OPEN ACCESS

ORIGINAL ARTICLE

Incremental value of T-SPOT.*TB* for diagnosis of active pulmonary tuberculosis in children in a high-burden setting: a multivariable analysis

Daphne I Ling,[1] Mark P Nicol,[2] Madhukar Pai,[1] Sandra Pienaar,[3] Nandini Dendukuri,[4] Heather J Zar[3]

[1]Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada
[2]Division of Medical Microbiology and Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa
[3]Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, University of Cape Town, Cape Town, South Africa
[4]Technology Assessment Unit, McGill University Health Centre, Montreal, Quebec, Canada

Correspondence to
Professor Heather J Zar, Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, 5th Floor, ICH Building, Cape Town 7700, South Africa; heather.zar@uct.ac.za

Received 3 December 2012
Revised 9 April 2013
Accepted 15 April 2013
Published Online First 14 May 2013

ABSTRACT

Introduction Interferon γ release assays (IGRAs) are increasingly used for tuberculosis (TB) infection, but their incremental value beyond patient demographics, clinical signs and conventional tests for active disease has not been evaluated in children.

Methods The incremental value of T-SPOT.*TB* was assessed in 491 smear-negative children from two hospitals in Cape Town, South Africa. Bayesian model averaging was used to select the optimal set of patient demographics and clinical signs for predicting culture-confirmed TB. The added value of T-SPOT.*TB* over and above patient characteristics and conventional tests was measured using statistics such as the difference in the area under the receiver operating characteristic curve (AUC), the net reclassification improvement (NRI) and the integrated discrimination improvement (IDI).

Results Cough longer than 2 weeks, fever longer than 2 weeks, night sweats, malaise, history of household contact and HIV status were the most important predictors of culture-confirmed TB. Binary T-SPOT.*TB* results did not have incremental value when added to the baseline model with clinical predictors, chest radiography and the tuberculin skin test. The AUC difference was 3% (95% CI 0% to 7%). Using risk cut-offs of <10%, 10–30% and >30%, the NRI was 7% (95% CI −8% to 31%) but the CI included the null value. The IDI was 3% (95% CI 0% to 11%), meaning that the average predicted probability across all possible cut-offs improved marginally by 3%.

Conclusions In a high-burden setting, the T-SPOT.*TB* did not have added value beyond clinical data and conventional tests for diagnosis of TB disease in smear-negative children.

Key messages

What is the key question?

- Interferon γ release assays (IGRAs) are used for diagnosis of tuberculosis (TB) infection, but their incremental value beyond patient demographics, clinical signs and conventional tests for active disease has not been evaluated in children.

What is the bottom line?

- Our study shows that the T-SPOT.*TB* does not have added value beyond patient demographics, clinical signs and conventional tests for diagnosis of active pulmonary TB (PTB) in smear-negative children.

Why read on?

- Our findings have great relevance for paediatric clinical practice in high-burden, low-resource settings and are consistent with the WHO recommendation against the use of IGRAs for diagnosis of active PTB in low and middle income countries.

INTRODUCTION

The diagnosis of pulmonary tuberculosis (PTB) in children may be challenging.[1 2] Microbiological confirmation with culture is often unavailable due to the difficulty of specimen collection; even when specimens are obtained, smear results are usually negative due to paucibacillary disease. The diagnosis therefore relies on clinical and radiological findings and evidence of TB infection as measured by the tuberculin skin test (TST).[3]

Interferon γ release assays (IGRAs) have been developed to replace the TST for diagnosis of TB infection. Meta-analyses of IGRA performance in children show that they have higher specificity but similar sensitivity compared with the TST for active TB, although IGRA sensitivity is lower in high-incidence versus low-incidence settings.[4 5]

While the intended use of IGRAs is not for active TB, they may be helpful in providing evidence for a rapid diagnosis in children when combined with other routine clinical and laboratory investigations. If conventional tests such as smears are negative, then a T-SPOT.*TB* (Oxford Immunotec, Abingdon, UK) result could add diagnostic information while culture results are pending. Two published studies on the incremental value of IGRAs in adult patients showed limited utility of IGRAs for active TB disease.[6 7]

Our study aim was to investigate the incremental value of T-SPOT.*TB* beyond patient demographics, clinical signs and conventional tests in the diagnostic workup of hospitalised children being evaluated for PTB in a high-burden setting. Using a multivariable approach, the incremental value was measured by the difference between the area under the receiver operating characteristic curve (AUC), net reclassification improvement (NRI) and integrated discrimination improvement (IDI).

To cite: Ling DI, Nicol MP, Pai M, *et al*. *Thorax* 2013;**68**:860–866.

METHODS

Study enrolment

The study enrolment process has been previously described.[8] Briefly, hospitalised children ≤15 years with suspected PTB were consecutively enrolled from two hospitals in Cape Town from February 2009 to July 2011. To be eligible, a child had to present with cough >2 weeks and at least one of the following: exposure to a household TB case within the past 3 months, weight loss or failure to thrive within the past 3 months, positive TST result or chest radiography (CXR) suggestive of PTB. After the parent or legal guardian provided informed consent, the child underwent a physical examination and extensive diagnostic testing. The study was approved by the Faculty of Health Sciences, Research Ethics Committee at the University of Cape Town.

Diagnostic testing

Two sequential induced sputum[8] and two nasopharyngeal aspirate specimens[9] were collected, after which concentrated fluorescent acid-fast smear microscopy and liquid culture were performed. A definite TB case was defined by at least one positive culture for *Mycobacterium tuberculosis* in any specimen.

An anterio-posterior and lateral CXR, taken on enrolment, were evaluated by a single experienced reviewer, blinded to other test results. The results were reported as consistent, inconsistent or inconclusive for PTB. When the result was inconclusive, a second reviewer read and classified the CXR; a third reviewer evaluated the radiograph when there was disagreement. The TST was performed using the Mantoux technique with 2 tuberculin units of purified protein derivative (PPD RT23, Statum Serum Institute) and read by a trained nurse at 48–72 h. The cut-off value for a positive TST was ≥5 mm in HIV-infected children and ≥10 mm for others. A HIV rapid test was performed on all children and confirmed by HIV ELISA in children ≥18 months or PCR in children <18 months.

The T-SPOT.*TB* was performed according to the manufacturer's guidelines. All of the tests were run within 3–4 h of a blood specimen taken by a single trained technologist. Peripheral blood mononuclear cells were separated, and the cell concentration was adjusted to 2.5×10^6 cells/ml. To each of four interferon γ capture antibody precoated wells, the following were added: 50 µl of Positive Control, Panel A (ESAT6), Panel B (CFP10), and Nil control. Then, 100 µl of cell suspension was added to each well. This was incubated overnight at 37°C in 5% CO_2. The test result was positive if Panel A minus Nil control and/or Panel B minus Nil control ≥6 spots. The test result was negative if both Panel A minus Nil control and Panel B minus Nil control ≤5 spots. A Nil control >10 spots was considered Indeterminate. Two observers, who were blinded to all other test results, reported the number of spots, and the mean number was used in the analysis. Indeterminate results were excluded from analysis. Physicians were blinded to the T-SPOT.*TB* results and did not use them in the clinical decision-making.

Estimation of incremental value

We used Bayesian model averaging (BMA) to identify the most parsimonious set of predictors in a logistic regression model for the outcome of culture-confirmed TB. Predictors for patient demographics and clinical signs were determined a priori and included passive smoke exposure, HIV status, history of household contact, weight for age Z score, night sweats, cough >2 weeks, fever >2 weeks, malaise, loss of appetite and weight loss. We used the BMA programme in the statistical software package R that considers all possible combinations of predictors. This programme performs a model selection procedure based on the Bayesian information criterion.[10] The advantage of this approach is that it does not lead to overfitting.[11] Predictors that were in the top six models, which together had an 88% probability of being the best model, were included in the clinical model.

Since children with positive smears are diagnosed with definite TB and started on treatment, the incremental value of T-SPOT.*TB* was evaluated only in smear-negative children. To reflect the diagnostic workup for active TB in children in our hospital setting, the clinical model was extended by the CXR result (consistent or inconsistent with PTB) and the TST result (positive or negative according to the recommended cut-off depending on HIV status). The clinical predictors, CXR and TST formed the baseline model. This model was further extended by the T-SPOT.*TB* result (positive or negative according to the manufacturer's recommended cut-off) to determine its incremental value. We also performed the same extension using the quantitative log-transformed spot-forming cells (SFCs) rather than the binary result. In the primary analysis using confirmed TB as the outcome, all culture-negative children were considered non-TB cases, even if they were clinically diagnosed with TB and started on treatment. In a secondary analysis, we estimated the incremental value of T-SPOT.*TB* using the children placed on empirical treatment for TB as the outcome.

Evaluating incremental value involves comparing prediction models, with and without the new test as a covariate. The difference in AUC for the two models is the most familiar statistic for estimating added value. The AUC is formally defined as the probability that for each randomly selected pair, consisting of one individual with and one without the event, the predicted probability of disease will be higher for the individual who has the event.[12]

More recently, Pencina and colleagues proposed the NRI and IDI statistics for comparing prediction models.[13] Both measures are based on quantifying appropriate reclassification by the addition of a new test to an existing model, or higher probability for patients with the event and lower probability for those without the event.

The NRI is the net proportion of individuals with the event who are reclassified into a higher risk category, plus the net proportion of individuals without the event who are reclassified into a lower risk category: $\mathrm{NRI}=(p_{up,events} - p_{down,events})+(p_{down,nonevents} - p_{up,nonevents})$, where lower-case p is the proportion of upward or downward movement into predefined risk categories. We used the risk categories of <10%, 10–30% and >30% based on the probability distribution of our models and the categories used by Metcalfe *et al*.[7]

The IDI, a continuous form of the NRI across all risk thresholds from 0% to 100%, is the sum of the increase in probability among patients with the event and the decrease in probability among patients without the event. It is formally defined by: $\mathrm{IDI} = (\overline{P}_{new,events} - \overline{P}_{old,events}) + (\overline{P}_{old,nonevents} - \overline{P}_{new,nonevents})$, where upper-case $\overline{P}$ is the average predicted probability of the outcome in the old model and the new extended model with the additional test.

Tenfold cross validation was performed on the data to avoid overestimating the incremental value. All 95% CIs for the AUC, NRI and IDI were obtained using 1000 bootstrap samples. An interval that did not cross the null value of 0 would represent statistically significant added value. All incremental value analyses were performed using the user-written package 'incrisk' in Stata, V.12.[14]

RESULTS

Patient characteristics

As shown in figure 1, 557 children with CXR and TST results had the T-SPOT.*TB* performed. In 29 (5%) children, the T-SPOT.*TB* was indeterminate (all due to high Nil control), and they were excluded from further analysis. The demographic and clinical characteristics of the 528 children are summarised in table 1. The median age was 22 months (IQR 12–53). The prevalence of HIV infection was 24%, with most children in the late clinical stages of disease as defined by WHO.[15] The median CD4 cell count was 489 cells/ml (IQR 226–832). Most children had findings consistent with PTB on CXR, while most were negative on TST and T-SPOT.*TB*. Culture-confirmed TB was diagnosed in 91 (17%) children, while only 37 (7%) were positive on smear microscopy. Based on clinical judgment, 296 (56%) children were diagnosed with PTB, for which they were started on treatment.

Incremental value

Predictors that were in the top six models according to the BMA procedure included cough >2 weeks, fever >2 weeks, night sweats, malaise, HIV infection, and history of household contact. These predictors comprised the clinical model. The Hosmer–Lemeshow goodness-of-fit test indicated no significant differences between the observed and predicted probability of TB (χ^2=7.84, p=0.45). The cross-validated odds ratios for the logistic regression models are presented in table 2.

The incremental value analysis was performed in 491 children with negative smear results. With culture-confirmed TB as the outcome, the AUC for the baseline model with clinical predictors, CXR and TST was 85% (95% CI 80% to 90%). Addition of T-SPOT.*TB* to this model did not substantially improve its discriminatory ability, as the AUC increased only to 88% (95% CI 85% to 92%) and 87% (95% CI 84% to 92%) when the binary result or quantitative result was added, respectively (table 3). When using the NRI measure, binary T-SPOT.*TB* results (NRI=7%; 95% CI −8% to 31%) and quantitative SFC results (NRI=8%; 95% CI −7% to 25%) reclassified patients appropriately. However, both CIs included the null value. Table 4 provides a detailed calculation of the NRI for comparing the baseline model with the expanded model with binary T-SPOT.*TB* results. The numbers along the diagonal represent children who remained in the same risk category even after addition of T-SPOT.*TB* results. In children with confirmed PTB, the T-SPOT.*TB* increased the probability of TB in 20% (11/54) of children, while incorrectly decreasing the probability in 17% (9/54) of children. Thus, the net improvement for confirmed PTB cases was 3% (95% CI −10% to 28%). In children without confirmed PTB, the T-SPOT.*TB* correctly decreased the probability of TB in 9% (38/437) of children but increased the probability in 5% (24/437) of children for a net improvement of 4% (95% CI −1% to 7%).

Considering all possible risk categories from 0% to 100%, the IDI showed that the T-SPOT.*TB* did not improve the average probability of TB (IDI=3% (95% CI 0% to 11%) for binary results and IDI=4% (95% CI 0% to 12%) for quantitative results). Figure 2 depicts the probability distributions for children with and without confirmed TB and the marginal improvements with addition of binary T-SPOT.*TB* results. As seen in this figure, adding T-SPOT.*TB* to the standard workup did not cause substantial increases in predicted probabilities for confirmed TB cases (ie, shifts to the right). For non-TB cases, the probabilities are appropriately concentrated to the left even without the addition of T-SPOT.*TB*.

When empirical treatment for active TB was used as the outcome of our model, T-SPOT.*TB* results increased the AUC marginally from 84% to 86% (AUC difference=2%; 95% CI 0% to 3%). A similar finding was seen with the IDI, in which the IDI was 2% (95% CI 1% to 6%) for binary results and 3% (95% CI 1% to 6%) for quantitative results. When risk categories were used for the NRI, the magnitude of improvement was larger in children who were not treated. The T-SPOT.*TB* correctly decreased the probability of disease in these children. However, the CI contained the null value. Overall, T-SPOT.*TB* results did not substantially reclassify smear-negative children into appropriate risk categories (table 3).

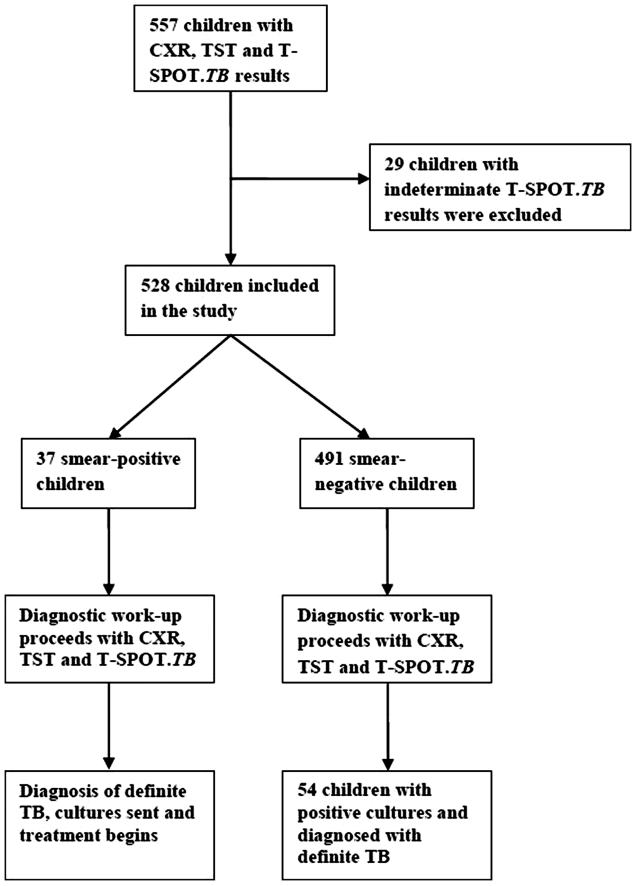

Figure 1 Flowchart of children included in the study. CXR, chest radiography; TB, tuberculosis; TST, tuberculin skin test.

DISCUSSION

Although meta-analyses have shown that IGRAs have suboptimal sensitivity for active PTB in children,[4 5] diagnostic tests such as the T-SPOT.*TB* are not performed alone. The diagnostic workup is inherently a sequential process that begins with patient demographics, then signs and symptoms, followed by simpler tests, and extending to more expensive and/or invasive tests. Unfortunately, diagnostic research that evaluates the added value of new TB tests beyond routinely measured variables and conventional tests is limited.

Our study is the first incremental value analysis of the IGRA in children, in whom diagnosis of active TB remains a challenge. We measured the added value of T-SPOT.*TB* in a high-burden setting using the difference in AUC and the NRI and IDI measures that have been more recently described. Our findings show that T-SPOT.*TB* does not have utility in supporting a diagnosis

Table 1 Demographic and clinical characteristics of 528 children at enrolment

Characteristic	Frequency (%) in all children (n=528)	Frequency (%) in culture-confirmed children (n=91)	Frequency (%) in treated children (n=296)
Median age in months (IQR)	22 (12 to 53)	28 (16 to 62)	25 (13 to 61)
Male	292 (55)	54 (59)	164 (55)
BCG vaccinated (based on presence of scar and vaccination records)	447 (85)	60 (66)	240 (81)
Night sweats	275 (52)	54 (59)	159 (54)
Malaise	245 (46)	56 (62)	153 (52)
Fever >2 weeks	85 (16)	24 (26)	59 (20)
Cough >2 weeks	169 (32)	45 (49)	118 (40)
History of household TB contact	321 (61)	61 (67)	185 (63)
HIV positive	124 (24)	15 (17)	71 (24)
WHO clinical stage 1	11 (9)	0	2 (3)
WHO clinical stage 2	35 (28)	5 (33)	15 (21)
WHO clinical stage 3	50 (40)	6 (40)	31 (44)
WHO clinical stage 4	28 (23)	4 (27)	23 (32)
Median CD4 cell count in HIV-positive children (IQR)*	489 (226 to 832)	333 (112 to 485)	471 (152 to 832)
Median weight for age Z score (IQR)	−1.56 (−2.84 to −0.56)	−1.62 (−3.07 to −0.78)	−1.61 (−3.01 to −0.68)
Median height for age Z score (IQR)†	−1.23 (−2.33 to −0.30)	−1.11 (−2.33 to −0.60)	−1.17 (−2.34 to −0.33)
Median weight for height Z score (IQR)‡	−0.66 (−2.06 to 0.42)	−1.39 (−3.05 to −0.48)	−0.77 (−2.40 to 0.23)
CXR consistent with pulmonary TB	372 (71)	81 (89)	217 (73)
TST positive (recommended cut-offs)§	213 (40)	68 (74)	193 (65)
T-SPOT.*TB* positive (manufacturer's cut-off)	175 (33)	75 (82)	157 (53)
Median number of SFCs on T-SPOT.*TB* (IQR)	3 (1 to 13)	25 (7 to 67)	7 (2 to33)
Smear positive	37 (7)	37 (41)	37 (13)

*CD4 cell count not available for nine HIV-positive children.
†Data available for 430 children.
‡Data available for 385 children.
§5 mm for HIV-infected children; 10 mm for others.
CXR, chest radiography; SFC, spot-forming cell; TB, tuberculosis; TST, tuberculin skin test.

of active PTB in hospitalised smear-negative children beyond the clinical history and standard tests. When culture-confirmed TB was used as the outcome in the analysis, T-SPOT.*TB* did not increase the discriminatory ability of the model according to the AUC. Furthermore, it did not correctly reclassify a substantial number of children, whether using prespecified risk categories with the NRI or continuous risk probabilities with the IDI. In a secondary analysis, the findings were similar when empirical TB treatment based on clinical judgment was used as the outcome.

Table 2 Cross-validated ORs (95% CIs) for culture-confirmed TB for the baseline model with clinical predictors, CXR and TST and the expanded model with binary T-SPOT.*TB* results (n=491)

Predictor	Baseline model without T-SPOT.*TB*	Expanded model with T-SPOT.*TB* (binary)
Night sweats	1.42 (0.72 to 2.80)	1.32 (0.66 to 2.69)
Malaise	1.82 (0.94 to 3.53)	2.08 (1.03 to 4.18)
Fever >2 weeks	1.90 (0.83 to 4.35)	1.86 (0.78 to 4.48)
Cough >2 weeks	1.52 (0.73 to 3.16)	1.42 (0.67 to 3.03)
History of TB contact	0.84 (0.43 to 1.65)	0.61 (0.29 to 1.25)
HIV positive	0.43 (0.14 to 1.34)	0.64 (0.19 to 2.14)
CXR consistent with TB	5.53 (2.03 to 15.03)	4.62 (1.65 to 12.94)
TST positive	19.30 (7.32 to 50.91)	7.54 (2.56 to 22.20)
T-SPOT.*TB*	–	5.93 (2.39 to 14.73)

CXR, chest radiography; TB, tuberculosis; TST, tuberculin skin test.

Two studies have been published thus far on the incremental value of IGRAs in smear-negative adult patients using culture-confirmed TB as the gold standard.[6 7] Both studies evaluated the QuantiFERON-TB (QFT, Cellestis/Qiagen, Australia) assay in adults. Metcalfe and colleagues found that quantitative interferon γ results significantly improved discrimination according to the AUC and appropriately reclassified patients using the NRI when added to patient demographic and clinical characteristics in a low-incidence setting.[7] The added value of the QFT, however, was lower when predictors based on subjective clinical judgment of risk (low, medium or high) were incorporated into their model. In the second study, Rangaka and colleagues evaluated the use of the QFT to screen for active TB in HIV-infected adults before starting isoniazid preventive therapy.[6] According to the AUC, the QFT had less added value compared with TST for improving upon a baseline clinical model. Based on evidence from these incremental value studies and those based on test accuracy,[16] WHO recommended against the use of IGRAs for diagnosing active TB disease in low and middle income countries.[17] Our findings are supportive of this negative recommendation in high-burden settings for children as well.

Strengths and limitations

In our methodological approach, we used a BMA procedure that considers all possible combinations of predictors. This model selection process and the cross-validation step avoid developing an overly optimistic model that would have worse performance when applied to other data. Furthermore, we assessed the incremental value using two newer measures that

Table 3 AUC, NRI and IDI results with culture-confirmed TB and empirical treatment for active TB as the outcome in smear-negative children (n=491). Comparisons are made between the baseline model with clinical predictors, CXR and TST and the expanded model with addition of T-SPOT.*TB*

	AUC expanded (95% CI)	AUC baseline (95% CI)	AUC difference (95% CI)	NRI events (95% CI)	NRI non-events (95%CI)	NRI sum (95% CI)	IDI events (95% CI)	IDI non-events (95% CI)	IDI sum (95% CI)
Culture-confirmed TB									
Clinical+CXR+TST+T-SPOT.*TB* (binary)	0.88 (0.85 to 0.92)	0.85 (0.80 to 0.90)	0.03 (0.002 to 0.07)	0.03 (−0.10 to 0.28)	0.04 (−0.01 to 0.07)	0.07 (−0.08 to 0.31)	0.02 (0.001 to 0.10)	0.01 (0.001 to 0.01)	0.03 (0.001 to 0.11)
Clinical+CXR+TST+T-SPOT.*TB* (quantitative)	0.87 (0.84 to 0.92)	0.85 (0.80 to 0.90)	0.02 (0.002 to 0.06)	0.07 (−0.09 to 0.22)	0.01 (−0.01 to 0.07)	0.08 (−0.07 to 0.25)	0.04 (0.002 to 0.10)	0.004 (0.001 to 0.01)	0.04 (0.002 to 0.12)
Empirical treatment for active TB									
Clinical+CXR+TST+T-SPOT.*TB* (binary)	0.86 (0.83 to 0.90)	0.84 (0.81 to 0.88)	0.02 (0.003 to 0.03)	0.01 (−0.02 to 0.04)	0.12 (−0.01 to 0.19)	0.13 (−0.001 to 0.20)	0.01 (0.002 to 0.03)	0.01 (0.003 to 0.03)	0.02 (0.01 to 0.06)
Clinical+CXR+TST+T-SPOT.*TB* (quantitative)	0.86 (0.83 to 0.90)	0.84 (0.81 to 0.88)	0.02 (0.003 to 0.04)	0.01 (−0.03 to 0.05)	0.08 (−0.01 to 0.22)	0.09 (−0.01 to 0.24)	0.01 (0.003 to 0.03)	0.02 (0.003 to 0.03)	0.03 (0.01 to 0.06)

AUC, area under the curve; CXR, chest radiography; IDI, integrated discrimination improvement; NRI, net reclassification improvement; TST, tuberculin skin test.

Table 4 Reclassification tables to calculate the NRI for binary T-SPOT.TB results in smear-negative children with and without culture-confirmed TB, using the risk categories <10%, 10–30% and >30%*

	Clinical+CXR+TST+T-SPOT.*TB* (expanded model)			
Clinical+CXR+TST (baseline model)	**<10%**	**10–30%**	**>30%**	**Total**
Confirmed TB (N=54)				
<10%	4	**5**	**0**	9
10–30%	4	9	**6**	19
>30%	1	4	21	26
Total	9	18	27	54
Not confirmed TB (N=437)				
<10%	318	12	0	330
10–30%	**26**	29	12	67
>30%	**2**	**10**	28	40
Total	346	51	40	437

Numbers in bold indicate correctly-reclassified patients that are used to calculate the NRI.
*The risk categories are based on the predicted probabilities from the logistic regression model. The T-SPOT.*TB* correctly reclassified 3% of culture-confirmed cases (11/54 above the diagonal minus 9/54 below the diagonal). The T-SPOT.*TB* correctly reclassified 4% of non-confirmed cases (38/437 below the diagonal minus 24/437 above the diagonal). Thus, the overall NRI is 7%.
CXR, chest radiography; NRI, net reclassification improvement; PTB, pulmonary tuberculosis; TST, tuberculin skin test.

are based on risk probabilities. One major criticism of the AUC as an overall measure of model performance has been its lack of direct clinical interpretation to individual patients. The NRI and IDI are more informative than the AUC in that they calculate the increase in appropriate reclassification separately for events and non-events.

Our study also had several limitations. Only results from one reader were used for most CXRs, so that a consensus read was not possible. Furthermore, we could not evaluate specific radiological findings, such as lobar consolidation or lymphadenopathy, in our analysis. In the secondary analysis in which empirical TB treatment based on clinical judgment was used as the outcome of the model, incorporation bias may be an issue. This type of bias would generally overestimate the performance of tests that are also used in assigning the final outcome. This problem is difficult to avoid in the context of childhood TB, as the absence of an adequate reference standard requires that the entire clinical picture be taken into account when making a final diagnosis.

Our study recruited children from secondary and tertiary-care hospitals, where disease is more severe and treating physicians are highly specialised and/or more prone to overtreatment. Thus, our findings may not be applicable to primary-care facilities with less experienced clinicians or in community and ambulatory settings where children would present with milder form of disease. Furthermore, the availability of resources in our hospital setting may not be representative of other settings in which certain diagnostic tests are not standard of care. In fact, smear microscopy is being phased out as the first-line diagnostic test in South Africa with the nationwide rollout of the rapid Xpert MTB/RIF test (Cepheid, USA).[18] Further analysis will be required to determine the incremental value of T-SPOT.*TB* in children in whom Xpert MTB/RIF is negative. Moreover, our study findings may not be generalisable to settings with low prevalence of HIV coinfection.

The lack of a gold standard to confirm the diagnosis further complicates conducting research studies for childhood TB.

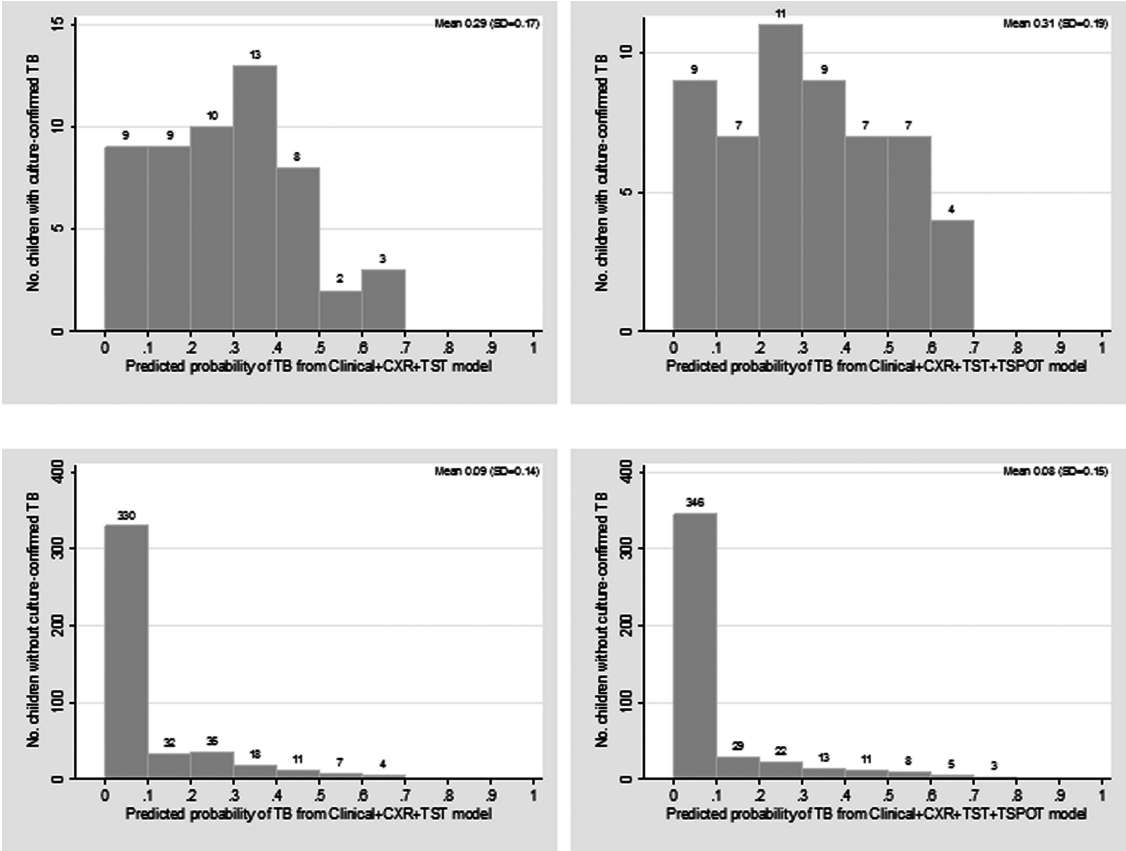

Figure 2 Distributions of the predicted probability of tuberculosis (TB) for children with (top) and without (bottom) culture-confirmed TB. Plots on the left represent the baseline model with clinical predictors, chest radiography (CXR) and tuberculin skin test (TST). Plots on the right represent this model extended by binary T-SPOT.*TB* results. The T-SPOT.*TB* increased marginally the average probability in confirmed cases from 29% to 31%, while for non-confirmed cases the average probability was reduced only from 9% to 8%.

Recently, efforts have been made to develop practical reference standards based on consensus clinical case definitions[19] and to suggest a standardised approach for reporting findings so that results are comparable across studies.[20] Thus, these are important steps forward for facilitating the inclusion of children in diagnostic studies that will target this important group as a priority for new TB technologies.

Acknowledgements We thank the clinical study and laboratory staff, the staff at Red Cross War Memorial Children's Hospital and New Somerset Hospital, and the children and their caregivers for participating. We also thank Dick Menzies for providing helpful advice on this study.

Contributors DIL: designed the study, performed the analysis and wrote the manuscript as first author. MPN: participated in the study design, provided input on the analysis and supervised the data collection. MP: participated in the study design and provided input on the analysis. SP: performed the T-SPOT.*TB* assay. ND: provided guidance on the methods used in the study. HJZ: designed the study, provided input on the analysis and supervised the study.

Funding This work was supported by the National Institutes of Health, USA (grant number 1R01HD058971-01), National Health Laboratory Services Research Trust, Medical Research Council of South Africa, and The Wellcome Trust (grant number 085251/B/08/Z). DL is funded by a doctoral research studentship from the Canadian Thoracic Society and the European and Developing Countries Clinical Trials Partnership (TBNEAT grant). MP is the recipient of a Canadian Institutes of Health Research New Investigator Award and a Fonds de Recherche Quebec Santé (FRQS) Salary Award. ND is supported by a FRQS Salary Award.

Competing interests None.

Ethics approval The study was approved by the Faculty of Health Sciences, Research Ethics Committee at the University of Cape Town.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES

1 Marais BJ, Pai M. New approaches and emerging technologies in the diagnosis of childhood tuberculosis. *Paediatr Respir Rev* 2007;8:124–33.
2 Zar HJ, Connell TG, Nicol M. Diagnosis of pulmonary tuberculosis in children: new advances. *Expert Rev Anti Infect Ther* 2010;8:277–88.
3 Hopewell PC, Pai M, Maher D, *et al*. International standards for tuberculosis care. *Lancet Infect Dis* 2006;6:710–25.
4 Machingaidze S, Wiysonge CS, Gonzalez-Angulo Y, *et al*. The utility of an interferon gamma release assay for diagnosis of latent tuberculosis infection and disease in children: a systematic review and meta-analysis. *Pediatr Infect Dis J* 2011;30:694–700.
5 Mandalakas AM, Detjen AK, Hesseling AC, *et al*. Interferon-gamma release assays and childhood tuberculosis: systematic review and meta-analysis. *Int J Tuberc Lung Dis* 2011;15:1018–32.
6 Rangaka MX, Gideon HP, Wilkinson KA, *et al*. Interferon release does not add discriminatory value to smear-negative HIV-tuberculosis algorithms. *Eur Respir J* 2012;39:163–71.
7 Metcalfe JZ, Cattamanchi A, Vittinghoff E, *et al*. Evaluation of quantitative IFN-gamma response for risk stratification of active tuberculosis suspects. *Am J Respir Crit Care Med* 2010;181:87–93.
8 Nicol MP, Workman L, Isaacs W, *et al*. Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study. *Lancet Infect Dis* 2011;11:819–24.
9 Zar HJ, Workman L, Isaacs W, *et al*. Rapid molecular diagnosis of pulmonary tuberculosis in children using nasopharyngeal specimens. *Clin Infect Dis* 2012;55:1088–95.

10 Raftery AE. Bayesian model selection in social research. *Sociolog Methodol* 1995;25:111–63.
11 Steyerberg EW, Bleeker SE, Moll HA, *et al*. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003;56:441–7.
12 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
13 Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, *et al*. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72; discussion 207–12.
14 Longton G, Pepe M. Risk Prediction Package. http://labs.fhcrc.org/pepe/dabs/ppsoft.html (accessed Nov 2012).
15 World Health Organization. *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children*, 2006. http://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf (accessed Nov 2012).
16 Metcalfe JZ, Everett CK, Steingart KR, *et al*. Interferon-gamma release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *J Infect Dis* 2011;204(Suppl 4): S1120–9.
17 World Health Organization. *Use of tuberculosis interferon-gamma release assays (IGRAs) in low- and middle-income countries: policy statement*. Geneva, Switzerland: WHO, 2011. http://whqlibdoc.who.int/publications/2011/9789241502672_eng.pdf (accessed Nov 2012).
18 Meyer-Rath G, Schnippel K, Long L, *et al*. The impact and Cost of scaling up GeneXpert MTB/RIF in South Africa. *PLoS One* 2012;7:e36966.
19 Graham SM, Ahmed T, Amanullah F, *et al*. Evaluation of tuberculosis diagnostics in children: 1. Proposed clinical case definitions for classification of intrathoracic tuberculosis disease. Consensus from an expert panel. *J Infect Dis* 2012;205(Suppl 2): S199–208.
20 Cuevas LE, Browning R, Bossuyt P, *et al*. Evaluation of tuberculosis diagnostics in children: 2. Methodological issues for conducting and reporting research evaluations of tuberculosis diagnostics for intrathoracic tuberculosis in children. Consensus from an expert panel. *J Infect Dis* 2012;205(Suppl 2):S209–15.