



# 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> and Karen R. Steingart<sup>5</sup>

**Affiliations:** <sup>1</sup>Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, Montreal. <sup>2</sup>Dept of Epidemiology and Biostatistics, McGill University, Montreal, and <sup>4</sup>Respiratory Division, Dept of Medicine, McGill University, Montreal, Canada. <sup>3</sup>University of Cape Town Lung Institute, Cape Town, South Africa. <sup>5</sup>Portland, OR, USA.

**Correspondence:** K.R. Steingart. E-mail: karen.steingart@gmail.com

**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 summarised 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.

13 studies were included in the review, nine of which were in low tuberculosis (TB) burden settings. No scoring system was based solely on radiographic findings. All studies used systems with various combinations of clinical and radiological features. 11 studies involved scoring systems that were used for making decisions concerning hospital respiratory isolation. None of the included studies reported data on intra- or inter-reporter reproducibility. Upper lobe infiltrates (pooled diagnostic OR 3.57, 95% CI 2.38–5.37, five studies) and cavities (diagnostic OR range 1.97–25.66, three 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.



@ERSpublications

Chest radiograph scoring systems are useful for ruling out PTB in hospital, but low specificity precludes ruling in PTB <http://ow.ly/lH9ah>

---

This article has supplementary material available from [www.erj.ersjournals.com](http://www.erj.ersjournals.com)

Received: July 12 2012 | Accepted after revision: October 11 2012 | First published online: Dec 06 2012

**Support statement:** This work was supported by the European and 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. L.M. Pinto is supported by a fellowship from the Shastri Indo-Canadian Institute. These agencies had no role in the analysis of data and decision to publish.

**Conflict of interest:** None declared.

Copyright ©ERS 2013

## Introduction

Despite early diagnosis being a key principle of tuberculosis (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 into 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 is dependent on the reader and not well standardised.

An analogous impediment to the use of chest radiography for the diagnosis of occupational lung diseases was overcome by the development of standardised 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 US 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 standardised 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 standardised scoring system for PTB also has the potential to be combined with newer nucleic acid amplification tests, such as Xpert MTB/RIF (Cepheid, Sunnyvale, CA, USA) [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 radiography 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 the 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].

We aimed to include randomised controlled trials and observational studies of all study designs (*i.e.* 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, true negative, false positive, and false negative values for determining sensitivity and specificity estimates, as well as studies that only reported summary measures of diagnostic accuracy (defined below).

Participants were patients with possible PTB who were  $\geq 15$  years old. 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 haematological and solid organ) and immune-mediated inflammatory disease, and patients on haemodialysis. We also excluded studies that investigated asymptomatic contacts of TB patients.

The index test was any chest radiograph scoring system, with the comparator being no chest radiograph scoring system.

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

We considered liquid or solid culture as the reference standard for active PTB.

A radiograph scoring system was defined as a system that assigned numerical weights to specific features of chest radiographs consistent with PTB (such as cavitory 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 or 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 August 30, 2012), EMBASE (1947 to August 30, 2012) and Web of Science (1899 to August 30, 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 supplementary material. We also reviewed the reference lists of included articles and review articles identified through the search, and hand-searched World Health Organization reports.

### **Selection of studies**

Initially, two review authors (L.M. Pinto and K.R. Steingart) independently scrutinised titles and abstracts in English, French and Spanish for eligibility. Citations deemed relevant by either reviewer were selected and the papers retrieved for full-text review. Next, each eligible article was independently assessed by two reviewers (L.M. Pinto and K.R. Steingart) 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 (L.M. Pinto and K.R. Steingart) 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 scored as “yes”, “no”, or “unclear”.

### **Data extraction**

Two review authors (L.M. Pinto and K.R. Steingart) independently extracted the data from each study, using a data extraction form that was piloted and then finalised 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]; two-by-two tables (with cells documenting true positives, false negatives, false positives and true negatives for the diagnostic test) of individual radiographic features; and details of the scoring systems and their performance characteristics. The data extraction form is included in the online supplementary material.

### **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 characteristic 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, was 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 9137 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 (fig. 1) [20–32].

### Included studies

We did not identify any scoring system that was based exclusively on radiographic criteria. Of the 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 visualised on the chest radiograph as part of the scoring system [35–39]. Table 1 lists the characteristics of the 12 included original

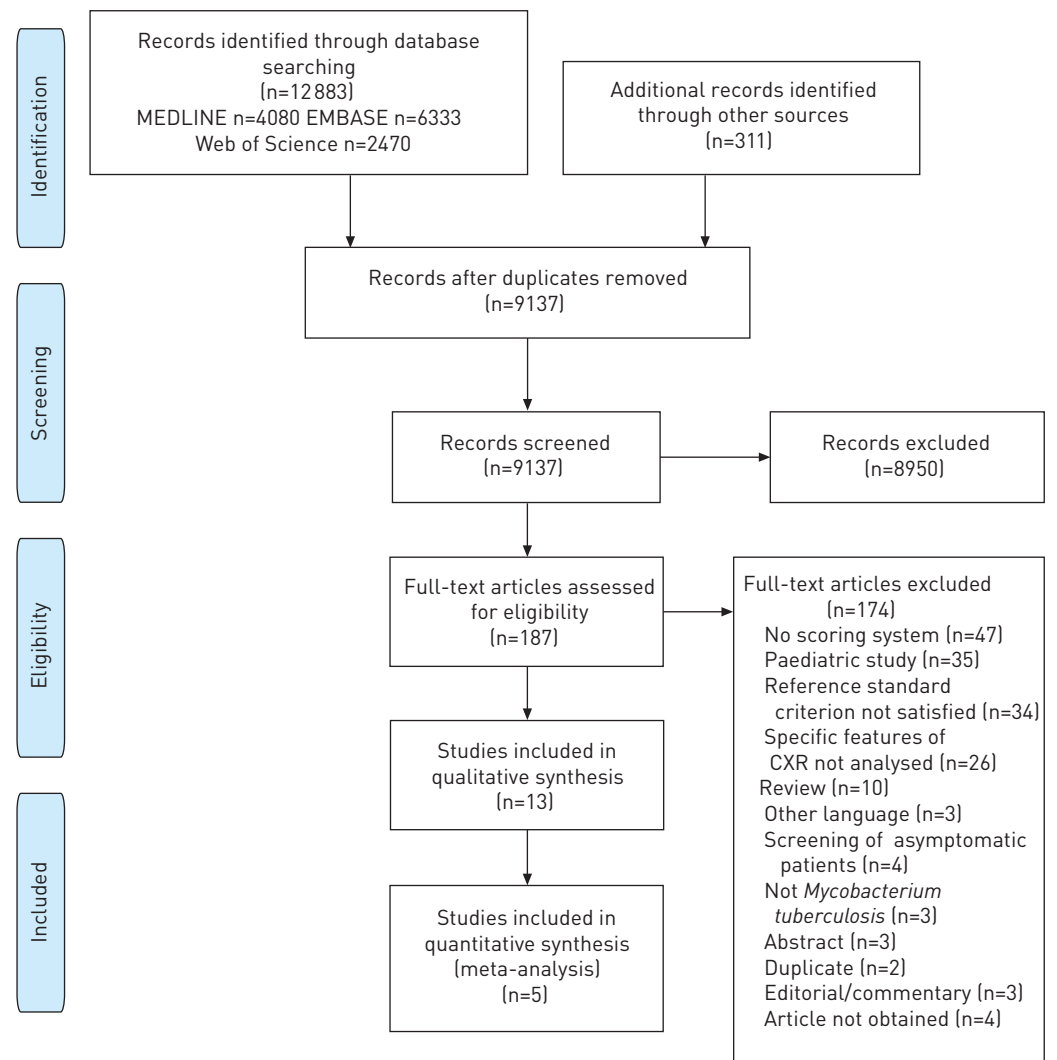


FIGURE 1 PRISMA flow diagram for included and excluded studies. CXR: chest radiograph.

TABLE 1 Characteristics of the included studies

Study	Country	Setting	Eligible patients with possible PTB n (%)	Study design	Inclusion criteria	Chest radiograph reader
<b>Studies that included all patients with possible TB</b>						
Bock <i>et al.</i> [20], 1996	USA	Inpatient	295 (78)	Cross-sectional, retrospective	Patients with active TB, Patients with TB in the differential diagnosis, AFB smears and cultures ordered, HIV positive with abnormal CXR	Radiologist
EL-SOLH <i>et al.</i> [21], 1997	USA	Inpatient	286 (100) <sup>#</sup>	Cross-sectional	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 <i>et al.</i> [22], 1999	USA	Inpatient	119 (100) <sup>#</sup>	Cross-sectional	All patients in whom AFB smear and culture were requested	Radiologist and pulmonologist
MORAN <i>et al.</i> [23], 2009	USA	Inpatient	2535 (91) <sup>#</sup>	Cross-sectional	Admission diagnosis of pneumonia or suspected TB	Emergency medicine resident
MYLOTTE <i>et al.</i> [24], 1997	USA	Inpatient	220 (100) <sup>#</sup>	Cross-sectional	All patients in whom an AFB smear and culture was requested by the admitting physician	Not reported
SOLARI <i>et al.</i> [25], 2008	Peru	Inpatient	345 (71)	Cross-sectional	Productive cough for >1 week or cough of any duration and: fever >3 weeks or weight loss of at least 3 kg in the previous month or night sweats or haemoptysis or differential diagnosis of PTB from attending physician	Internist, internal medicine resident
<b>Studies that only included smear-negative patients with possible PTB</b>						
LAGRANGE-XÉLOT <i>et al.</i> [26], 2010	France	Inpatient	134 (100)	Cross-sectional	Suspected TB, as recommended by French guidelines	Not reported
SOTO <i>et al.</i> [27], 2008	Peru	Inpatient	262 (100)	Cross-sectional	Cough ≥1 week and one or more of the following: fever, weight loss ≥4 kg in 1 month, breathlessness, constitutional symptoms (malaise or hyporexia for a minimum of 2 months)	Not reported
SOTO <i>et al.</i> [28], 2011	Peru	Outpatient	663 (97)	Cross-sectional	Cough ≥2 weeks and one or more of the following: fever, weight loss, breathlessness	1) General practitioner, 2) TB specialist, tie breaker: experienced radiologist
WISNIVESKY <i>et al.</i> [30], 2000	USA	Inpatient	112 (100)	Case-control	Cases: isolated TB patients. Controls: randomly selected from a log of patients who submitted smears and cultures matched on age (±3 years), sex and year of presentation, three smears negative, culture negative and isolated in a hospital	1) Radiologist 2) Radiologist
WISNIVESKY <i>et al.</i> [29], 2005	USA	Inpatient	516 (100)	Cross-sectional	Patients admitted and isolated because of suspicion of PTB	Not reported
<b>Study that included only HIV-uninfected patients with possible PTB</b>						
RAKOCZY <i>et al.</i> [31], 2008	USA	Inpatient	280 (100) <sup>#</sup>	Case-control	Cases: all TB inpatients. 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 <i>et al.</i> [32], 2011 <sup>#</sup>	Peru	Inpatient	345 (71)	Cross-sectional	Productive cough for >1 week or cough of any duration and: fever >3 weeks or weight loss of at least 3 kg in the previous month or night sweats or haemoptysis or differential diagnosis of PTB from attending physician	Internist, internal medicine resident

PTB: pulmonary TB; TB: tuberculosis; AFB: acid-fast bacilli; CXR: chest radiograph. <sup>#</sup>: The study by SOLARI *et al.* [32] applied various scoring systems to the same cohort of patients as the study by SOLARI *et al.* [25]; <sup>\*</sup>: Studies had derivation and validation cohorts. The number of patients with possible TB represents those in the validation cohorts.

studies, containing a total of 5767 participants, and the single 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 177–431).

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 supplementary material. When reported, the majority of patients were male. Eleven studies included PLWH, who represented 11–61% 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 hospitalised patients found to be sputum smear-negative [40, 41]. Both studies satisfied the 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.* [41] concluded that the “numbers were insufficient to develop a score for TB”, while the study by DAVIS *et al.* [40] stated that “after exhaustive testing, we were unable to identify any combination of factors which reliably predicted bacteriologically confirmed tuberculosis”.

We also excluded 13 studies that used automated computer-assisted diagnosis 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 visualised 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 visualised on a chest radiograph, as these studies did not use culture as a reference standard.

#### **Assessment of methodological quality**

As seen in figure 2, all the 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 who would be investigated further with culture, the reference standard. Seven (54%) of the 13 studies did not include a sample that was considered representative of the target population, as these studies did not enrol 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 4.42–10.01; five studies) (fig. 3a), and cavities (DOR range 2.11–10.08; three studies) (online supplementary material).

The details of the parameters included in the scores and their respective weights are summarised in table 2, along with the performance characteristics of the scoring system and the final rule to aid in decision-making. The 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 81–100%). For the five studies that reported specificity data, specificity estimates were low (median 42%, range 22–72%), suggesting a poor rule-in value for PTB.

Figure 4 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 2.38–5.37; five studies) (fig. 3b), and cavities (DOR range 1.97–25.66; three studies). We did not pool estimates based on scoring systems because of the differences in these scoring systems (online supplementary material).

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 93–98%). However, specificity estimates were low (median 35%, range 14–50%), again suggesting a poor rule-in value for PTB (fig. 5). We did not consider these scoring systems to be similar and therefore, did not pool the accuracy estimates.



Study	Representative spectrum?	Acceptable reference standard?	Acceptable delay?	Partial verification avoided?	Differential verification avoided?	Incorporation avoided?	Reference standard results blinded?	Index text results blinded?	Relevant clinical information available?	Uninterpretable results reported?	Withdrawals explained?
BOCK <i>et al.</i> [20], 1996											
EL-SOLH <i>et al.</i> [21], 1997											
EL-SOLH <i>et al.</i> [22], 1999											
MORAN <i>et al.</i> [23], 2009											
MYLOTTE <i>et al.</i> [24], 1997											
SOLARI <i>et al.</i> [25], 2008; [32], 2011											
LAGRANGE-XÉLOT <i>et al.</i> [26], 2010											
SOTO <i>et al.</i> [27], 2008											
SOTO <i>et al.</i> [28], 2011											
WISNIVESKY <i>et al.</i> [30], 2000											
WISNIVESKY <i>et al.</i> [29], 2005											
RAKOCZY <i>et al.</i> [31], 2008											

Yes
  No
  Unclear

FIGURE 2 Quality assessment of the included studies using the Quality Assessment of Diagnostic Accuracy Studies tool.

One study, performed in an inpatient setting, excluded PLWH (table 1) [31]. 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 2).

*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 been included in an earlier study [25] (this study is included in the current systematic review) 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 3.

*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 standardised radiographic scoring systems for the diagnosis of PTB, and whether standardisation 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 healthcare system [64] and in resource-limited settings, where isolation beds may be in short supply,

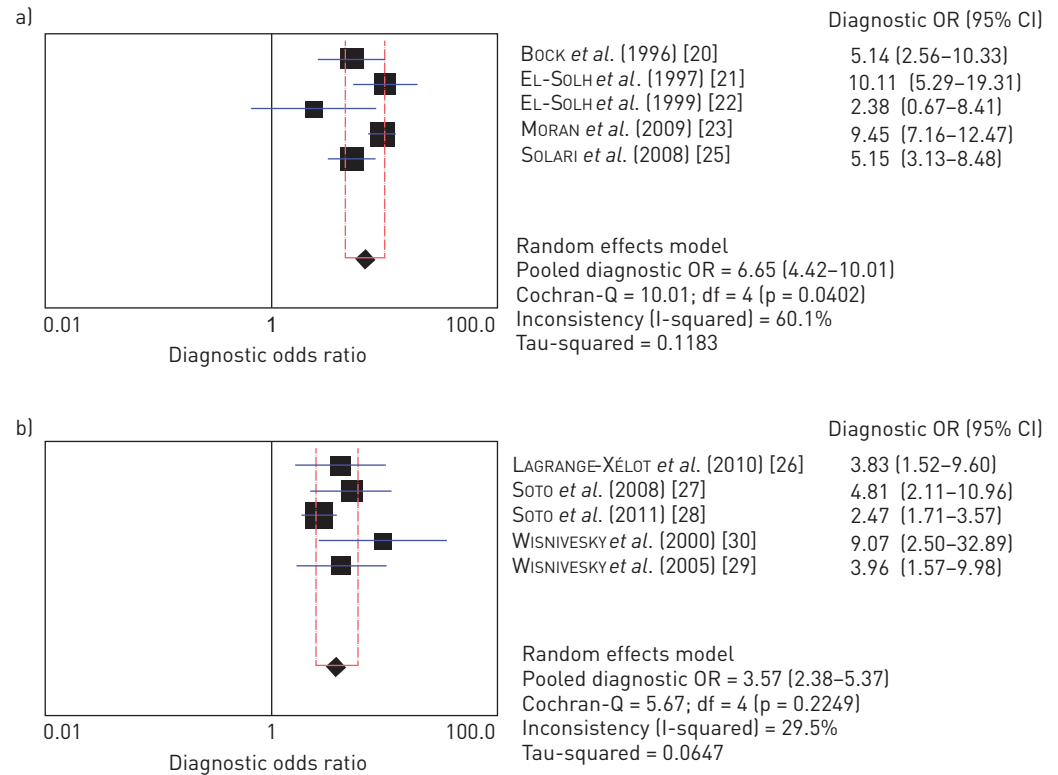


FIGURE 3 Diagnostic odds ratio for active pulmonary tuberculosis (PTB) with an upper lobe infiltrate visualised on the chest radiograph. a) Among all patients with possible PTB and b) among smear-negative patients with possible PTB. The size of each square is proportional to the sample size of the study, such that larger studies are represented by larger squares. Diamonds represent the pooled estimate for the diagnostic OR. The lines represent the confidence intervals around the respective estimates. df: degrees of freedom.

unnecessary isolation of some patients may preclude appropriate isolation of others. Scoring systems that improve the accuracy of decisions to subject patients with possible PTB to respiratory isolation can considerably improve the efficiency of healthcare systems and utilisation 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 seven studies) but a poor rule-in value (median PPV 22.5%, range 8–61%, values reported in eight 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 supplementary material). 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 *et al.* [30] 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 *et al.* [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 *et al.* [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].



TABLE 2 Details and performance characteristics of chest radiograph scoring systems

Study	Feature	Test details	Rule	Sensitivity % [95% CI]#	Specificity % [95% CI]#	Area under the curve (SE)
<b>Studies that included all patients with possible PTB</b> <b>Bock et al. [20], 1996</b>	1) CXR with upper lobe infiltrate 2) CXR with cavity 3) Knew someone with active TB 4) Self-report of positive tuberculin skin test in past 5) Self-report of isoniazid preventive therapy in the past	Logistic regression DOR (95%CI) 5 [2.38–10.51] 3.93 [1.06–14.62] 2.42 [1.1–5.32] 5.67 [1.57–22.01] 0.18 [0.04–0.82]	Any of 1–3 or 4 (in the absence of 5) considered test positive	81		
<b>EL-SOLH et al. [21], 1997</b>	Upper zone disease Weight loss Diabetes mellitus	Used classification and regression tree analysis	Upper zone disease absent and fever absent = test negative Upper zone disease absent, but fever present = test negative if no weight loss and CD4 <sup>+</sup> ≥ 200. Everyone else to be considered test positive	100 [78–100]	50 [44–57]	0.878 [0.029]
<b>EL-SOLH et al. [22], 1999</b>	Age, CD4 <sup>+</sup> counts, diabetes mellitus, HIV, tuberculin skin test positivity Chest pain, weight loss, cough, night sweats, fever, shortness of breath Upper lobe infiltrate, lower lobe infiltrate, upper lobe cavity, lower lobe cavity, adenopathy, unilateral pleural effusion, bilateral pleural effusion, pleural thickening, miliary pattern, normal	Used general regression neural network		100 [91–100]	72 [65–77]	0.947 [0.028]
<b>MORAN et al. [23], 2009</b>	1) Apical infiltrate 2) Cavitation 3) Immigrant 4) Weight loss 5) Positive TB history 6) Homeless 7) Incarcerated	Used Chi-squared recursive partitioning	Any of 1–7 present = test positive	96	49	
<b>MLOTTE et al. [24], 1997</b>	1) AFB positive smear 2) Localised CXR change 3) Correctional facility residence 4) History of weight loss	DOR (95% CI) 5.8 [3–11] 2.5 [1.3–4.9] 2.3 [1.2–4.4] 1.8 [1–3.2] DOR (95% CI) 0.97 [0.96–0.99]	Score ≥ 3 = test positive	88	63.2	0.86 [0.04]
<b>SOLARI et al. [25], 2008</b>	1) Age <35 years 2) Age 35–60 years 3) Age ≥ 61 years 4) Weight loss 5) History of PTB 6) Miliary pattern 7) Cavity 8) Upper lobe infiltrate	Score 3 2 2 1 Score 0 -1 -2 5 -3 10 5 9	Score ≥ 3 = test positive	93	42	0.809 [0.05]

TABLE 2 Continued

Study	Feature	Test details	Rule	Sensitivity % (95% CI) <sup>#</sup>	Specificity % (95% CI) <sup>#</sup>	Area under the curve (SE)
<b>Studies that only included smear-negative patients with possible PTB</b>						
<b>LAGRANGE-XÉLIOT et al. [26], 2010<sup>*</sup></b>	TB risk factors or chronic symptoms Self-report of positive tuberculin skin test in past Shortness of breath Temperature <38.5°C Temperature 38.5–39°C Temperature >39°C Crackles on physical examination Upper lobe disease on CXR	Score 4 5 -3 0 3 6 -3 6	≥1 = test positive	96.2	21.3	
<b>Soto et al. [27], 2008</b>	Haemoptysis Weight loss Age >45 years Expectoration Apical infiltrate Miliary infiltrate	DOR (95% CI) 3.24 (1.11–9.22) 2.35 (0.86–6.43) 2.01 (1.01–3.01) 0.35 (0.14–0.9) 4.29 (1.7–10.86) 9.31 (2.21–39.24)	Score 2 1 -1 -1 3 4 <0 = low probability >4 = high probability	At a cut-off <0 93	At a cut-off >4 50 92	At a cut-off of >2 0.83 (0.07)
<b>Soto et al. [28], 2011</b>	Same as above – validation study		≥5 = high probability	At a cut-off <0 97.8 (94.5–99.4) At a cut-off >5 23.9 (17.9–30.7)	At a cut-off <0 14 (11–17.4) At a cut-off >5 93.1 (90.5–95.2)	
<b>WISNIEWSKY et al. [30], 2000</b>	TB risk factors or chronic symptoms Self-report of positive tuberculin skin test in past Shortness of breath Temperature <38.5°C Temperature 38.5–39°C Temperature >39°C Crackles on physical examination Upper lobe disease on CXR	DOR (95% CI) 7.9 (4.4–24.2) 13.2 (4.4–40.7) 0.2 (0.1–0.5) 0.3 (0.1–0.5) 0.3 (0.1–0.5) 14.6 (3.7–57.5)	Score 4 5 -3 0 3 6 >1 = test positive	98 (95–100)	46 (33–59)	
<b>WISNIEWSKY et al. [29], 2005</b>	Same as above			95 (74–100)	35 (31–40)	
<b>Study that excluded people living with HIV</b>						
<b>Rakoczy et al. [31], 2008</b>	Chronic symptoms Immunosuppression <sup>+</sup> (other than HIV) Foreign birth CXR upper zone findings Shortness of breath Validation of the score developed by Wisniewsky	DOR (95% CI) 10.21 (2.95–35.4) 8.14 (2.08–31.8) 7.01 (2.1–23.8) 5.28 (1.6–17.2) 0.13 (0.04–0.45)	Score 6 4 2 2 -2	97	42	

CXR: chest radiograph; TB: tuberculosis; DOR: diagnostic odds ratio; AFB: acid-fast bacilli; PTB: pulmonary TB. <sup>#</sup>: 95% CIs are included when the published study either provided these, or the data were sufficient for the CIs to be calculated. <sup>\*</sup>: validation of the score developed by Wisniewsky et al. [30]; <sup>+</sup>: Calculated assuming a prevalence of 2%.

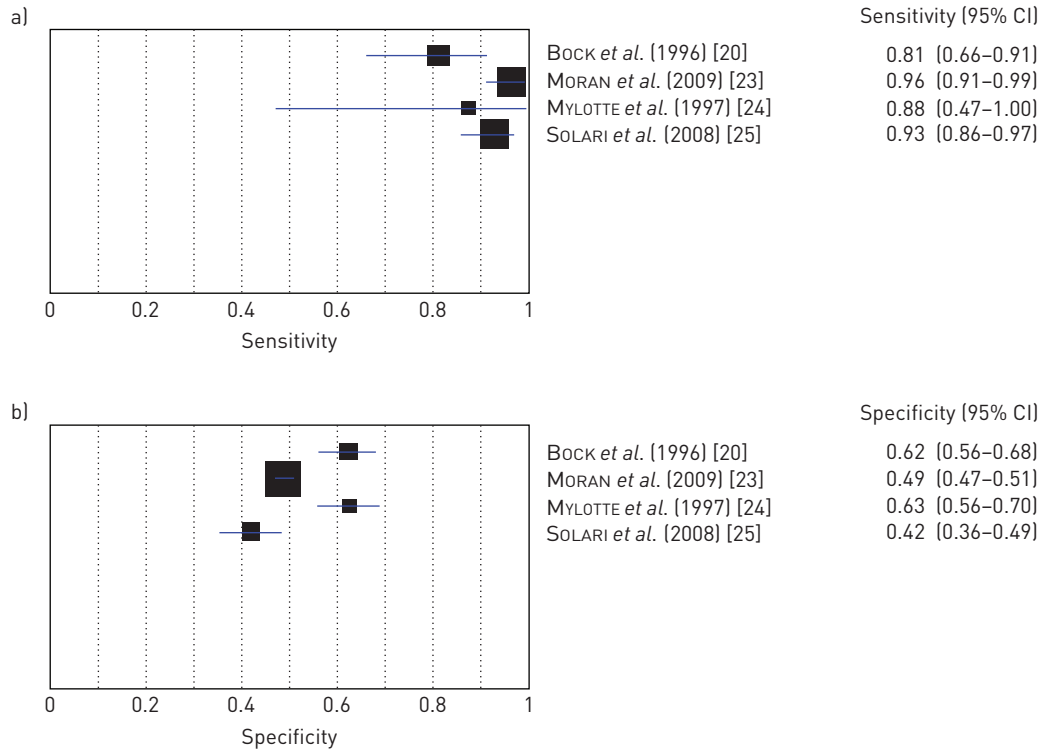


FIGURE 4 Scoring systems for studies that included all patients with possible pulmonary tuberculosis. a) The estimated sensitivity and b) the specificity of the study (black squares). The size of the square is proportional to the sample size of the study, such that larger studies are represented by larger squares. The lines represent the confidence intervals around the respective estimates.

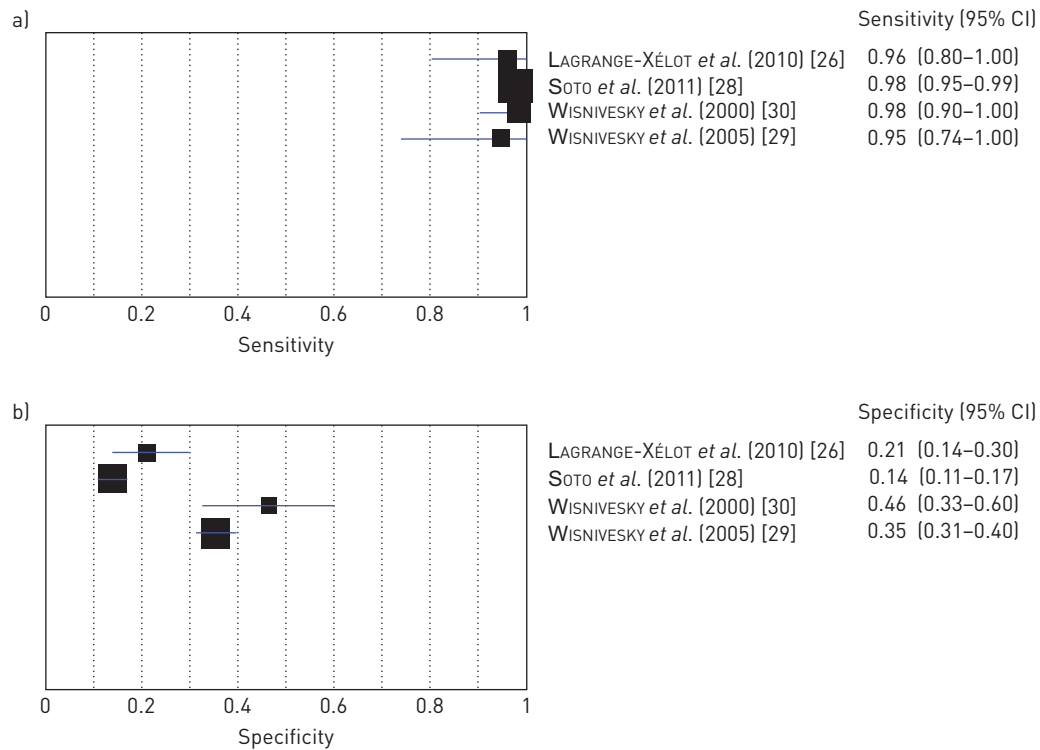


FIGURE 5 Scoring systems for studies that included smear-negative patients with possible pulmonary tuberculosis. a) The estimated sensitivity and b) the specificity of each scoring system (black squares). The size of the square is proportional to the sample size of the study, such that larger studies are represented by larger squares. The lines represent the confidence intervals around the respective estimates.

TABLE 3 Performance characteristics of six chest radiograph scoring systems in a retrospective validation study of 345 patients [32]

Scoring System	Sensitivity		Specificity	
	Original study cohort	Validation cohort	Original study cohort	Validation cohort
BOCK <i>et al.</i> [20], 1996	81	84	Not reported	42
EL-SOLH <i>et al.</i> [21], 1997	100	80	50	51
MORAN <i>et al.</i> [23], 2009	96	98	49	20
MYLOTTE <i>et al.</i> [24], 1997	88	89	63.2	69
WISNIVESKY <i>et al.</i> [30], 2000	98	93	46	20
RAKOCZY <i>et al.</i> [31], 2008	97	86	42	23

Automated computer-assisted diagnosis employs techniques such as texture analysis for reading digital chest radiographs, and appears to be a promising modality for standardising 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. First, 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 compared with 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 compared with 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 [15]. 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 the radiographic interpretation. Selection bias and absence of blinding are features of study design that have also been associated with inflated accuracy estimates [71, 72]. These limitations in the quality of the included studies need to be taken into consideration when interpreting the 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 standardise 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 computer-assisted diagnosis, need to be validated in well-designed studies to assess their utility.

### Acknowledgement

We thank L.A. Kloda (Life Sciences Library, McGill University, Montreal, Canada) for her guidance and assistance in formulating the search strategy.

## References

- 1 World Health Organization. Global Tuberculosis Control 2011. Geneva, WHO Press, 2011. Available from: [www.who.int/tb/publications/global\\_report/2011/gtbr11\\_full.pdf](http://www.who.int/tb/publications/global_report/2011/gtbr11_full.pdf) Date last accessed: April 4, 2012.
- 2 Steingart KR, Ng V, Henry M, *et al.* Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis* 2006; 6: 664–674.
- 3 Lerner BH. The perils of “X-ray vision”: how radiographic images have historically influenced perception. *Perspect Biol Med* 1992; 35: 382–397.
- 4 Koppaka R, Bock N. How reliable is chest radiography? In: Frieden T, ed. *Toman’s Tuberculosis: Case Detection, Treatment, and Monitoring – Questions and Answers*. 2nd Edn. Geneva, World Health Organization, 2004; pp. 51–60.
- 5 International Labour Office. Guidelines for the use of the ILO international classification of radiographs of pneumoconiosis. Revised edition 2000. Occupational Safety and Health Series, No. 22. Geneva, International Labour Office, 2002.
- 6 Bohlig H. UICC-Cincinnati-Klassifikation radiographischer Staublungenbefunde. Gemeinschaftsarbeit eines Komitees der Internationalen Union gegen den Krebs (UICC). [UICC-Cincinnati classification of radiographic findings in pneumoconiosis. Collective work of a subcommittee of the International Union Against Cancer (UICC)]. *Fortschr Geb Rontgenstr Nuklearmed* 1971; 115: 665–683.
- 7 Grum CM, Lynch JP 3rd. Chest radiographic findings in cystic fibrosis. *Semin Respir Infect* 1992; 7: 193–209.
- 8 Murray JF, Matthay MA, Luce JM, *et al.* An expanded definition of the adult respiratory distress syndrome. *Am Rev Respir Dis* 1988; 138: 720–723.
- 9 Boehme CC, Nabeta P, Hillemann D, *et al.* Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* 2010; 363: 1005–1015.
- 10 Wisnivesky JP, Serebrisky D, Moore C, *et al.* Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. *J Gen Intern Med* 2005; 20: 947–952.
- 11 Leeftang MM, Deeks JJ, Gatsonis C, *et al.* Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; 149: 889–897.
- 12 Macaskill P, Gatsonis C, Deeks JJ, *et al.* Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010. Available from: <http://srdta.cochrane.org/> Date last accessed: March 6, 2011.
- 13 Wilczynski NL, Haynes RB, Hedges Team. 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; 328: 1040.
- 15 Whiting P, Rutjes AW, Reitsma JB, *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
- 16 World Bank. World Bank List of Economies, 2011. <http://siteresources.worldbank.org/DATASTATISTICS/Resources/CLASS.XLS> Date last updated: November 9, 2011.
- 17 Zamora J, Abraira V, Muriel A, *et al.* Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006; 6: 31.
- 18 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–188.
- 19 Tatsioni A, Zarin DA, Aronson N, *et al.* Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med* 2005; 142: 1048–1055.
- 20 Bock NN, McGowan JE Jr, Ahn J, *et al.* Clinical predictors of tuberculosis as a guide for a respiratory isolation policy. *Am J Respir Crit Care Med* 1996; 154: 1468–1472.
- 21 El-Solh AA, Mylotte J, Sherif S, *et al.* Validity of a decision tree for predicting active pulmonary tuberculosis. *Am J Respir Crit Care Med* 1997; 155: 1711–1716.
- 22 El-Solh AA, Hsiao CB, Goodnough S, *et al.* Predicting active pulmonary tuberculosis using an artificial neural network. *Chest* 1999; 116: 968–973.
- 23 Moran GJ, Barrett TW, Mower WR, *et al.* Decision instrument for the isolation of pneumonia patients with suspected pulmonary tuberculosis admitted through US emergency departments. *Ann Emerg Med* 2009; 53: 625–632.
- 24 Mylotte JM, Rodgers J, Fassel M, *et al.* Derivation and validation of a pulmonary tuberculosis prediction model. *Infect Control Hosp Epidemiol* 1997; 18: 554–560.
- 25 Solari L, Acuna-Villaorduna C, Soto A, *et al.* A clinical prediction rule for pulmonary tuberculosis in emergency departments. *Int J Tuberc Lung Dis* 2008; 12: 619–624.
- 26 Lagrange-Xélot M, Porcher R, Gallien S, *et al.* Prevalence and clinical predictors of pulmonary tuberculosis among isolated inpatients: a prospective study. *Clin Microbiol Infect* 2011; 17: 610–614.
- 27 Soto A, Solari L, Agapito J, *et al.* Development of a clinical scoring system for the diagnosis of smear-negative pulmonary tuberculosis. *Braz J Infect Dis* 2008; 12: 128–132.
- 28 Soto A, Solari L, Díaz J, *et al.* Validation of a clinical-radiographic score to assess the probability of pulmonary tuberculosis in suspect patients with negative sputum smears. *PLoS One* 2011; 6: e18486.
- 29 Wisnivesky JP, Henschke C, Balentine J, *et al.* Prospective validation of a prediction model for isolating inpatients with suspected pulmonary tuberculosis. *Arch Intern Med* 2005; 165: 453–457.
- 30 Wisnivesky JP, Kaplan J, Henschke C, *et al.* Evaluation of clinical parameters to predict *Mycobacterium tuberculosis* in inpatients. *Arch Intern Med* 2000; 160: 2471–2476.
- 31 Rakoczy KS, Cohen SH, Nguyen HH. Derivation and validation of a clinical prediction score for isolation of inpatients with suspected pulmonary tuberculosis. *Infect Control Hosp Epidemiol* 2008; 29: 927–932.
- 32 Solari L, Acuna-Villaorduna C, Soto A, *et al.* Evaluation of clinical prediction rules for respiratory isolation of inpatients with suspected pulmonary tuberculosis. *Clin Infect Dis* 2011; 52: 595–603.
- 33 Tattevin P, Casalino E, Fleury L, *et al.* The validity of medical history, classic symptoms, and chest radiographs in predicting pulmonary tuberculosis: derivation of a pulmonary tuberculosis prediction model. *Chest* 1999; 115: 1248–1253.

- 34 Tessema TA, Bjune G, Assefa G, *et al.* An evaluation of the diagnostic value of clinical and radiological manifestations in patients attending the Addis Ababa tuberculosis centre. *Scand J Infect Dis* 2001; 33: 355–361.
- 35 Wang CS, Chen HC, Chong IW, *et al.* Predictors for identifying the most infectious pulmonary tuberculosis patient. *J Formos Med Assoc* 2008; 107: 13–20.
- 36 Aguilar J, Yang JJ, Brar I, *et al.* Clinical prediction rule for respiratory isolation of patients with suspected pulmonary tuberculosis. *Infect Dis Clin Pract* 2009; 17: 317–322.
- 37 Redd JT, Susser E. Controlling tuberculosis in an urban emergency department: a rapid decision instrument for patient isolation. *Am J Public Health* 1997; 87: 1543–1547.
- 38 Gaeta TJ, Webheh W, Yazji M, *et al.* Respiratory isolation of patients with suspected pulmonary tuberculosis in an inner-city hospital. *Acad Emerg Med* 1997; 4: 138–141.
- 39 Pegues CF, Johnson DC, Pegues DA, *et al.* Implementation and evaluation of an algorithm for isolation of patients with suspected pulmonary tuberculosis. *Infect Control Hosp Epidemiol* 1996; 17: 412–418.
- 40 Davis JL, Wordria W, Kitembo H, *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; 5: e9859.
- 41 Le Minor O, Germani Y, Chartier L, *et al.* Predictors of pneumocystosis or tuberculosis in HIV-infected Asian patients with AFB smear-negative sputum pneumonia. *J Acquir Immune Defic Syndr* 2008; 48: 620–627.
- 42 Hogeweg LE, Mol C, de Jong PA, *et al.* Rib suppression in chest radiographs to improve classification of textural abnormalities. *SPIE Proceedings* 2010; 7624.
- 43 Mouton A, Pitcher RD, Douglas TS. Computer-aided detection of pulmonary pathology in pediatric chest radiographs. *Med Image Comput Comput Assist Interv* 2010; 13: 619–625.
- 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; 42: 1768–1776.
- 45 Katsuragawa S, Doi K. Computer-aided diagnosis in chest radiography. *Comput Med Imaging Graph* 2007; 31: 212–223.
- 46 Le K. Chest X-Ray Analysis for Computer-Aided Diagnostic. *Advanced Computing* 2011; 133: 300–309.
- 47 Patil SA, Udupi VR, Kane CD, *et al.* Geometrical and Texture Features Estimation of Lung Cancer and TB Images Using Chest X-ray Database. *Int Conf Biomed Pharm Eng*, 2009: 1–7.
- 48 Mohd Rijal O, Mohd Noor N, Shaban H, *et al.* A statistical comparison of digital X-ray images for MTB patients. *Conf Proc IEEE Eng Med Biol Soc* 2005; 6: 6418–6421.
- 49 Rijal OM, Iqbal M, Yunus A, *et al.* Some Critical Remarks on the initial detection of lung ailments using clinical data and chest radiography. *Proc Wseas Int Conf Comput* 2009: 470–475.
- 50 Sarkar S, Chaudhuri S. Evaluation and progression analysis of pulmonary tuberculosis from digital chest radiographs. *Comput Med Imaging Graph* 1998; 22: 145–155.
- 51 Shen R, Cheng I, Basu A. A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. *IEEE Trans Biomed Eng* 2010; 57: 2646–2656.
- 52 van Ginneken B, ter Haar Romeny BM, 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; 685–688.
- 53 van Ginneken B, ter Haar Romeny BM. Automatic segmentation of lung fields in chest radiographs. *Med Phys* 2000; 27: 2445–2455.
- 54 van Ginneken B, Katsuragawa S, ter Haar Romeny BM, *et al.* Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Trans Med Imaging* 2002; 21: 139–149.
- 55 Caplin M, Grange JM, Morley S, *et al.* Relationship between radiological classification and the serological and haematological features of untreated pulmonary tuberculosis in Indonesia. *Tubercle* 1989; 70: 103–113.
- 56 Churchyard GJ, Fielding K, Roux S, *et al.* Twelve-monthly versus six-monthly radiological screening for active case-finding of tuberculosis: a randomised controlled trial. *Thorax* 2011; 66: 134–139.
- 57 Ralph AP, Ardian M, Wiguna A, *et al.* A simple, valid, numerical score for grading chest x-ray severity in adult smear-positive pulmonary tuberculosis. *Thorax* 2010; 65: 863–869.
- 58 Tuberculosis Chemotherapy Centre. A concurrent comparison of home and sanatorium treatment of pulmonary tuberculosis in South India. *Bull World Health Organ* 1959; 21: 51–144.
- 59 Wejse C, Gustafson P, Nielsen J, *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. *Scand J Infect Dis* 2008; 40: 111–120.
- 60 Den Boon S, Bateman ED, Enarson DA, *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; 9: 1088–1096.
- 61 Agizew T, Bachhuber MA, Nyirenda S, *et al.* Association of chest radiographic abnormalities with tuberculosis disease in asymptomatic HIV-infected adults. *Int J Tuberc Lung Dis* 2010; 14: 324–331.
- 62 Dawson R, Masuka P, Edwards DJ, *et al.* Chest radiograph reading and recording system: evaluation for tuberculosis screening in patients with advanced HIV. *Int J Tuberc Lung Dis* 2010; 14: 52–58.
- 63 Riley RL, Nardell EA. Clearing the air. The theory and application of ultraviolet air disinfection. *Am Rev Respir Dis* 1989; 139: 1286–1294.
- 64 Kellerman S, Tokars JL, Jarvis WR. The cost of selected tuberculosis control measures at hospitals with a history of Mycobacterium tuberculosis outbreaks. *Infect Control Hosp Epidemiol* 1997; 18: 542–547.
- 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; 131: 393–396.
- 66 Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. *Br J Radiol* 2005; 78 Spec1: S3–S19.
- 67 Banoo S, Bell D, Bossuyt P, *et al.* Evaluation of diagnostic tests for infectious diseases: general principles. *Nat Rev Microbiol* 2006; 4: S20–S32.
- 68 Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med* 2003; 138: 40–44.



- 69 Willis BH. Spectrum bias – why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract* 2008; 25: 390–396.
- 70 Laifer G, Widmer AF, Simcock M, *et al.* TB in a low-incidence country: differences between new immigrants, foreign-born residents and native residents. *Am J Med* 2007; 120: 350–356.
- 71 Lijmer JG, Mol BW, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061–1066.
- 72 Rutjes AW, Reitsma JB, Di Nisio M, *et al.* Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; 174: 469–476.