Tuberculosis diagnostics trials: do they lack methodological rigor?

Madhukar Pai & Richard O'Brien

Tuberculosis diagnostics trials: do they lack methodological rigor?

'Lack of methodological rigor in TB trials is a cause for concern, as it may prove to be an important hurdle for effective application of diagnostics in TB control.'

Madhukar Pai[†]
Richard O'Brien
[†]*Author for correspondence*
McGill University, Department of
Epidemiology, Biostatistics
& Occupational Health,
1020 Pine Avenue West, Montreal,
H3A 1A2, Canada
Tel.: +1 514 398 5422
Fax: +1 514 398 4503
madhukar.pai@mcgill.ca

Globally, tuberculosis (TB) is a problem of staggering proportions. Between 8 and 9 million people develop TB disease, and approximately 2 million die as a consequence every year [1]. Despite the widespread implementation of the WHO's DOTS strategy for TB control [101], case detection rates remain low, and the global targets for case detection have not been met [1]. Diagnostic algorithms in current use, particularly in developing countries, are based on tests that have been in clinical use for many decades [2,3]. Moreover, the two greatest challenges to TB control, the TB/HIV epidemic and the growing problem of multidrug-resistant TB, cannot be adequately addressed solely by sputum smear microscopy (the primary diagnostic tool of the DOTS strategy). Thus, the lack of accurate and rapid diagnostic tests for TB is an important impediment to global TB control.

However, the situation is slowly changing. In developing countries, there is a growing recognition of the need both to implement the diagnostic tests that are commonly used in industrialized countries and to promote the development of new tests. The Global Plan to Stop TB, 2006–2015 [102], recently issued by the Stop TB Partnership [103], calls for a significant increase in funding to meet these needs. Thanks to the involvement of global health agencies, donors, academic institutions, industries and public–private partnerships, substantial progress is being made in the quest to develop new tools for TB diagnosis, and several new tests are currently in the pipeline [3–5].

However, despite the progress made and the steady output of research on TB diagnostics, there is a concern that trials on the accuracy of TB diagnostics lack methodological rigor [2,6]. Consequently, there is a perception that new tests that reportedly perform well in clinical trials may turn out to be less useful in routine clinical practice [2,3]. Biased results from poorly designed studies can lead to premature adoption of diagnostics that may have little or no benefit, and result in adverse consequences for the patient. The situation is exacerbated by the fact that most high-burden countries have poor regulatory mechanisms for marketing and post-marketing surveillance of diagnostics [2,6]. For example, commercial serological tests for TB are marketed in many developing countries, despite lack of evidence on their accuracy and utility [5].

Is there evidence that trials of tuberculosis diagnostics lack methodological rigor?

Diagnostic accuracy trials are prone to several biases [7,8]. A methodologically rigorous trial would avoid or minimize such biases and produce valid results. There are several approaches to evaluate the methodological quality of diagnostic studies. TABLE 1 details one example, the quality assessment of diagnostic accuracy studies (QUADAS) tool, a validated quality assessment instrument specific for diagnostic studies [9,10]. To address the concern about quality of trials on TB diagnostics, we should, ideally, evaluate a sample of TB diagnostic trials using a tool such as QUADAS. However, given the large number of published diagnostic trials, this approach is not easy. A viable alternative is to exploit meta-analyses on TB diagnostics to determine the methodological quality of trials. Meta-analyses often include quality assessment as a key component of the systematic review process [11].

To identify meta-analyses on TB diagnostics, we searched PubMed (2000–2006) with 'tuberculosis' OR 'tuberculous' in the title or abstract, and combined this with 'meta-analysis' OR 'meta-regression' in the title or abstract. A total of ten eligible studies (in English) were identified [12–21]. To this list, we added two meta-analyses published as abstracts [22,23]. TABLE 2 lists the results of all 12 meta-analyses [12–23]. All were published within the past 5 years, and included a total of 513 diagnostic trials (with some overlap across meta-analyses). On average, each meta-analysis included 43 trials. There was great variation in the average number of patients or specimens in the individual trials, ranging from 42 to 493. Diagnostics included in these meta-analyses covered a wide spectrum: from smear microscopy to molecular-based tests, such as nucleic acid amplification tests (NATs). All meta-analyses conducted a quality assessment, although none covered all the QUADAS items.

TABLE 2 lists the quality elements that were most frequently reported in many of the meta-analyses. Only blinding was uniformly reported in all meta-analyses. On average, approximately 65% (range: 16–100%) of the trials used a prospective data collection design. However, only 33% (range: 0–53%) of the trials used a consecutive or random sampling method to recruit subjects. Approximately 72% (range: 61–85%) of the trials used a cross-sectional design, and the case–control approach was used in approximately a third of the studies. Any form of blinding was used in only 34% (range: 0–63%) of the trials. In most studies (94–100%), the index test results were verified by a reference standard test.

Sources of bias in diagnostic research

These data, although limited, do lend support to the concern that trials of TB diagnostics lack methodological rigor. Methodological problems, such as lack of a probabilistic patient sampling method, use of a case–control design and lack of blinding, appear to be fairly common in many of the trials. Of all the quality items reported, lack of blinding appears to be a frequently observed problem; only a third of all studies reported using any form of blinding. In addition, a sizeable

Table 1. The QUADAS tool, a validated quality assessment tool for diagnostic accuracy studies.

No.	QUADAS item	Potential bias if the criterion is not met
1	Was the spectrum of patients representative of the patients who will receive the test in practice?	Spectrum bias (commonly seen in case–control studies)
2	Were selection criteria clearly described?	Selection bias
3	Is the reference standard likely to correctly classify the target condition?	Misclassification of disease status due to an imperfect reference standard
4	Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Disease progression bias
5	Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?	Partial verification (work-up) bias
6	Did patients receive the same reference standard regardless of the index test result?	Differential verification (work-up) bias
7	Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?	Incorporation bias
8	Was the execution of the index test described in sufficient detail to permit replication of the test?	Source of variation, and likely to affect generalizability
9	Was the execution of the reference standard described in sufficient detail to permit its replication?	Source of variation, and likely to affect generalizability
10	Were the index test results interpreted without knowledge of the results of the reference standard?	Review bias due to lack of blinding
11	Were the reference standard results interpreted without knowledge of the results of the index test?	Review bias due to lack of blinding
12	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? For example, if clinical data (e.g., results of physical examination) will be available when the test is interpreted in clinical practice, then this should also be available when the test is evaluated.	Source of variation, and likely to affect generalizability
13	Were uninterpretable and/or intermediate test results reported?	Bias due to exclusion of uninterpretable and intermediate test results from the analyses
14	Were withdrawals from the study explained?	Bias due to differential loss to follow-up

Adapted with permission from [9].
QUADAS: Quality assessment of diagnostic accuracy studies.

Table 2. Methodological quality of studies on tuberculosis diagnostics in recently published meta-analyses.

Meta-analysis	No. of studies	Diagnostic test	Average size of each study	Prospective data collection (%)	Consecutive or random sampling of subjects (%)	Cross-sectional design (%)	Blinded interpretation of test results* (%)	Complete verification of index test results‡ (%)	Ref.
Sarmiento *et al.* (2003)	16	PCR on respiratory specimens for smear-negative pulmonary TB	NR	50	NR	NR	63	100	[12]
Goto *et al.* (2003)	40	ADA for TB pleural effusion	137	NR	NR	NR	0	NR	[13]
Pai *et al.* (2003)	49	NAT for TB meningitis	42	61	49	61	59	94	[14]
Greco *et al.* (2003)	44	ADA and IFN-γ tests for TB pleural effusion	135	NR	NR	NR	9	NR	[15]
Pai *et al.* (2004)	40	NAT for TB pleural effusion	60	63	53	70	55	100	[16]
Flores *et al.* (2005)	84	In-house PCR for pulmonary TB	149	NR	NR	71	34	NR	[17]
Kalantri *et al.* (2005)	13	Phage amplification tests for pulmonary TB	448	NR	NR	85	23	100	[18]
Pai *et al.* (2005)	21	Phage-based tests for rifampin resistance	85	NR	38	NR	57	100	[19]
Morgan *et al.* (2005)	15	Line probe assay for rifampin resistance	91	NR	0	NR	13	100	[20]
Greco *et al.* (2006)	63	Commercial NAT for pulmonary TB	410	16	32	NR	16	NR	[21]
Steingart *et al.* (2006)	45	Fluorescence versus conventional sputum smear microscopy for pulmonary TB	493	100	36	NR	49	NR	[22]
Steingart *et al.* (2006)	83	Direct versus concentrated sputum smear microscopy for pulmonary TB	256	100	21	NR	31	NR	[23]

*At least single blind. ‡By reference standard.
ADA: Adenosine deaminase; IFN: Interferon; NAT: Nucleic acid amplification test; NR: Not reported; TB: Tuberculosis.

proportion of the trials used retrospective methods (e.g., they used routinely collected laboratory data); missing data are a major concern with such analyses.

Do design flaws actually lead to biased estimates of accuracy? There is empirical evidence that methodological flaws can produce misleading estimates of diagnostic accuracy [8,24]. A recent, large, empirical study of 31 meta-analyses (with 487 primary diagnostic studies on a variety of diseases) found significantly higher estimates of diagnostic accuracy in studies with non-consecutive inclusion of patients and retrospective data collection. The estimates were highest in studies that compared severe cases and healthy controls [24].

Selection of appropriate control groups is a key issue in diagnostic trial design [8,9,24]. In case–control studies, researchers often recruit bacteriologically confirmed TB cases (often those with clear-cut, advanced smear-positive disease) and healthy

controls. This approach has been termed a two-gate design, because cases and controls are not sampled from the same study base; instead, they are sampled from two different populations [25]. For the same index test, studies with the two-gate design using healthy controls have been shown to produce higher estimates of diagnostic accuracy compared with trials that recruit a cohort of consecutive patients (single-gate design) in whom the test is clinically indicated [8,24,25]. This is because the two-gate design with healthy controls results in the selection of subjects from the extreme ends of the clinical spectrum (spectrum bias) [8,24,25]. For example, meta-analyses on NAT for TB meningitis [14] and pleuritis [16] found that, on average, case–control studies produced twofold higher estimates of diagnostic odds ratios than cross-sectional studies.

With respect to patient sampling, it is worth emphasizing that it is not only the sampling strategy, but also the population sampled that can potentially bias a study. For example, hospital-based studies are particularly prone to selection biases, even if consecutive or random sampling methods are employed. Therefore, prospective, community-based studies are valuable, especially in endemic areas where the disease burden is high and the selection biases imposed by studying only hospitalized (usually severely ill) patients is far more pronounced [26]. Recruitment of hospital patients can skew the disease spectrum of TB cases recruited, and potentially inflate sensitivity. It also has important implications for control selection; hospitalized patients tend to have multiple comorbid conditions that can affect test specificity.

Selection of an appropriate reference standard is another key determinant of trial quality. A key assumption in diagnostic trials is that the reference standard is 100% accurate. If the reference standard is imperfect, this introduces methodological problems. In TB trials, mycobacterial culture is often used as the reference standard; however, culture does not detect all TB, and false-positive cultures are relatively common. Thus, some amount of misclassification of the disease status is inevitable. Some studies have used smear microscopy (a test with modest sensitivity) as the reference test. Using an insensitive reference standard can lead to biased estimation of test accuracy [7].

The choice of reference standard is particularly difficult with childhood, extrapulmonary and latent TB infection (LTBI). The diagnosis of childhood TB is complicated by the absence of a practical gold standard [26]. In children, bacteriological confirmation is rarely achieved. Therefore, researchers often create composite reference standards and scoring systems, to classify patients as 'definite TB', 'probable TB', 'possible TB' and 'no TB'. However, none of these scoring systems have been adequately validated in high-burden settings [26]. This dilemma underscores the need to develop and validate symptom and case definitions for common clinical conditions [26].

Extrapulmonary TB (e.g., TB meningitis, pleuritis and lymphadenitis) are difficult to confirm using bacteriology. For example, meta-analyses on TB pleural effusion included studies that used a variety of reference tests, including clinical diagnoses,

'There is evidence that authors often fail to report all the critical components of a diagnostic trial.'

histopathology and response to anti-TB therapy [13,15,16]. In some situations, especially when using NATs, investigators often performed discrepant analyses because they believed that NAT was potentially more sensitive than the reference standard [14,16,17]. However, discrepant analysis introduces its own bias [27].

The evaluation of new tools that detect LTBI has been complicated by the lack of a reference standard for LTBI. The tuberculin skin test (TST), the conventional test for LTBI, has known accuracy limitations. Therefore, the TST is not useful as the reference standard. Although new tests, such as interferon-γ release assays (IGRAs), have demonstrated promise, their true sensitivity and specificity for LTBI is unknown [4,28]. Researchers have had to use indirect approaches (e.g., correlations between test results and markers of exposure to *Mycobacterium tuberculosis*) to determine whether IGRAs are superior to the TST [28]. Overall, for conditions with no reference standard or where the reference standard is known to be imperfect, novel epidemiological and statistical approaches may be necessary to determine test accuracy.

Poor reporting versus poor methodological quality

Although the meta-analyses suggest that TB trials lack methodological rigor, it is important to acknowledge that meta-analyses have their own limitations, and some of the findings could be explained by poor reporting rather than poor methodology. Poor study quality pertains to methodological flaws that lead to biased results. Poor reporting, on the other hand, refers to incomplete or inadequate reporting of the design, conduct, analysis and results of a study [29]. A poorly reported study may be well designed and executed; however, it is impossible to determine this without contacting the authors for additional information that is lacking in the published work [30].

There is evidence that authors often fail to report all the critical components of a diagnostic trial [24,30]. For example, in a meta-analysis on NAT for TB meningitis, 74% of 49 studies did not report on whether the NAT results were interpreted blindly, without the knowledge of culture results [14]. When study authors were contacted, the proportion with missing information on blinding was reduced from 74 to 31% [14,30]. Apparently, some authors had incorporated blinding in their trial design, but had failed to explicitly report it in their publications. Several of the meta-analyses listed in TABLE 2 found that authors frequently failed to report the type of study design, disease spectrum (i.e., clinical severity) and demographics of patients recruited, study direction (prospective or retrospective), sampling method, blinding, and whether all patients underwent the reference standard, irrespective of the index test results.

What can be done to improve the quality & reporting of diagnostic research?

It is clear that efforts are needed to improve both methodological quality and reporting of diagnostic trials. To improve the quality of reporting of diagnostic studies, the Standards for Reporting of Diagnostic Accuracy (STARD) initiative was launched by an

international consortium of investigators [29]. The objective is to improve the quality of reporting, and to encourage authors to use a more standardized and transparent format for preparing manuscripts of diagnostic accuracy studies. Several leading journals now require authors to format diagnostic trial manuscripts using the STARD template. In addition, the QUADAS tool, although designed for the assessment of study quality, appears to be a useful tool to improve study design and the quality of reporting [9].

To address the need for stricter controls on the introduction and use of diagnostic tests in national public health programs, and to provide specific guidance to researchers who conduct diagnostic trials in infectious diseases, the United Nations Children's Fund (UNICEF)/United Nations Development Programme (UNDP)/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) assembled a Diagnostics Expert Evaluation Panel (DEEP) [104]. This panel was charged with developing best practice guidelines (to be published later this year) for assessing the performance and operational characteristics of diagnostics for infectious diseases [J CUNNINGHAM, TDR, PERS. COMM.]. Efforts are also underway to prepare DEEP guidelines that are specific for TB. Others have recognized the need for guidelines specific to TB diagnostic trials [6].

Compliance with Good Clinical Practice (GCP) and Good Clinical Laboratory Practice (GCLP) is another approach to improve trial quality. Although not specific to diagnostic research, GCP is a standard for the design, conduct, performance, monitoring, recording, analysis and reporting of clinical trials [105]. GCLP is a newly proposed quality system for laboratories that undertake the analyses of samples from clinical trials [106]. GCLP is increasingly being adopted as the laboratory standard of choice for clinical and diagnostic trials. However, it is unclear whether standards such as GCP and GCLP are feasible in low-income countries with limited resources.

Finally, efforts are needed to strengthen research and laboratory capacities in developing countries. Quality research is unlikely to be produced in poorly functioning, under-staffed and overworked laboratories and institutions. International agencies, such as the WHO [101], TDR [104], the Stop TB Partnership [103], the American Thoracic Society (ATS) [107], and the International Union Against Tuberculosis and Lung Disease

> 'Quality research is unlikely to be produced in poorly functioning, under-staffed and overworked laboratories and institutions.'

(IUATLD) [108], have a key role to play in the strengthening of research capacity in developing countries. In fact, strengthening of health systems and promotion of research are important components of the new Stop TB strategy [109] and the Global Plan to Stop TB 2006–2015 [102]. The TDR has supported the training of hundreds of researchers in several developing countries, and has created useful research resources, such as the TB Specimen Bank [104]. The ATS and the IUATLD conduct courses in many developing countries in order to provide training in TB epidemiology, clinical research methods, operational and health systems research. In addition, collaborations between academic institutions in high- and low-income countries can also contribute to research capacity building.

If diagnostics are marketed based solely on accuracy studies, low-income countries may waste their scarce resources on inappropriate technologies. To address this concern, the Foundation for Innovative New Diagnostics, a nonprofit agency devoted to the development of new tools for the diagnosis of neglected diseases [110], has facilitated demonstration projects to assess the feasibility, applicability and cost-effectiveness of new diagnostics in high-burden countries [3]. These projects aim to go beyond diagnostic accuracy and tackle issues in the implementation of new tools in programmatic settings.

In conclusion, high-quality diagnostic studies are critical to evaluate new tools, to develop evidence-based policies on TB diagnostics, and, ultimately, for effective control of the global TB epidemic. Lack of methodological rigor in TB trials is a cause for concern, as it may prove to be an important hurdle for effective application of diagnostics in TB control. Several parallel initiatives, including those described above, are required to provide the much needed impetus to improve the methodological quality and reporting of trials on TB diagnostics.

References

1. WHO. *Global Tuberculosis Control. Surveillance, Planning, Financing. WHO Report 2006. WHO/HTM/TB/2006.362.* WHO, Geneva, Switzerland 1–242 (2006).

2. Small PM, Perkins MD. More rigour needed in trials of new diagnostic agents for tuberculosis. *Lancet* 356(9235), 1048–1049 (2000).

3. Perkins MD, Roscigno G, Zumla A. Progress towards improved tuberculosis diagnostics for developing countries. *Lancet* 367(9514), 942–943 (2006).

4. Pai M, Kalantri S, Dheda K. New tools and emerging technologies for the diagnosis of tuberculosis: Part 1. Latent tuberculosis. *Expert Rev. Mol. Diagn.* 6(3), 413–422 (2006).

5. Pai M, Kalantri S, Dheda K. New tools and emerging technologies for the diagnosis of tuberculosis: Part 2. Active tuberculosis and drug resistance. *Expert Rev. Mol. Diagn.* 6(3), 423–432 (2006).

6. Walsh A, McNerney R. Guidelines for establishing trials of new tests to diagnose tuberculosis in endemic countries. *Int. J. Tuberc. Lung Dis.* 8(5), 609–613 (2004).

7. Whiting P, Rutjes AW, Reitsma JB *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann. Intern. Med.* 140(3), 189–202 (2004).

8   Lijmer JG, Mol BW, Heisterkamp S *et al.*
    Empirical evidence of design-related bias in
    studies of diagnostic tests. *JAMA* 282(11),
    1061–1066 (1999).

9   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. 3, 25 (2003).

10 Whiting PF, Weswood ME, Rutjes AW
    et al. Evaluation of QUADAS, a tool for the
    quality assessment of diagnostic accuracy
    studies. *BMC Med. Res. Methodol.* 6, 9
    (2006).

11 Pai M, McCulloch M, Enanoria W,
    Colford JM Jr. Systematic reviews of
    diagnostic test evaluations: what's behind
    the scenes? *ACP J. Club* 141(1), A11–A13
    (2004).

12 Sarmiento OL, Weigle KA, Alexander J,
    Weber DJ, Miller WC. Assessment by
    meta-analysis of PCR for diagnosis of
    smear-negative pulmonary tuberculosis.
    J. Clin. Microbiol. 41(7), 3233–3240
    (2003).

13 Goto M, Noguchi Y, Koyama H *et al.*
    Diagnostic value of adenosine deaminase
    in tuberculous pleural effusion:
    a meta-analysis. *Ann. Clin. Biochem.*
    40(Pt 4), 374–381 (2003).

14 Pai M, Flores LL, Pai N *et al.*
    Diagnostic accuracy of nucleic acid
    amplification tests for tuberculous
    meningitis: a systematic review and
    meta-analysis. *Lancet Infect. Dis.* 3(10),
    633–643 (2003).

15 Greco S, Girardi E, Masciangelo R,
    Capoccetta GB, Saltini C.
    Adenosine deaminase and interferon γ
    measurements for the diagnosis of
    tuberculous pleurisy: a meta-analysis.
    Int. J. Tuberc. Lung Dis. 7(8), 777–786
    (2003).

16 Pai M, Flores LL, Hubbard A, Riley LW,
    Colford JM Jr. Nucleic acid amplification
    tests in the diagnosis of tuberculous
    pleuritis: a systematic review and
    meta-analysis. *BMC Infect. Dis.* 4(1), 6
    (2004).

17 Flores LL, Pai M, Colford JM Jr, Riley LW.
    In-house nucleic acid amplification tests for
    the detection of *Mycobacterium tuberculosis*
    in sputum specimens: meta-analysis and
    meta-regression. *BMC Microbiol.* 5, 55
    (2005).

18 Kalantri S, Pai M, Pascopella L, Riley L,
    Reingold A. Bacteriophage-based tests
    for the detection of *Mycobacterium*

    tuberculosis in clinical specimens:
    a systematic review and meta-analysis.
    BMC Infect. Dis. 5(1), 59 (2005).

19 Pai M, Kalantri S, Pascopella L, Riley LW,
    Reingold AL. Bacteriophage-based assays
    for the rapid detection of rifampicin
    resistance in *Mycobacterium tuberculosis*:
    a meta-analysis. *J. Infect.* 51(3), 175–187
    (2005).

20 Morgan M, Kalantri S, Flores L, Pai M.
    A commercial line probe assay for the rapid
    detection of rifampicin resistance in
    Mycobacterium tuberculosis: a systematic
    review and meta-analysis. *BMC Infect. Dis.*
    5, 62 (2005).

21 Greco S, Girardi E, Navarra S, Saltini C.
    The current evidence on diagnostic
    accuracy of commercial based nucleic acid
    amplification tests for the diagnosis of
    pulmonary tuberculosis. *Thorax* (2006)
    (In Press).

22 Steingart K, Henry M, Ng V *et al.*
    Fluorescence versus conventional sputum
    smear microscopy for tuberculosis:
    a systematic review. Presented at:
    *10th Annual Conference of the International
    Union Against Tuberculosis and Lung Disease
    (North American Region).* IL, USA,
    March 2–4 (2006).

23 Steingart KR, Ng V, Henry MC *et al.*
    Sputum processing methods to improve
    the sensitivity of smear microscopy for
    tuberculosis: a systematic review.
    Presented at: *10th Annual Conference of the
    International Union Against Tuberculosis and
    Lung Disease (North American Region).*
    IL, USA, March 2–4 (2006).

24 Rutjes AW, Reitsma JB, Di Nisio M *et al.*
    Evidence of bias and variation in diagnostic
    accuracy studies. *CMAJ* 174(4), 469–476
    (2006).

25 Rutjes AW, Reitsma JB, Vandenbroucke JP,
    Glas AS, Bossuyt PM. Case–control and two-
    gate designs in diagnostic accuracy studies.
    Clin. Chem. 51(8), 1335–1341 (2005).

26 Marais BJ, Gie RP, Schaaf HS *et al.*
    Childhood pulmonary tuberculosis: old
    wisdom and new challenges. *Am. J. Respir.
    Crit. Care Med.* 173(10), 1078–1090 (2006).

27 Hadgu A, Dendukuri N, Hilden J.
    Evaluation of nucleic acid amplification
    tests in the absence of a perfect
    gold-standard test: a review of the statistical
    and epidemiologic issues. *Epidemiology*
    16(5), 604–612 (2005).

28 Pai M, Riley LW, Colford JM Jr.
    Interferon-γ assays in the immunodiagnosis
    of tuberculosis: a systematic review.
    Lancet Infect. Dis. 4(12), 761–776 (2004).

29 Bossuyt PM, Reitsma JB, Bruns DE *et al.*
    Towards complete and accurate reporting of
    studies of diagnostic accuracy: the STARD
    Initiative. *Ann. Intern. Med.* 138(1), 40–44
    (2003).

30 Pai M, Flores LL, Hubbard A, Riley LW,
    Colford JM Jr. Quality assessment in
    meta-analyses of diagnostic studies:
    what difference does email contact with
    authors make? Presented at: *The XI
    Cochrane Colloquium*, Barcelona, Spain,
    October 26–31 (2003).

Websites

101 WHO: Stop TB Department
     www.who.int/tb

102 Global Plan to Stop TB 2006–2015
     www.stoptb.org/globalplan

103 Stop TB Partnership
     www.stoptb.org

104 UNICEF/UNDP/World Bank/WHO
     Special Programme for Research and
     Training in Tropical Diseases
     www.who.int/tdr

105 ICH Harmonised Tripartite Guideline:
     Guideline For Good Clinical Practice E6(R1)
     www.ich.org/LOB/media/MEDIA482.pdf

106 Good Clinical Laboratory Practice
     www.barqa.com/cms.php?pageid=645

107 American Thoracic Society
     www.thoracic.org

108 International Union Against Tuberculosis
     and Lung Disease
     www.iuatld.org

109 Stop TB Strategy
     www.who.int/tb/features_archive/stop_tb_
     strategy/en/index.html

110 Foundation for Innovative
     New Diagnostics
     www.finddiagnostics.org

Affiliations

•   *Madhukar Pai, MD, PhD
    Assistant Professor, McGill University,
    Department of Epidemiology, Biostatistics
    & Occupational Health, 1020 Pine Avenue
    West, Montreal, H3A 1A2, Canada
    Tel.: +1 514 398 5422
    Fax: +1 514 398 4503
    madhukar.pai@mcgill.ca*

•   *Richard O'Brien, MD
    Head, Product Evaluation, Foundation for
    Innovative New Diagnostics, 71,
    Av. Louis Casai, Case Postale 93,
    1216 Cointrin/Geneva, Switzerland
    Tel.: +41 22 710 0595
    Fax: +41 22 710 0599
    rick.obrien@finddiagnostics.org*