



**Cochrane**  
**Library**

Cochrane Database of Systematic Reviews

## Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status (Protocol)

van't Hoog AH, Langendam M, Mitchell E, Cobelens FG, Sinclair D, Leeflang MM, Lönnroth K

van't Hoog AH, Langendam M, Mitchell E, Cobelens FG, Sinclair D, Leeflang MM, Lönnroth K.

Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status.

*Cochrane Database of Systematic Reviews* 2014, Issue 1. Art. No.: CD010890.

DOI: 10.1002/14651858.CD010890.

[www.cochranelibrary.com](http://www.cochranelibrary.com)

## TABLE OF CONTENTS

HEADER . . . . .	1
ABSTRACT . . . . .	1
BACKGROUND . . . . .	2
Figure 1. . . . .	4
OBJECTIVES . . . . .	6
METHODS . . . . .	6
ACKNOWLEDGEMENTS . . . . .	11
REFERENCES . . . . .	11
APPENDICES . . . . .	15
CONTRIBUTIONS OF AUTHORS . . . . .	20
DECLARATIONS OF INTEREST . . . . .	20
SOURCES OF SUPPORT . . . . .	21

[Diagnostic Test Accuracy Protocol]

# Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status

Anna H van't Hoog<sup>1</sup>, Miranda Langendam<sup>2</sup>, Ellen Mitchell<sup>3</sup>, Frank G Cobelens<sup>1</sup>, David Sinclair<sup>4</sup>, Mariska M.G. Leeflang<sup>5</sup>, Knut Lönnroth<sup>6</sup>

<sup>1</sup>Department of Global Health, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands. <sup>2</sup>Dutch Cochrane Centre, Academic Medical Center, Amsterdam, Netherlands. <sup>3</sup>KNCV Tuberculosis Foundation, The Hague, Netherlands. <sup>4</sup>Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK. <sup>5</sup>Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, J1B-207-1, AMSTERDAM, Netherlands. <sup>6</sup>Stop TB Department, World Health Organization, Geneva, Switzerland

Contact address: Anna H van't Hoog, Department of Global Health, Academic Medical Centre, University of Amsterdam, Pietersbergweg 17, Amsterdam, 1105 BM, Netherlands. [a.h.vanthoog@aighd.org](mailto:a.h.vanthoog@aighd.org).

**Editorial group:** Cochrane Infectious Diseases Group.

**Publication status and date:** New, published in Issue 1, 2014.

**Citation:** van't Hoog AH, Langendam M, Mitchell E, Cobelens FG, Sinclair D, Leeflang MM, Lönnroth K. Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status. *Cochrane Database of Systematic Reviews* 2014, Issue 1. Art. No.: CD010890. DOI: 10.1002/14651858.CD010890.

Copyright © 2014 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

To assess the sensitivity and specificity of questioning for presence of one or more selected symptoms, or symptom combinations, or both; chest radiography; and combinations of those as screening tools for detecting bacteriologically confirmed active pulmonary TB in people considered eligible for TB screening who are HIV-negative or whose HIV-status is unknown.

If data allow, we will investigate heterogeneity in relation to:

- background epidemiology (prevalence of pulmonary TB and of HIV among the study population);
- risk groups targeted (for example, migrants, occupational, prisoners, or the general population);
- reference standard (culture, Xpert, smear microscopy);
- screen test definition;
- representativeness of the study design and population for intended screening practice (inclusion of people without any symptoms or CXR abnormalities);
- study participants characteristics (age, sex and HIV status);
- geographic area and economic region.

In the investigation of heterogeneity we intend to stratify for combinations of risk groups or specific populations and background epidemiology, that is, at different levels of TB prevalence among the screened population.

---

**Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status (Protocol)**

Copyright © 2014 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

We do not intend to do a formal comparison of the accuracy of screening tests as part of this review. As part of the TB screening guideline development process, we will compare diagnostic algorithms as described above.

## BACKGROUND

### Target condition being diagnosed

Tuberculosis (TB) is an important infectious cause of morbidity and mortality among adults worldwide. In 2011, there were 8.7 million new and 12 million prevalent cases of TB, almost one million TB deaths of HIV-uninfected people and an additional 0.43 million deaths among HIV-infected people (WHO 2012). An estimated one-third of the world's population is infected with *Mycobacterium tuberculosis*, the microorganism that causes TB. In humans, *M. tuberculosis* (MTB) infection usually affects the lungs and spreads by airborne transmission (Lawn 2011). Patients with infectious TB spread bacilli, most commonly through coughing. After initial infection, approximately 5% of infected people develop active tuberculosis, referred to as TB. Between 90 to 95% of infected people develop a latent TB infection (LTBI), which may reactivate at a later stage especially in the presence of conditions that affect immunity (including HIV infection, undernutrition, and old age) (Rieder 1999). It can take months to years for people to develop symptomatic and bacteriologically detectable TB. LTBI and TB are increasingly seen as two ends of a continuous spectrum. In between are early disease states that may be described as incipient TB and subclinical TB (Achkar 2011). In the absence of diagnosis and treatment, people with active TB may be infectious for prolonged time periods. In HIV-negative people with active TB, the average duration until self-cure or death is three years, and case fatality with no treatment is approximately 70% for smear-positive (that is, detectable with sputum smear microscopy) and 20% for smear-negative TB (Tiemersma 2011).

The decline in estimated global TB incidence, about 2% per year, is far below the average decline of 20% per year required to reach the elimination target of < one case per million population in 2050 (Raviglione 2012; WHO 2012). In 2011, only an estimated 66% of incident TB cases were detected globally (WHO 2012). Recent prevalence surveys have revealed a considerable burden of undiagnosed culture positive (that is, detectable with mycobacterial sputum culture) smear-negative TB, and a minority of those cases report classical symptoms (Ayles 2009; Corbett 2009; van't Hoog 2011a; MoH Myanmar 2012). Improving TB case detection to reduce the pool of infectious TB that contributes to transmission (Corbett 2010) is important to further reduce TB incidence, prevalence and mortality, and reach the goals of TB control (WHO 2006; Raviglione 2012). Most TB cases are detected passively, among symptomatic people seeking care (Golub 2005).

Passive case detection results in considerable delay in TB detection. (Sreeramareddy 2009) and at the time of diagnosis TB patients identified through passive case detection have more symptoms and signs of illness compared to patients found through active case detection (den Boon 2008; van't Hoog 2013). Thus, a large proportion of patients with infectious TB will go undiagnosed if only passive case detection is used. More active approaches are needed to increase case detection, and systematic screening for active TB is a possible means of achieving this (Raviglione 2012; Lonnroth 2013).

### Screening

The Strategic and Technical Advisory Group TB (STAG-TB) recommends that the World Health Organization (WHO), working with partners, develops guidelines on TB screening (Stop TB 2011). The WHO has defined screening as “the presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well people who probably have a disease from people who probably do not. Screening tests are not intended to be diagnostic. People with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment” (Wilson 1968). For the purpose of guideline development, TB screening is defined as “systematic identification, in a predetermined target group, of people with suspected active TB, by the application of tests, examinations, or other procedures which can be applied rapidly” and these people should be tested with a confirmative diagnostic test. Screening could be offered to both those who seek health care (with or without symptoms or signs compatible with TB) and those who do not. Screening is offered systematically to predetermined groups, and not only in response to a specific request or complaint by an individual seeking care (Lonnroth 2013; WHO 2013). The two main goals of systematic screening for active TB are (1) better health outcomes for people with TB, through earlier detection and treatment; and (2) more effective reduction of TB transmission and incidence through shortening the average duration of TB infectiousness (Lonnroth 2013; WHO 2013).

### Index test(s)

This review focuses on symptom and chest radiography (CXR) screening. In symptom screening, individuals are questioned about the presence of one or more symptoms considered suggestive of

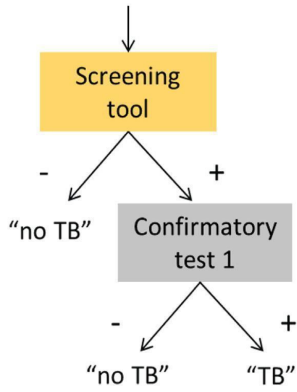
pulmonary TB, which are respiratory symptoms such as persistent cough and haemoptysis, and systemic symptoms including weight loss, fever, night sweats and fatigue (Maher 2009). Chest radiography as a screening tool involves having participants undergo one posterior-anterior CXR recording. Different technologies exist: conventional CXR (producing a 36 cm x 43 cm film), digital radiography and mass miniature radiography (MMR) (Kerley 1942). CXR classification systems may distinguish between any abnormality versus normal, or among abnormal CXRs only abnormalities suggestive of TB may qualify as a positive screen (den Boon 2006). The latter requires interpretation by specialist readers (usually radiologists or pulmonologists), while presence for any abnormality can more easily be interpreted by health workers with a general medical background (for example, medical officers, clin-

ical officers, radiographers) (WHO 2010; van't Hoog 2011). Screening may be done with either symptom or CXR screening, or with symptom and CXR screening combined in parallel or sequentially (Figure 1) (Hayen 2010). Sequential (or serial) screening means that in the first step people are screened for symptoms, and as a second step, CXR screening is offered only to symptom positives. Parallel screening implies that both symptom and CXR screening are offered, and people found to have symptoms, or abnormalities, or both on CXR are eligible for further bacteriological examination. This is for example practiced in TB prevalence surveys in order to have as high sensitivity as possible, while at the same time avoiding the need for laboratory investigation on all study subjects (WHO 2010).

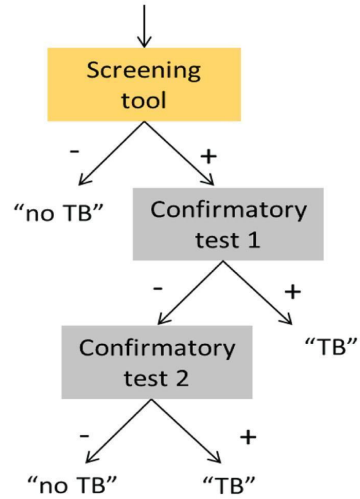
**Figure 1.**

Different screening and diagnostic algorithms.

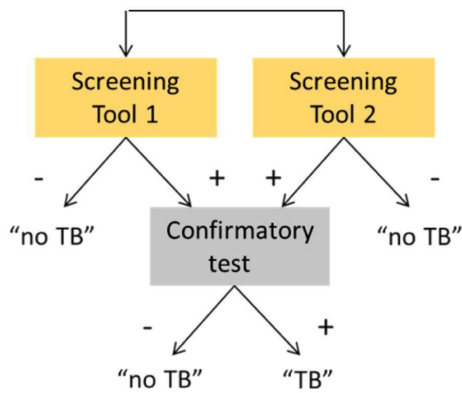
**A. One screening test followed by one confirmatory test**



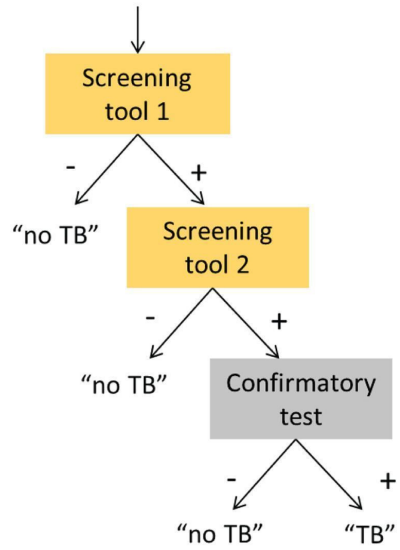
**B. One screening test followed by two sequential confirmatory tests**



**C. Two parallel screening tests**



**D. Two sequential screening tests**



## Clinical pathway

In a TB screening program, the screening test(s) are offered as part of a diagnostic algorithm that also includes one or more confirmatory tests. Individuals with a positive screen are offered further confirmatory testing to establish a TB diagnosis. True screen positives are people rightfully referred for confirmatory testing, and false screen positives are people who are referred for confirmatory testing while they do not have TB. They may or may not be ruled out by the confirmatory test. Individuals with a positive screen, but negative confirmatory test would not necessarily be declared disease-free, but may be advised on further examination or follow-up if warranted by the actual finding on screening (for example, severity of symptoms or the CXR finding (Okada 2012). People with a negative screen would not be further evaluated. This group includes both the true screen negatives who do not have TB, and false screen negatives, who will not be evaluated further although they do have TB. The confirmatory test may be sputum smear microscopy, the Xpert® MTB/RIF test (Cepheid, Sunnyvale, CA), and, in more resourceful settings, mycobacterial culture. These are also reference tests for the purpose of this review. People that have a negative result of the confirmatory test(s) available in their setting may be started on empirical TB treatment after further clinical evaluation and a trial of broad spectrum antibiotics, or chest radiography, or both. New reference tests may become available in the future.

## Reference tests

### Mycobacterial culture

The main goal of systematic TB screening is early detection of people who are infectious and can spread *M. tuberculosis*. For this condition, confirmation of mycobacterial growth in cultured sputum followed by mycobacterial speciation to demonstrate *M. tuberculosis* presence is considered the reference test. Culture on liquid medium is believed to be the most sensitive, although prior to the availability of automated reading of mycobacterial growth inhibitor tubes (MGIT culture), culture on solid medium (Löwenstein-Jensen (LJ)) has been the mainstay, and may still be the only available method in resource-constrained settings. MGIT culture increases the recovery of mycobacteria by 11 to 18% compared to LJ culture, but MGIT culture alone may have slightly lower specificity due to higher contamination rates (Hanna 1999; Chien 2000; Somoskövi 2000; Whitelaw 2009). The yield of mycobacterial culture also increases if two or three specimens per patient are tested (Monkongdee 2009).

### Sputum smear microscopy

Sputum smear microscopy is the most commonly available TB diagnostic test. Sputum smear microscopy detects acid fast bacilli (AFB) presence, which is considered indicative of *M. tuberculosis* in high TB-incidence settings. Compared to culture, sensitivity of the Ziehl-Neelsen method (ZN) shows wide variation, and is between 50 to 70% in a majority of studies (Steingart 2006a; Steingart 2006b). Direct ZN microscopy specificity is 98% (95% CI 97 to 99%) (Steingart 2006a; Steingart 2006b; Cattamanchi 2010). Smears may also be positive due to AFBs that are not *M. tuberculosis* or to artefacts. Auramine-stained fluorescence microscopy (FM) sensitivity is on average 10% higher than of ZN, but with slightly reduced specificity (Steingart 2006a). Processing sputum by centrifugation and various chemicals, including bleach and NaOH, show varying levels of increase in the sensitivity of microscopy compared with the direct smear method, and similar or slightly lower specificity (Steingart 2006b; Cattamanchi 2010).

### Nucleic acid amplification tests (NAAT)

The Xpert® MTB/RIF test (Xpert) is currently the only NAAT that is endorsed by WHO for large scale deployment (WHO 2011b). Compared to culture, Xpert has 92% sensitivity and 99% specificity in smear-positive and smear-negative patients combined in pilot studies (Boehme 2011), and a pooled sensitivity of 88% (95% CI 83 to 92%) and pooled specificity of 98% (95% CI 97 to 99%) in a systematic review (Steingart 2013) and is an acceptable reference test.

### Other

Other types of active TB are extra-pulmonary TB (EPTB), a condition that may affect almost every other organ and constitutes 13% of new TB cases in all ages globally (WHO 2012), and culture-negative active pulmonary TB, characterized by clinical disease and highly suggestive CXR abnormalities not explained by other causes (Maher 2009). Clinical diagnosis and start of empirical TB treatment, is commonly practiced in settings where mycobacterial culture is not part of routine diagnosis for people with suspected pulmonary TB who have negative sputum smears. Clinical algorithms that include trial of antibiotics and a CXR if the trial was not successful have generally very low sensitivity, while diagnosis based on CXR has low specificity (van Cleeff 2003; Soto 2011; Swai 2011). In this review, we do not consider clinically diagnosed TB as an acceptable reference test because of the lack of a uniform definition, poor and variable accuracy of clinical algorithms, and the varying ability to establish differential diagnostic causes across settings. EPTB and culture-negative active pulmonary TB may be detected earlier through active screening especially in high income countries, but are not a primary focus of active screening in other

settings due to diagnostic challenges and low probability of transmission. Also, we do not consider serological tests, which are not recommended for TB diagnosis (Steingart 2007), and other tests that are not endorsed by WHO for TB diagnosis as reference tests for this review.

## Rationale

This review aims to contribute to the development of TB screening guidelines which seek to provide guidance about if, when, whom and how to screen (WHO 2013). We will compile evidence about the accuracy of the most available screening tools, and if possible generate summary estimates of the sensitivity and specificity of symptoms, chest radiography (CXR) and combinations of those if used as TB screening tools. The accuracy of the screening tools and the confirmatory tests, as well as the TB prevalence in the screened population, will determine the potential yield of a screening program and the burden on individuals and the health service. The latter includes the required amount of confirmatory tests and possibly diagnostics and care for other conditions. In practice, screening initiatives may face lower yields if not all eligible individuals accept screening or confirmatory testing, or if some of the people diagnosed with TB as a result of the screening program do not initiate treatment. The literature on those challenges is summarized in other reviews (Lonnroth 2013; Kranzer 2013). The TB screening guidelines aim to provide guidance to decision-makers on the choice of diagnostic algorithms (combinations of one or more screening test(s) and confirmatory test(s)) in different populations and settings (Lonnroth 2013; WHO 2013). Therefore the yield, positive and negative predictive value, and requirements in terms of diagnostic tests of different diagnostic algorithms will be calculated for different levels of TB prevalence as part of the guideline development process. This information should help decision-makers choose the best diagnostic algorithm option for their specific setting, taking into account the TB prevalence, resource availability and logistical aspects (for example, availability of X-ray or Xpert equipment). The pooled estimates of sensitivity and specificity of symptom and CXR screening from this review will inform these calculations and recommendations.

This review includes TB screening of HIV-negative people and people with unknown HIV status (a proportion of whom may be HIV-infected). In regions with a generalized HIV-epidemic, the risk of developing active TB is 20 to 37 times greater in the presence of HIV-infection (Getahun 2010), and mortality in HIV-infected TB patients is high (Cox 2010; Kyeyune 2010). The sensitivity of sputum smear microscopy and Xpert is lower in HIV-infected individuals with presumed TB (Getahun 2007; Boehme 2011). Therefore people living with HIV should be systematically screened for active TB at each visit to a health facility, as outlined in the guidelines for intensified TB case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings (WHO 2011a). Individuals with a known HIV-positive

status should be referred for HIV-care and treatment if they are not yet enrolled. For those clinic settings, screening algorithms have already been defined based on a recent systematic review to determine a screening rule in HIV-infected people (Getahun 2011).

## OBJECTIVES

To assess the sensitivity and specificity of questioning for presence of one or more selected symptoms, or symptom combinations, or both; chest radiography; and combinations of those as screening tools for detecting bacteriologically confirmed active pulmonary TB in people considered eligible for TB screening who are HIV-negative or whose HIV-status is unknown.

### Secondary objectives

If data allow, we will investigate heterogeneity in relation to:

- background epidemiology (prevalence of pulmonary TB and of HIV among the study population);
- risk groups targeted (for example, migrants, occupational, prisoners, or the general population);
- reference standard (culture, Xpert, smear microscopy);
- screen test definition;
- representativeness of the study design and population for intended screening practice (inclusion of people without any symptoms or CXR abnormalities);
- study participants characteristics (age, sex and HIV status);
- geographic area and economic region.

In the investigation of heterogeneity we intend to stratify for combinations of risk groups or specific populations and background epidemiology, that is, at different levels of TB prevalence among the screened population.

We do not intend to do a formal comparison of the accuracy of screening tests as part of this review. As part of the TB screening guideline development process, we will compare diagnostic algorithms as described above.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

We will include cross-sectional studies or observational cohort studies where a series of participants are tested with symptom



screening, or chest radiography screening, or both, and the reference standard, or where participants are randomized to different screening tests and all participants are verified by the same reference standard. Also we will include studies conducted as part of a baseline of a cohort or randomized trial. In randomized studies comparing screening strategies, we will regard each arm as a separate cohort. Case control studies will not be included because of their high risk of bias in diagnostic accuracy studies (Rutjes 2006). Studies that collect data on one or more (potential) screening tools and also evaluate participants with a negative screen by a reference standard are not often conducted due to intense resource requirements. Studies with the primary objective of evaluating the accuracy of a screening tool are rare. Therefore we will not restrict the search and inclusion to 'screening studies' but will also look for studies that have a different primary objective but can potentially provide data that are relevant for our purpose. This applies, for instance, to community TB prevalence surveys for which the primary goal is measuring prevalence. We will include participants from these studies that are screened for symptoms, or abnormalities on CXR, or both, and are offered confirmatory testing. Also, baseline measurements of a TB incidence cohort or intervention trial in which people with prevalent TB need to be excluded at baseline may provide useful data for our purpose.

We will only include studies from which diagnostic two-by-two tables can be generated for a specific screen (symptom definition or chest radiography finding), that is, studies that report data from which we can extract true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). We will include studies in which not all participants are subjected to the reference standard. This applies to TB prevalence surveys whereby it is assumed that people without symptoms and without CXR abnormalities do not have active TB. The issue will be addressed in the quality rating (QUADAS-2 domain 4: Flow and timing).

We will exclude studies in which screening is applied, but the number of TB cases identified is zero. For cohort studies, we will only consider TB cases that were identified from investigations initiated at the time the screening was applied. Incident cases that arise after the screening will not be considered, unless the study evaluates screening methods to identify the incident cases, such as screening was applied at the time of case identification.

In this review, we will only include studies published in the past 20 years (from 1992 onwards) because from that timepoint onwards, the directly observed therapy (DOTS) strategy was implemented which has led to improvements in passive case detection and standardized treatment (Dye 1998). Prior to DOTS, case detection was generally lower and any screening would result in predominant detection of people with more advanced TB disease. Since this epidemiological situation differs from the current situation, the results from older studies are not as relevant. Moreover, older studies frequently screened using MMR which is rarely used nowadays and has lower sensitivity compared to standard chest radiography (Kerley 1942).

## Participants

Included participants will be individuals eligible for systematic screening and not known to have active TB at the time of screening. We will include all types of populations, so study populations may vary from the general population in an area with high TB rates (for example, in mass case finding or TB prevalence surveys) to specific target populations with much higher TB prevalence than the general population. Examples of specific populations are studies that target family members of a patient diagnosed with TB, studies in homeless populations, as well as studies about screening for immigration or occupational purposes (for example, in goldminers) where the goal may be to exclude people with active disease rather than early disease detection. We will include studies that screened participants for the first time or only once, as well as reports from populations enrolled in longitudinal screening programs with repeated rounds at predetermined intervals. These differences may affect the interpretation of a symptom or CXR screen and will be treated as a potential source of heterogeneity. We expect to report on the accuracy of screening tools in different populations subgroups.

The review will focus on adults (15 years and older) and include studies that combine adults and children if adults are a majority. We will exclude studies focusing on young children (0 to 5 years old) or paediatric TB only because the clinical presentation of TB in young children differs from the presentation in adults and older children. Extrapulmonary disease is more common. If the lungs are affected, young children more often have paucibacillary disease, and obtaining sputum is difficult. In children, clinical presentation including CXR findings are often part of the reference standard (Graham 2012; Luabeya 2012). We will exclude studies of HIV-infected people only, since this review covers TB screening of HIV-negative people and people with unknown HIV status.

We will exclude studies that evaluate symptoms, or CXR, or both in a typical passive case detection setting. This applies to clinical settings where patients report to a health facility due to illness and have symptoms and signs that warrant TB investigations according to regional or global guidelines for passive case detection (AARC 1993; Migliori 2006; WHO 2010a). We will include studies in an out-patient context among people who would not be considered a presumed TB case by such guidelines (for example, attendants of diabetic clinics, antenatal clinics). Prevalence surveys where people already on TB treatment who still have bacteriologically positive TB are also considered a case will be included and described, since the proportion of identified cases to whom this applies is usually small (Hoa 2010; van't Hoog 2011a).

## Index tests

For symptom screens, we will select studies that evaluate one or more author-defined symptoms or symptom combinations, and include all reported screens for data extraction and basic description.

For CXR screening we will include studies that used conventional radiography (such as, large films, chemical development), digital radiography, or computed radiography (which is an 'upgrade' that allows the production of digital radiographs by conventional X-ray equipment). We will exclude studies using MMR only, since this method is not expected to be used for future screening purposes and has generally lower sensitivity compared to conventional or digital radiography. With respect to the classification of abnormalities, we will include all author-defined classification systems.

We expect that prolonged cough, cough of any duration and any one out of a number of TB symptoms will be commonly reported symptom screens, and any CXR abnormality or abnormalities suggestive for TB the most common CXR screens. We will also include other symptom screen definitions or CXR classifications, and summarize them if they are reported in several studies and are sufficiently different from the ones already mentioned.

For screening programs, accuracy estimates of CXR screening may be informative separated for (i) the entire target population, and (ii) populations that are pre-screened with symptoms. Therefore we will include studies that report the accuracy of CXR in a population that was prescreened with symptoms, without reporting results on the accuracy of the symptom screening tool (for example, if first all people with a cough were selected, and the accuracy of CXR is only evaluated in this pre-selected group). In the analysis, we will consider studies evaluating CXR screening in a population in which there has been pre-screening with symptoms a subgroup, since there may be relevance in accuracy estimates of CXR as a second screen in a sequential screening algorithm. In such an algorithm, people would first be offered symptoms screening, and if positive, CXR screening. If the CXR screen is also positive, confirmatory testing would be offered.

### Target conditions

The target condition of TB screening is infectious pulmonary TB, characterized by the presence of *M. tuberculosis* in sputum. Studies that perform sputum culture on an entire population usually identify some people with a single positive sputum culture or smear at one point in time without any symptom or CXR abnormality. This may reflect an early stage of infectiousness along the TB spectrum but may however also reflect transient primary *MTB* infection or laboratory cross contamination (Corbett 2009; Lewis 2009). The latter two are not primary targets of screening programs and inclusion as a TB case in a screening tool evaluation would underestimate the sensitivity of the screening tool. To avoid the latter two, we will define the target condition as bacteriologically confirmed pulmonary TB with some suggestion of active disease, for example, a symptom, CXR abnormality or repeated positive sputum bacteriology or both. We will include studies that in their definition of a bacteriologically positive TB case allow for the inclusion of people with one positive sputum culture or smear or Xpert only but without symptoms or CXR abnormalities, since

the proportion of such cases is likely small and it may not be feasible to exclude them from the accuracy calculations. We will rate such studies differently in the assessment of methodological quality (Appendix 1) and exclude studies evaluating tests to investigate for the presence of latent TB infection only.

### Reference standards

The reference standard is defined as any author-defined combination of mycobacterial culture (on solid or liquid medium), or sputum smear microscopy, or Xpert or other NAATs. In the quality rating we will consider a number of bacteriological reference standards as equal, and of sufficient quality: 1) mycobacterial culture followed by mycobacterial speciation; 2) Xpert and 3) two positive smears but only in studies where participants were tested with sputum smear and culture and a small proportion ( $\leq 10\%$ ) of cases is defined based on two positive smears but contaminated or negative or missing culture results. We will include studies that use smear microscopy only as the reference standard, but considered at risk of bias in the QUADAS-2 tool (domain 3 - Reference standard), and be analysed as a subgroup. We will consider the (possible) inclusion among the TB cases of some people who had one positive sputum culture, or Xpert, or smear at one point in time without any symptom or CXR abnormality or confirmation at a second time point as an applicability concern, as explained above. Studies in which not all participants received the reference standard will be included and we will address the methodological limitation in the quality rating and analysis. This is a common design in TB prevalence surveys (WHO 2010) whereby it is assumed that people with a negative symptom screen and a negative CXR screen do not have TB. Bias from this assumption is likely small if a wide range of CXR abnormalities and symptoms is used for screening. If only prolonged cough and TB suggestive CXR abnormalities are used to screen, the proportion of TB cases identified in such surveys overestimate sensitivity and reflect a yield rather than sensitivity. The estimate of specificity from prevalence surveys with partial verification bias from this type of design is however still reliable, due to the large numbers of TB-negatives in prevalence surveys. We do not consider incorporation bias an issue because the reference standard requires evidence from microbiological tests, and cannot be based on symptoms, or radiograph findings, or both, alone.

### Search methods for identification of studies

#### Electronic searches

We will search the data bases MEDLINE, EMBASE, LILACS and HTA (Health Technology Assessment) from 1992 to 2013 to identify titles and abstracts of peer-reviewed papers using search terms listed in Appendix 2. We will include combinations of three

domains: (i) “tuberculosis” and related terms, (ii) terms related to “screening”, “survey”, “sensitivity”, “specificity”, and (iii) search terms related to the reference standard, “bacterial culture”, “microscopy” (Appendix 2). To identify all possible studies, we will not use a diagnostic search filter.

### Searching other resources

We will check reference lists of relevant reviews and studies, search websites of the WHO Stop TB department, and ask experts for relevant studies and still unpublished reports. Unpublished reports will be included if permission is granted by the investigators. We will perform forward and backward reference checking of the selected studies.

### Data collection and analysis

#### Selection of studies

We will include studies using broad inclusion criteria: (i) the publication was original research; and (ii) titles, abstracts, or key words suggested that symptom or CXR screening, or active case finding for TB took place in humans and data to determine accuracy of a screening tool may be available. Two authors will review all titles and abstracts for eligibility. Studies will be included if they meet the inclusion criteria. There will be no language restriction. We will develop a database of all articles, including full references and abstracts, in Reference Manager (v12) (Reference Manager 12). We will obtain full text articles of these studies and two authors will assess for study eligibility using the predefined inclusion and exclusion criteria. The authors will resolve any disagreements through discussion and, if necessary, with a third author.

#### Data extraction and management

We will develop a data-extraction form. To pilot the form, two independent reviewers (AvH and ML) will extract a set of data from a few studies. Based on the pilot, we will finalize the extraction form. One author will extract all relevant data from the included studies. A second author will check the data extraction. The two authors will discuss inconsistencies to obtain consensus. We will resolve any disagreements through either consensus or by a third author. We will enter data into Microsoft Excel (Excel) through a data entry screen.

The data extraction will include the following characteristics:

- Authors, publication year, journal;
- Details of study: participant’s country of residence (and classify country according to economic region - low versus middle versus high income); setting, including risk group (occupational, general population, immigrants, mass screening), urban or rural; study design; method of participant selection;

number of participants enrolled; number of participants for whom results were available;

- Study participants: age, sex, HIV-status, history of TB, % of smokers. On age we should record the mean+sd/median age and age range of the population included in the analysis. If results are presented for different age groups will we extract those as well. Similar for sex. On HIV-status: record the proportion of the population with known HIV-status, and % with positive HIV-status. Also add the (estimated) background HIV-prevalence in the study population, either from the publication or search from the UNAIDS reports. On history of TB: record proportion with history of TB, if available;

- First or one-off screening versus repeated screening at regular intervals. If the latter, we will record whether the same participant is included more than once in the analysis;

- Prevalence of target condition (pulmonary TB, smear-positive, bacteriologically-positive) in the population. In addition, we will also record (1) the TB case notification rate (per 100,000 population) in the study population, or if unavailable in the region or country, either from the introduction or methods section of the publication or search elsewhere; and (2) a measure of case detection (the patient detection rate, if known or can be calculated, or the case detection rate). In areas with poor case detection, one finds more cases of advanced TB disease during active case detection, which affects the sensitivity of the screening tools;

- Stage of infection: proportion of the true TB cases included in the report that are bacteriologically positive, but have no signs of active disease (are asymptomatic and have no CXR abnormalities, and no bacteriological confirmation at a second point in time);

- Treatment status: number and proportion of true TB cases included in the report who were already on TB treatment at the time of screening;

- Reference standard: culture and type of medium (solid or liquid), microscopy and type (light or fluorescence), Xpert; number of samples per individual tested, number of positive samples required for positive diagnosis, definition of positivity (including number of colonies, speciation method, smear grading), other criteria included in reference standard. We will also include an exact narrative of the definition and record the definitions of the classifications for each included publication;

- Index tests: For all reported screens or screen combinations that are meaningful we will extract the definition of symptom screen; total number of symptoms asked for; radiography: equipment type (conventional, digital, MMR), CXR classification system (any abnormality, suggestive of TB - if so, what is the definition);

- The definitions of the classifications for each included publication, type of reader (expert, radiologist, or pulmonologist; general medical officer; clinical officer; nurse; radiographer; other);

- QUADAS-2 items ([Appendix 1](#));
- Details of outcomes: the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN); number of participants missing or unavailable test results. If the analysis was adjusted for cluster sampling in the original report and results in wider confidence intervals (CIs) of a screen, we will extract the reported point estimate and CIs.

### Assessment of methodological quality

Two reviewers will assess the methodological quality of included studies using the modified Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) instrument ([Reitsma 2009](#); [Whiting 2011](#)). The tool with signalling questions tailored to this review is in [Appendix 1](#). We will assess each of the four domains - patient selection, index test(s), reference standard, flow and timing - in terms of risk of bias, and the first three domains in terms of concerns regarding applicability to the review's research question. We do not consider incorporation bias to be an issue because the reference standard requires evidence from microbiological tests, and cannot be based only on symptoms, or radiograph findings, or both.

We will classify each item as 'yes' (adequately addressed), 'no' (inadequately addressed), or 'unclear' when insufficient data are reported to permit a judgment. The data extraction form includes criteria for those scores. We will resolve any disagreements through consensus or through discussion with a third author. We will present results in text and graphs.

### Statistical analysis and data synthesis

We will first report the studies identified for inclusion in the analysis according to the number of studies found in different target populations, levels of TB prevalence and screen category. We will base the decision for which categories to collapse on the number of studies and screen definitions found in each category. Although this may potentially result in a large number of subgroups, in practice we expect that results will be retrieved for only a limited number of combinations.

### Screen definitions

We will categorize the reported symptom screens into at least two symptom screen definitions:

1. One of the symptom screens will focus on cough, for example, cough for two or more weeks. Prolonged cough is an important component of the definition of presumed TB in clinical guidelines ([Migliori 2006](#)). Closely related screens, for example prolonged productive cough, will be included in this category unless there are sufficient studies to do a separate analysis. At the data extraction stage we will record the exact definition of the reported screen.

2. Presence of at least one symptom positive out of a combination of at least three screening questions that, in addition to cough, also includes systemic symptoms such as fever, night sweats, and weight loss ([Getahun 2011](#)).

We will include more definitions of symptom combinations if they are frequently reported, are well defined and well distinguishable from the two symptom screen definitions given above.

Similarly, we will subdivide CXR screens by classification system, or type of reader, or both, if there are sufficient data to do so. At a minimum we will distinguish between any abnormality and abnormalities suggestive of TB or consistent with TB unless the heterogeneity in classification systems applied is too large. We do however expect heterogeneity regardless of the classification system because of inter-reader variation ([den Boon 2005](#); [van't Hoog 2011](#)).

We will subdivide the reference standards used for subgroup analysis according to assessment of evidence quality. We will distinguish between culture, culture and smear combined, or Xpert only, versus smear microscopy only. Since we expect a small number of eligible studies that apply Xpert only (or other NAATs), we will explore the effect of excluding those in sensitivity analyses.

Diagnostic two-by-two tables will be generated, from which we will calculate sensitivities and specificities for each index test with 95% CIs and present in paired forest plots for each study. In addition, we will use a Receiver Operating Characteristic (ROC) plot of sensitivity versus 1-specificity to display the data for each test. If studies show sufficient clinical homogeneity (for example, same index test, similar definition of the target condition or reference standard, similar screening population), we will perform meta-analysis of pairs of sensitivity and specificity by the use of bivariate random-effects methods ([Reitsma 2005](#)). The bivariate model is preferred because we deal with binary decisions for which an implicit threshold is assumed. We will develop the bivariate model in SAS<sup>®</sup> (SAS) or the Stata<sup>®</sup> (Stata) `metandi` command.

For some subgroups or screen definitions we may not be able to give meaningful summary estimates of sensitivity and specificity. We will evaluate them using descriptive methods.

### Investigations of heterogeneity

We will examine the forest plots and ROC plots for heterogeneity. If the data allow, we will analyse potential determinants or sources of heterogeneity as covariates in the models. We will include the following covariates, where appropriate:

- TB prevalence;
- Risk groups or subpopulations targeted, and combinations of those;
- First time versus repeatedly screened population; pre-screened;
- HIV-status (focus on HIV-negative), or background HIV-prevalence in study population;
- Age;

- Smoking prevalence;
- Reference standard (culture, Xpert or smear microscopy);
- Screening test definition consistency (duration of symptom(s); number of symptoms included; which CXR abnormalities are considered suggestive of TB);
  - Representativeness of the study population for intended screening practice (inclusion of people without any symptoms or CXR abnormalities);
  - Geographic area and economic region.

### Sensitivity analyses

To explore whether the results we find are robust for methodological challenges, we will perform a sensitivity analysis of the QUADAS-2 domains. We will assess the sensitivity of results to the inclusion and exclusion of studies with quality concerns.

### Assessment of reporting bias

We will not assess reporting bias in the included studies.

### Quality of the evidence

We will assess the quality of the evidence using the GRADE methodology (Schünemann 2008).

## ACKNOWLEDGEMENTS

We acknowledge the participants of the WHO screening meeting 23 to 24 May 2012 for their comments on the protocol, particularly those involved in groupwork 2: Noria Yamada, Ikushi Onozaki, Lucy Blok and Hannah Ryan. We thank René Spijker for his assistance with search strategy development. The editorial base for the Cochrane Infectious Diseases Group is funded by the UK Department for International Development (DFID) for the benefit of developing countries.

## REFERENCES

### Additional references

#### AARC 1993

American Association for Respiratory Care. AARC Clinical Practice Guidelines for Respiratory Care. *Respir Care* 1993; **38**(5):495–9.

#### Achkar 2011

Achkar JM, Jenny-Avital ER. Incipient and subclinical tuberculosis: defining early disease states in the context of host immune response. *Journal of Infect Diseases* 2011; **204** (Suppl 4):S1179–86.

#### Ayles 2009

Ayles H, Schaap A, Nota A, Sismanidis C, Tembwe R, de Haas P, et al. Prevalence of tuberculosis, HIV and respiratory symptoms in two Zambian communities: Implications for tuberculosis control in the era of HIV. *PLoS ONE* 2009; **4**(5):e5602.

#### Boehme 2011

Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet* 2011; **377**(9776): 1495–505.

#### Cattamanchi 2010

Cattamanchi A, Davis JL, Pai M, Huang L, Hopewell PC, Steingart KR. Does bleach processing increase the accuracy of sputum smear microscopy for diagnosing pulmonary tuberculosis?. *Journal of Clinical Microbiology* 2010; **48**(7): 2433–9.

#### Chien 2000

Chien HP, Yu MC, Wu MH, Lin TP, Luh KT. Comparison of the BACTEC MGIT 960 with Lowenstein-Jensen medium for recovery of mycobacteria from clinical specimens. *International Journal of Tuberculosis and Lung Disease* 2000; **4**(9):866–70.

#### Corbett 2009

Corbett EL, Bandason T, Cheung YB, Makamure B, Dauya E, Munyati SS, et al. Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control. *International Journal of Tuberculosis and Lung Disease* 2009; **13**(10):1231–7.

#### Corbett 2010

Corbett EL, Bandason T, Duong T, Dauya E, Makamure B, Churchyard GJ, et al. Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): a cluster-randomised trial. *Lancet* 2010; **376**(9748): 1244–53.

#### Cox 2010

Cox JA, Lukande RL, Lucas S, Nelson AM, Van Marck E, Colebunders R. Autopsy causes of death in HIV-positive individuals in sub-Saharan Africa and correlation with clinical diagnoses. *AIDS reviews* 2010; **12**(4):183–94.

#### den Boon 2005

den Boon S, Bateman ED, Enarson DA, Borgdorff MW, Verver S, Lombard CJ, et al. Development and evaluation of a new chest radiograph reading and recording system for epidemiological surveys of tuberculosis and lung disease.

- International Journal of Tuberculosis and Lung Disease* 2005; **9**(10):1088–96.
- den Boon 2006**  
den Boon S, White NW, van Lill SW, Borgdorff MW, Verver S, Lombard CJ, et al. An evaluation of symptom and chest radiographic screening in tuberculosis prevalence surveys. *International Journal of Tuberculosis and Lung Disease* 2006; **10**(8):876–82.
- den Boon 2008**  
den Boon S, Verver S, Lombard CJ, Bateman ED, Irusen EM, Enarson DA, et al. Comparison of symptoms and treatment outcomes between actively and passively detected tuberculosis cases: the additional value of active case finding. *Epidemiology and Infection* 2008; **136**(10):1342–9.
- Dye 1998**  
Dye C, Garnett GP, Sleeman K, Williams BG. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. Directly observed short-course therapy. *Lancet* 1998; **352**(9144):1886–91.
- Excel [Computer program]**  
Microsoft Corporation. Microsoft Excel. Version 2010. Redmond, Washington: Microsoft Corporation.
- Getahun 2007**  
Getahun H, Harrington M, O'Brien R, Nunn P. Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: informing urgent policy changes. *Lancet* 2007; **369**(9578):2042–9.
- Getahun 2010**  
Getahun H, Gunneberg C, Granich R, Nunn P. HIV infection-associated tuberculosis: the epidemiology and the response. *Clinical Infectious Diseases* 2010; **50**(Suppl 3):S201–7.
- Getahun 2011**  
Getahun H, Kittikraisak W, Heilig CM, Corbett EL, Ayles H, Cain KP, 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 Medicine* 2011; **8**(1):e1000391.
- Golub 2005**  
Golub JE, Mohan CI, Comstock GW, Chaisson RE. Active case finding of tuberculosis: historical perspective and future prospects. *International Journal of Tuberculosis and Lung Disease* 2005; **9**(11):1183–203.
- Graham 2012**  
Graham SM, Ahmed T, Amanullah F, Browning R, Cardenas V, Casenghi M, et al. Evaluation of tuberculosis diagnostics in children: 1. Proposed clinical case definitions for classification of intrathoracic tuberculosis disease. Consensus from an expert panel. *Journal of Infectious Diseases* 2012; **205**(Suppl 2):S199–208.
- Hanna 1999**  
Hanna BA, Ebrahimzadeh A, Elliott LB, Morgan MA, Novak SM, Rusch-Gerdes S, et al. Multicenter evaluation of the BACTEC MGIT 960 system for recovery of mycobacteria. *Journal of Clinical Microbiology* 1999; **37**(3):748–52.
- Hayen 2010**  
Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *Journal of Clinical Epidemiology* 2010; **63**(8):883–91.
- Hoa 2010**  
Hoa NB, Sy DN, Nhung NV, Tiemersma EW, Borgdorff MW, Cobelens FGJ. National survey of tuberculosis prevalence in Viet Nam. *Bulletin of the World Health Organization* 2010; **88**(4):273–80.
- Kerley 1942**  
Kerley P, Major T. Technique in mass miniature radiography. *British Journal of Radiology* 1942; **15**(180):346–7.
- Kranzer 2013**  
Kranzer K, Afnan-Holmes H, Tomlin K, Golub JE, Shapiro AE, Schaap A, et al. The benefits to communities and individuals of screening for active tuberculosis disease: a systematic review. *International Journal of Tuberculosis and Lung Disease* 2013; **17**(1815–7920 (Electronic), 1027–3719 (Linking), 4):432–46.
- Kyeyune 2010**  
Kyeyune R, den Boon S, Cattamanchi A, Davis JL, Worodria W, Yoo SD, et al. Causes of early mortality in HIV-infected TB suspects in an East African referral hospital. *Journal of Acquired Immune Deficiency Syndromes* 2010; **55**(4):446–50.
- Lawn 2011**  
Lawn SD, Zumla AI. Tuberculosis. *Lancet* 2011; **378**(9785):57–72.
- Lewis 2009**  
Lewis JJ, Charalambous S, Day JH, Fielding KL, Grant AD, Hayes RJ, et al. HIV infection does not affect active case finding of tuberculosis in South African gold miners. *American Journal of Respiratory and Critical Care Medicine* 2009; **180**(12):1271–8.
- Lonnroth 2013**  
Lönnroth K, Corbett E, Golub J, Godfrey-Faussett P, Uplekar M, Weil D, et al. Systematic screening for active tuberculosis: rationale, definitions and key considerations. *International Journal of Tuberculosis and Lung Disease* 2013; **17**(3):289–98.
- Luabeya 2012**  
Luabeya KK, Mulenga H, Moyo S, Tameris M, Sikhondze W, Geldenhuys H, et al. Diagnostic features associated with culture of *Mycobacterium tuberculosis* among young children in a vaccine trial setting. *Pediatric Infectious Disease Journal* 2012; **31**(1):42–6.
- Maher 2009**  
Maher D. Clinical features and index of suspicion in adults (HIV-negative and HIV-positive). *Tuberculosis. A comprehensive clinical reference*. Philadelphia: Saunders, 2009:164–8.

**Migliori 2006**

Migliori GB, Hopewell PC, Blasi F, Spanevello A, Raviglione MC. Improving the TB case management: The International Standards for Tuberculosis Care. *European Respiratory Journal* 2006;**28**(4):687–90.

**MoH Myanmar 2012**

Ministry of Health Department of Health Government of Myanmar. Report on National TB Prevalence Survey 2009-2010, Myanmar. [http://www.who.int/tb/advisory\\_bodies/impact\\_measurement\\_taskforce/prevalencesurveymyanmar\\_2009-10report.pdf](http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/prevalencesurveymyanmar_2009-10report.pdf) (accessed 05 Sep 2012).

**Monkongdee 2009**

Monkongdee P, McCarthy KD, Cain KP, Tasaneeyapan T, Nguyen HD, Nguyen TN, et al. Yield of acid-fast smear and mycobacterial culture for tuberculosis diagnosis in people with human immunodeficiency virus. *American Journal of Respiratory Critical Care Medicine* 2009;**180**(9):903–8.

**Okada 2012**

Okada K, Onozaki I, Yamada N, Yoshiyama T, Miura T, Saint S, et al. Epidemiological impact of mass tuberculosis screening: a 2-year follow-up after a national prevalence survey. *International Journal of Tuberculosis and Lung Disease* 2012;**16**(12):1619–24.

**Raviglione 2012**

Raviglione M, Marais B, Floyd K, Lönnroth K, Getahun H, Migliori GB, et al. Scaling up interventions to achieve global tuberculosis control: progress and new developments. *Lancet* 2012;**379**(9829):1902–13.

**Reference Manager 12 [Computer program]**

Thomson Reuters. Reference Manager [Reference Manager]. Version 12. Philadelphia, PA: Thomson Reuters, December 2010.

**Reitsma 2005**

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**(10):982–90.

**Reitsma 2009**

Reitsma JB, Rutjes AWS, Whiting P, Vlassov V, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors) editor(s). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration, 2009.

**Rieder 1999**

Rieder HL. Tuberculosis. Etiologic epidemiology: risk factors for disease given that infection has occurred. *Epidemiologic basis of tuberculosis control*. 1st Edition. Paris: International Union Against Tuberculosis and Lung Disease, 1999:63–119.

**Rutjes 2006**

Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;**174**(4):469–76.

**SAS [Computer program]**

SAS Institute Inc.. Statistical Analysis Software. Version 9. Cary, NC: SAS Institute Inc..

**Schünemann 2008**

Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;**336**(7653):1106–10.

**Somoskövi 2000**

Somoskövi A, Ködmön C, Lantos A, Bártfai Z, Tamási L, Füzy J, et al. Comparison of recoveries of mycobacterium tuberculosis using the automated BACTEC MGIT 960 system, the BACTEC 460 TB system, and Lowenstein-Jensen medium. *Journal of Clinical Microbiology* 2000;**38**(6):2395–7.

**Soto 2011**

Soto A, Solari L, Gotuzzo E, Acinelli R, Vargas D, Van der Stuyft P. Performance of an algorithm based on WHO recommendations for the diagnosis of smear-negative pulmonary tuberculosis in patients without HIV infection. *Tropical Medicine and International Health* 2011;**16**(4):424–30.

**Sreeramareddy 2009**

Sreeramareddy CT, Panduru KV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infectious Diseases* 2009;**9**:91.

**Stata [Computer program]**

StataCorp LP. Stata: Statistical Software. Version Release 11.. College Station, TX: StataCorp LP, 2010.

**Steingart 2006a**

Steingart KR, Henry M, Ng V, Hopewell PC, Ramsay A, Cunningham J, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infectious Diseases* 2006;**6**(9):570–81.

**Steingart 2006b**

Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, et al. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *Lancet Infectious Diseases* 2006;**6**(10):664–74.

**Steingart 2007**

Steingart KR, Henry M, Laal S, Hopewell PC, Ramsay A, Menzies D, et al. Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. *PLoS Medicine* 2007;**4**(6):e202.

**Steingart 2013**

Steingart KR, Sohn H, Schiller I, Kloda LA, Boehme CC, Pai M, et al. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database of Systematic Reviews* 2013;**1**:CD009593.

**Stop TB 2011**

Stop TB Department World Health Organization. Scoping meeting for the development of guidelines on screening for active TB, 31st May to 1 June 2011, Geneva, Switzerland. Draft meeting report. <http://www.who.int/tb/TBscreeningmeetingreport2011.pdf>. World Health Organization, Geneva, Switzerland.

**Swai 2011**

Swai HF, Mugusi FM, Mbwanbo JK. Sputum smear negative pulmonary tuberculosis: sensitivity and specificity of diagnostic algorithm. *BMC Research Notes* 2011;**4**:475.

**Tiemersma 2011**

Tiemersma EW, van der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJ. Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLoS One* 2011;**6**(4):e17601.

**van Cleeff 2003**

van Cleeff MR, Kivihya-Ndugga L, Githui W, Nganga L, Odhiambo J, Klatsner PR. A comprehensive study of the efficiency of the routine pulmonary tuberculosis diagnostic process in Nairobi. *International Journal of Tuberculosis and Lung Disease* 2003;**7**(2):186–9.

**van't Hoog 2011**

van't Hoog AH, Meme HK, van Deutekom H, Mithika AM, Olunga C, Onyino E, et al. High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. *International Journal of Tuberculosis and Lung Disease* 2011;**15**(10):1308–14.

**van't Hoog 2011a**

van't Hoog AH, Laserson KF, Githui WA, Meme HK, Agaya JA, Odeny LO, et al. High prevalence of pulmonary tuberculosis and inadequate case finding in rural western Kenya. *American Journal of Respiratory and Critical Care Medicine* 2011;**183**(9):1245–53.

**van't Hoog 2013**

van't Hoog AH, Marston BJ, Ayisi JG, Agaya JA, Muhenje O, Odeny LO, et al. Risk factors for inadequate TB case finding in Rural Western Kenya: a comparison of actively and passively identified TB patients. *PLoS One* 2013;**8**(4):e61162.

**Whitelaw 2009**

Whitelaw A, Sturm WA. Microbiological testing for *Mycobacterium tuberculosis*. In: Schaaf HS, Zumla AI editor(s). *Tuberculosis. A comprehensive clinical reference*. Philadelphia: Saunders, 2009:164–8.

**Whiting 2011**

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36.

**WHO 2006**

World Health Organization. The Stop TB Strategy: Building on and enhancing DOTS to meet the TB-related

Millennium Development Goals. [http://whqlibdoc.who.int/hq/2006/WHO\\_HTM\\_STB\\_2006.368\\_eng.pdf](http://whqlibdoc.who.int/hq/2006/WHO_HTM_STB_2006.368_eng.pdf). Geneva: World Health Organization (WHO/HTM/TB/2006.368), 2006.

**WHO 2010**

World Health Organization. *Tuberculosis Prevalence Surveys: a handbook* (accessed at [http://www.who.int/tb/advisory\\_bodies/impact\\_measurement\\_taskforce/resources\\_documents/thelimebook/en/index.html](http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/resources_documents/thelimebook/en/index.html)). 2nd Edition. Geneva: World Health Organization, 2010.

**WHO 2010a**

World Health Organization. Treatment of tuberculosis: guidelines for national programmes 4th edition. [http://whqlibdoc.who.int/publications/2010/9789241547833\\_eng.pdf](http://whqlibdoc.who.int/publications/2010/9789241547833_eng.pdf). Geneva: World Health Organization, 2010, issue WHO/HTM/TB/2009.420: 1–160.

**WHO 2011a**

World Health Organization. Guidelines for intensified tuberculosis case-finding and isoniazid preventive therapy for people living with HIV in resource-constrained settings. [http://whqlibdoc.who.int/publications/2011/9789241500708\\_eng.pdf](http://whqlibdoc.who.int/publications/2011/9789241500708_eng.pdf). Geneva: World Health Organization, 2011:1–187.

**WHO 2011b**

World Health Organization. Policy Statement: automated real-time nucleic acid amplification technology for rapid and simultaneous detection of tuberculosis and rifampicin resistance: Xpert MTB/RIF system. [http://whqlibdoc.who.int/publications/2011/9789241501545\\_eng.pdf](http://whqlibdoc.who.int/publications/2011/9789241501545_eng.pdf). Geneva, Switzerland: World Health Organization, 2011, issue WHO/HTM/TB/2011.4:1–36.

**WHO 2012**

World Health Organization. Global Tuberculosis Control: WHO report 2012. [http://www.who.int/tb/publications/global\\_report/2012/en/index.html](http://www.who.int/tb/publications/global_report/2012/en/index.html). Geneva: World Health Organization (WHO/HTM/TB/2012.6), 2012.

**WHO 2013**

World Health Organization. Systematic screening for active tuberculosis. Principles and recommendations. [http://apps.who.int/iris/bitstream/10665/84971/1/9789241548601\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/84971/1/9789241548601_eng.pdf). Geneva: World Health Organization., 2013:1–146.

**Wilson 1968**

Wilson JMG, Jungner G. Principles and practice of screening for disease. *Originally appearing in: Commission on Chronic Illness (1957) Chronic illness in the United States: Volume I. Prevention of Chronic Illness, Cambridge, Massachusetts, Harvard University Press, p.45* (accessed at [http://whqlibdoc.who.int/php/WHO\\_PHP\\_34.pdf](http://whqlibdoc.who.int/php/WHO_PHP_34.pdf)). Geneva, Switzerland: WHO, 1968.

\* Indicates the major publication for the study



## APPENDICES

### Appendix I. QUADAS-2 tool

Key questions	Signaling questions
<b>Domain 1: Patient selection</b>	
Risk of bias: Could the selection of patients have introduced bias?	<ol style="list-style-type: none"> <li>1. Did the study enrol a consecutive or random sample of patients? <ul style="list-style-type: none"> <li>• Yes: if all eligible patients were enrolled; or if the authors reported that the patients were either a consecutive series or randomly selected;</li> <li>• No: if the authors report that the selection was based on clinical judgement of health workers, or participation of randomly selected people in the study was low;</li> <li>• Unclear: if there is discrepancy between the numbers of eligible people and the number of included people, but no reasons given for that, or the selection procedure is not clearly described.</li> </ul> </li> <li>2. Was a case-control design avoided? <ul style="list-style-type: none"> <li>• Yes: if a case-control design was avoided;</li> <li>• No: if a case-control design was not avoided;</li> <li>• Unclear: if not reported or insufficient information is provided to decide.</li> </ul> </li> <li>3. Did the study avoid inappropriate exclusions? <ul style="list-style-type: none"> <li>• Yes: if no study participants were excluded after inclusion;</li> <li>• No: if study participants were excluded (for example, participants with mild or severe symptoms or signs);</li> <li>• Unclear: if insufficient information is provided to decide.</li> </ul> </li> </ol>
Applicability: Are there concerns that the included patients and setting do not match the review question?	<ul style="list-style-type: none"> <li>• High concern: if the study population does not resemble a population that would be considered for a screening TB screening program in practice;</li> <li>• Low concern: if the study population does resemble a population that would be considered for a screening TB screening program in practice;</li> <li>• Unclear: if not reported or insufficient information is provided to decide.</li> </ul>
<b>Domain 2: Index test</b>	
Risk of bias: Could the conduct or interpretation of the index test have introduced bias?	<ol style="list-style-type: none"> <li>1. Were the index test results interpreted without knowledge of the results of the reference standard? <ul style="list-style-type: none"> <li>• Yes: if the screening test was performed without knowing whether the person had infectious TB.</li> <li>• No: if symptom questions were asked after the results of the</li> </ul> </li> </ol>

(Continued)

	<p>reference test were known, or the CXR was interpreted with knowledge of the results of the reference test.</p> <ul style="list-style-type: none"> <li>• Unclear: if insufficient information is provided to decide. For example, if it was unclear whether the CXR reader was blinded to the results of the reference test.</li> </ul> <p>2. If a threshold was used, was it pre-specified?</p> <ul style="list-style-type: none"> <li>• This question was not applicable for our review question.</li> </ul>
<p>Applicability: Are there concerns that the index test, its conduct or its interpretation differ from the review question?</p>	<ul style="list-style-type: none"> <li>• High concern: if the symptom questions or CXR classification were intended as a diagnostic rather than a screening tool; or if part of the population was screened with MMR;</li> <li>• Low concern: if the symptom questions or CXR assessment were done with the intention to screen;</li> <li>• Unclear: if insufficient information is provided to decide.</li> </ul>
<p><b>Domain 3: Reference standard</b></p>	
<p>Risk of bias: Could the reference standard, its conduct or its interpretation have introduced bias?</p>	<p>1. Is the reference standard likely to correctly classify the target condition?</p> <ul style="list-style-type: none"> <li>• Yes: if the reference standard was an author-defined combination of mycobacterial culture (on solid or liquid medium) and possibly sputum smear microscopy, or Xpert, or both, and cases defined by sputum microscopy only are limited to a small proportion (<math>\leq 10\%</math>) in whom culture was contaminated or negative or missing but smears were positive;</li> <li>• No: if the reference standard was not an author-defined combination of mycobacterial culture (on solid or liquid medium) and possibly sputum smear microscopy, or Xpert, or both. This includes studies where sputum smear microscopy was the only reference test;</li> <li>• Unclear: if insufficient information is provided to decide.</li> </ul> <p>2. Were the reference standard results interpreted without knowledge of the results of the index test?</p> <ul style="list-style-type: none"> <li>• Yes: if the screening test results were not known to the people interpreting the reference standard results;</li> <li>• No: if the screening test results were known to the people interpreting the reference standard results;</li> <li>• Unclear: if insufficient information is provided to decide.</li> </ul>
<p>Applicability: Are there concerns that the target condition as defined by the reference standard does not match the question?</p>	<ul style="list-style-type: none"> <li>• High concern: if there was a high probability that a considerable proportion of the TB cases identified in the study did not have bacteriologically confirmed TB or did not have active TB;</li> <li>• Low concern: (i) if the TB cases in the study have TB symptoms or CXR abnormalities in addition to a positive culture, or positive smear microscopy, or both; or (ii) if they have at least two different samples positive on culture, or on smear</li> </ul>

(Continued)

	<p>microscopy, or both.</p> <ul style="list-style-type: none"><li>• Moderate concern: Because we perceive a large contrast between “low” and “high” we added a category “moderate” for the applicability sections. We applied the “moderate” category if the TB cases in the study could include people with one positive sputum culture or Xpert, NAAT or smear only, without the presence or symptoms or CXR abnormalities;</li><li>• Unclear: if insufficient information is provided to decide.</li></ul>
<b>Domain 4: Flow and timing</b>	
Risk of bias: Could the patient flow have introduced bias?	<ol style="list-style-type: none"><li>1. Was there an appropriate interval between the index test and reference standard?<ul style="list-style-type: none"><li>• Yes: if the screening test and reference standard were applied (or samples taken) at the same time or within 1 week;</li><li>• No: if the time between the screening test and reference standard (sample collection) was more than 1 week;</li><li>• Unclear: if insufficient information is provided to decide.</li></ul></li><li>2. Did all patients receive the same reference standard?<ul style="list-style-type: none"><li>• Yes: if all participants were evaluated with the reference standard, and if all or a large majority of participants were evaluated with the same test(s);</li><li>• No: if not all participants were evaluated with the reference standard, or participants received different tests (for example, some smear only, some culture, or different numbers of samples were submitted for testing);</li><li>• Unclear: if insufficient information is provided to decide.</li></ul></li><li>3. Were all patients included in the analysis?<ul style="list-style-type: none"><li>• Yes: if all participants were included;</li><li>• No: if participants who participated were excluded. For instance because they did not provide sputum for a reference test;</li><li>• Unclear: if insufficient information is provided to decide.</li></ul></li></ol>

## Appendix 2. Search strategy

### A. MEDLINE search strategy

Platform: OvidSP

Database: MEDLINE(R) In-Process & Other Non-Indexed Citations

Limits: no limits were used

Methodological filters: none

1	exp Mycobacterium/
2	mycobacterium.ti,ab.
3	tuberculosis/
4	peritonitis, tuberculous/
5	exp tuberculoma/
6	tuberculosis, bovine/
7	exp tuberculosis, cardiovascular/
8	exp tuberculosis, central nervous system/
9	tuberculosis, cutaneous/
10	erythema induratum/
11	tuberculosis, endocrine/
12	tuberculosis, gastrointestinal/
13	tuberculosis, hepatic/
14	exp tuberculosis, lymph node/
15	tuberculosis, miliary/
16	tuberculosis, multidrug-resistant/
17	tuberculosis, ocular/
18	tuberculosis, oral/
19	tuberculosis, osteoarticular/
20	tuberculosis, pleural/
21	tuberculosis, pulmonary/
22	tuberculosis, splenic/
23	tuberculosis, urogenital/
24	(tuberculo* or TB or scrofuloderma).ti,ab.

(Continued)

25	or/1-24
26	(case adj finding).ti,ab.
27	screen*.ti,ab.
28	Mass Screening/ or Mass Chest X-ray/
29	exp Population Surveillance/
30	(disease adj3 surveillance).ti,ab.
31	(case adj detection).ti,ab.
32	Contact Tracing/
33	(contact adj tracing).ti,ab.
34	exp Health Surveys/
35	survey.ti,ab.
36	exp "Sensitivity and Specificity"/
37	(false adj negative).ti,ab.
38	odds.mp.
39	((ROC or HSROC or SROC) adj2 (curve* or analys?s or plot*1)).ti,ab
40	(predictive adj3 value).ti,ab.
41	specificit*.ti,ab.
42	accuracy.ti,ab.
43	or/36-42
44	prevalence.mp. or Prevalence/
45	Cross-Sectional Studies/ or cross sectional.mp.
46	44 or 45
47	34 or 35 or 46
48	(mycobacteri\$ adj2 culture).ti,ab.

(Continued)

49	(microscopy adj2 (sputum smear or ZN or Ziehl-neelsen or FM or fluorescence)).ti,ab
50	lowenstein-jensen.ti,ab.
51	(LJ adj2 medium).ti,ab.
52	“mycobacteria growth incubator tube”.ti,ab.
53	mgit.ti,ab.
54	Xpert.ti,ab.
55	(auramine adj2 staining).ti,ab.
56	((culture or smear) adj positiv*).ti,ab.
57	or/48-56
58	25 and 47 and 57
59	or/26-35
60	25 and 59 and 43
61	25 and 59 and 57
62	58 or 60 or 61
63	Limit 62 to ed=19920101-20130801

We will use the same approach for the EMBASE and LILACS searches.

## CONTRIBUTIONS OF AUTHORS

Anja van't Hoog and Miranda Langendam wrote the protocol.

Miranda Langendam, Mariska Leeftang and Dave Sinclair provided methodological advice.

Knut Lonnoth, Ellen Mitchell and Frank Cobelens provided advice on content. All authors edited the protocol and agreed with the final draft of the protocol.

## **DECLARATIONS OF INTEREST**

None declared.

## **SOURCES OF SUPPORT**

### **Internal sources**

- No sources of support supplied

### **External sources**

- USAID - TB Care, USA.

Financial support through the WHO