

# Systematic review of interferon-gamma release assays in tuberculosis: focus on likelihood ratios

Kwok Chiu Chang, Chi Chiu Leung

► Supplementary tables are published online only at <http://thx.bmj.com/content/vol65/issue3>.

Tuberculosis and Chest Service, Department of Health, Hong Kong SAR, China

## Correspondence to

Dr Kwok Chiu Chang, Wanchai Chest Clinic, 1st Floor, Wanchai Polyclinic, 99 Kennedy Road, Wanchai, Hong Kong, China; [kc\\_chang@dh.gov.hk](mailto:kc_chang@dh.gov.hk)

Received 7 September 2009  
Accepted 23 November 2009

## ABSTRACT

**Background** Clinical roles of QuantiFERON-TB Gold (QFT-G)/Gold in-Tube (QFT-G-IT) and T-SPOT.TB in tuberculosis require clarification.

**Methods** MEDLINE and EMBASE were searched for relevant English papers. Summary estimates of likelihood ratios (LR) of QFT-G/QFT-G-IT and T-SPOT.TB for latent tuberculosis infection (LTBI) and tuberculosis disease in adults were obtained by bivariate and univariate random effects meta-analyses after assessing heterogeneity. Probable ranges of prevalence for LTBI and tuberculosis disease were estimated. Critical values of positive LR (PLR) and negative LR (NLR) corresponding to a 90% certainty threshold were calculated over probable prevalence ranges. It was considered reliable to rule in when the best estimated PLR exceeds the corresponding critical value and to rule out when the best estimated NLR is less than the corresponding critical value.

**Results** 35 studies involving predominantly immunocompetent adults were identified. Based on bivariate meta-analysis, PLR (95% CI) for LTBI were 7.9 (3.6 to 17.3) for T-SPOT.TB and 48.1 (19.7 to 117.6) and 10.8 (5.3 to 21.8) for QFT-G/QFT-G-IT based on Japanese and other studies, respectively. Corresponding NLR (95% CI) were 0.10 (0.06 to 0.18), 0.11 (0.07 to 0.18) and 0.23 (0.16 to 0.32). PLR (95% CI) for tuberculosis disease were 3.6 (2.3 to 5.6) for QFT-G, 2.1 (1.1 to 4.0) for QFT-G-IT and 4.7 (2.4 to 9.1) and 2.3 (1.3 to 4.0) for T-SPOT.TB based on studies with mean or median age >47.1 years and ≤47.1 years, respectively. Corresponding NLR (95% CI) were 0.18 (0.12 to 0.27), 0.38 (0.22 to 0.68), 0.11 (0.06 to 0.20) and 0.20 (0.10 to 0.40). Estimated prevalence ranges were 10–55% for LTBI and 40–60% for tuberculosis disease.

**Conclusions** At a 90% certainty threshold, LTBI is best diagnosed by QFT-G/QFT-G-IT and excluded by T-SPOT.TB or QFT-G/QFT-G-IT; none can diagnose tuberculosis disease, whereas, T-SPOT.TB can exclude tuberculosis disease among middle-aged and older patients.

Tuberculosis (TB) is an ancient disease that has re-emerged as a major public health concern. While rapid diagnosis and treatment of patients with infectious TB remain the cornerstone in TB control,<sup>1</sup> targeted screening and treatment of high-risk subjects with latent TB infection (LTBI) has been recognised as an important control measure, especially in low-burden countries.<sup>2–3</sup> Unlike many viral diseases, the diagnosis of TB disease still relies heavily on the isolation of *Mycobacterium tuberculosis* complex in culture, which is often achievable in only 60% of cases.<sup>4</sup> The tuberculin skin test (TST), which has been the gold standard for the diagnosis of LTBI until recent years, is cross-reactive to Bacille

Calmette Guérin (BCG) and many non-tuberculous mycobacteria. While a high cut-off value such as 15 mm may enhance the specificity of the TST at the expense of sensitivity, the latter remains suboptimal regardless of the cut-off values among infants, young children, elderly subjects, severely malnourished subjects and those who are immunocompromised.<sup>2–5–6</sup>

With advances in mycobacteriology, *M tuberculosis*-specific region of difference 1 antigens such as culture filtrate protein 10 and early secretory antigen target 6 were discovered. Absent in BCG and most environmental mycobacteria, these antigens form the basis for interferon-gamma release assays (IGRA), which assess the presence of TB infection by detecting the in vitro release of interferon-gamma upon stimulation from previously sensitised T cells. Commercially available IGRA formats include QuantiFERON-TB Gold (QFT-G), QuantiFERON-TB Gold in-Tube (QFT-G-IT) (Cellestis Ltd, Carnegie, Victoria, Australia) and the T-SPOT.TB test (Oxford Immunotec, Oxford, UK).

The clinical roles of QFT-G, QFT-G-IT and T-SPOT.TB in TB require clarification. A number of systematic reviews have examined the test characteristics of QFT-G, QFT-G-IT and T-SPOT.TB with a focus on pooled estimates of sensitivity and specificity rather than the actual predictive value of a positive or negative test.<sup>7–9</sup> The positive predictive value (PPV) or negative predictive value (NPV), defined as respective proportions of true positive results among test-positive subjects and true negative results among test-negative subjects, are highly dependent on prevalence and can be estimated only in appropriate cross-sectional studies. As such, predictive values cannot be effectively combined across different settings in meta-analyses. Summary estimates of sensitivity and specificity obtained by separate pooling cannot be used for estimating predictive values owing to lack of consideration for variations between studies. This is possible with likelihood ratios (LR) that incorporate both sensitivity and specificity of the same study. The positive likelihood ratio (PLR) tells how much the odds of a condition are increased by a positive test, while the negative likelihood ratio (NLR) tells how much they are decreased by a negative test. The post-test odds (the odds of a condition after applying the test) equal LR multiplied by the pre-test odds (the odds of a condition before applying the test). For example, a PLR of 10 gives post-test odds that are 10 times the pre-test odds, whereas an NLR of 0.1 gives post-test odds that are 0.1 times the pre-test odds. It has been proposed that LR should not be pooled directly in systematic reviews.<sup>10</sup> The hierarchical SROC

model<sup>11</sup> and the bivariate random effects model<sup>12</sup> have been proposed to tackle the above problem.

An evaluation of the clinical roles of IGRA for TB would be incomplete without considering both LTBI and active TB disease. A distinction must be made between these two conditions as active TB disease develops among only one-tenth of immuno-competent subjects infected with the tubercle bacillus after an incubation period that ranges from a few weeks to a few decades.<sup>2</sup> The LTBI state is completely free of clinical manifestations, and there is no golden standard for its diagnosis. Bacteriologically-confirmed cases of TB are often used as surrogates for LTBI in the assessment of sensitivity of a diagnostic test for LTBI. A very low prevalence of LTBI is expected among unexposed subjects in an area with a low incidence of active TB disease (and hence low background transmission risk). Healthy controls without contact exposure or other risk factors for TB are therefore used as surrogates for the absence of LTBI to assess the specificity of such a test. For targeted screening of LTBI, the focus is to rule in LTBI with reasonable certainty among subjects at a high risk of progressing to active disease. IGRA is sometimes used as an adjunctive test for the diagnosis of active TB disease because infection must precede active disease. To assess the sensitivity and specificity of a test for active TB disease among patients with clinical manifestations, confirmed cases of active TB disease (by bacteriology, histology and/or appropriate response to treatment) are used to estimate sensitivity, whereas cases with unconfirmed or alternative diagnoses are used to estimate specificity. For TB disease a delay in treatment may be complicated by considerable morbidity or mortality. A diagnostic test with high NPV will therefore be required to rule out the condition.

The primary objective of the current review is to clarify the clinical roles of IGRA in LTBI and TB disease by focusing on summary measures of LR generated by bivariate random effects models. The secondary objective is to examine the feasibility of arriving at similar conclusions by pooling LR using univariate random effects models.

## METHODS

A literature search was performed through the OvidSP platform to browse MEDLINE, EMBASE and other non-indexed citations to 11 July 2009 for non-review non-animal English papers by combining (using the Boolean operator “and”) articles identified by three search phrases containing Medical Subject Headings or key words in titles or abstracts for: (1) interferon-gamma or QuantiFERON or ELISPOT or TSPOT.TB; (2) tuberculosis; and (3) cut-off values or sensitivity coexisting with specificity. The literature search was supplemented by relevant studies from a recent systematic review.<sup>9</sup> Only studies with concurrent data on sensitivity and specificity of QFT-G, QFT-G-IT, T-SPOT.TB and pre-commercial versions of these IGRA using similar criteria for interpretations were included. All studies examined the sensitivity of IGRA among subjects with TB disease. Studies with fewer than five subjects for evaluating either sensitivity or specificity were excluded. Indeterminate results and data on IGRA applied in body fluids other than blood were excluded from analysis. One reviewer abstracted data and the other double-checked the data.

Before meta-analysis, sources of heterogeneity within data grouped by QFT-G/QFT-G-IT, T-SPOT.TB and TST at respective cut-off values were examined by unweighted meta-regression analysis using the Moses–Shapiro–Littenberg method, which regressed the log diagnostic OR against a measure of diagnostic threshold, to identify heterogeneous subgroups.<sup>13–16</sup> Significant heterogeneity for a covariate was considered present when

$p \leq 0.05$ . The following covariates were considered: country of origin (for countries contributing at least two sets of data), estimated TB incidence, proportion of culture-proven cases, proportion of co-morbidity (among cases only when controls were healthy low-risk subjects), proportion of HIV infection (among cases only when controls were healthy low-risk subjects), mean or median age (separately for cases and controls when controls were healthy low-risk subjects), proportion of men (separately for cases and controls when controls were healthy low-risk subjects) and, if applicable, QFT-G versus QFT-G-IT or TST with cut-off at 10 mm versus 15 mm. Except for the country of origin, all covariates were rated as either above the median value or not.

Summary estimates of PLR and NLR were generated by two different methods of meta-analysis in the presence of at least four sets of data grouped by QFT-G/QFT-G-IT, T-SPOT.TB and TST at specified cut-off values: (1) the bivariate random effects model using SAS proc mixed procedure<sup>12</sup>; and (2) pooling by a univariate random effects model using the DerSimonian–Laird method in the absence of significant threshold effect,<sup>17</sup> which was assessed by the Spearman correlation coefficient between sensitivity and specificity and denoted as significant by  $p$  values  $\leq 0.05$ .

The most probable range of pre-test odds was determined by reference to a recent systematic review<sup>18</sup> and a few early studies<sup>19–21</sup> for LTBI, and from relevant studies included in the current review for TB disease. Critical values of PLR ( $PLR_{crit}$ ) and NLR ( $NLR_{crit}$ ) corresponding to PPV and NPV of 90%, respectively, could be calculated over the probable range of prevalence by the following equations: (1) pre-test odds = prevalence/(1–prevalence); (2)  $PPV = PLR_{crit} \times \text{pre-test odds} / (1 + PLR_{crit} \times \text{pre-test odds})$ , and (3)  $NPV = 1 - (NLR_{crit} \times \text{pre-test odds} / (1 + NLR_{crit} \times \text{pre-test odds}))$ . At a 90% certainty threshold, it was considered reliable to rule in when the best-estimated PLR exceeds the critical value and to rule out when the best-estimated NLR is less than the corresponding critical value.

Funnel plot asymmetry was examined by a regression of the natural log diagnostic odds ratio against standard error<sup>22</sup> separately for QFT-G/QFT-G-IT, T-SPOT.TB and TST in the respective context of LTBI and TB disease. Significant asymmetry is denoted by  $p \leq 0.05$ .

MetaDiSc V.1.4,<sup>23</sup> SAS Enterprise Guide 3.0, SSPS V.10 and OpenOffice.org 3.0 were used for statistical analysis.

## RESULTS

### Studies included

A total of 218 articles were identified after adding two articles identified only by a recent systematic review<sup>9</sup> but not by the literature search through the OvidSP platform. A total of 35 adult studies (see tables E1 and E2 in online supplement) were included in the current review after excluding 183 articles for the following reasons: irrelevant ( $n=95$ ), no data on both sensitivity and specificity ( $n=46$ ), no data on specificity ( $n=24$ ), no data on sensitivity ( $n=14$ ), no data for sensitivity for one test plus no data for specificity for another test ( $n=1$ ) and paediatric study ( $n=3$ ). The majority involved subjects that were predominantly BCG-vaccinated. Among 24 studies involving non-TB patients as controls (E1, E4, E5, E13, E16, E19–E37), the proportion of HIV infection exceeded 20% among either cases or controls in three (13%) studies (E21, E24, E28).

### Adult studies on LTBI

Among adult studies using healthy low-risk subjects as controls, the following data were available for different tests: 16 sets of data for QFT-G (E1–E12)/QFT-G-IT (E3, E6, E13, E14), six sets for

**Table 1** Summary estimates of test characteristics of interferon-gamma release assays and tuberculin skin tests based on adult studies using healthy low-risk subjects as controls

Tests	Sets of data	Cases	Controls	Method of meta-analysis	Mean sensitivity % (95% CI)	Mean specificity % (95% CI)	Mean PLR (95% CI)	Mean NLR (95% CI)	Mean DOR (95% CI)
QFT-G (E4–E7, E12)/ QFT-G-IT (E6) (based on Japanese studies)	6	443	718	Bivariate random effects model	89 (84 to 93)	98 (96 to 99)	48.1 (19.7 to 117.6)	0.11 (0.07 to 0.18)	436.5 (159.5 to 1194.8)
QFT-G (E1–E3, E8–11)/ QFT-G-IT (E3, E13, E14) (based on other studies)	10	433	498	Univariate random effects model Bivariate random effects model	89 (85 to 93) 79 (72 to 84)	98 (97 to 99) 93 (86 to 96)	43.4 (24.8 to 75.9) 10.8 (5.3 to 21.8)	0.11 (0.08 to 0.16) 0.23 (0.16 to 0.32)	460.0 (237.8 to 889.7) 47.6 (21.5 to 105.3)
T-SPOT.TB (E7, E8, E11, E12, E15, E16)	6	364	363	Univariate random effects model Bivariate random effects model	79 (75 to 84) 91 (85 to 94)	90 (84 to 96) 88 (78 to 94)	11.6 (5.3 to 25.6) 7.9 (3.6 to 17.3)	0.24 (0.18 to 0.31) 0.10 (0.06 to 0.18)	46.9 (23.2 to 94.9) 75.6 (29.0 to 197.3)
TST-BCG (5 mm) (E4, E8, E11, E12)	4	282	341	Univariate random effects model Bivariate random effects model	92 (86 to 97) 74 (64 to 83)	87 (83 to 92) 40 (22 to 60)	6.7 (4.8 to 9.4) 1.2 (0.6 to 2.4)	0.09 (0.04 to 0.19) 0.65 (0.33 to 1.29)	116.7 (41.6 to 327.0) 1.9 (0.7 to 5.0)
TST-BCG (10 or 15 mm) (E5, E8, E10, E11)	7	474	604	Univariate random effects model Bivariate random effects model Univariate random effects model	75 (63 to 86) 66 (56 to 74) 64 (52 to 75)	43 (16 to 69) 70 (55 to 81) 71 (58 to 85)	1.3 (0.8 to 2.2) 2.2 (1.1 to 4.1) 2.2 (1.4 to 3.4)	0.70 (0.28 to 1.75) 0.49 (0.32 to 0.76) 0.51 (0.36 to 0.73)	1.9 (0.4 to 8.5) 4.4 (2.0 to 9.4) 4.6 (2.0 to 10.4)

DOR, diagnostic odds ratio; NLR, negative likelihood ratio; NA, not applicable; PLR, positive likelihood ratio; QFT-G, QuantiFERON-TB Gold; QFT-G-IT, QuantiFERON-TB Gold in-Tube; TST-BCG (x mm), tuberculin skin test with cut-off at x mm among predominantly BCG-vaccinated subjects.

T-SPOT.TB (E7, E8, E11, E12, E15, E16), four sets for TST with cut-off at 5 mm (E4, E8, E11, E12), seven sets for TST with cut-off at 10 mm (E5, E8, E10, E11) or 15 mm (E8, E10, E11). Table E1 summarises major findings of included adult studies using healthy low-risk subjects as controls. Unweighted meta-regression analysis showed no significant heterogeneity except for Japanese versus other studies within data grouped by QFT-G/QFT-G-IT.

### Adult studies on TB disease

Among adult studies using non-TB patients as controls, the following data were available for different tests: 16 sets of data for QFT-G (E1, E4, E5, E19–E26)/QFT-G-IT (E13, E27–E30), 12 sets for T-SPOT.TB (E16, E20, E21, E27, E31–E37) and six sets for TST using 10 mm as the cut-off (E5, E26, E34, E35, E37). Table E2 in the online supplement summarises the major findings of included adult studies using non-TB patients as controls. Data on TST were obtained from predominantly BCG-vaccinated subjects. Non-TB diseases that might mimic TB disease included tumours, bronchiectasis, congestive heart failure, sarcoidosis and infection due to non-tuberculous mycobacteria and other pathogens. Unweighted meta-regression analysis showed no significant heterogeneity except for QFT-G versus QFT-G-IT within data grouped by QFT-G/QFT-G-IT, and mean or median age >47.1 years versus ≤47.1 years within data grouped by T-SPOT.TB.

### Summary estimates from meta-analysis

Tables 1 and 2 show summary estimates of test characteristics of IGRA and TST based on adult studies using healthy low-risk subjects and non-TB patients as controls, respectively. Both methods of meta-analysis gave similar best-estimated LR. Tables E3 and E4 in the online supplement compare different diagnostic tools by sensitivity, specificity and diagnostic odds ratios using the bivariate random effects model.

### Probable range of pre-test odds

Based on a recent systematic review<sup>18</sup> which showed that the prevalence of LTBI controlled for overdispersion among close contacts during TB contact investigation in low- and middle-income countries was 51.4% (95% CI 50.6% to 52.2%) and a few early studies,<sup>19–21</sup> it was assumed that the prevalence of LTBI in the context of TB contact investigation probably ranged from 10% to 55%. The corresponding range of pre-test odds would be 1/9 to 11/9. Based on all studies (n=24) used in the current review (E1, E4, E5, E13, E16, E19–E37), the pooled estimate of prevalence of TB disease corrected for over-dispersion was 49% (95% CI 43% to 56%). Thus, it was assumed that the prevalence of TB disease among patients with clinical manifestations probably ranged from 40% to 60%. The corresponding range of pre-test odds would be 2/3 to 3/2.

### Tests with ≥90% certainty for LTBI and TB disease

Tables 3 and 4 show that, under the most probable range of pre-test odds, LTBI among largely immunocompetent adults is best ruled in by QFT-G/QFT-G-IT (based on Japanese studies) and ruled out by T-SPOT.TB or QFT-G/QFT-G-IT (based on Japanese studies) at a 90% certainty threshold.

Over the probable prevalence range for TB disease (40–60%), critical values of PLR corresponding to PPV of at least 90% would lie in the range of 6.0–13.5. Table 2 shows that, under the probable prevalence range, neither IGRA evaluated (QFT-G, QFT-G-IT, T-SPOT.TB) nor TST (cut-off at 10 mm) is suitable for ruling in TB diseases at a 90% certainty threshold. On the other hand, T-SPOT.TB can rule out TB disease among

**Table 2** Summary estimates of test characteristics of interferon-gamma release assays and tuberculin skin tests based on adult studies using non-tuberculosis patients as controls

Tests	Sets of data	Cases	Controls	Method of meta-analysis	Mean sensitivity % (95% CI)	Mean specificity % (95% CI)	Mean PLR (95% CI)	Mean NLR (95% CI)	Mean DOR (95% CI)
QFT-G (E1, E4, E5, E19–E26)	11	551	679	Bivariate random effects model	86 (81 to 90)	76 (66 to 83)	3.6 (2.3 to 5.6)	0.18 (0.12 to 0.27)	19.6 (10.7 to 36.2)
QFT-G-IT (E13, E27–E30)	5	307	266	Univariate random effects model	86 (82 to 91)	77 (70 to 85)	3.6 (2.5 to 5.1)	0.19 (0.13 to 0.28)	21.3 (11.4 to 39.8)
				Bivariate random effects model	75 (64 to 84)	64 (47 to 78)	2.1 (1.1 to 4.0)	0.38 (0.22 to 0.68)	5.5 (2.3 to 13.1)
T-SPOT.TB (based on studies with mean or median age >47.1 years) (E16, E31, E34, E36, E37)	6	236	194	Univariate random effects model	75 (64 to 86)	66 (55 to 78)	2.0 (1.3 to 3.1)	0.38 (0.23 to 0.63)	5.7 (2.6 to 12.3)
				Bivariate random effects model	91 (85 to 95)	80 (67 to 89)	4.7 (2.4 to 9.1)	0.11 (0.06 to 0.20)	43.9 (17.2 to 111.9)
T-SPOT.TB (based on studies with mean or median age ≤47.1 years) (E20, E21, E27, E32, E33, E35)	6	179	218	Univariate random effects model	93 (88 to 98)	77 (67 to 87)	4.4 (2.7 to 7.4)	0.11 (0.06 to 0.20)	52.0 (26.1 to 103.6)
				Bivariate random effects model	87 (79 to 93)	61 (46 to 75)	2.3 (1.3 to 4.0)	0.20 (0.10 to 0.40)	11.1 (4.6 to 26.8)
TST-BCG (10 mm) (E5, E26, E34, E35, E37)	6	193	280	Univariate random effects model	87 (85 to 90)	57 (47 to 66)	2.2 (1.7 to 2.9)	0.24 (0.16 to 0.36)	10.0 (5.2 to 19.3)
				Bivariate random effects model	57 (44 to 68)	68 (54 to 80)	1.8 (1.0 to 3.3)	0.63 (0.39 to 1.03)	2.8 (1.3 to 6.2)
				Univariate random effects model	59 (48 to 69)	62 (46 to 79)	1.8 (1.2 to 2.7)	0.66 (0.54 to 0.80)	2.8 (1.6 to 5.2)

DOR, diagnostic odds ratio; NLR, negative likelihood ratio; NA, not applicable; PLR, positive likelihood ratio; QFT-G, Quantiferon-TB Gold; QFT-G-IT, Quantiferon-TB Gold in-Tube; TST-BCG (x mm), tuberculin skin test with cut-off at x mm among predominantly BCG-vaccinated subjects.

immunocompetent middle-aged and older adults at a 90% certainty threshold (table 4). Table E4 in the online supplement shows that better NPV of T-SPOT.TB for TB disease among older patients with clinical manifestations can be attributed to higher specificity rather than higher sensitivity.

### Funnel plot asymmetry

Among studies using healthy low-risk controls, regression analysis showed significant funnel plot asymmetry for studies of QFT-G/QFT-G-IT among non-Japanese studies, but not for studies of QFT-G/QFT-G-IT among Japanese studies, studies of T-SPOT.TB and those of TST with cut-off at 5 mm, and at 10 mm or 15 mm.

Among studies using non-TB patients as controls, regression analysis showed no significant funnel plot asymmetry for studies of QFT-G/QFT-G-IT after controlling for QFT-G versus QFT-G-IT, studies of T-SPOT.TB after controlling for age groups, and those of TST with cut-off at 10 mm.

### DISCUSSION

Besides updating the test performance of IGRA with recent publications, the current review is unique by its focus on summary estimates of LR, rather than those of sensitivity and specificity, over the most probable range of pre-test odds for LTBI and TB disease. The current review has further clarified the clinical roles of IGRA and demonstrated no significant difference between univariate random effects models using the DerSimonian–Laird method and bivariate random effects models as far as best-estimated values of LR are concerned.

### Significant heterogeneity between Japanese and non-Japanese studies

The current review demonstrated significant heterogeneity in the test performance of QFT-G/QFT-G-IT for LTBI between Japanese and non-Japanese adult studies. Unweighted meta-regression analysis failed to identify any other independent confounding factor; this is corroborated by a comparison of estimates of sensitivity and specificity using the bivariate random effects model (see table E3 in online supplement). The difference might be partly explained by publication or other forms of selection bias, as significant asymmetry was found in the funnel plot for non-Japanese studies but not Japanese studies. Other sources of variations might include measurement errors, interpretation bias and genetic differences. Further investigations are warranted to explore the reason(s) underlying the observed heterogeneity.

### Method of meta-analysis

It has been proposed that LR should not be directly pooled in systematic reviews as pooling may result in values of sensitivity and specificity of >1 or <0.<sup>10</sup> Notwithstanding often wider CI for the bivariate random effects model, the current review showed that best-estimated LR obtained by a univariate random effects model using the DerSimonian–Laird method closely approximated those obtained by the bivariate random effects model. This implies that diagnostic test evaluation by meta-analysis with clinical focus on best-estimated LR might be accomplished by simpler pooling methods. However, a more thorough comparison of sensitivity, specificity and diagnostic odds ratio between different assays would probably require bivariate meta-analysis using software such as SAS proc mix procedure (see tables E3 and E4 in online supplement).



**Table 3** Tests with best-estimated positive predictive values of at least 90% for latent tuberculosis infection among adults over the probable range of prevalence

Prevalence (%)	Pre-test odds	Critical PLR*	Bivariate random effects model		Univariate random effects model
			Test, PLR (95% CI)		Test, PLR (95% CI)
10–20	1/9–1/4	81.0–36.0	QFT-G/QFT-G-IT (Japan) 48.1 (19.7 to 117.6)		QFT-G/QFT-G-IT (Japan) 43.4 (24.8 to 75.9)
20–30	1/4–3/7	36.0–21.0	—		—
30–40	3/7–2/3	21.0–13.5	—		—
40–50	2/3–1/1	13.5–9.0	QFT-G/QFT-G-IT (non-Japan) 10.8 (5.3 to 21.8)		QFT-G/QFT-G-IT (non-Japan) 11.6 (5.3 to 25.6)
50–55	1.0–11/9	9.0–7.4	T-SPOT.TB 7.9 (3.6 to 17.3)		

Probable pre-test odds range from 1/9 to 11/9 for latent tuberculosis infection during tuberculosis contact investigation. Japan and non-Japan in brackets denote results based on Japanese and non-Japanese studies, respectively. Tests that are suitable in settings with lower pre-test odds also work in settings with higher pre-test odds for better positive predictive values. TST with cut-off at 5 mm and 10 mm or 15 mm among predominantly BCG-vaccinated subjects are not included under tests for LTBI as their best-estimated PLR (95% CI) by bivariate meta-analysis are all below critical PLR given probable pre-test odds: cut-off at 5 mm, 1.2 (0.6 to 2.4); cut-off at 10 or 15 mm, 2.2 (1.1 to 4.1).

\*Corresponding to positive predictive values of at least 90% (see mathematical equations in the Methods section).

LTBI, latent tuberculosis infection; PLR, positive likelihood ratio; NLR, negative likelihood ratio; QFT-G, QuantiFERON-TB Gold; QFT-G-IT, QuantiFERON-TB Gold in-Tube; TB, tuberculosis.

**Table 4** Tests with best-estimated negative predictive values of at least 90% for latent tuberculosis infection or tuberculosis disease among adults over the probable range of prevalence

Prevalence (%)	Pre-test odds	Critical NLR*	LTBI among adults		TB disease among adults	
			Bivariate random effects model Test, NLR (95% CI)	Univariate random effects model Test, NLR (95% CI)	Bivariate random effects model Test, NLR (95% CI)	Univariate random effects model Test, NLR (95% CI)
60–50	3/2–1/1	0.07–0.11	T-SPOT.TB 0.10 (0.06 to 0.18)	T-SPOT.TB 0.09 (0.04 to 0.19)	T-SPOT.TB based on studies with mean or median age >47.1 years 0.11 (0.06 to 0.20)	T-SPOT.TB based on studies with mean or median age >47.1 years 0.11 (0.06 to 0.20)
50–40	1/1–2/3	0.11–0.17	—	—	—	—
40–30	2/3–3/7	0.17–0.26	QFT-G/QFT-G-IT (non-Japan) 0.23 (0.16 to 0.32)	QFT-G/QFT-G-IT (non-Japan) 0.24 (0.18 to 0.31)	—	—
30–20	3/7–1/4	0.26–0.44	—	—	—	—
20–10	1/4–1/9	0.44–1.00	TST-BCG (10 or 15 mm) 0.49 (0.32 to 0.76)	TST-BCG (10 or 15 mm) 0.51 (0.36 to 0.73)	—	—
			TST-BCG (5 mm) 0.65 (0.33 to 1.29)	TST-BCG (5 mm) 0.70 (0.28 to 1.75)	—	—

Probable pre-test odds range from 1/9 to 11/9 for latent tuberculosis infection during tuberculosis contact investigation, and 2/3 to 3/2 for tuberculosis disease among patients with clinical manifestations. Japan and non-Japan in brackets denote results based on Japanese and non-Japanese studies, respectively. Tests that are suitable in settings with higher pre-test odds also work in settings with lower pre-test odds for better negative predictive values.

\*Corresponding to negative predictive values of at least 90% (see mathematical equations in the Methods section).

LTBI, latent tuberculosis infection; PLR, positive likelihood ratio; NLR, negative likelihood ratio; QFT-G, QuantiFERON-TB Gold; QFT-G-IT, QuantiFERON-TB Gold in-Tube; TB, tuberculosis; TST-BCG (x mm), tuberculin skin test with cut-off at x mm among predominantly BCG-vaccinated subjects.

## Limitations

There are a number of limitations. First, although the literature search among English papers may have been reasonably thorough, exclusion of non-English papers from the current review could have introduced publication bias. Second, the quality of the current review is inevitably affected by intrinsic errors arising from the use of TB disease as a surrogate marker for LTBI and the assumption of no LTBI among healthy low-risk subjects. Errors due to inclusion of false TB cases might be modest as most cases of TB had been confirmed by acid-fast bacilli smear, culture, polymerase chain reaction for *M tuberculosis* or histopathology. Third, the use of median values in meta-regression analysis could have reduced the statistical power in identifying heterogeneity. Fourth, differences between QFT-G and QFT-G-IT for LTBI could have been missed owing to the small number of studies. Lastly, the predominance of immunocompetent subjects in the current review and the exclusion of indeterminate results from analysis may render findings of the current review less applicable to severely immunocompromised subjects. As IGRA depend on host immunity, indeterminate and false-negative results of IGRA will inevitably increase among hosts with impaired immunity, notably HIV-infected subjects and those with low CD4 count, especially <200 cells/ $\mu$ L.<sup>24–28</sup> Indeterminate results may be less frequent for T-SPOT.TB than QFT-G as a result of the choice of cut-off value.<sup>24–29</sup>

In conclusion, at a 90% certainty threshold, LTBI is best diagnosed by QFT-G/QFT-G-IT (based on Japanese studies) and excluded by T-SPOT.TB or QFT-G/QFT-G-IT (based on Japanese studies); none can diagnose TB disease, whereas T-SPOT.TB can exclude TB disease among middle-aged and older patients.

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

1. **World Health Organization.** *Global tuberculosis control: epidemiology, strategy, financing: WHO report WHO/HTM/TB/2009.411.* Geneva: World Health Organization, 2009.
2. **Anon.** Targeted tuberculin testing and treatment of latent tuberculosis infection. *Am J Respir Crit Care Med* 2000;**161**:S221–47.
3. **Richeldi L.** An update on the diagnosis of tuberculosis infection. *Am J Respir Crit Care Med* 2006;**174**:736–42.
4. **Davies PDO, Pai M.** The diagnosis and misdiagnosis of tuberculosis. *Int J Tuberc Lung Dis* 2008;**12**:1226–34.
5. **Stead WW, To T.** The significance of the tuberculin skin test in elderly persons. *Ann Intern Med* 1987;**107**:837–42.
6. **Liebeschuetz S, Bamber S, Ewer K, et al.** Diagnosis of tuberculosis in South African children with a T-cell-based assay: a prospective cohort study. *Lancet* 2004;**364**:2196–203.
7. **Pai M, Riley LW, Colford JMJ.** Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. *Lancet Infect Dis* 2004;**4**:761–76.
8. **Menzies D, Pai M, Comstock G.** Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. *Ann Intern Med* 2007;**146**:340–54.

9. **Pai M**, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. *Ann Intern Med* 2008;**149**:177–84.
10. **Zwinderman AH**, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. *Stat Med* 2008;**27**:687–97.
11. **Rutter CM**, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;**20**:2865–84.
12. **Reitsma JB**, Glas AS, Rutjes AWS, *et al*. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;**58**:982–90.
13. **Rutter CM**, Gatsonis CA. Regression methods for meta-analysis of diagnostic test data. *Acad Radiol* 1995;**2**(Suppl 1):S48–56 [discussion S65–7, S70–1 pas].
14. **Lijmer JG**, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002;**21**:1525–37.
15. **Moses LE**, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993;**12**:1293–316.
16. **Irwig L**, Macaskill P, Glasziou P, *et al*. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30 [discussion 131–2].
17. **Deeks JJ**. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**:157–62.
18. **Morrison J**, Pai M, Hopewell PC. Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis. *Lancet Infect Dis* 2008;**8**:359–68.
19. **Shaw JB**, Wynn-Williams N. Infectivity of pulmonary tuberculosis in relation to sputum status. *Am Rev Tuberc* 1954;**69**:724–32.
20. **van Geuns HA**, Meijer J, Styblo K. Results of contact examination in Rotterdam, 1967–1969. *Bull Int Union Tuberc* 1975;**50**:107–21.
21. **Grzybowski S**, Barnett GD, Styblo K. Contacts of cases of active pulmonary tuberculosis. *Bull Int Union Tuberc* 1975;**50**:90–106.
22. **Sterne JA**, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53**:1119–29.
23. **Zamora J**, Abraira V, Muriel A, *et al*. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Method* 2006;**6**:31.
24. **Kobashi Y**, Sugiu T, Shimizu H, *et al*. Clinical evaluation of the T-SPOT.TB test for patients with indeterminate results on the QuantiFERON TB-2G test. *Intern Med* 2009;**48**:137–42.
25. **Mandalakas AM**, Hesselink AC, Chegou NN, *et al*. High level of discordant IGRA results in HIV-infected adults and children. *Int J Tuberc Lung Dis* 2008;**12**:417–23.
26. **Raby E**, Moyo M, Devendra A, *et al*. The effects of HIV on the sensitivity of a whole blood IFN-gamma release assay in Zambian adults with active tuberculosis. *PLoS ONE* 2008;**3**:e2489.
27. **Talati NJ**, Seybold U, Humphrey B, *et al*. Poor concordance between interferon-gamma release assays and tuberculin skin tests in diagnosis of latent tuberculosis infection among HIV-infected individuals. *BMC Infect Dis* 2009;**9**:15.
28. **Syed Ahamed Kabeer B**, Sikhamani R, Swaminathan S, *et al*. Role of interferon gamma release assay in active TB diagnosis among HIV infected individuals. *PLoS One* 2009;**4**:e5718.
29. **Lee JY**, Choi HJ, Park I, *et al*. Comparison of two commercial interferon-gamma assays for diagnosing Mycobacterium tuberculosis infection. *Eur Respir J* 2006;**28**:24–30.

## Lung alert

### Updating prognostic parameters in COPD: the updated BODE index and ADO

The BODE index was devised to better reflect the multisystem effects of chronic obstructive pulmonary disease (COPD), allowing for better prognostic estimation than that provided by forced expiratory volume in 1 s (FEV<sub>1</sub>) alone. The grading system, based on four parameters—that is, body mass index, airflow obstruction, Medical Research Council (MRC) dyspnoea score and the 6 min walk distance—was verified by population studies.

This study set out to assess whether the BODE index could match the observed mortality in different populations of patients with COPD. The authors assessed the calibration of the BODE index, updated it to reflect any changes in calibration and subsequently developed a simplified index for use in Primary Care.

Two populations of patients with COPD were observed for their 3-year mortality as opposed to those predicted by the BODE index. The populations included patients in the Swiss Barmelweid and the Spanish Phenotype and Course COPD cohorts. In both cohorts they compared the observed 3-year risk of all-cause mortality with the risk predicted by the BODE index.

The authors found a poor calibration of the BODE index, with relative underprediction of the 3-year risk of mortality in the Swiss cohort (3-year predicted mortality risk of 21.7% vs 34.1% observed mortality), and an overprediction of the mortality risk in the Spanish cohort (predicted 16.7% vs 12% observed mortality). They concluded that the BODE index does not reflect all-cause mortality in the different populations. Subsequently they performed further regression analysis and updated the index with a greater emphasis on the 6 min walk distance. They also devised a new predictor of all-cause mortality by the ADO index, using age, airflow obstruction and MRC dyspnoea score.

The authors conclude that the updated BODE and ADO indices provide better prognostic assessment of patients with COPD as measured in the named populations. They hope that the identification of baseline risks through prognostic studies may aid in guideline development, and be followed by targeted therapies to alter the risks.

- Puhan MA, Garcia-Aymerich J, Frey M, *et al*. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009;**374**:704–11.

#### Farid Bazari

**Correspondence to** Dr Farid Bazari, Kingston Hospital, Kingston upon Thames, UK; Farid.Bazari@kingstonhospital.nhs.uk  
*Thorax* 2010;**65**:276. doi:10.1136/thx.2009.132894