

**Scoring systems using chest radiographic features for the diagnosis of pulmonary tuberculosis in adults: A systematic review**

Lancelot M Pinto<sup>1,2</sup>, Madhukar Pai<sup>1,2</sup>, Keertan Dheda<sup>3</sup>, Kevin Schwartzman<sup>1,4</sup>, Dick Menzies<sup>1,2,4</sup>, Karen R Steingart<sup>5</sup>

<sup>1</sup>Respiratory Epidemiology & Clinical Research Unit, Montreal Chest Institute, Montreal, Canada

<sup>2</sup>Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada

<sup>3</sup>University of Cape Town Lung Institute, Cape Town, South Africa

<sup>4</sup>Respiratory Division, Department of Medicine, McGill University, Montreal, Canada

<sup>5</sup>Department of Health Services, University of Washington School of Public Health, Seattle, USA

**Correspondence**

**Dr. Karen R Steingart, MD, MPH**

Department of Health Services  
University of Washington School of Public Health  
Seattle, USA

Tel: +1-646-243-9043  
Fax: +1-775-254-3124  
email: [karenst@uw.edu](mailto:karenst@uw.edu)

## **ABSTRACT**

Chest radiography for the diagnosis of active pulmonary tuberculosis (PTB) is limited by poor specificity and reader inconsistency. Scoring systems have been employed successfully for improving the performance of chest radiography for various pulmonary diseases. We conducted a systematic review to assess the diagnostic accuracy and reproducibility of scoring systems for PTB.

We searched multiple databases for studies that evaluated the accuracy and reproducibility of chest radiograph scoring systems for PTB. We summarized results for specific radiographic features and scoring systems associated with PTB. Where appropriate, we estimated pooled performance of similar studies using a random effects model.

Thirteen studies were included in the review, nine of which were in low TB burden settings. No scoring system was based solely on radiographic findings. All studies used systems with various combinations of clinical and radiologic features. Eleven studies involved scoring systems that were used for making decisions concerning hospital respiratory isolation. None of the included studies reported data on intra-reporter or inter-reporter reproducibility. Upper lobe infiltrates (pooled diagnostic odds ratio [DOR] 3.57, 95% CI 2.38 to 5.37, five studies) and cavities (DOR range, 1.97 to 25.66, 3 studies) were significantly associated with PTB. Sensitivities of the scoring systems were high (median 96%, IQR 93-98%), but specificities were low (median 46%, IQR 35-50%).

Chest radiograph scoring systems appear useful in ruling-out PTB in hospitals, but their low specificity precludes ruling-in PTB. There is a need to develop accurate scoring systems for people living with HIV and for outpatient settings, especially in high TB burden settings.

**Keywords:** tuberculosis, radiography, scoring system, systematic review

## **INTRODUCTION**

Despite early diagnosis being a key principle of TB control, in 2010, the global case detection rate for all forms of TB was only 65% (1). Limitations of existing diagnostic tests are considered to contribute to the low case detection rate (1). Sputum smear microscopy and chest radiography are two of the most commonly used tests for TB in most high TB burden countries. Smear microscopy has low sensitivity and fails to detect nearly half of all TB cases (2). Smear microscopy is of particularly limited value in extrapulmonary TB, children, and people living with HIV (PLWH).

Chest radiography is a rapid test that has been used for over a century to diagnose pulmonary TB (PTB) (3). Chest radiography can be performed at the point-of-care and incorporated in screening and diagnostic algorithms, usually as an add-on test to smear microscopy among persons with possible PTB, or as a screening test among specific populations such as immigrants from high-burden TB settings. While chest radiography is acknowledged to have high sensitivity for detecting pulmonary abnormalities, its use for diagnosing PTB has been limited by modest specificity, and high inter- and intra-observer differences in reporting of radiographs (4). Consequently, the probability of diagnosing active PTB based on a chest radiograph reading is dependent on the reader and not well standardized.

An analogous impediment to the use of chest radiography for the diagnosis of occupational lung diseases was overcome by the development of standardized methods for the reading of chest radiographs, a system that is now employed successfully by the

International Union Against Cancer, the International Labour Organization, and the National Institute for Occupational Safety and Health (5, 6). Scoring systems have also been developed for grading the severity and extent of pulmonary disease among patients with cystic fibrosis (7) and form part of the lung injury score for assessing the severity of adult respiratory distress syndrome (8). Similarly, a standardized scoring system for PTB that assigns weights to specific features of chest radiographs consistent with PTB, if accurate and reproducible, could potentially augment TB case detection rates using largely pre-existing resources. Such a standardized scoring system for PTB also has the potential to be combined with newer nucleic acid amplification tests such as Xpert MTB/RIF<sup>®</sup> (Cepheid, Sunnyvale, CA) (9), either as a triage test to reduce costs or as an add-on test in Xpert MTB/RIF negative persons.

A systematic review of clinical prediction rules for isolating inpatients with suspected PTB was published in 2005 (10), but to our knowledge, no previous systematic reviews have assessed the performance of radiograph scoring systems for PTB. Therefore, we carried out a systematic review to estimate the diagnostic accuracy of scoring systems using chest radiograph features for active PTB in patients with possible disease. A secondary objective was to assess the reproducibility of chest radiograph scoring systems for PTB.

## **METHODS**

We followed guidelines for systematic reviews of diagnostic test accuracy recommended by the Cochrane Collaboration Diagnostic Test Accuracy Working Group, including writing a detailed protocol before starting the review (11, 12).

**Types of studies:** We aimed to include randomized controlled trials and observational studies of all study designs (e.g. cross-sectional, case-control, and cohort) that assessed the performance of radiographic scoring systems for the diagnosis of PTB. We included studies that reported data from which we could extract true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values for determining sensitivity and specificity estimates, as well as studies that only reported summary measures of diagnostic accuracy (defined below).

**Participants:** Participants were patients with possible pulmonary TB who were 15 years of age and older. We restricted studies to those that included a minimum of 10 patients with TB. With the aim of evaluating patients similar to those who present in routine clinical practice, we excluded studies that exclusively involved specific patient groups such as patients with pneumoconioses, malignancies (both hematological and solid organ), and immune-mediated inflammatory disease, and patients on hemodialysis. We also excluded studies that investigated asymptomatic contacts of TB patients.

**Index test:** Any chest radiograph scoring system

**Comparator:** No chest radiograph scoring system

**Target condition:** TB of the pulmonary parenchyma, pleura, intrathoracic lymph nodes. We included miliary TB if the disease involved either pulmonary parenchyma or multiple sites, one of which was the lung.

**Reference standards:** We considered liquid or solid culture as the reference standards for active PTB.

**Definitions:** A *radiograph scoring system* was defined as a system that assigned numeric weights to specific features of chest radiographs consistent with PTB (such as cavitary lesions), with or without the presence of clinical findings.

*Sensitivity* is the proportion of patients with PTB who are correctly identified by the scoring system. *Specificity* is the proportion of patients without PTB who are correctly identified by the scoring system. *Positive predictive value (PPV)* is the proportion of patients considered positive by the scoring system who are correctly diagnosed with PTB. *Negative predictive value (NPV)* is the proportion of patients considered negative by the scoring system who are correctly diagnosed without PTB.

*Diagnostic Odds Ratio (DOR)* is the odds of a patient with PTB having a specific clinical or radiographic feature divided by the odds of a participant without PTB having the same clinical or radiographic feature.

*Reproducibility* refers to agreement of the scoring system when a chest radiograph is read more than once. Agreement could either be ‘intra-reader’, when the same person reads the chest radiograph more than once, blinded to his/her previous reading, or ‘inter-reader’, when two or more people read the same chest radiograph. Agreement is a

reflection of the repeatability of a test (the scoring system) and is independent of the accuracy of the test.

### ***Search methods for identification of studies***

We searched MEDLINE (1946 to 30 August 2012), EMBASE (1947 to 30 August 2012) and Web of Science (1899 to 30 August 2012) for relevant articles, using published filters for diagnostic tests to improve sensitivity (13, 14). We used the terms sensitiv\*[tw] OR diagnos\*[tw] OR di [fs] AND radiograph\*[MeSH] OR chest xray[tw] OR mass chest x-Ray[MeSH] OR photofluorograph\*[tw] OR scor\*[tw] AND tuberculosis(sub-headings: lymph node/miliary/multidrug-resistant/Pleural/ Pulmonary) [MeSH] OR Mycobacterium tuberculosis [MeSH]. The detailed search strategy can be found in the online supplement. We also reviewed reference lists of included articles and review articles identified through the search and hand-searched World Health Organization reports.

### ***Selection of studies***

Initially, two reviewer authors (LMP and KRS) independently scrutinized titles and abstracts in English, French and Spanish for eligibility. Citations deemed relevant by either reviewer were selected and papers retrieved for full-text review. Next, each eligible article was independently assessed by two reviewers (LMP and KRS) against the selection criteria. Disagreements were resolved by discussion between the reviewers. A list of excluded studies with their reasons for exclusion was maintained.



### ***Assessment of study quality***

Two reviewers (LMP and KRS) independently assessed study quality using the core set of 11 items from Quality Assessment of Diagnostic Accuracy Studies (QUADAS), a validated tool to evaluate the presence of bias and variation in diagnostic accuracy studies (15). As recommended, each item was be scored as ‘yes’, ‘no’, or ‘unclear’.

### ***Data extraction***

Two reviewer authors (LMP and KRS) independently extracted data from each study using a data extraction form that was piloted and then finalized based on the experience gained from the pilot. Disagreements were resolved by discussion. Data were extracted for various characteristics, including the following: author; publication year; study design; country income status classified by the World Bank List of Economies World Bank 2012 data (16) for two-by-two tables of individual radiographic features; details of the scoring systems and their performance characteristics. The data extraction form is included in the online depository.

### ***Statistical analysis***

Data from the two-by-two tables were used to calculate sensitivity and specificity estimates for the scoring system for individual studies along with their 95% confidence intervals at cut-offs determined by the study authors (mostly based on optimal sensitivities and specificities using receiver operator curves). Forest plots were generated to display sensitivity and specificity estimates using Meta-Disc (version 1.4) (17). DORs for specific radiographic features associated with PTB were determined when data were

provided. Heterogeneity was assessed by visual inspection of forest plots and the degree of heterogeneity measured by the *I-squared* statistic ( $I^2$ ). DORs for specific radiographic features were pooled only if the radiographic features and patient populations were similar across studies and  $I^2 \leq 75\%$ . Meta-analysis was performed using a random effects approach to account for the variability across studies and to derive conservative assessments of the uncertainty in the estimates (18), using Meta-Disc (version 1.4) (17). Formal assessment of publication bias using methods such as funnel plots or regression tests were not performed because such techniques have not been found to be useful for diagnostic data (19). An estimation of language bias was attempted by retrieving citations from the search strategy with and without a language filter, and the “filtered” citations were reported as a percentage of the overall citations retrieved.

## **RESULTS**

We identified 12,883 citations, of which 9,137 unique articles were identified after exclusion of duplicate articles. We conducted the search with and without the language filters to assess the degree of bias, and found that our search strategy with language filters included 81% of all studies. After screening titles and abstracts, 187 articles were found to satisfy the criteria for further review and their full-texts were retrieved. After full-text review, 174 articles were excluded for various reasons, and 13 articles (all observational studies) were included in the systematic review (20-32), Fig.1

### *Included studies*

We did not identify any scoring system that was based exclusively on radiographic criteria. Of the total 13 included studies, 12 studies involved scoring systems that

combined clinical and radiographic features (20-31) and one study involved validation of 13 clinical prediction rules for inpatient respiratory isolation (32). This study was included as it retrospectively applied prediction rules from six studies identified by our systematic review to a group of eligible participants with symptoms suggestive of PTB. This study also tested rules that were derived from seven other studies which were excluded from our systematic review as they did not satisfy our reference standard (33, 34) or did not use specific features visualized on the chest radiograph as part of the scoring system (35-39). Table 1 lists the characteristics of the 12 included original studies, containing a total of 5,767 participants, and the one study that validated six of these scoring systems (validated in 345 participants). The median number of individuals with possible PTB included in the studies was 283 (interquartile range 177 to 431).

Seven studies included all patients with possible PTB (20-25, 32), five studies included patients with possible PTB who were found to have negative sputum smears (26-30), and one study specifically excluded PLWH (31). Nine studies were performed in high-income countries. Five studies involved radiologists; two studies included pulmonologists and five studies did not report the specialty of the radiograph reader. The demographic characteristics of the patients are provided in the online depository. When reported, the majority of patients were male. Eleven studies included PLWH, who represented 11% to 61% of eligible patients.

#### *Excluded studies*

We identified two studies that were designed with the aim of deriving a clinical-radiographic scoring system for PLWH among hospitalized patients found to be sputum smear-negative (40, 41). Both studies satisfied a majority of our inclusion criteria and found the presence of mediastinal adenopathy and cavities to be significantly associated with PTB in univariate analysis. However, neither of the studies derived a score for PTB. The study by Le Minor et al. concluded that the “numbers were insufficient to develop a score for TB” (41), while the study by Davis et al. stated that “After exhaustive testing, we were unable to identify any combination of factors which reliably predicted bacteriologically confirmed tuberculosis” (40).

We also excluded 13 studies that used automated computer-assisted diagnosis (CAD) as none of these studies used culture as a reference standard, a criterion for inclusion in this review (42-54). Five studies that involved the grading of chest radiographs were excluded as these studies were designed to grade the severity of PTB based on the extent of abnormalities visualized on the chest radiograph, and not the diagnostic accuracy of scoring systems (55-59). We also excluded three studies that used the Chest Radiograph Reading and Recording System (CRRS) (60-62), despite these studies demonstrating the CRRS tool to have good reliability for features of PTB visualized on a chest radiograph, as these studies did not use culture as a reference standard.

#### *Assessment of methodological quality*

As seen in Table 2, all studies suffered from verification bias, as the results of the chest radiograph and/or the clinical components of the scoring system played a role in the

selection of those patients would who be investigated further with culture, the reference standard. Seven (54%) of the total 13 studies did not include a sample that was considered representative of the target population, as these studies did not enroll all individuals with possible PTB in a consecutive or random manner, often only including admitted patients without a suitable control group. Six (46%) studies did not specify whether the person assigning scores to the patients for the various components of the scoring system was blinded to the results of the reference standard.

### *Findings*

#### *Studies that included all patients with possible PTB*

We identified six studies that included all patients with possible PTB (20-25). All studies were performed in an inpatient setting. All studies were aimed at deriving optimal prediction scores to identify patients who were likely to have PTB and require respiratory isolation, Table 1. In univariate analyses, the most common radiographic features across studies found to be significantly associated with PTB were upper lobe infiltrates [pooled DOR 6.65, 95% CI 4.42 to 10.01, five studies], Figure 2A, and cavities (DOR range, 2.11 to 10.08, three studies) (online supplement).

The details of the parameters included in the scores and their respective weights are summarized in Table 3, along with the performance characteristics of the scoring system and the final rule to aid in decision-making. Studies used several different methods to derive weights for the scoring system: logistic regression of the parameters found significant by univariate analysis (three studies); classification and regression tree

analysis (one study) (21); general regression neural network analysis (one study) (22); and chi-squared recursive partitioning (one study) (23). All six studies achieved a sensitivity of the scoring system greater than 80% (median 95%, range, 81% to 100%). For the five studies that reported specificity data, specificity estimates were low (median 42%, range 22% to 72%), suggesting a poor rule-in value for PTB.

Figure 3 shows forest plots of sensitivities and specificities of the scoring systems reported in individual studies. We did not pool estimates because of the differences in the scoring systems.

*Studies that only included patients with possible PTB who were found to be sputum smear negative*

We identified five studies in this category (26-30). Four studies were conducted in an inpatient setting for the purpose of determining a clinical rule for respiratory isolation (26, 27, 29, 30), while one study was performed in an outpatient setting (28), Table 1. As with the previously described set of studies that included all patients with possible PTB, in the univariate analysis, the most common radiographic features across studies found to be associated with PTB were upper lobe infiltrates (pooled DOR 3.57, 95% CI 2.38 to 5.37, five studies), Figure 2B, and cavities (DOR range, 1.97 to 25.66, three studies). We did not pool estimates based on scoring systems because of the differences in these scoring systems (online supplement).

To derive weights for the scoring system, two studies used logistic regression of the parameters found significant in the univariate analysis (27, 30), while three studies involved validation of previous studies (26, 28, 29). One of the validation studies used bootstrapping, which is a resampling method aimed at improving the internal validity of the data (27). All studies achieved a sensitivity of the scoring system greater than 93% (median 96%, range, 93% to 98%). However, specificity estimates were low (median 35%, range 14% to 50%), again suggesting a poor rule-in value for PTB, Figure 4. We did not consider these scoring systems to be similar and therefore, did not pool the accuracy estimates.

One study, performed in an inpatient setting, excluded PLWH (31), Table 1. Logistic regression was used to derive weights for the score, but the study also validated the score derived by Wisnivesky et al. (30) This study found a sensitivity of 97% and specificity of 42%, Table 3.

#### *Study that validated various clinical prediction rules*

One study evaluated 13 different clinical prediction rules (identified through a comprehensive literature search) for the respiratory isolation of inpatients with suspected PTB (32). As noted above, six of the 13 prediction rules met criteria for inclusion in our review. The authors applied the various rules retrospectively to emergency room patients who had earlier been included in a study (this study is included in the current systematic review) (25) to derive a scoring system for PTB. Similar to the original studies, the validation study found that most scoring systems had poor specificity for PTB.. A

comparison of the performance characteristics of the six scoring systems in their respective derivation studies, and in the validation study is provided in Table 4.

### *Reproducibility*

None of the included studies reported data on intra-reporter or inter-reporter reproducibility.

## **DISCUSSION**

We conducted this systematic review with the aim of assessing the diagnostic accuracy of standardized radiographic scoring systems for the diagnosis of PTB, and whether standardization improves the performance of chest radiography. Our review failed to find any study that exclusively relied on radiographic features to derive a score, and all the included studies combined defined radiographic criteria with different clinical criteria. While the aim of the review was to assess the utility of radiographic scoring systems as diagnostic tools, especially in low-income, high-TB burden outpatient settings where such systems would be extremely beneficial if accurate, there appears to be a dearth of such studies. Most of the included studies were hospital-based, decision-to-isolate studies in high-income, low-TB burden settings.

Patients with PTB can generate up to 44 quanta of TB bacilli per hour (one quantum is defined as the infectious dose) (63), highlighting the necessity for rapid respiratory isolation of patients with PTB in the hospital setting. Yet, the unnecessary respiratory isolation of patients considerably increases costs to the health care system (64), and in



resource-limited settings where isolation beds may be in short supply, unnecessary isolation of some patients may preclude appropriate isolation of others. Scoring systems that improve the accuracy of the decisions to subject those patients with possible PTB to respiratory isolation can considerably improve the efficiency of healthcare systems and utilization of resources. The scores developed suffered from low specificity, and had a high rule-out value (median NPV 99%, range 93-100% values reported in 7 studies) but a poor rule-in value (median PPV 22.5%, range 8-61%, values reported in 8 studies) for PTB. The validation of six of these scores in a separate study reflected the same lack of specificity. A stratified analysis by country income status reflected better performance of scoring systems in high-income countries compared with middle- and low- income countries, but the improvement was unlikely to be clinically significant (online depository). However, such scores may still be useful for limiting the number of patients for whom further investigations would be warranted (as compared to the use of subjective assessments of chest radiographs that are known to have extremely poor specificity, thereby warranting further testing for a greater number of patients suspected of having PTB), especially among patients who are smear-negative.

The prediction rule developed by Wisnivesky and colleagues was validated in four studies, two of which were conducted in patients who had negative sputum smears. The scoring system consistently demonstrated sensitivity higher than 92%, but had poor specificity. As a rule-out test, this scoring system appears to be validated in multiple studies. The study by Soto and colleagues (28) was a validation study of a score derived by the same research group in an earlier study (27). Although the cut-off for the score

was modified in the validation cohort, the scoring system performed well in a subgroup of patients with no prior history of TB. Four studies considered the patient to be positive for PTB if they had any one of the features present (20, 23, 26, 29, 30), making these systems resemble checklists rather than weighted scoring systems.

We identified only one study that assessed a clinical-radiographic scoring system for outpatients. Our systematic review also failed to identify a clinical radiographic scoring system for PLWH with possible PTB. Bock and colleagues (20) performed a subgroup analysis in PLWH, but found no radiographic feature to be significantly associated with PTB in this subgroup, a finding that is consistent with the atypical nature of radiographic manifestations of PTB described among PLWH (65).

Automated computer-assisted diagnosis (CAD) employs techniques such as texture analysis for reading digital chest radiographs, and appears to be a promising modality for standardizing and improving the diagnostic performance of digital chest radiography (66). However, our review suggested a lack of methodologically high-quality studies. Researchers in this field could help advance their techniques better by following published guidelines for conducting and reporting their work to ensure that their efforts contribute to a high-quality evidence base (67, 68).

The strength of our systematic review is in the extensive review of the literature, with two reviewers independently performing screening, quality assessment, and data extraction. However, three caveats need to be acknowledged while interpreting the results of the

systematic review. Firstly, 12 of the 13 studies were conducted among inpatients, a population in whom the manifestations of TB disease are likely to be subject to a spectrum bias as compared to outpatients (69). Secondly, nine of the 13 studies were conducted in high-income, low TB burden settings, and the radiographic manifestations of the disease, pre-test probabilities, and technical quality of radiographs are likely to be different in such settings as compared to low-income, high-burden settings (70). Lastly, the purpose of most of the studies was to assess the likelihood of PTB for purposes of respiratory isolation in hospitals, and the derivation of such scores could have different aims (and consequently different cut-offs) from scoring systems derived with diagnostic purposes in mind. We restricted our search to articles written in English, French and Spanish, but an assessment for language bias suggested that we included a high proportion of the available literature. However, we may have inadvertently failed to include articles in other languages, and acknowledge this as a shortcoming of the review. As assessed with QUADAS, we judged the studies to be at risk of bias for several items. In all studies, the results of the chest radiograph and/or the clinical components of the scoring system played a role in deciding which patients would receive a culture (verification bias). This bias may have led to overestimates of diagnostic accuracy (71). This is especially true for studies in which radiographic findings were used to guide the decision for requesting sputum cultures for TB. Seven (54%) studies were not considered to include a representative sample (selection bias). Six (46%) studies did not provide adequate information about blinding of radiographic interpretation. Selection bias and absence of blinding are features of study design that have also been associated with

inflated accuracy estimates(72, 73). These limitations in the quality of the included studies need to be taken into consideration when interpreting results.

## **CONCLUSIONS**

Our systematic review did not identify a scoring system for PTB based solely on radiographic features. The development and validation of such a system could help standardize the interpretation of chest radiographs. The review identified clinical-radiographic scoring systems, for predicting the likelihood of PTB among patients admitted to hospitals. Such scoring systems are intended for assessing the need for respiratory isolation. Most of these systems have high sensitivity but low specificity for PTB. There is a pressing need to derive accurate scoring systems for PLWH and outpatients, especially in low-resource settings. Technological advances in the interpretation of chest radiographs, such as CAD, need to be validated in well-designed studies to assess their utility.

**Acknowledgement:** We thank Lorie A. Kloda at the Life Sciences Library, McGill University for her guidance and assistance in formulating the search strategy.

This work was supported by the European & Developing Countries Clinical Trials Partnership (TB-NEAT grant) and the Canadian Institutes of Health Research (CIHR) (MOP-89918). MP is supported by salary awards from CIHR and Fonds de recherche du Québec – Santé. These agencies had no role in the analysis of data and decision to publish.

## REFERENCES

1. World Health Organization. Global tuberculosis control 2011. [http://www.who.int/tb/publications/global\\_report/2011/gtbr11\\_full.pdf](http://www.who.int/tb/publications/global_report/2011/gtbr11_full.pdf) (accessed April 4, 2012).
2. 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 Infect Dis*. 2006 Oct;6(10):664-74.
3. Lerner BH. The perils of 'x-ray vision': How radiographic images have historically influenced perception. *Perspectives in Biology and Medicine*. 1992;35(3):382-97.
4. Koppaka R, Bock N. How reliable is chest radiography? Toman's tuberculosis: case detection, treatment, and monitoring – questions and answers 2nd ed. ed. Geneva: World Health Organization; 2004. p. 51-60.
5. International Labour Office. Guidelines for the use of the ILO international classification of radiographs of pneumoconiosis. Revised ed 2000. (Occupational Safety and Health Series, No. 22). Geneva, Switzerland: ILO, 2002.
6. Bohlig H. [UICC-Cincinnati classification of radiographic findings in pneumoconioses. Collective work of a subcommittee of the International Union Against Cancer (UICC)]. *Fortschr Geb Rontgenstr Nuklearmed*. 1971 Nov;115(5):665-83.
7. Grum CM, Lynch JP, 3rd. Chest radiographic findings in cystic fibrosis. *Semin Respir Infect*. 1992 Sep;7(3):193-209.

8. Murray JF, Matthay MA, Luce JM, Flick MR. An expanded definition of the adult respiratory distress syndrome. *Am Rev Respir Dis.* 1988 Sep;138(3):720-3.
9. Boehme CC, Nabeta P, Hillemann D, Nicol MP, Shenai S, Krapp F, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med.* 2010 Sep 9;363(11):1005-15.
10. Wisnivesky JP, Serebrisky D, Moore C, Sacks HS, Iannuzzi MC, McGinn T. Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. *Journal of General Internal Medicine.* 2005 Oct;20(10):947-52.
11. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic reviews of diagnostic test accuracy. *Ann Intern Med.* 2008 Dec 16;149(12):889-97.
12. Macaskill P GC, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and Presenting Results. Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy, version 090* London: The Cochrane Collaboration. Available: <http://srdta.cochrane.org/> Accessed 6 March 2011.
13. Wilczynski NL, Haynes RB. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. *BMC Med.* 2005;3:7.
14. Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. *BMJ.* 2004 May 1;328(7447):1040.
15. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003 Nov 10;3:25.

16. World Bank. World Bank List of Economies  
<http://siteresources.worldbank.org/DATASTATISTICS/Resources/CLASS.XL>. 2011.
17. Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC medical research methodology*. 2006;6:31.
18. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986 Sep;7(3):177-88.
19. Tatsioni A, Zarin DA, Aronson N, Samson DJ, Flamm CR, Schmid C, et al. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med*. 2005 Jun 21;142(12 Pt 2):1048-55.
20. Bock NN, McGowan Jr JE, Ahn J, Tapia J, Blumberg HM. Clinical predictors of tuberculosis as a guide for a respiratory isolation policy. *American Journal of Respiratory and Critical Care Medicine*. 1996;154 (5):1468-72.
21. El-Solh A, Mylotte J, Sherif S, Serghani J, Grant BJ. Validity of a decision tree for predicting active pulmonary tuberculosis. *American Journal of Respiratory & Critical Care Medicine*. 1997 May;155(5):1711-6.
22. El-Solh AA, Hsiao CB, Goodnough S, Serghani J, Grant BJ. Predicting active pulmonary tuberculosis using an artificial neural network. *Chest*. 1999 Oct;116(4):968-73.
23. Moran GJ, Barrett TW, Mower WR, Krishnadasan A, Abrahamian FM, Ong S, et al. Decision Instrument for the Isolation of Pneumonia Patients With Suspected Pulmonary Tuberculosis Admitted Through US Emergency Departments. *Annals of Emergency Medicine*. 2009 May;53(5):625-32.

24. Mylotte JM, Rodgers J, Fassl M, Seibel K, Vacanti A. Derivation and validation of a pulmonary tuberculosis prediction model. *Infection Control & Hospital Epidemiology*. 1997 Aug;18(8):554-60.
25. Solari L, Acuna-Villaorduna C, Soto A, Agapito J, Perez F, Samalvides F, et al. A clinical prediction rule for pulmonary tuberculosis in emergency departments. *International Journal of Tuberculosis and Lung Disease*. 2008 Jun;12(6):619-24.
26. Lagrange-Xelot M, Porcher R, Gallien S, Wargnier A, Pavie J, de Castro N, et al. Prevalence and clinical predictors of pulmonary tuberculosis among isolated inpatients: a prospective study. *Clinical Microbiology and Infection*. 2011 Apr;17(4):610-4.
27. Soto A, Solari L, Agapito J, Acuna-Villaorduna C, Lambert ML, Gotuzzo E, et al. Development of a clinical scoring system for the diagnosis of smear-negative pulmonary tuberculosis. *Brazilian Journal of Infectious Diseases*. 2008 Apr;12(2):128-32.
28. Soto A, Solari L, Diaz J, Mantilla A, Matthys F, van der Stuyft P. Validation of a Clinical-Radiographic Score to Assess the Probability of Pulmonary Tuberculosis in Suspect Patients with Negative Sputum Smears. *Plos One*. 2011 Apr;6(4).
29. Wisnivesky JP, Henschke C, Balentine J, Willner C, Deloire AM, McGinn TG. Prospective validation of a prediction model for isolating inpatients with suspected pulmonary tuberculosis. *Archives of Internal Medicine*. 2005 Feb;165(4):453-7.
30. Wisnivesky JP, Kaplan J, Henschke C, McGinn TG, Crystal RG. Evaluation of clinical parameters to predict *Mycobacterium tuberculosis* in inpatients. *Archives of Internal Medicine*. 2000 Sep 11;160(16):2471-6.



31. Rakoczy KS, Cohen SH, Nguyen HH. Derivation and Validation of a Clinical Prediction Score for Isolation of Inpatients With Suspected Pulmonary Tuberculosis. *Infection Control and Hospital Epidemiology*. 2008 Oct;29(10):927-32.
32. Solari L, Acuna-Villaorduna C, Soto A, Van Der Stuyft P. Evaluation of clinical prediction rules for respiratory isolation of inpatients with suspected pulmonary tuberculosis. *Clinical Infectious Diseases*. 2011 01 Mar;52 (5):595-603.
33. Tattevin P, Casalino E, Fleury L, Egmann G, Ruel M, Bouvet E. The validity of medical history, classic symptoms, and chest radiographs in predicting pulmonary tuberculosis: derivation of a pulmonary tuberculosis prediction model. *Chest*. 1999 May;115(5):1248-53.
34. Tessema TA, Bjune G, Assefa G, Bjorvatn B. An evaluation of the diagnostic value of clinical and radiological manifestations in patients attending the Addis Ababa Tuberculosis Centre. *Scandinavian Journal of Infectious Diseases*. 2001;33(5):355-61.
35. Wang CS, Chen HC, Chong IW, Hwang JJ, Huang MS. Predictors for identifying the most infectious pulmonary tuberculosis patient. *J Formos Med Assoc*. 2008 Jan;107(1):13-20.
36. Aguilar J. Clinical prediction rule for respiratory isolation of patients with suspected pulmonary tuberculosis. *Infectious Diseases in Clinical Practice*. 2009 September;17(5):317-22.
37. Redd JT, Susser E. Controlling tuberculosis in an urban emergency department: a rapid decision instrument for patient isolation. *Am J Public Health*. 1997 Sep;87(9):1543-7.

38. Gaeta TJ, Webheh W, Yazji M, Ahmed J, Yap W. Respiratory isolation of patients with suspected pulmonary tuberculosis in an inner-city hospital. *Academic Emergency Medicine*. 1997 Feb;4(2):138-41.
39. Pegues CF, Johnson DC, Pegues DA, Spencer M, Hopkins CC. Implementation and evaluation of an algorithm for isolation of patients with suspected pulmonary tuberculosis. *Infection Control & Hospital Epidemiology*. 1996 Jul;17(7):412-8.
40. Davis JL, Worodria W, Kitembo H, Metcalfe JZ, Cattamanchi A, Kawooya M, et al. Clinical and Radiographic Factors Do Not Accurately Diagnose Smear-Negative Tuberculosis in HIV-infected Inpatients in Uganda: A Cross-Sectional Study. *Plos One*. 2010 Mar;5(3).
41. Le Minor O, Germani Y, Chartier L, Lan NH, Lan NT, Duc NH, et al. Predictors of pneumocystosis or tuberculosis in HIV-infected Asian patients with AFB smear-negative sputum pneumonia. *Journal of Acquired Immune Deficiency Syndromes: JAIDS*. 2008 Aug 15;48(5):620-7.
42. Hogeweg LE, Mol C, de Jong PA, van Ginneken B. Rib suppression in chest radiographs to improve classification of textural abnormalities. *Medical Imaging 2010: Computer - Aided Diagnosis*. 2010;7624.
43. Mouton A, Pitcher RD, Douglas TS. Computer-Aided Detection of Pulmonary Pathology in Pediatric Chest Radiographs. *Medical Image Computing and Computer-Assisted Intervention - Miccai 2010, Pt Iii*. 2010;6363:619-25.
44. Arzhaeva Y, Tax DMJ, van Ginneken B. Dissimilarity-based classification in the absence of local ground truth: Application to the diagnostic interpretation of chest radiographs. *Pattern Recognition*. 2009 Sep;42(9):1768-76.

45. Katsuragawa S, Doi K. Computer-aided diagnosis in chest radiography. *Computerized Medical Imaging and Graphics*. 2007 Jun;31 (4-5):212-23.
46. Le K. Chest X-Ray Analysis for Computer-Aided Diagnostic. *Advanced Computing, Pt Iii*. 2011;133:300-9.
47. Patil SA, Udipi VR, Kane CD, Wasif AI, Desai JV, Jadhav AN. Geometrical and Texture Features Estimation of Lung Cancer and TB Images Using Chest X-ray Database. *2009 International Conference on Biomedical and Pharmaceutical Engineering*. 2009:9-15.
48. Rijal OM, Noor NM, Shaban H, Teng SL. A statistical comparison of digital X-ray images for MTB patients. *2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2005:6418-21.
49. Rijal OM, Iqbal M, Yunus A, Noor NM. Some Critical Remarks on the Initial Detection of Lung Ailments Using Clinical Data and Chest Radiography. *Proceedings of the 13th Wseas International Conference on Computers*. 2009:470-5.
50. Sarkar S, Chaudhuri S. Evaluation and progression analysis of pulmonary tuberculosis from digital chest radiographs. *Computerized Medical Imaging and Graphics*. 1998 Mar-Apr;22(2):145-55.
51. Shen R, Cheng I, Basu A. A Hybrid Knowledge-Guided Detection Technique for Screening of Infectious Pulmonary Tuberculosis From Chest Radiographs. *Ieee Transactions on Biomedical Engineering*. 2010 Nov;57(11):2646-56.
52. van Ginneken B, Romeny BMT, Viergever MA. Automatic segmentation and texture analysis of PA chest radiographs to detect abnormalities related to interstitial

disease and tuberculosis. *Cars 2002: Computer Assisted Radiology and Surgery, Proceedings*. 2002:685-8.

53. van Ginneken B, Romeny BMT. Automatic segmentation of lung fields in chest radiographs. *Medical Physics*. 2000 Oct;27(10):2445-55.

54. van Ginneken B, Katsuragawa S, Romeny BMT, Doi K, Viergever MA. Automatic detection of abnormalities in chest radiographs using local texture analysis. *Ieee Transactions on Medical Imaging*. 2002 Feb;21(2):139-49.

55. Caplin M, Grange JM, Morley S, Brown RA, Kemp M, Gibson JA, et al. Relationship between radiological classification and the serological and haematological features of untreated pulmonary tuberculosis in Indonesia. *Tubercle*. 1989 Jun;70(2):103-13.

56. Churchyard GJ, Fielding K, Roux S, Corbett EL, Chaisson RE, De Cock KM, et al. Twelve-monthly versus six-monthly radiological screening for active case-finding of tuberculosis: a randomised controlled trial. *Thorax*. 2011 Feb;66(2):134-9.

57. Ralph AP, Ardian M, Wiguna A, Maguire GP, Becker NG, Drogumuller G, et al. A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis. *Thorax*. 2010 Oct;65(10):863-9.

58. Tuberculosis Research Centre, Madras. A concurrent comparison of home and sanatorium treatment of pulmonary tuberculosis in South India. *Bulletin of the World Health Organization*. 1959;21(1):51-144.

59. Wejse C, Gustafson P, Nielsen J, Gomes VF, Aaby P, Andersen PL, et al. TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have

predictive value and may be used to assess clinical course. *Scandinavian Journal of Infectious Diseases*. 2008;40(2):111-20.

60. 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. *Int J Tuberc Lung Dis*. 2005 Oct;9(10):1088-96.

61. Agizew T, Bachhuber MA, Nyirenda S, Makwaruzi VZ, Tedla Z, Tallaksen RJ, et al. Association of chest radiographic abnormalities with tuberculosis disease in asymptomatic HIV-infected adults. *Int J Tuberc Lung Dis*. 2010 Mar;14(3):324-31.

62. Dawson R, Masuka P, Edwards DJ, Bateman ED, Bekker LG, Wood R, et al. Chest radiograph reading and recording system: evaluation for tuberculosis screening in patients with advanced HIV. *Int J Tuberc Lung Dis*. 2010 Jan;14(1):52-8.

63. Riley RL, Nardell EA. Clearing the air. The theory and application of ultraviolet air disinfection. *Am Rev Respir Dis*. 1989 May;139(5):1286-94.

64. Kellerman S, Tokars JI, Jarvis WR. The cost of selected tuberculosis control measures at hospitals with a history of *Mycobacterium tuberculosis* outbreaks. *Infect Control Hosp Epidemiol*. 1997 Aug;18(8):542-7.

65. Pitchenik AE, Rubinson HA. The radiographic appearance of tuberculosis in patients with the acquired immune deficiency syndrome (AIDS) and pre-AIDS. *Am Rev Respir Dis*. 1985 Mar;131(3):393-6.

66. Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. *Br J Radiol*. 2005;78 Spec No 1:S3-S19.

67. Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, Poole F, et al. Evaluation of diagnostic tests for infectious diseases: general principles. *Nat Rev Microbiol.* 2006 Dec;4(12 Suppl):S20-32.
68. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med.* 2003 Jan 7;138(1):40-4.
69. Willis BH. Spectrum bias - Why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract.* 2008;25(5):390-6.
70. Laifer G, Battegay MFU. TB in a Low-Incidence Country: Differences Between New Immigrants, Foreign-Born Residents and Native Residents. *Am J Med.* 2007 April;120(4):350-6.
71. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003 Nov 10;3:25.
72. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *Jama.* 1999 Sep 15;282(11):1061-6.
73. 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 Feb 14;174(4):469-76.

**Table 1. Characteristics of Included Studies**

| Study  | Country | Setting   | No. of Eligible Patients with Possible PTB (%) | Study Design                          | Inclusion Criteria   | Chest Radiograph Reader               |
|--|---------|-----------|--|---------------------------------------|--|---------------------------------------|
| <b>Studies that included all patients with possible TB</b> |         |           |  |                                       |  |                                       |
| Bock et al (1996) <sup>20</sup>                            | USA     | Inpatient | 295 (78)                                       | Cross-sectional, <u>retrospective</u> | 1. Patients with active TB<br>2. Patients with TB in the differential diagnosis<br>3. AFB smears and cultures ordered  | Radiologist                           |
| El-Solh et al (1997) <sup>21</sup>                         | USA     | Inpatient | 286 (100)*                                     | Cross-sectional                       | 4. HIV + with abnormal CXR<br>All isolated patients, based on symptoms, prior history of TB exposure, HIV status, medical and social risk factors, and radiographic findings | Radiologist and pulmonologist         |
| El-Solh et al (1999) <sup>22</sup>                         | USA     | Inpatient | 119 (100)*                                     | Cross-sectional                       | All patients in whom AFB smear and culture were requested  | Radiologist and pulmonologist         |
| Moran et al (2009) <sup>23</sup>                           | USA     | Inpatient | 2535 (91)*                                     | Cross-sectional                       | Admission diagnosis of pneumonia or suspected TB   | Emergency medicine resident           |
| Mylotte et al (1997) <sup>24</sup>                         | USA     | Inpatient | 220 (100)*                                     | Cross-sectional                       | All patients in whom an AFB smear and culture was requested by the admitting physician   | Not reported                          |
| Solari et al (2008) <sup>25</sup>                          | Peru    | Inpatient | 345 (71)                                       | Cross-sectional                       | Productive cough for > 1 week or Cough of any duration and<br>1. Fever > 3 weeks or<br>2. Weight loss of at least 3kg in previous month or<br>3. Night sweats or hemoptysis  | Internist, internal medicine resident |

| Studies that only included smear-negative subjects with possible PTB |         |            |   |                 |   |  |
|--|---------|------------|---|-----------------|---|--|
| Study  | Country | Setting    | No. of Eligible Patients with Possible TB (%) | Study Design    | Inclusion Criteria  | Chest Radiograph Reader  |
| Lagrange-Xelot et al (2010) <sup>26</sup>                            | France  | Inpatient  | 134 (100)                                     | Cross-sectional | Suspected TB, as recommended by French guidelines   | Not reported   |
| Soto et al (2008) <sup>27</sup>                                      | Peru    | Inpatient  | 262 (100)                                     | Cross-sectional | Cough $\geq$ one week AND one or more of the following:<br>1.Fever<br>2.Weight loss $\geq$ 4kg in 1 month<br>3.Breathlessness<br>4.Constitutional symptoms (malaise or hyporexia for a minimum of 2 months)   | Not reported   |
| Soto et al (2011) <sup>28</sup>                                      | Peru    | Outpatient | 663 (97)                                      | Cross-sectional | Cough $\geq$ 2 weeks AND one or more of the following:<br>1.Fever<br>2.Weight loss<br>3.Breathlessness  | 1.General practitioner<br>2.TB specialist<br>Tie breaker:<br>Experienced radiologist |
| Wisnivesky et al (2000) <sup>30</sup>                                | USA     | Inpatient  | 112 (100)                                     | Case-control    | Cases - isolated TB patients<br>controls - randomly selected from a log of patients who submitted smears and cultures matched on age ( $\pm$ 3 years), sex and year of presentation, 3 smears negative, culture negative and isolated in a hospital | 1.Radiologist<br>2.Radiologist   |

or differential diagnosis of PTB from attending physician



|   |                |                |  |                     |   |                                       |
|---|----------------|----------------|--|---------------------|---|---------------------------------------|
| Wisnivesky et al (2005) <sup>29</sup>   | USA            | Inpatient      | 516 (100)  | Cross-sectional     | Patients admitted and isolated because of suspicion of PTB  | Not reported                          |
| <b>Study that included only HIV-uninfected participants with possible PTB</b> |                |                |  |                     |   |                                       |
| <b>Study</b>  | <b>Country</b> | <b>Setting</b> | <b>No. of Eligible Patients with Possible TB (%)</b> | <b>Study Design</b> | <b>Inclusion Criteria</b>   | <b>Chest Radiograph Reader</b>        |
| Rakoczy et al (2008) <sup>31</sup>  | USA            | Inpatient      | 280 (100)*   | Case-control        | Cases: all TB inpatients<br>Controls: all inpatients placed under airborne precautions with negative smears and cultures matched with cases on time of admission (+/- 6 days)   | Not reported                          |
| <b>Study that validated six of the scoring systems described above</b>        |                |                |  |                     |   |                                       |
| Solari et al (2011) <sup>32**</sup>   | Peru           | Inpatient      | 345 (71)   | Cross-sectional     | Productive cough for > 1 week or cough of any duration and<br>1. Fever > 3 weeks or<br>2. Weight loss of at least 3kg in previous month or<br>3. Night sweats or hemoptysis or differential diagnosis of PTB from attending physician | Internist, internal medicine resident |

\*Studies had derivation and validation cohorts. The number of patients with possible TB represents those in the validation cohorts  
PTB, pulmonary TB

\*\*The study by Solari et al (2011)<sup>32</sup> applied various scoring systems to the same cohort of patients as the study by Solari et al (2008)<sup>25</sup>.

**Table 2. Quality assessment of the included studies using the QUADAS tool**

| Item   | Representative sample? | Acceptable reference standard? | Acceptable delay? | Partial verification avoided? | Differential verification avoided? | Incorporation avoided? | Reference standard blinded? | Index results blinded? | Relevant clinical information available? | Uninterruptable results reported? | Withdrawals explained? |
|--|------------------------|--------------------------------|-------------------|-------------------------------|------------------------------------|------------------------|-----------------------------|------------------------|--|-----------------------------------|------------------------|
| Bock et al (1996) <sup>20</sup>                        |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| El-Solh et al (1997) <sup>21</sup>                     |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| El-Solh et al (1999) <sup>22</sup>                     |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Moran et al (2009) <sup>23</sup>                       |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Mylotte et al (1997) <sup>24</sup>                     |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Solari et al (2008 <sup>25</sup> , 2011) <sup>32</sup> |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Lagrange-Xelot et al (2010) <sup>26</sup>              |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Soto et al (2008) <sup>27</sup>                        |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Soto et al (2011) <sup>28</sup>                        |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Wisnivesky et al (2000) <sup>30</sup>                  |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Wisnivesky et al (2005) <sup>29</sup>                  |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |
| Rakoczy et al (2008) <sup>31</sup>                     |                        |                                |                   |                               |                                    |                        |                             |                        |  |                                   |                        |

Yes
  No
  Unclear

**Table 3 Details and performance characteristics of chest radiograph scoring systems**

| Study   | Feature   | Test Details  | Rule  | Sensitivity<br>%<br>(95%CI) <sup>#</sup> | Specificity<br>%<br>(95%CI) <sup>#</sup> | Area<br>under the<br>Curve<br>(SE) |
|---|---|---|---|--|--|------------------------------------|
| <b>Studies that included all patients with possible PTB</b> |   |   |   |  |  |                                    |
|   |   | <b>Logistic regression DOR<br/>(95%CI)</b>  |   |  |  |                                    |
| Bock et al<br>(1996) <sup>20</sup>                          | 1.CXR with upper lobe infiltrate<br>2.CXR with cavity<br>3.Knew someone with active TB<br>4.Self-report of positive tuberculin skin test in past<br>5.Self-report of isoniazid preventive therapy in the past | 5 (2.38,10.51)<br>3.93 (1.06,14.62)<br>2.42 (1.1,5.32)<br>5.67 (1.57,22.01)<br>0.18 (0.04,0.82) | Any of 1-3 or 4(in the absence of 5) considered test positive   | 81                                       |  |                                    |
| El-Solh et al<br>(1997) <sup>21</sup>                       | Upper zone disease<br>Weight loss<br>Diabetes mellitus  | Used classification and regression tree analysis  | Upper zone disease absent and fever absent = test negative<br><br>Upper zone disease absent, but fever present = test negative if no weight loss and CD4+ ≥ 200<br><br>Everyone else to be considered test positive | 100<br>(78,100)                          | 50 (44,57)                               | 0.878<br>(0.029)                   |

|                                       |  |  |  |                                   |               |                        |
|---------------------------------------|--|--|--|-----------------------------------|---------------|------------------------|
| El-Solh et al<br>(1999) <sup>22</sup> | Age, CD4+counts,<br>diabetes mellitus, HIV,<br>tuberculin skin test<br>positivity<br><br>Chest pain, weight loss,<br>cough, night sweats, fever,<br>shortness of breath<br><br>Upper lobe infiltrate,<br>lower lobe infiltrate, upper<br>lobe cavity, lower lobe<br>cavity, adenopathy,<br>unilateral pleural effusion,<br>bilateral pleural effusion,<br>pleural thickening, miliary<br>pattern, normal | Used general regression<br>neural network  |  | 100<br>(91,100)                   | 72<br>(65,77) | 0.947<br>(0.028)       |
| Moran et al<br>(2009) <sup>23</sup>   | 1.Apical infiltrate<br>2.Cavitation<br>3.Immigrant<br>4.Weight loss<br>5.Positive TB history<br>6.Homeless<br>7.Incarcerated   | Used chi-squared recursive<br>partitioning   | Any of 1-7<br>present = test<br>positive | 96                                | 49            |                        |
| Mylotte et al<br>(1997) <sup>24</sup> | 1.AFB positive smear<br>2.Localized CXR change<br>3.Correctional facility<br>residence<br>4.History of weight loss   | <b>DOR (95% CI)</b><br>5.8 (3,11)<br>2.5 (1.3,4.9)<br>2.3 (1.2,4.4)<br>1.8 (1,3.2) | <b>Score</b><br>3<br>2<br>2<br>1         | Score $\geq$ 3 = test<br>positive | 88            | 63.2<br>0.86<br>(0.04) |

|  |   |  |                   |   |                            |      |                 |                   |    |
|--|---|--|-------------------|---|----------------------------|------|-----------------|-------------------|----|
| Solari et al<br>(2008) <sup>25</sup>   | 1.Age < 35  | <b>DOR (95% CI)</b><br>0.97 (0.96, 0.99) | <b>Score</b><br>0 | Score $\geq$ 3 = test<br>positive                     | 93                         | 42   | 0.809<br>(0.05) |                   |    |
|  | 2.Age 35-60   |  |                   |   |                            |      |                 | -1                |    |
|  | 3.Age $\geq$ 61   |  |                   |   |                            |      |                 | -2                |    |
|  | 3.Weight loss   |  |                   |   |                            |      |                 | 2.79 (1.51,5.18)  | 5  |
|  | 4.History of PTB  |  |                   |   |                            |      |                 | 0.51 (0.28,0.95)  | -3 |
|  | 5.Miliary pattern                                       |  |                   |   |                            |      |                 | 8.04 (2.79,23.16) | 10 |
|  | 6.Cavity  |  |                   |   |                            |      |                 | 2.54 (1.4,4.62)   | 5  |
|  | 7.Upper lobe infiltrate                                 |  |                   |   |                            |      |                 | 5.64 (3.2,9.93)   | 9  |
| <b>Studies that only included smear-negative patients with possible PTB</b>                    |   |  |                   |   |                            |      |                 |                   |    |
| <b>Score</b>   |   |  |                   |   |                            |      |                 |                   |    |
| Lagrange-<br>Xelot et al<br>(2010) <sup>26</sup>   | TB risk factors or chronic<br>symptoms                  |  | 4                 | $\geq$ 1 = test positive                              | 96.2                       | 21.3 |                 |                   |    |
| <b>(validation of<br/>score<br/>developed by<br/>Wisnivesky<br/>et al [2000]<sup>30</sup>)</b> | Self-report of positive<br>tuberculin skin test in past |  | 5                 |   |                            |      |                 |                   |    |
|  | Shortness of breath                                     |  | -3                |   |                            |      |                 |                   |    |
|  | Temperature < 38.5 deg C                                |  | 0                 |   |                            |      |                 |                   |    |
|  | Temperature 38.5 -39 deg<br>C                           |  | 3                 |   |                            |      |                 |                   |    |
|  | Temperature > 39 deg C                                  |  | 6                 |   |                            |      |                 |                   |    |
|  | Crackles on physical<br>examination                     |  | -3                |   |                            |      |                 |                   |    |
|  | Upper lobe disease on<br>CXR                            |  | 6                 |   |                            |      |                 |                   |    |
| Soto et al<br>(2008) <sup>27</sup>   | Hemoptysis  | <b>DOR (95% CI)</b><br>3.24 (1.11, 9.22) | <b>Score</b><br>2 | < 0 = low<br>probability<br>> 4 = high<br>probability | 93                         | 50   | 0.83<br>(0.07)  |                   |    |
|  | Weight loss   |  |                   |   |                            |      |                 | 2.35 (0.86,6.43)  | 1  |
|  | Age > 45  |  |                   |   |                            |      |                 | 2.01 (1.01,3.01)  | -1 |
|  | Expectoration   |  |                   |   |                            |      |                 | 0.35 (0.14,0.9)   | -1 |
|  | Apical infiltrate                                       |  |                   |   |                            |      |                 | 4.29 (1.7,10.86)  | 3  |
|  | Miliary infiltrate                                      |  |                   |   |                            |      |                 | 9.31 (2.21,39.24) | 4  |
|  |   |  |                   |   | <b>At a cut-off &lt; 0</b> |      |                 |                   |    |
|  |   |  |                   |   | <b>At a cut-off &gt; 4</b> | 92   |                 |                   |    |

|   |  |                     |              |                        |   |                     |
|---|--|---------------------|--------------|------------------------|---|---------------------|
| Soto et al (2011) <sup>28</sup>                   | Same as above – validation study                     |                     |              | ≥ 5 = high probability | <b>At a cut-off &lt; 0</b><br>97.8<br>(94.5,99.4) | 14<br>(11,17.4)     |
|   |  |                     |              |                        | <b>At a cut-off &gt; 5</b><br>23.9<br>(17.9,30.7) | 93.1<br>(90.5,95.2) |
|   |  | <b>DOR (95% CI)</b> | <b>Score</b> |                        |   |                     |
| Wisnivesky et al (2000) <sup>30</sup>             | TB risk factors or chronic symptoms                  | 7.9 (4.4,24.2)      | 4            | > 1 = test positive    | 98<br>(95,100)                                    | 46<br>(33,59)       |
|   | Self-report of positive tuberculin skin test in past | 13.2 (4.4,40.7)     | 5            |                        |   |                     |
|   | Shortness of breath                                  | 0.2 (0.1,0.5)       |              |                        |   |                     |
|   | Temperature < 38.5 deg C                             | 2.8 (1.1,8.3)       | -3           |                        |   |                     |
|   | Temperature 38.5 - 39 deg C                          |                     | 0            |                        |   |                     |
|   | Temperature > 39 deg C                               |                     | 3            |                        |   |                     |
|   | Crackles on physical examination                     | 0.3 (0.1,0.5)       | -3           |                        |   |                     |
|   | Upper lobe disease on CXR                            | 14.6 (3.7,57.5)     | 6            |                        |   |                     |
| Wisnivesky et al (2005) <sup>29</sup>             | Same as above  |                     |              |                        | 95<br>(74,100)                                    | 35<br>(31,40)       |
| <b>Study that excluded people living with HIV</b> |  |                     |              |                        |   |                     |
|   |  | <b>DOR (95% CI)</b> | <b>Score</b> |                        |   |                     |
| Rakoczy et al (2008) <sup>31</sup>                | Chronic symptoms                                     | 10.21(2.95,35.4)    | 6            | > 4 = test positive    | 97  | 42                  |
|   | Immunosuppression* (other than HIV)                  | 8.14(2.08,31.8)     | 4            |                        |   |                     |
|   | foreign birth  | 7.01(2.1,23.8)      | 2            |                        |   |                     |
|   | CXR upper zone findings                              | 5.28(1.6,17.2)      | 2            |                        |   |                     |
|   | shortness of breath                                  | 0.13(0.04,0.45)     | -2           |                        |   |                     |
|   | Validation of the score developed by Wisnivesky      |                     |              |                        | 96  | 18                  |

\*Calculated assuming a prevalence of 2%

# 95% CIs are included when the published study either provided these, or the data were sufficient for the CIs to be calculated  
CXR, chest x-ray; NPV, negative predictive value; PPV, positive predictive value

**Table 4. Performance characteristics of six chest radiograph scoring systems in a retrospective validation study of 345 patients<sup>32</sup>**

| Scoring System                        | Sensitivity    |            | Specificity    |            |
|---------------------------------------|----------------|------------|----------------|------------|
|                                       | Original Study | Validation | Original Study | Validation |
|                                       | Cohort         | Cohort     | Cohort         | Cohort     |
| Bock et al (1996) <sup>20</sup>       | 81             | 84         | Not reported   | 42         |
| El-Solh et al(1997) <sup>21</sup>     | 100            | 80         | 50             | 51         |
| Moran et al (2009) <sup>23</sup>      | 96             | 98         | 49             | 20         |
| Mylotte et al (1997) <sup>24</sup>    | 88             | 89         | 63.2           | 69         |
| Wisnivesky et al (2000) <sup>30</sup> | 98             | 93         | 46             | 20         |
| Rakoczy et al (2008) <sup>31</sup>    | 97             | 86         | 42             | 23         |

Figure 1

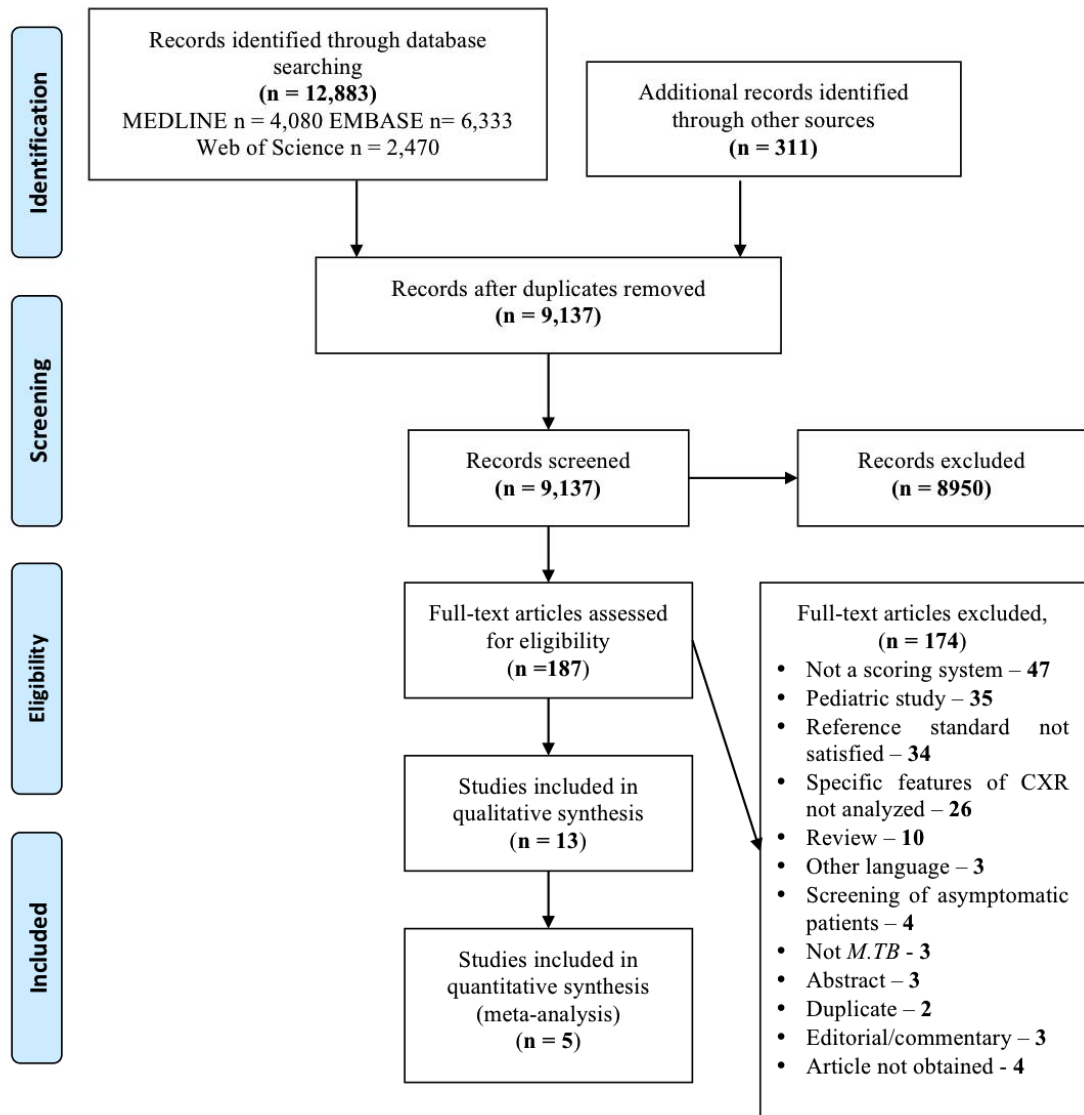
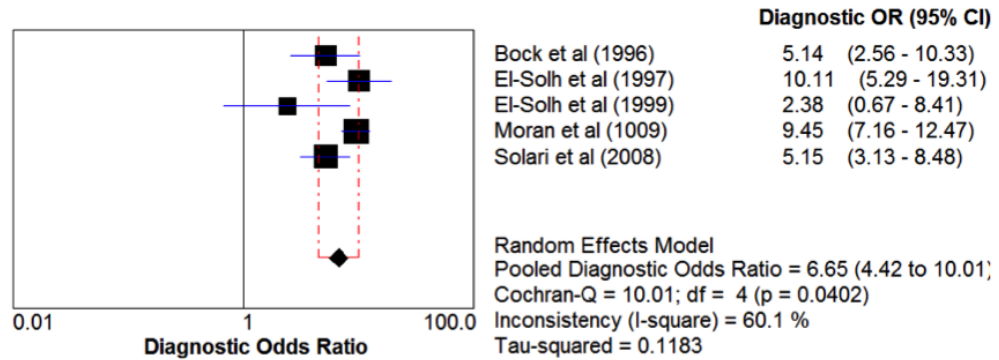


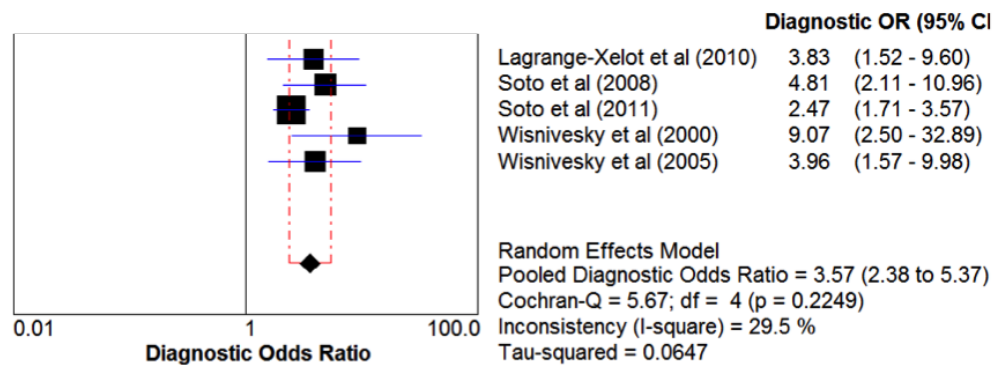


Figure 2

**(A) Among all patients with possible PTB**



**(B) Among smear-negative patients with possible PTB**



The size of the square is proportional to the sample size of the study such that larger studies are represented by larger squares.

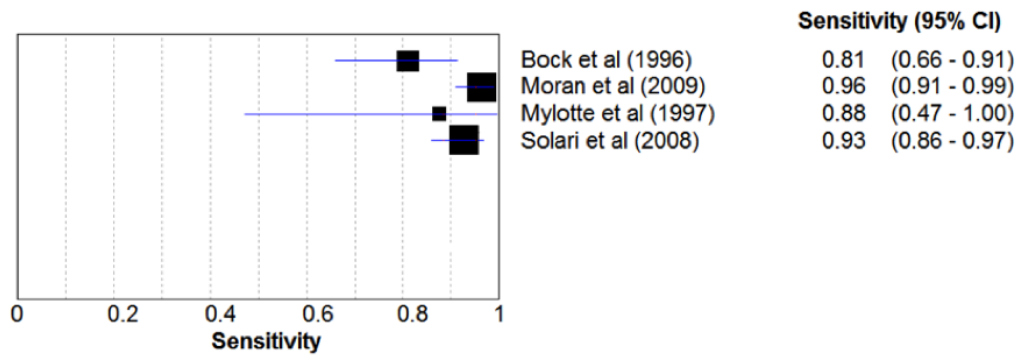
The diamond represents the pooled estimate for the diagnostic odds ratio.

The lines represent the confidence intervals around the respective estimates.

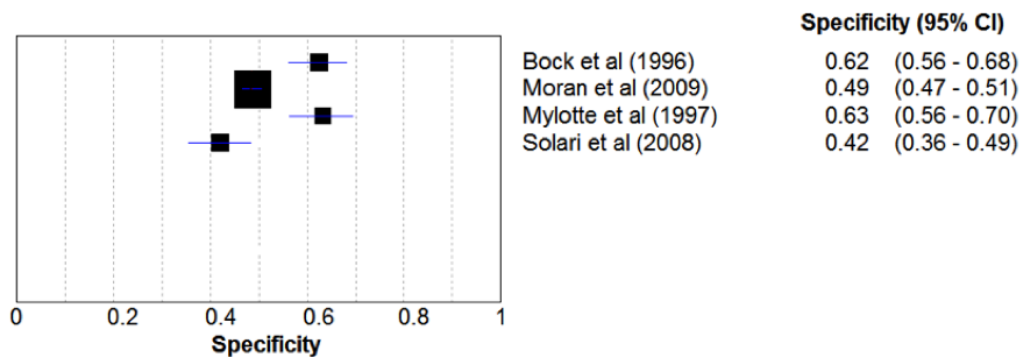
Figure 3

The figures show the estimated sensitivity (A) and specificity (B) of the study (black square) and its 95% confidence interval (blue horizontal line).

(A)



(B)

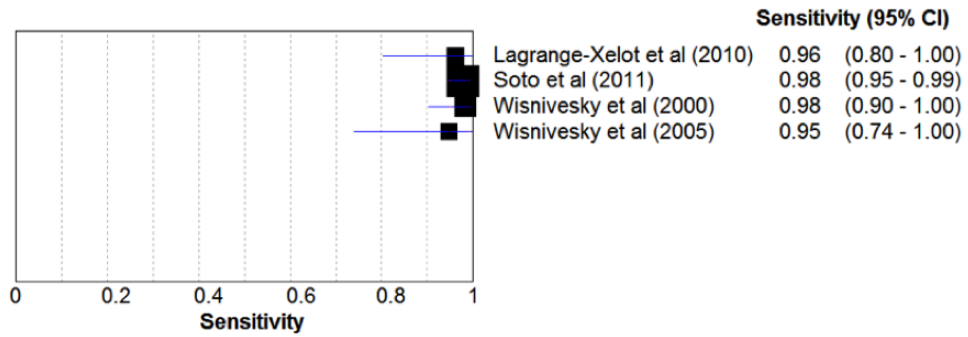


The size of the square is proportional to the sample size of the study such that larger studies are represented by larger squares.

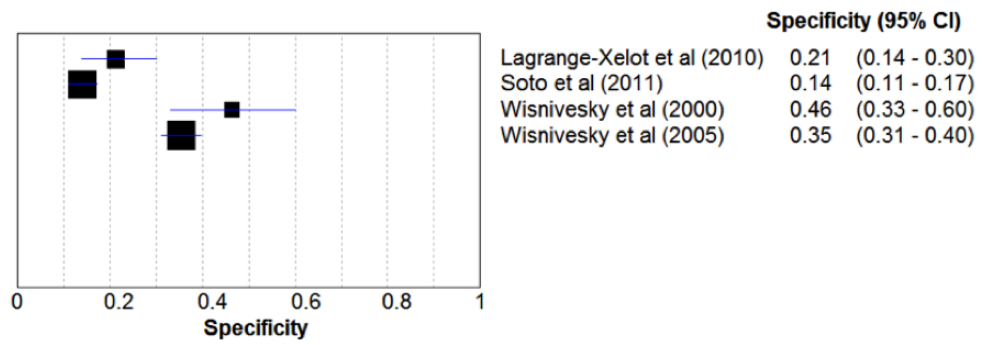
The diamond represents the pooled estimate for the diagnostic odds ratio.

The lines represent the confidence intervals around the respective estimates.

Figure 4  
(A)



(B)



The size of the square is proportional to the sample size of the study such that larger studies are represented by larger squares.

The diamond represents the pooled estimate for the diagnostic odds ratio.

The lines represent the confidence intervals around the respective estimates.