinical officer and an abnormality classified by at least one of the experts (after the consensus panel)

<table>
<thead>
<tr>
<th>Clinical officer</th>
<th>n</th>
<th>Normal-normal</th>
<th>Normal-abnormal</th>
<th>Abnormal-normal</th>
<th>Abnormal-abnormal</th>
<th>κ (95% CI)</th>
<th>Mean pair agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>424</td>
<td>245</td>
<td>30</td>
<td>42</td>
<td>107</td>
<td>0.62</td>
<td>(0.54-0.70)</td>
</tr>
<tr>
<td>2</td>
<td>432</td>
<td>240</td>
<td>35</td>
<td>52</td>
<td>105</td>
<td>0.55</td>
<td>(0.47-0.64)</td>
</tr>
<tr>
<td>3</td>
<td>284</td>
<td>165</td>
<td>39</td>
<td>23</td>
<td>57</td>
<td>0.49</td>
<td>(0.38-0.60)</td>
</tr>
<tr>
<td>All 3</td>
<td>1031</td>
<td>648</td>
<td>103</td>
<td>102</td>
<td>178</td>
<td>0.50</td>
<td>(0.44-0.56)</td>
</tr>
</tbody>
</table>

* Expert classification is considered abnormal if either one or both experts classified the CXR as abnormal.

CXR = Chest radiograph; CI = Confidence interval
Of 780 CXRs selected for quality control of the clinical officer field reading, 655 (84%) had rereading results of all three clinical officers, of which 649 (99%) could be linked to the field-reading results. The κ for inter-reader agreement between each pair of the three clinical officers ranged from 0.57 to 0.63 (Table 4).

### Table 4. Inter-reader agreement between the clinical officers and intra-reader agreement for each clinical officer on any abnormality.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Matched</th>
<th>Normal - normal</th>
<th>Normal - abnormal</th>
<th>Abnormal - normal</th>
<th>Abnormal - abnormal</th>
<th>κ (95% CI)</th>
<th>Mean pair agreement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-reader agreement</td>
<td>1 vs. 2</td>
<td>655</td>
<td>371</td>
<td>64</td>
<td>46</td>
<td>174</td>
<td>0.63 (0.57-0.69)</td>
</tr>
<tr>
<td></td>
<td>1 vs. 3</td>
<td>655</td>
<td>383</td>
<td>52</td>
<td>70</td>
<td>150</td>
<td>0.57 (0.51-0.65)</td>
</tr>
<tr>
<td></td>
<td>2 vs. 3</td>
<td>655</td>
<td>377</td>
<td>40</td>
<td>76</td>
<td>162</td>
<td>0.60 (0.54-0.67)</td>
</tr>
<tr>
<td>Intra-reader agreement</td>
<td>1</td>
<td>226</td>
<td>122</td>
<td>16</td>
<td>26</td>
<td>62</td>
<td>0.60 (0.49-0.71)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>210</td>
<td>117</td>
<td>18</td>
<td>22</td>
<td>53</td>
<td>0.58 (0.46-0.70)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>197</td>
<td>112</td>
<td>16</td>
<td>33</td>
<td>36</td>
<td>0.43 (0.30-0.56)</td>
</tr>
</tbody>
</table>

CI=Confidence Interval
Sensitivity and specificity
The sensitivity of the clinical officer readings (all three combined) in classifying CXRs of bacteriologically confirmed TB cases as abnormal was 95% (95%CI 89-98). For the expert readings prior to the consensus panel these were 83% (95%CI 75-89) and 81% (95% CI 73-88) respectively when any abnormality was considered. Considering only abnormalities consistent with TB further reduced the sensitivity, but had the highest specificity: respectively 87% (95%CI 85-89) and 92% (95%CI 90-94; Table 5). The sensitivity differed slightly by HIV status but the 95%CIs overlapped. The sensitivity of the clinical officers’ readings was 94% (95%CI 82-99) in HIV-infected and 100% (95%CI 92-100) in non-HIV-infected TB cases. The lowest expert values for any abnormality were respectively 83% (95%CI 69-92) and 80% (95%CI 66-91).

Table 5. Sensitivity and specificity of clinical officers and of each expert’s CXR classifications for ‘any abnormality’ and ‘abnormality consistent with TB’ (before consensus panel) in identifying bacteriologically confirmed TB

<table>
<thead>
<tr>
<th></th>
<th>Before consensus panel</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity %</td>
<td>Specificity %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(95% CI)</td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>Any abnormality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical officers (all 3) field reading</td>
<td>95 (89-98)</td>
<td>73 (70-76)</td>
<td></td>
</tr>
<tr>
<td>Expert 1</td>
<td>83 (75-89)</td>
<td>72 (69-75)</td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>81 (73-88)</td>
<td>80 (77-82)</td>
<td></td>
</tr>
<tr>
<td>Abnormality consistent with TB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 1</td>
<td>77 (68-84)</td>
<td>87 (85-89)</td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>74 (65-82)</td>
<td>92 (90-94)</td>
<td></td>
</tr>
</tbody>
</table>

CXR = Chest radiograph; TB = tuberculosis; CI = Confidence interval.

Estimating missed cases by clinical officers
Of the 652 CXRs of non-suspects in the randomly selected CXRs, 24 (3.7%; 95%CI 2.2-5.1) were classified by at least one of the experts as having an abnormality consistent with TB and 13 (2.0%; 95% CI 0.9-3.1) by both experts, after the consensus panel. Among 3850 suspects in the total survey who did not report symptoms suggestive of TB, had negative sputum smears, and whose CXR was classified as abnormal by the clinical officers, 43 had a positive sputum culture (prevalence 1.12/100, 95%CI 0.67-1.56). Among the 12 041 non-suspects in whom CXR abnormalities might have been missed, this would
translate to \( \frac{24}{652} \times \frac{43}{3850} \times 12041 = 5.0 \) possibly missed cases (95%CI 3.5-6.4) considering CXRs with an abnormality consistent with TB according to at least one expert and thus a prevalence estimate under-reported by 3.9% (95%CI 2.8-5.0). Considering CXRs classified by both experts as having an abnormality consistent with TB, \( \frac{13}{652} \times \frac{43}{3850} \times 12041 = 2.7 \) prevalent cases would have been missed (95%CI 1.9-3.5), reflecting underreporting by 2.1% (95%CI 1.5-2.8). If expert classifications before the consensus panel were used, the estimated prevalence underestimation would range between 7.7% (95%CI 5.6-9.8) and 0.5% (95%CI 0.4-0.6, calculations not shown).

Among the 112 CXRs of confirmed TB cases, the clinical officers classified six as normal, and missed one (0.9%) abnormality consistent with TB (according to both experts) in a smear-positive case, while one CXR showed a non-TB related abnormality according to one expert in a smear-negative case. Thirteen (11%) CXRs were classified as normal by both experts, and as abnormal by the clinical officers. Of these TB cases, five were HIV-infected, seven were non-HIV-infected two were smear-positive. Ten reported ≥1 TB symptom while three smear-negative non-HIV-infected cases did not report TB symptoms. Four CXRs were considered normal by all readers. All four TB cases had ≥2 TB symptoms, two were HIV-infected with low CD4 cell counts, and HIV status was unavailable for two.

**DISCUSSION**

The sensitivity of the clinical officers’ detection of abnormalities on CXRs of bacteriologically confirmed cases was 95%. The underestimation of prevalence due to CXR screening by clinical officers was in the range of 1.5-5.0%, based on a review of a random sample of CXRs by two experts followed by a consensus debate.

The uncertainty range of the prevalence underestimation decreased after the consensus panel, and could have been further reduced by a third expert referee reader and by a larger random sample of CXRs for expert reading. However the uncertainty range of 1.5%-5.0% is small compared to the 95% CI of the overall prevalence estimate of the survey, determined by the overall survey design, which was ±23% of the point estimate.

The high sensitivity of the clinical officers in classifying CXRs of confirmed TB cases as abnormal was reassuring, as were the levels of inter-reader agreement.\(^{11, 15}\) The lower sensitivity of the experts is in line with other reports, also considering an expected reduced sensitivity in this high HIV prevalence population.\(^{16-19}\) The clinical officers were
trained to recognize common abnormalities, consider a CXR abnormal in case of doubt, and to consider all abnormalities regardless of their possible relevance for TB. This resulted in low specificity (71%), also reported in mass screening campaigns and in the clinical TB diagnostic process. Consequently, a large proportion of survey participants requires culture, which has considerable logistical and cost implications. From our data, CXR screening by experts would have increased specificity above 90% if only those abnormalities consistent with TB were considered. However, sensitivity would decrease to below 80%, resulting in a larger underestimation of prevalence.

As the availability of physicians for TB diagnosis in clinical settings is limited in several African countries, CXRs reading by clinical officers or other cadres of health worker could also be applied to select suspects for further bacteriological examination with newer rapid, sensitive, but still expensive diagnostic test like GenXpert (Cepheid, Sunnyvale, CA, USA). A potential concern is confusion between sensitive reading for screening purposes with the more specific approach for clinical diagnosis of smear-negative TB in absence of additional bacteriological testing, to limit false-positive diagnoses; this limitation could, however, be addressed by training, supervision, quality assurance and operations research.

Our method of estimating missed cases has not been evaluated in a survey where all participants had sputum culture and CXRs read by multiple experts and clinical officers. We may have overestimated sensitivity and underestimated missed cases, due either to cases being missed by the study design, as smear-negative culture-positive TB cases with negative symptoms and CXR on screening were not captured by our survey, or to false-positive TB diagnosis in asymptomatic participants with an ‘over-read’ CXR abnormality. There were few of the latter, however. CXR reading by experts had 97% sensitivity (95%CI 90–100) in a survey conducted in Cape Town, where all participants were eligible for CXR and sputum culture. Conversely, our assumption that, among participants with a negative symptom screen, the prevalence of smear-negative culture-positive TB was similar in participants whose abnormalities were missed and in those whose abnormalities were not missed may overstate the estimate of missed cases. Missed abnormalities may be less severe and less likely be associated with bacteriologically active TB than more pronounced abnormalities. The interpretation of the reported κ statistic requires some caution, as it is influenced by the prevalence of the assessed condition.
Other aspects that could be improved include reading of a standard set of CXRs by experts, as is done in the case of the International Labour Organization. The clinical officers were trained with a set of CXRs prepared by the radiologist, but the experts did not review the same set. The optimal reporting form also remains to be determined. The CRRS form, developed for community surveys, allows for detailed reporting, which may increase specificity, but also requires more time than the simpler classification used by the clinical officers. We used conventional radiography in our survey. Digital equipment could make training and supervision easier, in particular if communication with off-site experts is available. Digital equipment also allows the evaluation of computerised systems to assess CXR abnormalities for their usefulness in prevalence surveys.

Ultimately, a simple diagnostic test would eliminate the need for screening CXRs. However, in the interim, an acceptable level of CXR screening for ‘any abnormality’ can be achieved by clinical officers, with sufficient training and standard operating procedures to maintain quality of CXR reading. The degree to which prevalence is underestimated can be assessed by review of a sample of CXRs by two experts. We recommend this approach for TB prevalence surveys planned in similar settings and further evaluation of clinical applicability in combination with new bacteriological tests.
ACKNOWLEDGEMENTS

We acknowledge the staff of the KEMRI/CDC Research and Public Health Collaboration, the KEMRI Centre for Respiratory Diseases Research, the study participants and the study team in general for their contributions. In particular we thank the radiographers Moses Kojo and Agneta Monari for taking the CXRs, the data staff Lazarus Odeny, Elizabeth Kendi and John Hongo and study coordinator Janet Agaya for facilitating the expert reading exercise. This paper is published with the approval of the Director, KEMRI.

Sources of support

This assessment was part of the operations of the TB prevalence survey, supported by US President’s Emergency Plan for AIDS Relief and by the United States Agency for International Development through John’s Hopkins University.

Disclaimer

The views expressed in this manuscript do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices or organizations imply endorsement by the US government.
REFERENCES


