

# Tuberculosis screening among ambulatory people living with HIV: a systematic review and individual participant data meta-analysis



Ashar Dhana, Yohhei Hamada, Andre P Kengne, Andrew D Kerkhoff, Molebogeng X Rangaka, Tamara Kredo, Annabel Baddeley, Cecily Miller, Satvinder Singh, Yasmeen Hanifa, Alison D Grant, Katherine Fielding, Dissou Affolabi, Corinne S Merle, Ablo Prudence Wachinou, Christina Yoon, Adithya Cattamanchi, Christopher J Hoffmann, Neil Martinson, Eyongetah Tabenyang Mbu, Melissa S Sander, Taye T Balcha, Sten Skogmar, Byron W P Reeve, Grant Theron, Gcobisa Ndlangalavu, Surbhi Modi, Joseph Cavanaugh, Susan Swindells, Richard E Chaisson, Faiz Ahmad Khan, Andrea A Howard, Robin Wood, Swe Swe Thit, Mar Mar Kyi, Josh Hanson, Paul K Drain, Adrienne E Shapiro, Tendesayi Kufa, Gavin Churchyard, Duc T Nguyen, Edward A Graviss, Stephanie Bjerrum, Isik S Johansen, Jill K Gersh, David J Horne, Sylvia M LaCourse, Haider Abdulrazaq Abed Al-Darraj, Adeeba Kamarulzaman, Russell R Kempker, Nestani Tukvadze, David A Barr, Graeme Meintjes, Gary Maartens

## Summary

**Background** The WHO-recommended tuberculosis screening and diagnostic algorithm in ambulatory people living with HIV is a four-symptom screen (known as the WHO-recommended four symptom screen [W4SS]) followed by a WHO-recommended molecular rapid diagnostic test (eg Xpert MTB/RIF [hereafter referred to as Xpert]) if W4SS is positive. To inform updated WHO guidelines, we aimed to assess the diagnostic accuracy of alternative screening tests and strategies for tuberculosis in this population.

**Methods** In this systematic review and individual participant data meta-analysis, we updated a search of PubMed (MEDLINE), Embase, the Cochrane Library, and conference abstracts for publications from Jan 1, 2011, to March 12, 2018, done in a previous systematic review to include the period up to Aug 2, 2019. We screened the reference lists of identified pieces and contacted experts in the field. We included prospective cross-sectional, observational studies and randomised trials among adult and adolescent (age  $\geq 10$  years) ambulatory people living with HIV, irrespective of signs and symptoms of tuberculosis. We extracted study-level data using a standardised data extraction form, and we requested individual participant data from study authors. We aimed to compare the W4SS with alternative screening tests and strategies and the WHO-recommended algorithm (ie, W4SS followed by Xpert) with Xpert for all in terms of diagnostic accuracy (sensitivity and specificity), overall and in key subgroups (eg, by antiretroviral therapy [ART] status). The reference standard was culture. This study is registered with PROSPERO, CRD42020155895.

**Findings** We identified 25 studies, and obtained data from 22 studies (including 15 666 participants; 4347 [27.7%] of 15 663 participants with data were on ART). W4SS sensitivity was 82% (95% CI 72–89) and specificity was 42% (29–57). C-reactive protein ( $\geq 10$  mg/L) had similar sensitivity to (77% [61–88]), but higher specificity (74% [61–83];  $n=3571$ ) than, W4SS. Cough (lasting  $\geq 2$  weeks), haemoglobin ( $< 10$  g/dL), body-mass index ( $< 18.5$  kg/m<sup>2</sup>), and lymphadenopathy had high specificities (80–90%) but low sensitivities (29–43%). The WHO-recommended algorithm had a sensitivity of 58% (50–66) and a specificity of 99% (98–100); Xpert for all had a sensitivity of 68% (57–76) and a specificity of 99% (98–99). In the one study that assessed both, the sensitivity of sputum Xpert Ultra was higher than sputum Xpert (73% [62–81] vs 57% [47–67]) and specificities were similar (98% [96–98] vs 99% [98–100]). Among outpatients on ART (4309 [99.1%] of 4347 people on ART), W4SS sensitivity was 53% (35–71) and specificity was 71% (51–85). In this population, a parallel strategy (two tests done at the same time) of W4SS with any chest x-ray abnormality had higher sensitivity (89% [70–97]) and lower specificity (33% [17–54];  $n=2670$ ) than W4SS alone; at a tuberculosis prevalence of 5%, this strategy would require 379 more rapid diagnostic tests per 1000 people living with HIV than W4SS but detect 18 more tuberculosis cases. Among outpatients not on ART (11 160 [71.8%] of 15 541 outpatients), W4SS sensitivity was 85% (76–91) and specificity was 37% (25–51). C-reactive protein ( $\geq 10$  mg/L) alone had a similar sensitivity to (83% [79–86]), but higher specificity (67% [60–73];  $n=3187$ ) than, W4SS and a sequential strategy (both test positive) of W4SS then C-reactive protein ( $\geq 5$  mg/L) had a similar sensitivity to (84% [75–90]), but higher specificity than (64% [57–71];  $n=3187$ ), W4SS alone; at 10% tuberculosis prevalence, these strategies would require 272 and 244 fewer rapid diagnostic tests per 1000 people living with HIV than W4SS but miss two and one more tuberculosis cases, respectively.

**Interpretation** C-reactive protein reduces the need for further rapid diagnostic tests without compromising sensitivity and has been included in the updated WHO tuberculosis screening guidelines. However, C-reactive protein data were scarce for outpatients on ART, necessitating future research regarding the utility of C-reactive protein in this group. Chest x-ray can be useful in outpatients on ART when combined with W4SS. The WHO-recommended algorithm has suboptimal sensitivity; Xpert for all offers slight sensitivity gains and would have major resource implications.

**Lancet Infect Dis 2021**

Published Online  
November 17, 2021  
[https://doi.org/10.1016/S1473-3099\(21\)00387-X](https://doi.org/10.1016/S1473-3099(21)00387-X)

Department of Medicine (A Dhana MBBCh, Prof G Meintjes PhD, Prof G Maartens MMed), Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine (M X Rangaka PhD, Prof G Meintjes, Prof G Maartens), Faculty of Health Sciences (Prof R Wood DSc), University of Cape Town, Cape Town, South Africa; Centre for International Cooperation and Global Tuberculosis Information, The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan (Y Hamada MD); Institute for Global Health, University College London, London, UK (Y Hamada, M X Rangaka); Non-communicable Diseases Research Unit (Prof A P Kengne PhD) and Cochrane South Africa (T Kredo PhD), South African Medical Research Council, Cape Town, South Africa; Division of HIV, Infectious Diseases and Global Medicine, Zuckerberg San Francisco General Hospital and Trauma Center (A D Kerkhoff MD) and Department of Medicine, Division of Pulmonary and Critical Care Medicine, Center for Tuberculosis (C Yoon MD, Prof A Cattamanchi MD), University of California, San Francisco, CA, USA; Division of Clinical Pharmacology, Department of

Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa (T Kreda PhD); Global Tuberculosis Programme (A Baddeley MPH, C Miller PhD) and Global HIV, Hepatitis and STIs Programme (S Singh MD), World Health Organization, Geneva, Switzerland; TB Centre, London School of Hygiene & Tropical Medicine, London, UK (Y Hanifa PhD, Prof A Grant PhD, Prof K Fielding PhD); Africa Health Research Institute, School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa (Prof A Grant); School of Public Health (Prof A Grant, T Kufa PhD, Prof G Churchyard MBBCh) and Perinatal HIV Research Unit (N Martinson MBBCh), University of the Witwatersrand, Johannesburg, South Africa; Laboratoire de Référence des Mycobactéries, Cotonou, Benin (Prof D Affolabi PhD); UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, Geneva, Switzerland (C S Merle PhD); National Hospital for Tuberculosis and Pulmonary Diseases, Cotonou, Benin (A P Wachinou MD); Johns Hopkins University School of Medicine, Baltimore, MD, USA (C J Hoffmann MD); Johns Hopkins University Center for Tuberculosis Research, Baltimore, MD, USA (N Martinson, Prof R E Chaisson MD); Bamenda Regional Hospital HIV Treatment Center, Bamenda, Cameroon (E T Mbu MD); Tuberculosis Reference Laboratory Bamenda, Bamenda, Cameroon (M S Sander PhD); Clinical Infection Medicine (T T Balcha MD), Department of Translational Medicine (T T Balcha, S Skogmar PhD), Lund University, Malmö, Sweden; DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch

Funding World Health Organization.

**Copyright** © 2021 World Health Organization; licensee Elsevier. This is an Open Access article published under the CC BY 3.0 IGO license which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In any use of this article, there should be no suggestion that WHO endorses any specific organisation, products or services. The use of the WHO logo is not permitted. This notice should be preserved along with the article's original URL.

## Introduction

Tuberculosis is the leading cause of death among people living with HIV and often goes undiagnosed.<sup>1,2</sup> One approach to reduce this tuberculosis burden involves systematic screening as part of an intensified

case-finding strategy. WHO recommends a tuberculosis screening and diagnostic algorithm in people living with HIV at each clinical encounter using the WHO-recommended four-symptom screen (W4SS; comprising any one of current cough, fever, night sweats, or weight

## Research in context

### Evidence before this study

Tuberculosis is common and often goes undiagnosed in people living with HIV. The WHO-recommended four-symptom screen (W4SS; comprising any one of current cough, fever, night sweats, or weight loss) was developed after a 2011 individual participant data meta-analysis to rule out active tuberculosis before initiating tuberculosis preventive therapy. WHO recommends that ambulatory people living with HIV be screened for tuberculosis at each clinical encounter with the W4SS followed by a WHO-recommended molecular rapid diagnostic test (eg, Xpert MTB/RIF [Xpert] or Xpert MTB/RIF Ultra [Xpert Ultra]) for those with a positive W4SS. In 2018, WHO commissioned an updated systematic review and meta-analysis because the earlier meta-analysis primarily comprised people living with HIV not on antiretroviral therapy (ART). The updated review showed that the W4SS had a specificity of only 27% in people living with HIV not on ART. Thus, large numbers of people ultimately undergo unnecessary and expensive WHO-recommended rapid diagnostic testing. The updated review also found that the W4SS had a sensitivity of only 51% in people living with HIV on ART (vs 89% for those not on ART). Furthermore, the entire WHO-recommended algorithm (W4SS then Xpert) might be suboptimal because its sensitivity depends on both the W4SS and Xpert. Alternative screening tests to the W4SS need to be explored. Several recent studies have shown that C-reactive protein might have improved diagnostic accuracy compared with W4SS. The 2018 WHO-commissioned systematic review assessed the addition of chest x-ray to W4SS, but identified few studies, particularly among those on ART. Other screening tests (eg, haemoglobin and body-mass index [BMI]) are also known to be associated with tuberculosis, but their diagnostic accuracy is unclear. Finally, some experts have argued for an Xpert for all strategy (as opposed to the WHO-recommended algorithm) to improve sensitivity.

### Added value of this study

To inform an update to WHO guidelines on tuberculosis screening among ambulatory people living with HIV regardless of signs and symptoms of tuberculosis, we did an individual participant data meta-analysis of 22 studies and did pre-specified subgroup analyses, notably by ART status. We found that the

W4SS had a specificity of 37% in people not on ART, and a sensitivity of 53% in people on ART (vs 85% in people not on ART). We found that C-reactive protein ( $\geq 10$  mg/L cutoff) had comparable sensitivity to W4SS but higher specificity, leading to fewer rapid diagnostic tests being needed. Chest x-ray had lower sensitivity than W4SS in studies that directly compared both tests, making it unsuitable as a standalone screening test. Cough (lasting  $\geq 2$  weeks), haemoglobin ( $< 10$  g/dL), body-mass index ( $< 18.5$  kg/m<sup>2</sup>), and lymphadenopathy had high specificities, but their low sensitivities also made them unsuitable as screening tests. Xpert for all slightly improved sensitivity compared with the WHO-recommended algorithm (W4SS followed by Xpert). In one study, Xpert Ultra improved sensitivity over Xpert (73% vs 57%). ART status had a major effect on the diagnostic accuracy of W4SS and C-reactive protein, both of which had lower sensitivities but higher specificities among outpatients on ART. Among outpatients on ART, the best performing screening strategy to improve sensitivity was a parallel strategy (two screening tests offered at the same time) of W4SS with any chest x-ray abnormality. Among outpatients not on ART, the best performing screening strategy to improve specificity was C-reactive protein ( $\geq 10$  mg/L) as a standalone test or a sequential strategy (second screening test offered only if first screening test is positive) of W4SS then C-reactive protein ( $\geq 5$  mg/L).

### Implications of all the available evidence

Compared with W4SS, C-reactive protein reduces the need for further rapid diagnostic tests without compromising sensitivity, and it has been included in the updated WHO tuberculosis screening guidelines. However, data on use of C-reactive protein in outpatients on ART were scarce. In outpatients on ART, chest x-ray could be used in parallel with W4SS, depending on available resources, because this strategy detects more tuberculosis cases than W4SS alone. The current WHO-recommended algorithm (W4SS followed by Xpert) is insufficiently sensitive to identify all tuberculosis cases. Xpert for all would offer slight gains over this strategy in terms of sensitivity, but would be resource intensive. Future research is needed to assess the utility of C-reactive protein in outpatients on ART and Xpert Ultra in all people living with HIV.

loss) followed by confirmatory testing using a WHO-recommended molecular rapid diagnostic test such as Xpert MTB/RIF (referred to hereon as Xpert) or Xpert MTB/RIF Ultra (referred to hereon as Xpert Ultra) for those with a positive W4SS.<sup>3,4</sup> However, the W4SS has low specificity, meaning many people require unnecessary and expensive confirmatory testing with a rapid diagnostic test.<sup>3,5</sup> Furthermore, the W4SS has reduced sensitivity in specific subgroups (eg, those who are on antiretroviral therapy [ART], are pregnant, or have high CD4 counts).<sup>3,5,6</sup> The entire algorithm might also have low sensitivity,<sup>7</sup> because overall sensitivity depends on the combined sensitivity of the W4SS and the rapid diagnostic test.

Alternative screening tests to the W4SS need to be explored. According to WHO, a screening test should have a sensitivity of more than 90% and a specificity of more than 70%.<sup>8</sup> Several studies have shown that C-reactive protein has improved diagnostic accuracy compared with W4SS.<sup>9–11</sup> C-reactive protein assays as point-of-care assays are easy to use, inexpensive (approximately US\$2 per test), and provide rapid results (<3 min). One study among people living with HIV initiating ART found that replacing the W4SS with C-reactive protein (10 mg/L) could halve the number of Xpert tests performed.<sup>7</sup> Chest x-ray might also be useful for tuberculosis screening, especially when combined with the W4SS in people living with HIV on ART;<sup>5</sup> however, it is often unavailable and resource intensive. Haemoglobin, body-mass index (BMI), and lymphadenopathy are other predictors of tuberculosis,<sup>12,13</sup> but their diagnostic accuracy is unclear. The authors of some studies among people living with HIV initiating ART have argued that Xpert for all, rather than Xpert only for those who are positive on the W4SS, should be the preferred strategy.<sup>14,15</sup> This approach could optimise diagnostic yield, but cost and capacity issues could restrict its implementation in resource-poor settings. The Alere Determine TB-LAM (AlereLAM) Ag lateral flow urine assay for screening outpatients living with HIV has been recently reviewed and has a sensitivity of 31% and specificity of 95%.<sup>16</sup> Next-generation assays based on detection of lipoarabinomannan (eg, Fujifilm SILVAMP TB-LAM) have higher sensitivity (eg, 71%).<sup>17</sup> WHO recommends the AlereLAM assay if an outpatient has a positive W4SS, CD4 count of 100 cells per  $\mu$ L or lower, is WHO clinical stage 3 or 4, or has a WHO-defined danger sign.<sup>18</sup>

We did a systematic review and individual participant data meta-analysis to provide a more detailed and precise analysis of the accuracy of different tuberculosis screening tests and strategies compared with W4SS among ambulatory people living with HIV, including key subgroups. We also assessed the accuracy of the WHO-recommended screening and diagnostic algorithm (W4SS followed by Xpert) and compared its accuracy with Xpert for all as the first screening test.

## Methods

### Search strategy and selection criteria

In this systematic review and individual participant data meta-analysis, we updated the systematic review done by Hamada and colleagues,<sup>5</sup> who searched PubMed (MEDLINE), Embase, the Cochrane Library, and conference abstracts (from the Conference on Retroviruses and Opportunistic Infections, AIDS/International AIDS Society, and International Union Against TB and Lung Diseases conferences) without language or geographical restrictions from Jan 1, 2011, to March 12, 2018. The start date restrictions correspond to the year WHO issued recommendations on the W4SS. We rescreened all potential full texts identified via Hamada and colleagues' search to identify eligible studies. Additionally, we applied the same search strategy to the same databases for publications between March 12, 2018, and Aug 2, 2019. We also screened reference lists of reviews and included articles and contacted field experts. Detailed search terms are in the appendix (p 2).

Two authors (AD and YHam) independently screened titles and abstracts from the search and subsequently screened the full texts of potentially eligible articles. For abstracts that were not in English, we used Google Translate to translate the abstracts before screening. We included prospective cross-sectional studies, prospective observational studies, and randomised trials that collected at least one sputum sample for tuberculosis culture from adult and adolescent (ie, aged  $\geq 10$  years) ambulatory people living with HIV regardless of signs and symptoms of tuberculosis. We excluded case-control studies, general community or household contact-screening studies, and studies that involved people living with HIV who were already on tuberculosis treatment or had a current tuberculosis diagnosis.

The target condition was active tuberculosis (ie, we excluded articles on latent tuberculosis infections). The reference standard for confirmed tuberculosis was bacteriological confirmation of *Mycobacterium tuberculosis* using culture of a sputum sample or other samples, or both.

We included primary datasets that had sufficient data to allow us to compare the W4SS with alternative screening tests or strategies and the WHO-recommended algorithm (W4SS followed by Xpert) with Xpert for all. We examined several systematically performed screening tests: C-reactive protein, chest x-ray, Xpert or Xpert Ultra, haemoglobin, BMI, lymphadenopathy (on examination), and cough (lasting  $\geq 2$  weeks). A positive chest x-ray was defined by the authors of the included studies and categorised as any abnormality or abnormality suggestive of tuberculosis. We were primarily interested in any abnormality on chest x-ray because identification of features suggestive of tuberculosis on chest x-ray requires a skilled reader. For C-reactive protein, we primarily focused on the 10 mg/L threshold, which is considered the upper limit of normal.<sup>19,20</sup> We also explored a 5 mg/L

University, Cape Town, Western Cape, South Africa (B W P Reeve PhD, Prof G Theron PhD, G Ndlangalavu MSc); US Centers for Disease Control and Prevention, Atlanta, GA, USA (S Modi MD, J Cavanaugh MD); University of Nebraska Medical Center, Omaha, NE, USA (Prof S Swindells MBBS); McGill International Tuberculosis Centre, Research Institute of the McGill University Health Centre, Montreal, QC, Canada (F Ahmad Khan MD); ICAP at Columbia University, New York, NY, USA (A A Howard MD); Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA (A A Howard); Department of Medicine, University of Medicine 2, Yangon, Yangon Division, Myanmar (Prof S S Thit FRCP, Prof M M Kyi FRCP); The Kirby Institute, University of New South Wales, Sydney, NSW, Australia (J Hanson FRACP); Department of Global Health (Prof P K Drain MD, A E Shapiro MD), Department of Medicine (Prof P K Drain, A E Shapiro), and Department of Epidemiology (Prof P K Drain), University of Washington, Seattle, WA, USA; Centre for HIV and STIs, National Institute for Communicable Diseases, Johannesburg, South Africa (T Kufa PhD); The Aurum Institute, Parktown, South Africa (Prof G Churchyard); Department of Pathology and Genomic Medicine, Houston Methodist Research Institute, Houston, TX, USA (D T Nguyen PhD, E A Graviss PhD); Department of Clinical Research, Infectious Diseases (S Bjerrum PhD), and Research Unit for Infectious Diseases, Odense University Hospital (Prof I S Johansen PhD), University of Southern Denmark, Odense, Denmark; Department of Medicine (D J Horne MD, S M LaCourse MD), and Department of Global Health, Division of Infectious Diseases (S M LaCourse MD), University of Washington, Seattle, WA, USA (J K Gersh MD); Centre of Excellence for Research in AIDS,



University of Malaya, Kuala Lumpur, Malaysia (H A A Al-Darraj MBChB, Prof A Kamarulzaman FRACP); Department of Medicine, Division of Infectious Diseases, Emory University School of Medicine, Atlanta, GA, USA (R R Kempker MD); National Center for Tuberculosis and Lung Diseases, Tbilisi, Georgia (N Tukvadze MD); Institute of Infection and Global Health, University of Liverpool, Liverpool, UK (D A Barr MBChB)

Correspondence to: Prof Gary Maartens, Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town 7935, South Africa [gary.maartens@uct.ac.za](mailto:gary.maartens@uct.ac.za)

See Online for appendix

threshold to maximise sensitivity and an 8 mg/L threshold because a previous study found that this cutoff met WHO's minimum sensitivity ( $\geq 90\%$ ) and specificity ( $\geq 70\%$ ) targets.<sup>8,11</sup> Finally, we examined several parallel strategies (two screening tests offered at the same time) to improve sensitivity and sequential strategies (second screening test offered only if first screening test is positive) to improve specificity.

We have reported our findings according to the PRISMA-IPD and PRISMA-DTA statements.<sup>21,22</sup> This study was registered with PROSPERO (CRD42020155895).

### Data extraction, study quality, and individual participant data synthesis

Using a standardised data extraction form, two authors (AD and YHam) independently extracted study-level information on first author, publication year, study period, country, setting (eg, HIV clinic, hospital clinic, prison clinic), exclusion criteria, study design, type of participants (eg, all people living with HIV, only pregnant people), and method of tuberculosis diagnosis. Two authors independently (AD and YHam) assessed study quality using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.<sup>23</sup>

We invited authors of eligible datasets by email to contribute individual participant data. We prespecified variables to be collected after consultation with WHO and our study group (appendix p 3). We standardised individual participant data, then synthesised a single dataset with study-level data. Study participants younger than 10 years were excluded, and contaminated cultures were considered negative. To ensure integrity of the individual participant data, we checked information against study publications and did checks on each dataset for missing, duplicate, invalid, and implausible items.<sup>24,25</sup> We resolved discrepancies by contacting the corresponding author.

### Statistical analysis

We did analyses overall and in key subgroups, comprising outpatient clinic attendees (on ART vs not on ART), CD4 count ( $\leq 200$  vs  $> 200$  cells per  $\mu\text{L}$ ), and pregnancy. To analyse individual participant data we used a two-stage approach. Individual participant data were first analysed separately in each study using an appropriate statistical method (accounting for the design of data collection) and reduced to aggregate data, which were then synthesised using meta-analytical techniques.

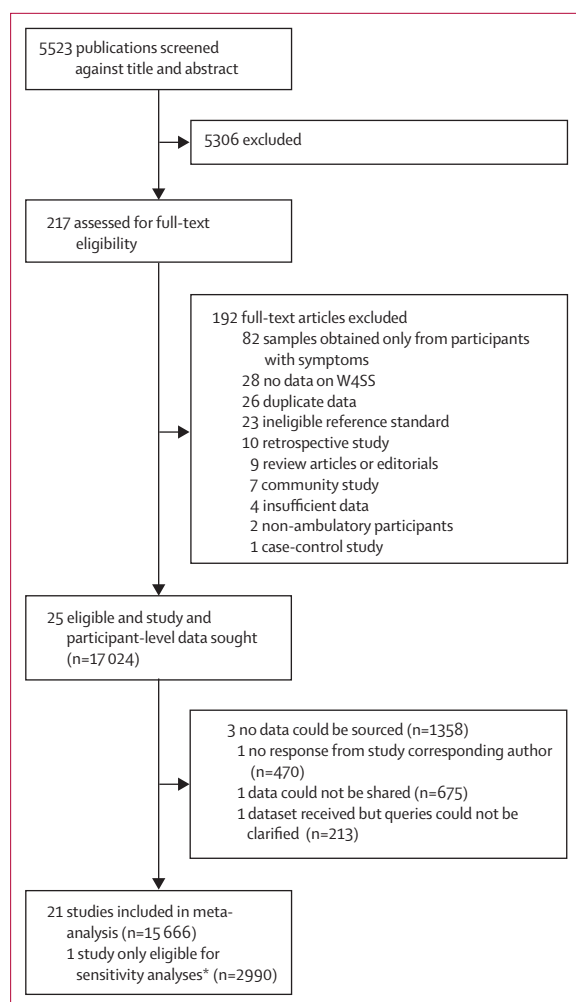
In the first stage, we estimated tuberculosis prevalence, positivity rate (proportion of screen-positive participants), and measures of diagnostic performance (including sensitivity and specificity) by screening test or strategy. In the second stage, we pooled tuberculosis prevalence and positivity rates using a generalised linear mixed model with logit transformation<sup>26</sup> in preference to the protocol specified DerSimonian and Laird random effects model for proportions with variance stabilisation by applying the Freeman-Tukey double arcsine transformation. We

assessed heterogeneity using Cochran's Q test and the  $I^2$  statistic.<sup>27</sup> We pooled absolute accuracy measures (sensitivity, specificity) in a bivariate generalised linear mixed model.<sup>28</sup> In the case of non-convergence, we assumed no correlation between measures of sensitivity and specificity to simplify the model.<sup>29</sup> When data were sparse, we did not do a meta-analysis (eg, for C-reactive protein [n=62] and lymphadenopathy [n=34] in pregnant participants). We illustrated the absolute pooled sensitivity and specificity using summary receiver-operating characteristic (ROC) curves.<sup>30</sup> To compare the accuracy of screening tests and strategies, we did both indirect and direct comparisons. Direct comparisons were based on studies that assessed both tests of interest; indirect comparisons were based on all studies that assessed at least one test of interest. We did a bivariate meta-regression with test type as a covariate and used likelihood ratio tests to assess the significance of differences in sensitivity and specificity. We explored study-level characteristics (tuberculosis prevalence and reference standard) as potential sources of heterogeneity. Accounting for the variation of tuberculosis prevalence across studies and their pooled values, we applied pooled accuracy estimates to a hypothetical cohort of 1000 individuals to show the consequences of using each screening test and strategy, which included calculating negative and positive predictive values using Bayes' theorem. We also calculated predictive values using a trivariate generalised linear mixed model that jointly models predictive values and test prevalence.<sup>31</sup>

We did several sensitivity analyses. We assessed diagnostic accuracy using a prespecified second reference standard of culture or Xpert. This analysis included one additional study of outpatients living with HIV (not on ART and on ART) that did not meet our primary reference standard criterion.<sup>13</sup> We also assessed diagnostic accuracy using a reference standard of Xpert alone because it is one of the molecular rapid diagnostic tests recommended by WHO. Finally, we did a direct comparison of the accuracy of W4SS followed by Xpert with the accuracy of C-reactive protein ( $\geq 10$  mg/L) followed by Xpert.

We assessed publication bias with funnel plots (for analyses with ten or more studies) and applied Egger's test. Although Deeks' test might be more appropriate, most methods to test for publication bias in studies of test accuracy have limitations.<sup>32</sup> Therefore, we also applied the trim-and-fill method to provide bias-adjusted estimates.<sup>33</sup>

We selected a p value threshold of 0.05 to characterise statistically significant findings. We did all meta-analyses using *lme*, *altmeta*, *meta*, *metafor*, and *mada* packages in R (version 3.6.1). The substantive protocol deviations were that we did not perform a leave-one-out sensitivity analysis and did not compare individual participant data results with aggregate data for which individual participant data were not obtained because we obtained more than 90% of requested data.



**Figure 1: Study selection**

W4SS=WHO-recommended four symptom screen. \*One study (Hanifa and colleagues<sup>33</sup>) was incorporated into sensitivity analyses because the study's reference standard made it ineligible for the main analyses.

### Role of the funding source

The funder had a role in the study design, data collection, data analysis, data interpretation, and writing of the report.

### Results

Of 5523 potentially eligible publications, 25 were eligible (figure 1). Individual participant data were provided for 22 studies (including one study<sup>13</sup> that was eligible only for sensitivity analyses).<sup>6,10–12,34–50</sup> Individual participant data were not provided for three studies.<sup>51–53</sup> Hence, we obtained individual participant data for 15 666 (92%) of 17 024 participants identified. The characteristics of included studies are shown in the appendix (pp 4–5). The studies collected data from 2007 to 2020. 18 studies were done in sub-Saharan Africa. Two studies included only pregnant women, and one study included only people living in prison. Overall, we

All participants (n=15 666)	
Clinical setting	
Outpatient	15 541 (99.2%)
Other setting*	125 (0.8%)
Age, years	34 (28–42)
Sex	
Female	10 388 (66.3%)
Male	5 278 (33.7%)
ART status	
On ART	4 347/15 663 (27.8%)
Outpatients on ART	4 328/15 538 (27.9%)
Not on ART	11 316/15 663 (72.2%)
Outpatients not on ART	11 210/15 538 (72.1%)
CD4 count, cells per $\mu$ L	
n	15 281
Median	269 (142–439)
History of tuberculosis	1 955/11 148 (17.5%)
W4SS	8 028/15 652 (51.3%)
Cough	4 629/15 623 (29.6%)
Fever	3 391/15 631 (21.7%)
Weight loss	5 575/15 602 (35.7%)
Night sweats	3 270/15 630 (20.9%)
Cough lasting $\geq$ 2 weeks	2 205/10 919 (20.2%)
Lymphadenopathy	3 74/2 394 (15.6%)
Chest x-ray	
Suggestive of tuberculosis	1 296/6 177 (21.0%)
Any abnormality	2 158/6 222 (34.7%)
Xpert positive†	6 16/8 625 (7.1%)
Body-mass index, kg/m <sup>2</sup>	
n	12 704
Median	22 (19–26)
C-reactive protein, mg/L‡	
n	3 582
Median	4 (2–21)
$\geq$ 10 mg/L	1 259 (35.1%)
Haemoglobin, g/dL	
n	5 118
Median	12 (10–13)
<10 g/dL	1 093 (21.4%)

Data are median (IQR), n (%), or n/N (%). If data were not available for the full cohort, the revised denominator or count is provided. ART=antiretroviral therapy. W4SS=WHO four-symptom screening. \*One study was among people living in prison. †Sputum or non-sputum sample Xpert result, or both. ‡Measured with a point-of-care assay (n=2 695) or laboratory assay (n=887).

**Table 1: Summary of main characteristics for all participants**

judged studies as low risk of bias in most QUADAS-2 domains (appendix pp 54–55), but six studies had high applicability concerns for participant selection (eg, selected only people living with HIV with advanced immunosuppression). Missing data by study are shown in the appendix (p 6).

Participant characteristics overall are shown in table 1 and by study are shown in the appendix (pp 7–10). 10 388 (66.3%) of 15 666 participants were female, and

	Number of studies	Number of participants	Number of tuberculosis cases	Prevalence (95% CI)*	Heterogeneity		p value for publication bias†	p value for between-subgroup heterogeneity‡
					I <sup>2</sup> (95% CI)	p value		
All	21	15 611	1347	7.7% (5.7–10.4)	95 (94–96)	<0.0001	0.024	..
Setting and ART status	21	15 608	1347	7.7% (5.7–10.4)	95 (94–96)	<0.0001	0.025	..
Outpatients (on ART)§	9	4309	137	3.3% (2.2–4.8)	81 (65–90)	<0.0001	0.79	<0.0001
Outpatients (not on ART)	20	11 174	1195	9.3% (7.0–12.1)	92 (89–94)	<0.0001	0.050	..
Other setting¶	1	125	15	12.0% (7.4–19.0)	..	..	..	..
CD4 count	21	15 227	1320	7.8% (5.8–10.4)	95 (94–96)	<0.0001	0.024	..
≤200 cells per µL	21	5622	866	13.7% (11.1–16.7)	84 (77–89)	<0.0001	0.035	<0.0001
>200 cells per µL	21	9605	454	4.9% (3.6–6.6)	88 (84–92)	<0.0001	0.22	..
Pregnancy status	21	10 351	701	6.4% (4.7–8.7)	91 (88–94)	<0.0001	0.15	..
Pregnant	8	1938	53	2.7% (2.1–3.6)	0 (0–60)	<0.0001	0.038	<0.0001
Not pregnant	19	8413	648	7.3% (5.4–9.8)	90 (85–93)	<0.0001	0.21	..

ART=antiretroviral therapy. \*Calculated using meta-analysis of proportions. †Egger's test. ‡Cochran's Q test (based on random effects model). §p value for between-subgroup heterogeneity compares outpatients (on ART) with outpatients (not on ART). ¶One study was among a prison population. ||Pregnancy status was unavailable for some studies, and so female participants in those studies were categorised as not pregnant.

**Table 2: Prevalence of tuberculosis in all participants and by subgroup (using culture as a reference standard)**

4347 (27.8%) of 15663 with available data were on ART. W4SS was positive in 8028 (51.3%) of 15625 participants, and C-reactive protein was elevated (≥10 mg/L) in 1259 (35.1%) of 3582 participants. C-reactive protein was measured with a point-of-care assay (2695 participants) or laboratory assay (887 participants) in five studies. The median CD4 count was 269 cells per µL (IQR 142–439; in 15 281 participants).

The pooled tuberculosis prevalence was 7.7% (95% CI 5.7–10.4) using culture as a reference standard (table 2). The pooled prevalence of tuberculosis in outpatients not on ART was 9.3% (7.0–12.1) compared with 3.3% (2.2–4.8) among outpatients on ART. For participants with a CD4 count of 200 cells per µL or less, the prevalence of tuberculosis was 13.7% (11.1–16.7) and among those with a CD4 count of more than 200 cells per µL it was 4.9% (3.6–6.6; table 2). Heterogeneity of tuberculosis prevalence was high. The pooled tuberculosis prevalences were slightly higher using a reference standard of either culture or Xpert than with a reference standard of culture alone, but subgroup comparisons remained qualitatively similar (appendix p 11).

Plots of sensitivity and specificity for each test in all participants and each subgroup are shown in the appendix (pp 56–61). Indirect comparisons between each test and W4SS in all participants are shown in table 3 and each subgroup are shown in the appendix (pp 12–16). Among 15 597 participants with available culture results, the sensitivity of W4SS was 82% (95% CI 72–89) and specificity was 42% (29–57; table 3; appendix p 56). The sensitivity of C-reactive protein (≥10 mg/L) was similar to, and its specificity was higher than, that of W4SS (sensitivity 77% [95% CI

61–88; p=0.71], specificity 74% [61–83; p=0.041]; table 3; figure 2). The sensitivity of chest x-ray (with any abnormality) was 72% (65–78) and specificity was 62% (51–71; table 3; appendix p 56). Cough (lasting ≥2 weeks), haemoglobin (<10 g/dL), BMI (<18.5 kg/m<sup>2</sup>), and lymphadenopathy had high specificities but low sensitivities, making them unsuitable to be explored further as screening tests.

Parallel strategies that combined W4SS with either chest x-ray (with any abnormality) or C-reactive protein (≥10 mg/L) had higher sensitivities and lower specificities than W4SS alone (table 3). A sequential strategy of W4SS followed by C-reactive protein (≥5 mg/L) had a lower sensitivity but higher specificity than W4SS alone. A sequential strategy of W4SS followed by chest x-ray (with any abnormality) had a sensitivity of 63% (54–71) and specificity of 73% (62–82); we did not assess this strategy further because of reduced sensitivity compared with W4SS alone.

The sensitivity of W4SS followed by Xpert was 58% (95% CI 50–66; table 3). The sensitivity of Xpert for all was 68% (95% CI 57–76). The specificities of both strategies—W4SS followed by Xpert and Xpert for all—were 99% (table 3). The sensitivity of sputum Xpert Ultra was higher than that of sputum Xpert (73% [95% CI 62–81] vs 57% [47–67]) and specificities were similar (98% [96–98] and 99% [98–100]) in the only study (unpublished) that compared both tests.<sup>50</sup>

Direct and indirect comparisons of individual tests were largely similar (appendix p 17); however, the lower sensitivity and higher specificity of chest x-ray (with any abnormality) than with W4SS were more pronounced in the direct comparison. Forest plots and summary ROC curves for all tests and screening strategies are provided

	Number of studies	Number of participants	Sensitivity (95% CI)	Specificity (95% CI)	Difference from W4SS*	
					Sensitivity (p value)	Specificity (p value)
W4SS	21	15 597	82% (72–89)	42% (29–57)	..	..
C-reactive protein						
≥10 mg/L	5	3571	77% (61–88)	74% (61–83)	0.71	0.041
≥8 mg/L	5	3571	81% (68–89)	70% (57–81)	0.91	0.071
≥5 mg/L	5	3571	87% (77–93)	60% (48–71)	0.51	0.27
Chest x-ray						
With any abnormality	8	6195	72% (65–78)	62% (51–71)	0.26	0.13
Suggestive of tuberculosis	8	6150	63% (57–70)	78% (67–86)	0.071	0.0049
Cough						
Any	21	15 568	56% (48–63)	72% (65–79)	<0.0001	0.0006
Lasting ≥2 weeks	17	10 906	38% (29–49)	84% (77–90)	<0.0001	<0.0001
Haemoglobin						
<10 g/dL	9	5116	43% (33–54)	80% (73–85)	0.0006	0.0013
<8 g/dL	9	5116	12% (9–16)	96% (93–97)	<0.0001	<0.0001
BMI (<18.5 kg/m <sup>2</sup> )	18	12 650	29% (22–38)	89% (84–92)	<0.0001	<0.0001
Lymphadenopathy	4	2391	31% (14–55)	90% (75–96)	0.0023	0.0018
Parallel strategies†						
W4SS and C-reactive protein (≥10 mg/L)	5	3571	88% (63–97)	31% (13–57)	0.358	0.46
W4SS and chest x-ray with abnormal findings	8	6186	94% (89–97)	20% (10–37)	0.0077	0.066
Sequential strategies†						
W4SS then C-reactive protein (≥5 mg/L)	5	3571	70% (31–92)	75% (53–88)	0.55	0.0405
W4SS then Xpert‡§	12	8557	58% (50–66)	99% (98–100)	..	..
Xpert for all‡§	12	8570	68% (57–76)	99% (98–99)	0.094¶	0.40¶

Indirect comparisons are based on all studies that assessed at least one of the W4SS or relevant screening tests. BMI=body-mass index. W4SS=WHO-recommended four-symptom screen. \*For Xpert for all, the comparator is W4SS then Xpert. †For parallel strategies, two screening tests are offered at the same time; for sequential strategies, a second screening test is offered only if the first screening test is positive. ‡Accuracy measures for entire algorithm using total Xpert (sputum or non-sputum sample Xpert result, or both); alternative algorithms are W4SS then single sputum Xpert (12 studies; 8556 participants; sensitivity 55% [95% CI 48–63], specificity 99% [99–100]) and single sputum Xpert alone (12 studies; 8569 participants; sensitivity 64% [53–74], specificity 99% [98–99]). §One study assessed Xpert and Xpert Ultra among 733 participants; sputum Xpert sensitivity was 57% (95% CI 47–67) and specificity was 99% (98–100), sputum Xpert Ultra sensitivity was 73% (62–81) and specificity was 98% (96–98), urine Xpert Ultra sensitivity was 27% (19–38) and specificity was 98% (96–99), and sputum and urine Xpert Ultra sensitivity was 75% (65–83) and specificity was 95% (94–97). ¶Bivariate model did not converge; results from a model assuming no correlation between sensitivity and specificity.

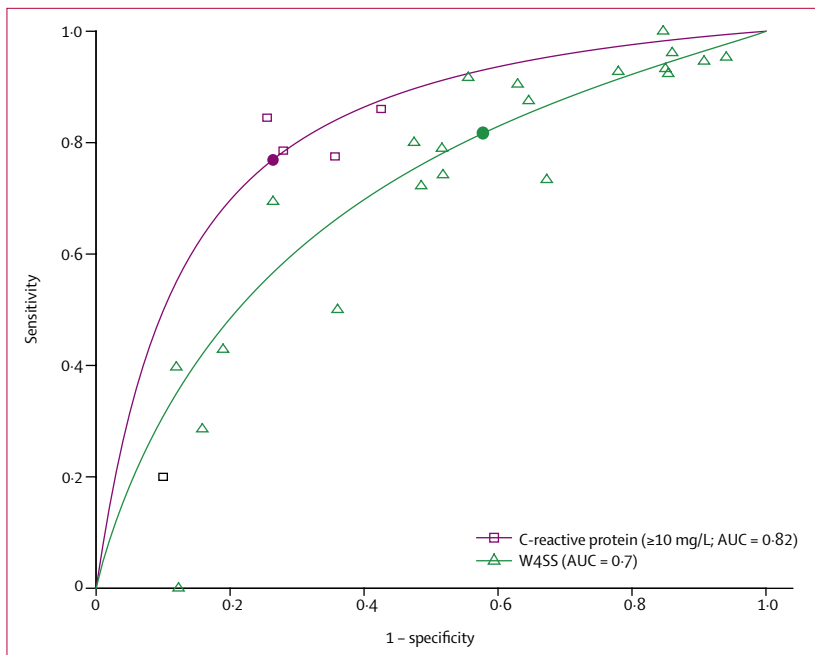
**Table 3: Indirect comparisons between each test and W4SS for the detection of tuberculosis in all participants (using culture as a reference standard)**

in the appendix (pp 62–146). The point estimates for the specificities of C-reactive protein (≥10 mg/L cutoff) were numerically higher than those of W4SS in each individual study that had these data (appendix pp 62–63). Additional diagnostic accuracy measures are shown in the appendix (pp 23–28).

We assessed how estimates for each test or strategy affected detection rates in a hypothetical cohort of 1000 people living with HIV at different tuberculosis prevalences (appendix pp 29–46). At a tuberculosis prevalence of 10%, the W4SS would result in 604 rapid diagnostic tests being needed; C-reactive protein (≥10 mg/L) would reduce the number of rapid diagnostic tests needed by 293 but miss five additional tuberculosis cases, and chest x-ray (with any abnormality) would reduce the number of rapid diagnostic tests needed by 190, but miss ten additional tuberculosis cases

(figure 3; appendix pp 29–31). At 10% prevalence, the WHO-recommended algorithm (W4SS followed by Xpert) would result in 604 Xpert tests, and Xpert for all would increase the number of Xpert tests needed by 396 (ie, because all 1000 people would receive an Xpert test), but it would detect ten additional tuberculosis cases (appendix pp 29–31).

Indirect comparisons by ART status are shown in the appendix (pp 12–13, 57–58). Most tests, except chest x-ray and haemoglobin, had lower sensitivity and higher specificity in outpatients on ART than in outpatients not on ART. In outpatients on ART, a parallel strategy of W4SS and chest x-ray (with any abnormality) had higher sensitivity than W4SS alone (89% [95% CI 70–97] vs 53% [35–71]) but lower specificity (33% [17–54] vs 71% [51–85]; appendix p 12). In a hypothetical cohort of 1000 outpatients on ART with 5% tuberculosis prevalence,



**Figure 2: Summary ROC curves comparing C-reactive protein ( $\geq 10$  mg/L) with W4SS in all participants\***  
 AUC=area under the ROC. ROC=receiver operating characteristic. W4SS=WHO-recommended four-symptom screen. \*Data were extrapolated beyond observed datapoints.

this strategy would increase the number of rapid diagnostic tests needed by 378 compared with W4SS alone but detect 18 additional tuberculosis cases (figure 3; appendix pp 32–34).

In outpatients not on ART, sensitivities for C-reactive protein ( $\geq 10$  mg/L) alone (83% [95% CI 79–86]) and a sequential strategy of W4SS then C-reactive protein ( $\geq 5$  mg/L; 84% [75–90]) were similar to the sensitivity of W4SS alone (85% [76–91]), but their specificities were higher (67% [60–73] for C-reactive protein alone; 64% [57–71] for sequential strategy) than with W4SS alone (37% [25–51]; appendix p 13). In a hypothetical cohort of 1000 outpatients not on ART with 10% tuberculosis prevalence, compared with use of W4SS alone, use of C-reactive protein ( $\geq 10$  mg/L) would reduce the number of rapid diagnostic tests needed by 272 but miss two additional tuberculosis cases, and use of the sequential strategy of W4SS then C-reactive protein ( $\geq 5$  mg/L) would reduce the number of rapid diagnostic tests needed by 244 but miss one additional tuberculosis case (figure 3; appendix pp 35–37).

Indirect comparisons between each test and W4SS by CD4 cell count are shown in the appendix (pp 14–15, 59–60). Most tests, except chest x-ray, had lower sensitivity and higher specificity in participants with CD4 counts of more than 200 cells per  $\mu\text{L}$  than those with CD4 counts of 200 cells per  $\mu\text{L}$  or lower. Similarly, most tests had lower sensitivity and higher specificity in pregnant women living with HIV than in the overall population (appendix pp 16, 61); however, these estimates had suboptimal precision.

Indirect and direct comparisons for the subgroups were largely similar (appendix pp 12–16, 18–22). However, among outpatients on ART, the slightly higher sensitivity of chest x-ray (both with any abnormality and suggestive of tuberculosis) than of W4SS alone in indirect comparisons was attenuated in direct comparisons (appendix p 18). Only one study ( $n=381$ ) among outpatients on ART assessed C-reactive protein ( $\geq 10$  mg/L), for which there was a similar sensitivity and specificity compared with W4SS alone (appendix p 18).<sup>49</sup>

We did sensitivity analyses using two alternative reference standards: culture or Xpert, and Xpert alone (appendix pp 47–52). Results were largely similar to the main analyses, although sensitivities were slightly higher for the reference standard of Xpert alone than for the main reference standard of culture. In sensitivity analyses directly comparing W4SS followed by Xpert with C-reactive protein ( $\geq 10$  mg) followed by Xpert, both strategies had similar sensitivities and specificities (appendix p 53).

Egger's test and meta-regression results are provided in the appendix (pp 23–28), as well as funnel plots (pp 147–154). We found no evidence of publication bias (Egger's test  $p>0.05$ ) for most tests. Meta-regression showed that prevalence explained some heterogeneity in the analyses for several tests, but reference standard type generally did not.

## Discussion

In this systematic review and individual participant data meta-analysis, we found that the sensitivity of C-reactive protein ( $\geq 10$  mg/L) was similar to that of W4SS alone, but its specificity was higher (74% vs 42%). Chest x-ray (with any abnormality) had lower sensitivity than W4SS alone in direct comparisons, making it less suitable than a standalone screening test. Cough (lasting  $\geq 2$  weeks), haemoglobin ( $<10$  g/dL), BMI ( $<18.5$  kg/m<sup>2</sup>), and lymphadenopathy had high specificities ( $>80\%$ ), but their low sensitivities also made them less suitable as screening tests than W4SS. The WHO-recommended algorithm of W4SS then Xpert had a sensitivity of only 58% (95% CI 50–66), and Xpert for all had a slightly higher sensitivity of 68% (57–76). In one unpublished study, Xpert Ultra improved sensitivity over Xpert (73% [62–81] vs 57% [47–67]).<sup>50</sup>

Among outpatients on ART, the sensitivity of a parallel strategy of W4SS with chest x-ray (any abnormality) was higher than that of W4SS alone, but its specificity was lower. At 5% tuberculosis prevalence, this strategy was estimated to require more than double the number of rapid diagnostic tests needed compared with W4SS alone but would detect 70% more tuberculosis cases. Among outpatients not on ART, the sensitivities of C-reactive protein ( $\geq 10$  mg/L) and a sequential strategy of W4SS then C-reactive protein ( $\geq 5$  mg/L) were similar to W4SS alone, but specificities were higher. At 10% tuberculosis prevalence, these strategies would reduce



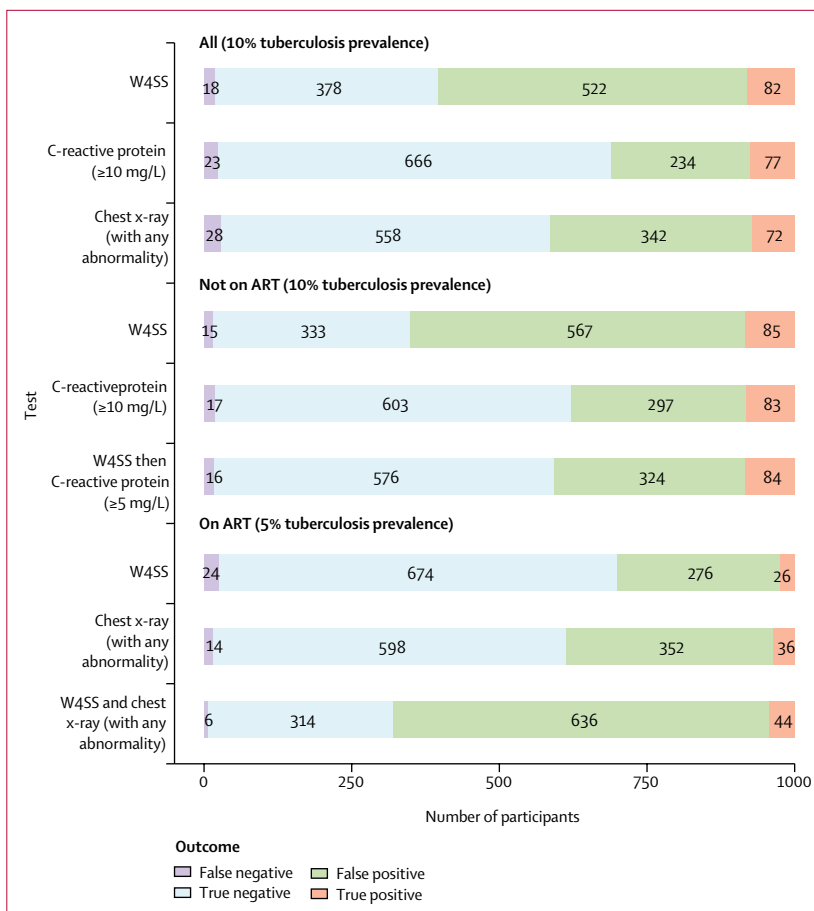
the number of rapid diagnostic tests needed by 42% for C-reactive protein ( $\geq 10$  mg/L) and 37% for W4SS then C-reactive protein ( $\geq 5$  mg/L) compared with W4SS alone, but would miss a similar number of tuberculosis cases.

We found that C-reactive protein ( $\geq 10$  mg/L) approached the WHO-defined minimum thresholds for a screening test (with 83% sensitivity and 67% specificity vs WHO's thresholds of 90% sensitivity and 70% specificity) for outpatients not on ART.<sup>8</sup> Efforts to scale-up of access to WHO-recommended, molecular, rapid diagnostic tests have been slow, particularly in decentralised locations.<sup>54,55</sup> C-reactive protein testing could allow for broader implementation of rapid diagnostic tests because its greater specificity means that screening using C-reactive protein would require fewer subsequent rapid diagnostic tests than screening with W4SS. The need for fewer tests could also reduce laboratory processing time; Xpert can provide a result in less than 2 h, but a result often takes several days in the real world.<sup>56</sup> The high specificity of C-reactive protein would reduce the time to start tuberculosis preventive therapy in people living with HIV. Current C-reactive protein point-of-care assays have differing complexities, ranging from qualitative lateral-flow assays that do not require a power source or refrigeration to quantitative assays that require a small machine.<sup>57</sup> C-reactive protein point-of-care assays can cost approximately US\$2 per test, provide results in less than 3 min, and be performed easily with minimal expertise (blood collected by finger prick). Thus, available point-of-care assays have the potential for affordable scale-up.

The sensitivity of a parallel strategy incorporating W4SS and chest x-ray was higher than the sensitivity of other tests or strategies in those on ART; however, the higher number of rapid diagnostic tests needed might pose a substantial cost burden. Furthermore, a 2016 survey of 14 countries with high HIV-associated tuberculosis burdens found that chest x-ray as a screening tool was available at only 14% of primary health-care centres.<sup>55</sup> We found that the sensitivity of chest x-ray was not increased in those not on ART and at lower CD4 cell counts of 200 per  $\mu$ L or lower; the most likely explanation for these findings is that normal chest x-ray images in patients with pulmonary and extra-pulmonary tuberculosis occur more frequently in those with advanced immunosuppression than in other people living with HIV.<sup>58,59</sup>

The low sensitivities of haemoglobin, BMI, and lymphadenopathy make them unsuitable as screening tests. However, haemoglobin levels below 10 g/dL, a BMI of less than 18.5 kg/m<sup>2</sup>, and lymphadenopathy in ambulatory people living with HIV should prompt a thorough search for tuberculosis, given their high specificities and known association with mortality.<sup>60,61</sup>

We found that the WHO-recommended strategy (W4SS followed by Xpert) would miss approximately 40% of tuberculosis cases. The low yield is a result of the inadequate sensitivities of both the W4SS and Xpert.



**Figure 3:** Screening outcomes for selected screening tests and strategies in a hypothetical cohort of 1000 people living with HIV at 10% (all and not on ART) and 5% (on ART) tuberculosis prevalence. ART=antiretroviral therapy. W4SS=WHO-recommended four-symptom screen.

Approximately 20% of people living with HIV with tuberculosis will be missed with W4SS and thus have subclinical tuberculosis, 56–75% of whom will probably progress to symptomatic disease.<sup>62,63</sup> Although Xpert for all would still miss approximately 33% of tuberculosis cases, Xpert Ultra showed improved sensitivity over Xpert in one study.<sup>50</sup> Xpert Ultra costs the same as Xpert, and the point-of-care GeneXpert Omni platform might allow its use at decentralised locations. Further research is needed to assess this approach.

Our study has limitations. First, we did not have adequate precision in some analyses for outpatients on ART and pregnant people living with HIV. Specifically, we had little data on C-reactive protein in people living with HIV on ART. Furthermore, there was a paucity of data on countries other than South Africa, where almost half of all included studies were done, and which might be more urbanised than other low-income and middle-income countries. Second, we largely excluded participants who were unable to produce a sputum sample, meaning our findings might not generalise to this group. Few studies also systematically included extra-pulmonary

tuberculosis samples, meaning our results are more applicable to pulmonary tuberculosis. However, pulmonary tuberculosis probably comprises most tuberculosis cases in an ambulatory screening setting. Third, we used an imperfect reference standard, because sputum culture, which was all that was done in most of the included studies, should ideally comprise multiple samples collected in the early morning to maximise sensitivity, but this was not done in any of our included studies. Fourth, although direct comparison minimises confounding, these analyses involved fewer studies and reduced precision. Fifth, we were unable to obtain individual participant data from three studies. However, these studies comprised only approximately 8% of data. Sixth, only one study assessed Xpert Ultra,<sup>50</sup> and we did not assess non-Xpert nucleic acid amplification tests. Seventh, our study findings might not be generalisable to children with HIV and they might not be generalisable to all settings because most included studies were done in settings with high tuberculosis prevalence. Test performance might also vary in the context of regular screening. Finally, although calculations based on a hypothetical cohort give insight into consequences of testing, they were often based on heterogenous results.

Findings from this study have informed the updated 2021 WHO tuberculosis screening guidelines in people living with HIV.<sup>64</sup> Compared with W4SS, C-reactive protein reduces the need for additional rapid diagnostic tests without compromising sensitivity, but there was a paucity of data for outpatients on ART. In outpatients not on ART, C-reactive protein assays could be used as a standalone screening test or combined with W4SS in a sequential strategy. In outpatients on ART, chest x-ray could be used in parallel with W4SS, depending on available resources, because this strategy detects more tuberculosis cases than does W4SS alone. Overall, the WHO-recommended screening and diagnostic algorithm (W4SS followed by Xpert) has suboptimal sensitivity; Xpert for all offers small improvements in sensitivity and would be resource intensive. Future research is needed to assess the utility of C-reactive protein screening in outpatients on ART and Xpert Ultra in all people living with HIV, and to investigate the cost-effectiveness of different screening tests and strategies. Because no test or strategy met both WHO-defined minimum sensitivity and specificity thresholds, improved screening tests for tuberculosis need to be developed for this population.

#### Contributors

AD, YHam, APK, ADK, MXR, TKr, AB, CM, SSi, DAB, GMe, and GMa designed the study and protocol and interpreted the results. GMa supervised the study. AD and YHam did the systematic review. ADK, MXR, YHam, ADG, KF, DA, CSM, APW, CY, AC, CJH, NM, ETM, MSS, TTB, SSk, BWPR, GT, GN, SM, JC, SSw, REC, FAK, AAH, RW, SST, MMK, JH, PKD, AES, TKu, GC, DTN, EAG, SB, ISJ, JKG, DJH, SML, HAAA, AK, RRR, NT, and GMe contributed data to the meta-analysis. AD analysed the data with assistance from APK, DAB, and YHam. AD and GMa wrote the first draft of the manuscript, which was revised based on comments from co-authors. AD, DAB, and YHam accessed

and verified the data. All authors approved the final version of the manuscript.

#### Declaration of interests

AC reports grants from National Institutes of Health (NIH), Global Health Labs, and Stop TB Partnership, and consulting fees from the US Centers for Disease Control and Prevention (CDC). AK reports grants from Sanofi. FAK reports grants from WHO, Canadian Institutes of Health Research, Fonds de Recherche Quebec, and McGill Interdisciplinary Initiative on Infection and Immunity. GT reports receipt of consumables and equipment from Boditech and Cepheid. NM reports grants from Pfizer and Roche. REC reports grants from NIH, CDC, and Unitaid, and consulting fees from Sanofi. SML reports grants from NIH and CDC. SSk reports grants from Swedish Heart-Lung Foundation. TKr reports consulting fees from WHO. All other authors declare no competing interests.

#### Data sharing

The aggregate datasets and analysis code are available online. The study investigators of the original studies retain ownership of their data. Any requests for access to individual participant data should be made directly to study investigators of the original studies.

#### Acknowledgments

This work was supported by WHO. AD received training in research that was supported by the Fogarty International Center of the National Institutes of Health under award number D43 TW010559 and the National Research Foundation (NRF) of South Africa. GT acknowledges funding from the South African Medical Research Council (SAMRC Flagship Project MRC-RFA-IFSP-01-2013), the EDCTP2 programme supported by the EU (grant SF1401, OPTIMAL DIAGNOSIS), and the Faculty of Medicine and Health Sciences, Stellenbosch University. GN acknowledges that the work reported herein was made possible through funding by the South African Medical Research Council (SAMRC) through its Division of Research Capacity Development, under the Internship Scholarship Programme, from funding received from the South African National Treasury. The content of this paper is the sole responsibility of the authors and does not necessarily represent the official views of the SAMRC or the funders. GMe and GMa were supported by core funding from the Wellcome Centre for Infectious Diseases Research in Africa (203135/Z/16/Z). GMe was supported by the Wellcome Trust (214321/Z/18/Z) and the South African Research Chairs Initiative of the Department of Science and Technology and NRF of South Africa (grant number 64787). AB, CM, SSi, and CSM work for WHO; the authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions, policy, or views of WHO. The findings and conclusions in this manuscript are those of the author(s) and do not necessarily represent the official position of the CDC.

#### References

- Gupta RK, Lucas SB, Fielding KL, Lawn SD. Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings: a systematic review and meta-analysis. *AIDS* 2015; **29**: 1987–2002.
- WHO. Global tuberculosis report 2020. Geneva: World Health Organization, 2020.
- Getahun H, Kittikraisak W, Heilig CM, et al. Development of a standardized screening rule for tuberculosis in people living with HIV in resource-constrained settings: individual participant data meta-analysis of observational studies. *PLoS Med* 2011; **8**: e1000391.
- WHO. WHO consolidated guidelines on tuberculosis: tuberculosis preventive treatment: module 1: prevention: tuberculosis preventive treatment. Geneva: World Health Organization, 2020.
- Hamada Y, Lujan J, Schenkel K, Ford N, Getahun H. Sensitivity and specificity of WHO's recommended four-symptom screening rule for tuberculosis in people living with HIV: a systematic review and meta-analysis. *Lancet HIV* 2018; **5**: e515–23.
- Rangaka MX, Wilkinson RJ, Glynn JR, et al. Effect of antiretroviral therapy on the diagnostic accuracy of symptom screening for intensified tuberculosis case finding in a South African HIV clinic. *Clin Infect Dis* 2012; **55**: 1698–706.

For the aggregate datasets and analysis code see [https://github.com/ashdha/ambulatory\\_tb\\_ipd\\_meta\\_analysis](https://github.com/ashdha/ambulatory_tb_ipd_meta_analysis)

- 7 Yoon C, Semitala FC, Asege L, et al. Yield and efficiency of novel intensified tuberculosis case-finding algorithms for people living with HIV. *Am J Respir Crit Care Med* 2019; **199**: 643–50.
- 8 WHO. High priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting, 28–29 April 2014, Geneva, Switzerland. Geneva: World Health Organization, 2014.
- 9 Lawn SD, Kerkhoff AD, Vogt M, Wood R. Diagnostic and prognostic value of serum C-reactive protein for screening for HIV-associated tuberculosis. *Int J Tuberc Lung Dis* 2013; **17**: 636–43.
- 10 Shapiro AE, Hong T, Govere S, et al. C-reactive protein as a screening test for HIV-associated pulmonary tuberculosis prior to antiretroviral therapy in South Africa. *AIDS* 2018; **32**: 1811–20.
- 11 Yoon C, Semitala FC, Atuhumuza E, et al. Point-of-care C-reactive protein-based tuberculosis screening for people living with HIV: a diagnostic accuracy study. *Lancet Infect Dis* 2017; **17**: 1285–92.
- 12 Balcha TT, Skogmar S, Sturegård E, et al. A clinical scoring algorithm for determination of the risk of tuberculosis in HIV-infected adults: a cohort study performed at Ethiopian health centers. *Open Forum Infect Dis* 2014; **1**: ofu095.
- 13 Hanifa Y, Fielding KL, Chihota VN, et al. A clinical scoring system to prioritise investigation for tuberculosis among adults attending HIV clinics in South Africa. *PLoS One* 2017; **12**: e0181519.
- 14 Floridia M, Ciccacci F, Andreotti M, et al. Tuberculosis case finding with combined rapid point-of-care assays (Xpert MTB/RIF and Determine TB LAM) in HIV-positive individuals starting antiretroviral therapy in Mozambique. *Clin Infect Dis* 2017; **65**: 1878–83.
- 15 Lawn SD, Brooks SV, Kranzer K, et al. Screening for HIV-associated tuberculosis and rifampicin resistance before antiretroviral therapy using the Xpert MTB/RIF assay: a prospective study. *PLoS Med* 2011; **8**: e1001067.
- 16 Bjerrum S, Schiller I, Dendukuri N, et al. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV. *Cochrane Database Syst Rev* 2019; **10**: CD011420.
- 17 Broger T, Nicol MP, Székely R, et al. Diagnostic accuracy of a novel tuberculosis point-of-care urine lipoarabinomannan assay for people living with HIV: a meta-analysis of individual in- and outpatient data. *PLoS Med* 2020; **17**: e1003113.
- 18 WHO. WHO operational handbook on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection. Geneva: World Health Organization, 2020.
- 19 Claus DR, Osmand AP, Gewurz H. Radioimmunoassay of human C-reactive protein and levels in normal sera. *J Lab Clin Med* 1976; **87**: 120–28.
- 20 Shine B, de Beer FC, Pepys MB. Solid phase radioimmunoassays for human C-reactive protein. *Clin Chim Acta* 1981; **117**: 13–23.
- 21 Stewart LA, Clarke M, Rovers M, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD statement. *JAMA* 2015; **313**: 1657–65.
- 22 McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. *JAMA* 2018; **319**: 388–96.
- 23 Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36.
- 24 Tierney JF, Vale C, Riley R, et al. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med* 2015; **12**: e1001855.
- 25 Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions: Cochrane book series. The Cochrane Collaboration, 2008.
- 26 Lin L, Chu H. Meta-analysis of proportions using generalized linear mixed models. *Epidemiology* 2020; **31**: 713–17.
- 27 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–60.
- 28 Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; **59**: 1331–32.
- 29 Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017; **26**: 1896–911.
- 30 Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**: 2865–84.
- 31 Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Stat Med* 2009; **28**: 2384–99.
- 32 Adelman MW, Tsegaye M, Kempker RR, et al. Intensified tuberculosis case finding among HIV-infected persons using a WHO symptom screen and Xpert MTB/RIF. *Int J Tuberc Lung Dis* 2015; **19**: 1197–203.
- 33 Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; **56**: 455–63.
- 34 Ahmad Khan F, Verkuil S, Parrish A, et al. Performance of symptom-based tuberculosis screening among people living with HIV: not as great as hoped. *AIDS* 2014; **28**: 1463–72.
- 35 Al-Darraj HA, Abd Razak H, Ng KP, Altice FL, Kamarulzaman A. The diagnostic performance of a single GeneXpert MTB/RIF assay in an intensified tuberculosis case finding survey among HIV-infected prisoners in Malaysia. *PLoS One* 2013; **8**: e73717.
- 36 Bjerrum S, Kenu E, Lartey M, et al. Diagnostic accuracy of the rapid urine lipoarabinomannan test for pulmonary tuberculosis among HIV-infected adults in Ghana—findings from the DETECT HIV-TB study. *BMC Infect Dis* 2015; **15**: 407.
- 37 Hanifa Y, Fielding KL, Charalambous S, et al. Tuberculosis among adults starting antiretroviral therapy in South Africa: the need for routine case finding. *Int J Tuberc Lung Dis* 2012; **16**: 1252–59.
- 38 Hoffmann CJ, Variava E, Rakgokong M, et al. High prevalence of pulmonary tuberculosis but low sensitivity of symptom screening among HIV-infected pregnant women in South Africa. *PLoS One* 2013; **8**: e62211.
- 39 Kempker RR, Chkhartishvili N, Kinkladze I, et al. High yield of active tuberculosis case finding among HIV-infected patients using Xpert MTB/RIF testing. *Open Forum Infect Dis* 2019; **6**: ofz233
- 40 Kerkhoff AD, Wood R, Lowe DM, Vogt M, Lawn SD. Blood neutrophil counts in HIV-infected patients with pulmonary tuberculosis: association with sputum mycobacterial load. *PLoS One* 2013; **8**: e67956.
- 41 Kufa T, Mngomezulu V, Charalambous S, et al. Undiagnosed tuberculosis among HIV clinic attendees: association with antiretroviral therapy and implications for intensified case finding, isoniazid preventive therapy, and infection control. *J Acquir Immune Defic Syndr* 2012; **60**: e22–28.
- 42 LaCourse SM, Cranmer LM, Matemo D, et al. Tuberculosis case finding in HIV-infected pregnant women in Kenya reveals poor performance of symptom screening and rapid diagnostic tests. *J Acquir Immune Defic Syndr* 2016; **71**: 219–27.
- 43 Mbu ET, Sauter F, Zoufaly A, et al. Tuberculosis in people newly diagnosed with HIV at a large HIV care and treatment center in northwest Cameroon: burden, comparative screening and diagnostic yields, and patient outcomes. *PLoS One* 2018; **13**: e0199634.
- 44 Modi S, Cavanaugh JS, Shirraishi RW, et al. Performance of clinical screening algorithms for tuberculosis intensified case finding among people living with HIV in western Kenya. *PLoS One* 2016; **11**: e0167685.
- 45 Nguyen DT, Bang ND, Hung NQ, Beasley RP, Hwang LY, Graviss EA. Yield of chest radiograph in tuberculosis screening for HIV-infected persons at a district-level HIV clinic. *Int J Tuberc Lung Dis* 2016; **20**: 211–17.
- 46 Swindells S, Komarow L, Tripathy S, et al. Screening for pulmonary tuberculosis in HIV-infected individuals: AIDS Clinical Trials Group Protocol A5253. *Int J Tuberc Lung Dis* 2013; **17**: 532–39.
- 47 Thit SS, Aung NM, Htet ZW, et al. The clinical utility of the urine-based lateral flow lipoarabinomannan assay in HIV-infected adults in Myanmar: an observational study. *BMC Med* 2017; **15**: 145.
- 48 Affolabi D, Wachinou AP, Bekou W, et al. Screening tuberculosis in HIV infected patients: which algorithms work best? A multicountry survey in Benin, Guinea and Senegal (RAFAScreen project). 49th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease; The Hague, The Netherlands; Oct 24–27, 2018: OA09-262-25 (abstr).
- 49 Gersh JK, Barnabas RV, Matemo D, et al. Pulmonary tuberculosis screening in anti-retroviral treated adults living with HIV in Kenya. *BMC Infect Dis* 2021; **21**: 218.

- 50 Reeve B, Ndlangalavu G, Palmer Z, et al. Accuracy of Xpert Ultra and Xpert MTB/RIF in people living with HIV initiating antiretroviral treatment who have minimal TB symptoms. 50th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease; Hyderabad, India; Oct 30, to Nov 2, 2019: S115: SOA-01-1010-31 (abstr).
- 51 Calnan M. Developing strategies for TB screening among HIV-infected and HIV-uninfected pregnant and postpartum women in Swaziland. 47th World Conference on Lung Health of the International Union Against Tuberculosis and Lung Disease; Liverpool; Oct 26–29, 2016.
- 52 Drain PK, Losina E, Coleman SM, et al. Rapid urine lipoarabinomannan assay as a clinic-based screening test for active tuberculosis at HIV diagnosis. *BMC Pulm Med* 2016; **16**: 147.
- 53 Telisinghe L, Fielding KL, Malden JL, et al. High tuberculosis prevalence in a South African prison: the need for routine tuberculosis screening. *PLoS One* 2014; **9**: e87262.
- 54 Médecins Sans Frontières. Out of step: TB policies in 29 countries. Geneva: MSF Access Campaign, 2017.
- 55 Huddart S, MacLean E, Pai M. Location, location, location: tuberculosis services in highest burden countries. *Lancet Glob Health* 2016; **4**: e907–08.
- 56 Cohen GM, Drain PK, Noubary F, Cloete C, Bassett IV. Diagnostic delays and clinical decision making with centralized Xpert MTB/RIF testing in Durban, South Africa. *J Acquir Immune Defic Syndr* 2014; **67**: e88–93.
- 57 van Griensven J, Cnops L, De Weggheleire A, Declercq S, Bottieau E. Point-of-care biomarkers to guide antibiotic prescription for acute febrile illness in sub-Saharan Africa: promises and caveats. *Open Forum Infect Dis* 2020; **7**: ofaa260.
- 58 Perlman DC, el-Sadr WM, Nelson ET, et al. Variation of chest radiographic patterns in pulmonary tuberculosis by degree of human immunodeficiency virus-related immunosuppression. *Clin Infect Dis* 1997; **25**: 242–46.
- 59 Wasserman S, Meintjes G. The diagnosis, management and prevention of HIV-associated tuberculosis. *S Afr Med J* 2014; **104**: 886–93.
- 60 Hanrahan CF, Golub JE, Mohapi L, et al. Body mass index and risk of tuberculosis and death. *AIDS* 2010; **24**: 1501–08.
- 61 Kerkhoff AD, Wood R, Cobelens FG, Gupta-Wright A, Bekker LG, Lawn SD. The predictive value of current haemoglobin levels for incident tuberculosis and/or mortality during long-term antiretroviral therapy in South Africa: a cohort study. *BMC Med* 2015; **13**: 70.
- 62 Lawn SD, Kerkhoff AD, Wood R. Progression of subclinical culture-positive tuberculosis to symptomatic disease in HIV-infected individuals. *AIDS* 2011; **25**: 2190–91.
- 63 Oni T, Burke R, Tsekela R, et al. High prevalence of subclinical tuberculosis in HIV-1-infected persons without advanced immunodeficiency: implications for TB screening. *Thorax* 2011; **66**: 669–73.
- 64 WHO. WHO operational handbook on tuberculosis: module 2: screening: systematic screening for tuberculosis disease. Geneva: World Health Organization, 2021.