American Journal of Epidemiology Advance Access published October 13, 2016

American Journal of Epidemiology

DOI: 10.1093/aje/kww094

## Practice of Epidemiology

# Diagnostic Test Accuracy in Childhood Pulmonary Tuberculosis: A Bayesian Latent Class Analysis

**Samuel G. Schumacher*, Maarten van Smeden*, Nandini Dendukuri, Lawrence Joseph, Mark P. Nicol, Madhukar Pai, and Heather J. Zar**

* Correspondence to Dr. Samuel G. Schumacher, FIND, Campus Biotech, Building B2, Level 0, 9 Chemin des Mines 1202 Geneva, Switzerland (e-mail: samuel.schumacher@mail.mcgill.ca); or Dr. Maarten van Smeden, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands (e-mail: M.vanSmeden@umcutrecht.nl).

*Initially submitted November 5, 2015; accepted for publication February 25, 2016.*

Evaluation of tests for the diagnosis of childhood pulmonary tuberculosis (CPTB) is complicated by the absence of an accurate reference test. We present a Bayesian latent class analysis in which we evaluated the accuracy of 5 diagnostic tests for CPTB. We used data from a study of 749 hospitalized South African children suspected to have CPTB from 2009 to 2014. The following tests were used: mycobacterial culture, smear microscopy, Xpert MTB/RIF (Cepheid Inc.), tuberculin skin test (TST), and chest radiography. We estimated the prevalence of CPTB to be 27% (95% credible interval (CrI): 21, 35). The sensitivities of culture, Xpert, and smear microscopy were estimated to be 60% (95% CrI: 46, 76), 49% (95% CrI: 38, 62), and 22% (95% CrI: 16, 30), respectively; specificities of these tests were estimated in accordance with prior information and were close to 100%. Chest radiography was estimated to have a sensitivity of 64% (95% CrI: 55, 73) and a specificity of 78% (95% CrI: 73, 83). Sensitivity of the TST was estimated to be 75% (95% CrI: 61, 84), and it decreased substantially among children who were malnourished and infected with human immunodeficiency virus (56%). The specificity of the TST was 69% (95% CrI: 63%, 76%). Furthermore, it was estimated that 46% (95% CrI: 42, 49) of CPTB-negative cases and 93% (95% CrI: 82; 98) of CPTB-positive cases received antituberculosis treatment, which indicates substantial overtreatment and limited undertreatment.

childhood pulmonary tuberculosis; diagnosis; latent class analysis; overtreatment; sensitivity; specificity

Abbreviations: CPTB, childhood pulmonary tuberculosis; CrI, credible interval; HIV, human immunodeficiency virus; PTB, pulmonary tuberculosis; TST, tuberculin skin test.

Tuberculosis in children is an important global health problem. There are an estimated 0.5 to 1 million new cases each year (1, 2), with childhood pulmonary tuberculosis (CPTB) being the most common form. One of the major challenges in diagnosing CPTB is the lack of sensitive diagnostic tests (3–6). In clinical practice, the diagnosis of CPTB therefore relies on a combination of imperfect tests, which gives rise to unknown degrees of under- or overtreatment (7, 8).

In recent years, new tests for CPTB have been developed, and their accuracy has been evaluated using mycobacterial culture as a reference standard (4, 9). Although culture is currently considered the best available reference standard, its sensitivity for detecting CPTB is acknowledged to be imperfect (3, 4, 10). The culture reference standard thus inevitably leads to true CPTB case patients being misclassified as being negative for CPTB. If these misclassifications by the reference standard are ignored, then the assessment of the test accuracy can be biased (11–14).

To address the problem of the lack of an accurate reference standard, multivariable diagnostic algorithms for CPTB have been proposed to combine information from multiple imperfect diagnostic tests (including tests for tuberculosis infection and clinical data) in a systematic manner. Although more

than a dozen of these algorithms have been described to date, estimates of CPTB prevalence derived from them vary widely (15, 16). None of these algorithms has relied on statistical modeling approaches that take into account the imperfect nature and relative weight of each of the diagnostic tests.

In the present study, we re-analyzed the results of a study of hospitalized children suspected to have CPTB in which data on commonly used tests for CPTB had been prospectively collected (9). The tests include 3 microbiological tests, the tuberculin skin test (TST), and chest radiography. We used Bayesian latent class analysis to simultaneously estimate the accuracy of the 5 tests with regard to the detection of CPTB, the prevalence of CPTB, and the degree of under- and overtreatment in the cohort. Latent class analysis has successfully been used in other studies of the accuracy of diagnostic tests in the absence of a gold standard (12, 17–19). However, here we present one of the first applications of latent class analysis to prospectively collected data on CPTB.

## METHODS

Data were obtained from a study of hospitalized South African children who were suspected to have CPTB (9). Details on the design of the study are available from the original publications (9, 20). Briefly, between February 2009 and June 2014, children were consecutively enrolled when they presented to a hospital in Cape Town, South Africa, with signs or symptoms suggestive of pulmonary tuberculosis (PTB). Inclusion criteria were: 1) cough and at least 1 additional factor suggestive of CPTB (9, 20); 2) age younger than 15 years; and 3) a parent or legal guardian who provided informed consent. Children were excluded if: 1) they had received tuberculosis treatment or prophylaxis for more than 72 hours or 2) their place of residence precluded follow-up. Patient characteristics are shown in Table 1. In total, 749 children were included in our analysis.

Written informed consent for enrollment in the study was obtained from a parent or legal guardian. The Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town, approved the study. Renewed approval for the current analysis was not required because anonymized data were used.

### Study procedures

Up to 3 induced sputum samples per child were each tested with 3 different microbiological tests: liquid culture (mycobacterial growth indicator, BACTEC MGIT, Becton Dickinson Microbiology Systems, Cockeysville, Maryland; hereafter referred to as culture), a molecular nucleic acid amplification test (Xpert MTB/RIF, Cepheid Inc., Sunnyvale, California; hereafter referred to as Xpert), and sputum smear microscopy. A TST was administered and read according to standard procedures by measuring transverse induration in response to purified protein derivative (2TU, PPD RT23, Staten Serum Institute, Denmark, Copenhagen). Based on a standardized reporting format, radiographs of the chest were judged as "consistent with CPTB" or "not consistent with CPTB" by 2 independent readers who were blinded to all other investigations. Disagreement between readers was resolved by a third reader. The studied tests are complementary; none of these tests in isolation is expected to yield perfect diagnostic accuracy (Table 2).

Testing for human immunodeficiency virus (HIV) (HIV rapid test in all children, followed by a confirmatory polymerase chain reaction for children younger than 18 months or HIV enzyme-linked immunosorbent assay for children 18 months of age or older) was done in all children with unknown HIV infection status. The weight of the child was transformed to a standardized score ($z$ score) as a measure of malnutrition according to World Health Organization Child Growth Standards (21). Parents provided information about the child's date of birth and any household contact who was treated for tuberculosis in the past 3 months. Antituberculosis treatment decisions were at the discretion of the treating doctor based on all available routinely collected information.

**Table 1.** Characteristics of 749 Children Suspected to Have Pulmonary Tuberculosis, South Africa, 2009–2014

| Characteristic | Median (IQR) | No. | % |
|---|---|---|---|
| Female sex | | 347 | 46 |
| Age, months | 22 (12 to 50) | 1–120[a] | |
| Infected with HIV | | 154 | 21 |
| Weight, kg | 10 (8 to 14) | | |
| Weight, $z$ score[b] | −1.1 (−2.2 to 0.2) | | |
| Malnutrition[c] | | 211 | 28 |
| Diagnostic test positive | | | |
| Liquid culture | | 122 | 16 |
| Xpert MTB/RIF | | 106 | 14 |
| Microscopy | | 42 | 6 |
| Radiography | | 249 | 33 |
| TST | | 321 | 43 |
| Household tuberculosis contact | | 409 | 55 |
| Treated for PTB | | 436 | 58 |

Abbreviations: HIV, human immunodeficiency virus; IQR, interquartile range; PTB, pulmonary tuberculosis; TST, tuberculin skin test.

[a] Value is expressed as median (range).

[b] Weight for age $z$ score, calculated according to World Health Organization Child Growth Standards (21).

[c] Malnutrition was defined as having a weight-for-age $z$ score lower than −2.

### CPTB model

Before undertaking our statistical analyses, we defined a heuristic CPTB model that represented our assumptions about the pathophysiology of CPTB and the biological mechanisms that are believed to have given rise to the test results. This model is graphically depicted in Figure 1.

The mechanisms of the 3 microbiological tests under study (culture, Xpert, and microscopy) are based on highly similar

**Table 2.** Diagnostic Tests for Childhood Pulmonary Tuberculosis and Their Expected Accuracy

| Test and Accuracy Measure | Expected Accuracy |
|---|---|
| Microbiological[a] | |
| Sensitivity | Not perfect. Concentration of *Mycobacterium tuberculosis* is usually very low in sputum of children; true cases of childhood pulmonary tuberculosis can be missed. |
| Specificity | Nearly perfect for all 3 tests. Xpert MTB/RIF and microscopy might give rare false positive results due to detection of dead bacilli or in children with bacillus Calmette-Guérin disease. |
| Radiography | |
| Sensitivity | Not perfect. Limited pathology might not be visible; subject to inter- and intra-observer variability. |
| Specificity | Not perfect. Positive reading could be caused by other respiratory diseases and past pulmonary tuberculosis infection; subject to inter- and intra-observer variability. |
| Tuberculin skin test | |
| Sensitivity | Not perfect because of anergy (HIV or malnutrition) or other reasons related to limited immune response. In adults, decreased in cases with severe disease. |
| Specificity | Not perfect. False positives expected because of latent tuberculosis infection, bacillus Calmette-Guérin immunization, and nontuberculous Mycobacterial infections. |

Abbreviation: HIV, human immunodeficiency virus.
[a] Liquid culture, Xpert MTB/RIF, and microscopy.

biological principles: directly visualizing tuberculosis bacilli (microscopy), detecting their growth (culture), or amplifying and detecting bacterial DNA (Xpert). Among children with CPTB, we anticipated a positive relation between bacillary burden and the probability of a positive test outcome. This is because a higher bacillary burden is more easily detected by all 3 microbiological tests, whereas a very low bacillary burden is more likely to be missed by all tests.

We also anticipated conditional dependence between the TST results and results from the microbiological tests. With regard to tuberculosis in adults, it has been reported that the TST may be less sensitive in persons with severe disease, which can in turn be associated with high bacillary burden (22–24). Because little is known about the exact functional form of this relationship in children, we allowed for the possibility that this association was nonlinear.

Further, based on the literature and the clinical expertise of our team members, we expected that certain covariates would influence the sensitivity and specificity of the different tests and CPTB prevalence. For example, we expected the sensitivity of the TST to be systematically lower for HIV-infected children than for children not infected with HIV (3, 8). We also expected that some covariates would influence CPTB prevalence. Table 3 lists the covariates and associations of interest. A distinction is made between those associations that are well established and those that remain to be studied.

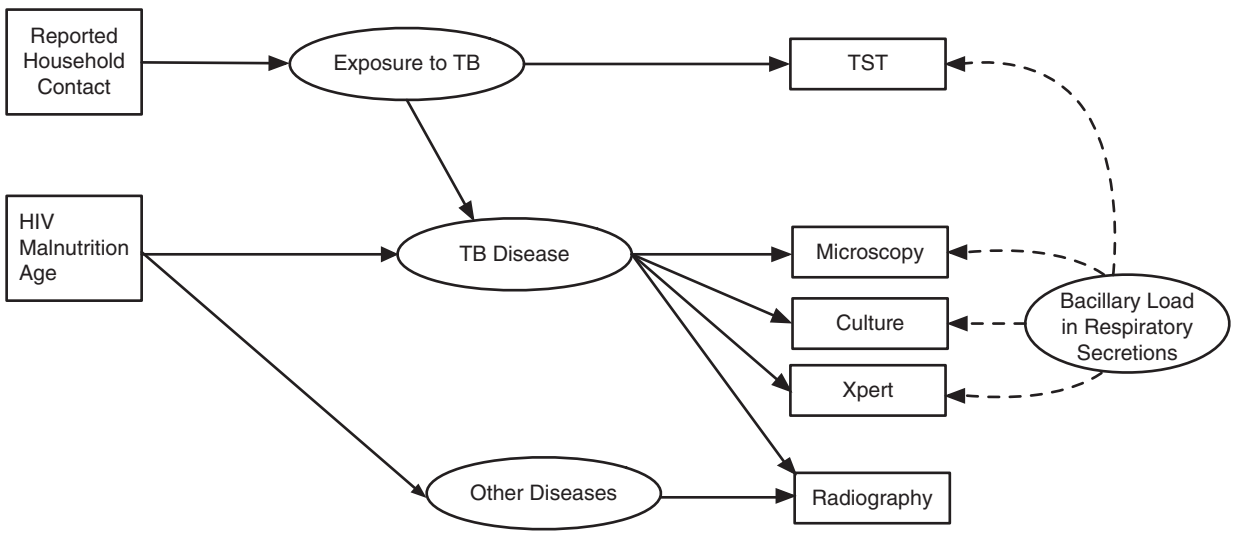


**Figure 1.** Heuristic model specifying prior beliefs of relations between latent and manifest variables. HIV, human immunodeficiency virus; TB, tuberculosis; TST, tuberculin skin test.

**Table 3.** Covariates Potentially Associated With Test Accuracy and Childhood Pulmonary Tuberculosis Prevalence Parameters[a]

| Test and Parameter | Covariate | | | | Bacillary Burden[b] |
|---|---|---|---|---|---|
| | HIV | Age | Malnutrition | Household Tuberculosis Contact | |
| PTB prevalence | Strong | Strong | Strong | Strong | |
| Liquid culture | | | | | |
| Sensitivity | Weak | Weak | | | Strong |
| Specificity | | | | | |
| Xpert MTB/RIF | | | | | |
| Sensitivity | Weak | Weak | | | Strong |
| Specificity | | | | | |
| Microscopy | | | | | |
| Sensitivity | Weak | Weak | | | Strong |
| Specificity | | | | | |
| Radiography | | | | | |
| Sensitivity | Weak | Weak | Weak | | |
| Specificity | Weak | | | | |
| TST | | | | | |
| Sensitivity | Strong | | Strong | | Weak |
| Specificity | | | | | |

Abbreviations: HIV, human immunodeficiency virus; PTB, pulmonary tuberculosis; TST, tuberculin skin test.

[a] Clinical experts indicated "strong" or "weak" belief that the covariate should be included in the model. Empty cells indicate that clinical experts provided no reason to believe that covariate should be included.

[b] Bacillary burden is represented by a random effect.

### Latent class models

Based on the heuristic CPTB model, we determined that the available tests would allow us to classify subjects into 1 of 2 latent classes representing true CPTB-positive and true CPTB-negative subjects. It should be noted that CPTB-negative subjects include those who have different respiratory diseases and CPTB-positive subjects may also include children with co-infections.

We assumed conditional independence of test outcomes for CPTB-negative subjects. Our analyses proceeded stepwise through 4 different latent class models of increasing complexity. Our purpose was to study improvement in model fit and changes in parameter estimates as we proceeded to the model that most closely represented the model illustrated in Figure 1.

Latent class model 1 is a 2-class latent class model based on the assumption of conditional independence of test outcomes within both classes. In all other latent class models that we considered, this conditional independence assumption is relaxed. Latent class model 2 adds to model 1 a continuous random effect representing the unobserved true bacillary burden and its association with the microbiological tests. The sensitivities of the microbiological tests are expressed as functions of the random effect and are therefore allowed to vary with it (see Web Appendix 1 and Web Table 1, available at http://aje.oxfordjournals.org/). Latent class model 3 adds 9 established covariate associations with the sensitivity, specificity, and/or class prevalence (i.e., CPTB prevalence) to model 2. It also includes the association between the random effect and the sensitivity of the TST. We consider model 3 to be our main model.

We considered 1 more elaborate latent class model (model 4) that included the 10 additional covariate associations with test sensitivity, specificity, and prevalence that are not well established but are of potential interest (Table 3). As explained further in the Results section, we considered model 4 to be an exploratory analysis, given the large number of covariate associations considered (19 in total) relative to the available data. Details about the specification of these 4 latent class models are in Web Appendix 1.

Similar to what was done in earlier studies (25, 26), we assume that the random effect is a Gaussian random variable with equal magnitudes of association with the sensitivity of each of the 3 microbiological tests. The covariates and the random effect influence the sensitivities and specificities of individual tests through a probit model (see Web Appendix 1 for details). In order to model the possibly nonlinear relation between TST sensitivity and bacillary burden, we use a quadratic function; all other associations with covariates and random effects are assumed to be linear and additive. The latent class prevalence parameter is allowed to vary with covariates (household contact, age, HIV status, and malnutrition) through a linear binary logistic function. To simplify modeling, age of the child was dichotomized at 24 months of age to separate very young and young children.

We estimated the probability of having CPTB for each combination of test results. The model was also used to

estimate the proportion of overtreatment (the proportion of those who received antituberculosis treatment in the latent class but were CPTB disease negative) and undertreatment (proportion who did not receive antituberculosis treatment in the latent class but were CPTB disease positive) in the cohort.

### Bayesian estimation

We fitted the latent class models to the data using a Bayesian approach. Web Appendix 1 provides details of the form of the likelihood and prior distributions. We used informative prior distributions only on the specificity parameters of the culture test and Xpert. The specificities of these tests are widely acknowledged to be nearly perfect (3, 4). We selected hyperparameters that let the 95% prior credible interval of the specificity parameters for the culture test and Xpert range from 99% to 100% and 98% to 100%, respectively. For all other parameters, we used noninformative prior distributions (Web Appendix 1).

The analytical form of the joint posterior distribution or the marginal posterior distributions of individual parameters cannot be obtained for any of the 4 models considered. Therefore, a Markov chain Monte Carlo approach was used to sample from these distributions. Using the statistical software package JAGS (27) in R (28) for all models, we ran 3 parallel Markov chain Monte Carlo simulations, each with 50,000 iterations. The first 10,000 iterations of each chain were discarded. Convergence was assessed by visual inspection and by examining the Brooks-Gelman-Rubin statistic (29). No convergence problems were identified. To avoid label switching problems between chains (30), the parameters associated with random effect were constrained to positive values among the microbiological tests. R code is available in Web Appendix 2.

Estimation of exploratory model 4 with informative priors on only the sensitivity of Xpert and culture test (as in models 1–3) yielded at least 1 covariate parameter estimated with extremely wide credible intervals. This suggests that with the current priors some parameters of model 4 are not identifiable or only weakly identifiable with these data. In a Bayesian context, defining additional informative priors may help overcome this problem (31–33). We therefore adopted a Bayesian Lasso approach (34) to estimate the 10 additional covariate associations in model 4 (Table 3). This is implemented by placing zero-centered Laplace prior distributions with a diffuse prior on the scale parameter (for details, see Web Appendix 1). The shrinkage is adaptive in the sense that it is proportional to the variance of the parameter estimate, such that parameters that are estimated with poor precision are more likely to be shrunk to the null value.

### Sensitivity analyses

We conducted a series of sensitivity analyses to explore alternative modeling choices. First, to consider the impact of an alternative conditional dependence structure, we fit a 3-class latent class model resulting from treating the random effect as a dichotomous rather than a Gaussian variable. In this model, children with CPTB belong to 1 of 2 groups: "CPTB disease with tuberculosis detectable in respiratory secretions" or "CPTB disease with tuberculosis not detectable in respiratory secretions." In the latter group, we assume the sensitivity of each of the 3 microbiological tests to be 0%, and the sensitivity of the TST is assumed equal for the 2 CPTB disease classes. We compared this model (model 2B) with model 2 for differences in sensitivity, specificity, and prevalence estimates.

Further, the informative prior distributions for the specificity parameters of the culture and Xpert tests were replaced by noninformative priors. To study the consequence of relaxing the assumption that the random effects equally affect the sensitivity of each of the 3 microbiological tests, we conducted an analysis in which this constraint was removed.

## RESULTS

### Model fit

Figure 2 shows the pairwise residual correlations (25) between the test outcomes for the 4 latent class models considered. For models 1 (the conditional independence model) and 2 (conditional dependence assumed only between microbiological tests), substantial residual correlation was found. In comparison, models 3 and 4 had low residual correlation. From Table 4, we can see that for models 1 and 2, the expected frequencies of test outcome patterns substantially deviated from the observed frequencies. For model 3, the expected frequencies of test outcome patterns are close to the observed frequencies; together with low residual correlation, this suggests satisfactory fit of model 3 to the data. The expected frequencies of test outcome patterns for exploratory model 4 were similar to those for model 3 (not shown), despite the addition of 10 covariate associations to the latent class model.

### Estimates of disease prevalence and diagnostic test accuracy

Table 5 summarizes the estimates of test accuracy and CPTB prevalence based on models 1–3. Estimates from model 3 were marginalized over the covariates and random effect. The corresponding estimates for model 4 were very similar to those from model 3 and are not shown.

When ignoring the conditional dependence (i.e., model 1), the sensitivity and specificity of the culture test were estimated to be close to 100%; the estimates for the other tests and CPTB prevalence were therefore close to those obtained from a naïve analysis in which we assumed culture is a perfectly accurate reference standard. Accounting for conditional dependence between the microbiological tests (model 2) provided lower estimates of culture sensitivity. Adjustment for the conditional dependence between the TST and the microbiological tests caused the estimate of TST sensitivity to increase.

From Table 5, we see that by model 3, prevalence of CPTB was estimated (posterior median) at 26.7% (95% credible interval (CrI): 20.8, 35.2). The average sensitivities of the microbiological tests were 60.0% (95% CrI: 45.7, 75.5) for culture, 49.4% (95% CrI: 37.7, 62.2) for Xpert, and 22.3% (95% CrI: 15.6, 30.3) for microscopy. These sensitivities strongly depend on the random effect that

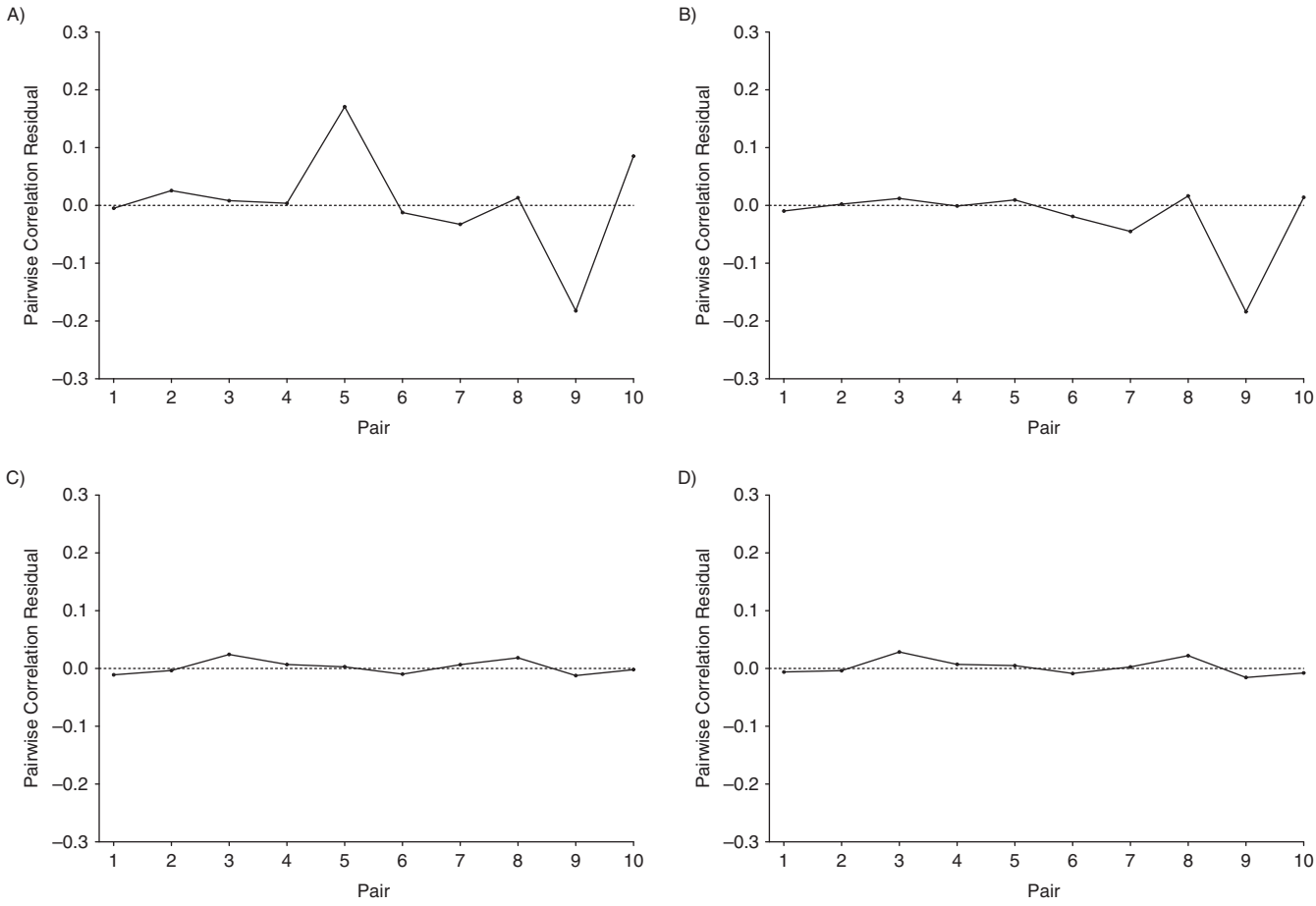


**Figure 2.** Residual correlation plots for models 1–4 (A–D), South Africa, 2009–2014. Residual correlations are computed as the difference between observed and model-predicted correlations between each pair of tests: pair 1, liquid culture and Xpert MTB/RIF; pair 2, liquid culture and microscopy; pair 3, liquid culture and radiography; pair 4, liquid culture and tuberculin skin test; pair 5, Xpert MTB/RIF and microscopy; pair 6, Xpert MTB/RIF and radiography; pair 7, Xpert MTB/RIF and tuberculin skin test; pair 8, microscopy and radiography; pair 9, microscopy and tuberculin skin test; and pair 10, radiography and tuberculin skin test.

represents the unobserved bacillary burden in the sputum of the child (Figure 3A). In accordance with our prior beliefs, the specificities of culture and Xpert were estimated to be nearly 100%; the specificity of microscopy was estimated to be 99.7% (95% CrI: 99.0, 100). The sensitivity of diagnosis by radiography was estimated to be 64.2% (95% CrI: 54.9, 72.8), and the specificity was estimated to be 78.0% (95% CrI: 73.4, 83.4). For the TST, the overall sensitivity was estimated to be 75.2% (95% CrI: 61.2, 83.8), and the specificity was estimated to be 69.3% (95% CrI: 63.2, 75.9).

## Sensitivity analyses

Results of our sensitivity analyses are shown in Web Table 2. Replacing the Gaussian random effects model (model 2) between the microbiological tests with a 3-class latent class model (model 2B, defined in Web Appendix 1) did not affect the results substantially. Therefore, we concluded that our results seemed robust to the choice of the conditional dependence structure and retained the Gaussian random effect in more complex models 3 and 4. Also, relaxing the equal random effects assumption (model 3A) and replacing the informative priors on culture and Xpert specificity parameters with noninformative priors (models 3B and 3C) had little effect on the model parameters.

## Covariate associations

The estimated coefficients under model 3 showed lower TST sensitivity in HIV-infected children. A graphical presentation of the relations of TST sensitivity with the covariate associations and random effect is found in Figure 3B. Sensitivity of the TST dropped significantly at higher and low levels of the random effect and for malnourished and HIV-infected children. Average sensitivity of the TST under model 3 for children who were both malnourished and HIV infected was estimated to be 55.8% (95% CrI: 30.8, 79.2);

**Table 4.** Posterior Median Expected Frequency of Each Combination of Test Results for Models 1–3 and Predicted Probability of Childhood Pulmonary Tuberculosis Based on Model 3, South Africa, 2009–2014

| Test Outcome Pattern | | | | | Observed Frequency | Posterior Median Expected Frequency | | | Predicted Probability of CPTB Based on Model 3 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cu | Xp | Mi | Ra | TS | | Model 1 | Model 2 | Model 3 | % | 95% CrI |
| 0 | 0 | 0 | 0 | 0 | 296 | 278 | 292 | 294 | 2 | 0, 7 |
| 0 | 0 | 0 | 0 | 1 | 149 | 168 | 155 | 151 | 16 | 5, 33 |
| 0 | 0 | 0 | 1 | 0 | 87 | 102 | 90 | 89 | 9 | 0, 34 |
| 0 | 0 | 0 | 1 | 1 | 78 | 62 | 73 | 77 | 52 | 26, 74 |
| 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 11 | 0, 100 |
| 0 | 1 | 0 | 0 | 0 | 5 | 5 | 4 | 4 | 4 | 0, 40 |
| 0 | 1 | 0 | 0 | 1 | 7 | 3 | 4 | 4 | 56 | 0, 100 |
| 0 | 1 | 0 | 1 | 0 | 2 | 2 | 2 | 1 | 12 | 0, 100 |
| 0 | 1 | 0 | 1 | 1 | 2 | 2 | 4 | 4 | 88 | 50, 100 |
| 1 | 0 | 0 | 0 | 0 | 3 | 3 | 4 | 2 | 23 | 0, 100 |
| 1 | 0 | 0 | 0 | 1 | 8 | 5 | 7 | 10 | 93 | 62, 100 |
| 1 | 0 | 0 | 1 | 0 | 1 | 4 | 6 | 1 | 54 | 0, 100 |
| 1 | 0 | 0 | 1 | 1 | 20 | 9 | 13 | 17 | 99 | 90, 100 |
| 1 | 1 | 0 | 0 | 0 | 1 | 6 | 5 | 3 | 100 | 100, 100 |
| 1 | 1 | 0 | 0 | 1 | 17 | 14 | 12 | 15 | 100 | 100, 100 |
| 1 | 1 | 0 | 1 | 0 | 4 | 12 | 10 | 5 | 100 | 100, 100 |
| 1 | 1 | 0 | 1 | 1 | 27 | 26 | 22 | 26 | 100 | 100, 100 |
| 1 | 1 | 1 | 0 | 0 | 8 | 3 | 4 | 10 | 100 | 100, 100 |
| 1 | 1 | 1 | 0 | 1 | 5 | 7 | 10 | 5 | 100 | 100, 100 |
| 1 | 1 | 1 | 1 | 0 | 21 | 6 | 8 | 18 | 100 | 100, 100 |
| 1 | 1 | 1 | 1 | 1 | 7 | 13 | 18 | 9 | 100 | 100, 100 |

Abbreviations: CPTB, childhood pulmonary tuberculosis; CrI, credible interval; Cu, liquid culture; Mi, microscopy; Ra, radiography; TST, tuberculin skin test; Xp, Xpert MTB/RIF.

for those who were only HIV positive, it was estimated to be 61.7% (95% CrI: 41.5, 84.9); for those who were only malnourished, it was estimated to be 74.3% (95% CrI: 58.6, 86.1); and for those who were not malnourished and not infected with HIV, it was estimated to be 80.9% (95% CrI: 69.0, 89.5).

The (marginal) covariate associations estimated by model 4 are tabulated in Web Table 3. Based on this exploratory analysis, only the association between age and sensitivity of radiography stands out. For children older than 24 months of age, the sensitivity of radiography was estimated to be 52.5% (95% CrI: 39.4, 66.0), whereas for children younger than 24 months, it was 75.0% (95% CrI: 62.6, 85.8).

### Posterior probability of CPTB

In addition to the estimates of prevalence and test accuracy, the Bayesian latent class model was used to estimate the posterior probability of CPTB for a given set of test outcomes. The estimated posterior probability of CPTB per test pattern under model 3 is given in Table 4. Not surprisingly, test patterns that included a positive culture generally had a predicted probability of CPTB of 1, with very high precision. Two other patterns associated with a greater than 50% predicted probability were those in which Xpert and radiology alone were positive and in which radiology and TST were positive. However, these estimates were accompanied by wide credible intervals, illustrating the difficulty in diagnosing individual culture-negative children based on the 4 other tests we have considered.

### Under- and overtreatment

Based on model 3, we evaluated potential overtreatment and undertreatment in the cohort. Details about these calculations can be found in Web Appendix 1. Subjects were classified into quintiles based on their posterior probability of CPTB. Within each quintile, we estimated the mean probability of CPTB and the proportion who received antituberculosis treatment. The relationship between these 2 variables is depicted by Figure 4. The steep initial rise of the curve reflects the low treatment threshold applied by clinicians, suggesting that the probability of receiving antituberculosis treatment exceeds 80% even among subjects with probability of CPTB as low as 30%.

**Table 5.** Posterior Median Estimates of Marginalized Sensitivity, Specificity, and Childhood Pulmonary Tuberculosis Prevalence for Models 1–3, South Africa, 2009–2014

| Test and Parameter | Model | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| | Posterior Median Estimate | 95% CrI | Posterior Median Estimate | 95% CrI | Posterior Median Estimate | 95% CrI |
| CPTB prevalence | 16.6 | 15.6, 18.0 | 28.7 | 22.2, 36.3 | 26.7 | 20.8, 35.2 |
| Liquid culture | | | | | | |
| Sensitivity | 96.7 | 87.8, 99.8 | 57.2 | 44.8, 73.5 | 60.0 | 45.7, 75.5 |
| Specificity | 99.8 | 98.9, 100.0 | 99.9 | 99.3, 100.0 | 99.6 | 98.7, 100.0 |
| Xpert MTB/RIF | | | | | | |
| Sensitivity | 74.4 | 66.0, 82.2 | 46.7 | 37.1, 59.1 | 49.4 | 37.7, 62.2 |
| Specificity | 98.3 | 97.0, 99.4 | 98.9 | 97.3, 99.9 | 98.6 | 97.3, 99.5 |
| Microscopy | | | | | | |
| Sensitivity | 33.3 | 25.3, 42.1 | 20.4 | 14.6, 27.9 | 22.3 | 15.6, 30.3 |
| Specificity | 99.8 | 99.2, 100.0 | 99.7 | 99.0, 100.0 | 99.7 | 99.0, 100.0 |
| Radiography | | | | | | |
| Sensitivity | 65.4 | 56.5, 73.8 | 64.7 | 56.0, 73.0 | 64.2 | 54.9, 72.8 |
| Specificity | 73.1 | 69.6, 76.6 | 79.4 | 74.2, 84.9 | 78.0 | 73.4, 83.4 |
| TST | | | | | | |
| Sensitivity | 69.0 | 60.5, 76.7 | 69.3 | 61.1, 76.8 | 75.2 | 61.2, 83.8 |
| Specificity | 62.4 | 58.5, 66.1 | 67.8 | 62.6, 73.4 | 69.3 | 63.2, 75.9 |
| Deviance[a] | 2,366.1 | | 2,141.8 | | 2,013.9 | |

Abbreviations: CrI, credible interval; TST, tuberculin skin test.
[a] Posterior median model deviance.

The proportion of CPTB-positive children within the group of children who received antituberculosis treatment was estimated to 45.8% (95% CrI: 42.4, 48.7), which reflects the level of overtreatment; the proportion of children with CPTB within the group who did not receive antituberculosis treatment was estimated to be 7.0% (95% CrI: 1.8, 17.8), which reflects the level of undertreatment. Additionally, the probability of a CPTB-negative child not receiving treatment was estimated as 42.4% (95% CrI: 34.4, 51.9); conversely, the probability of a CPTB-positive child receiving treatment was 95.5% (95% CrI: 85.6, 99.0), which suggests that nearly all CPTB-positive children did receive treatment.

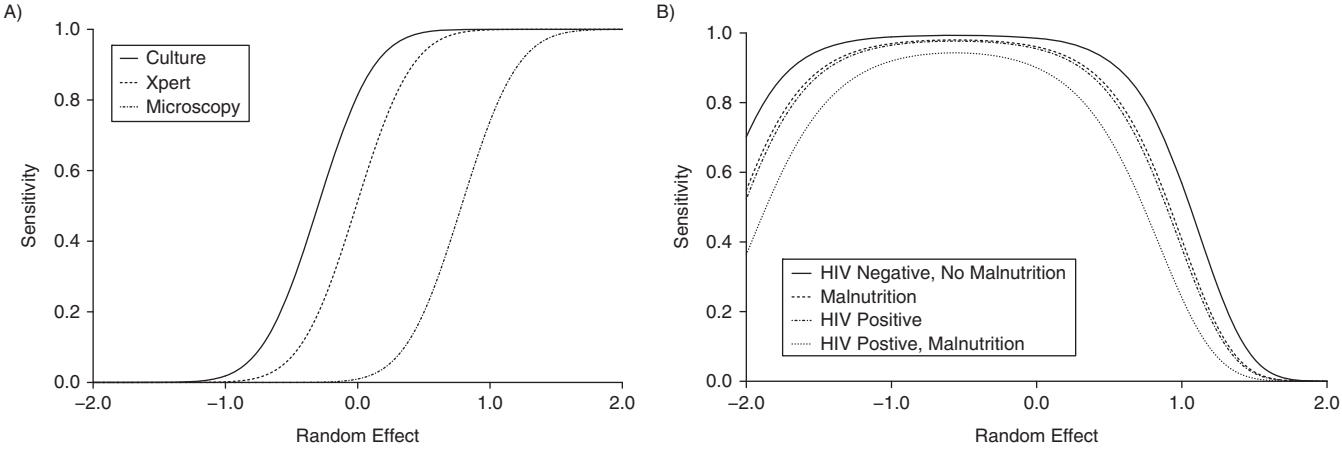


**Figure 3.** Estimated sensitivity as a function of the random effect (latent class model 3), South Africa, 2009–2014. A) Microbiological tests; B) tuberculin skin test. HIV, human immunodeficiency virus.

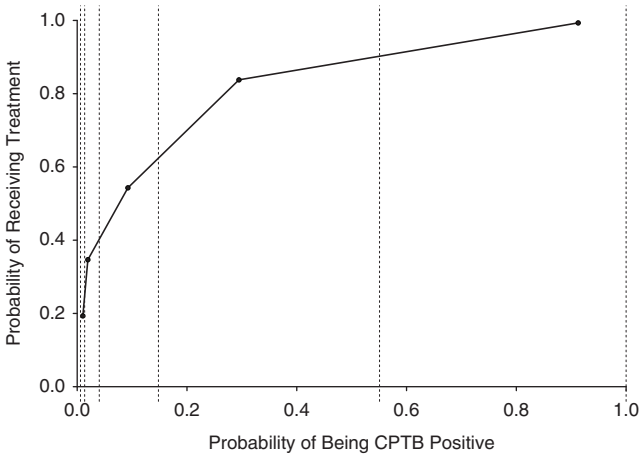


**Figure 4.** The probability of childhood pulmonary tuberculosis (CPTB) and the probability of treatment were estimated within quantiles of the posterior mean probabilities of CPTB for each child based on latent class model 3, South Africa, 2009–2014. Dashed lines are boundaries of quantiles.

## DISCUSSION

We presented a Bayesian latent class analysis in the context of CPTB. Using prospectively collected data from hospitalized children in South Africa who were suspected to have CPTB, we estimated the accuracy of 5 commonly used diagnostic tests and provided estimates of under- and overtreatment in this cohort. The predefined latent class models that incorporated conditional dependence between the 3 microbiological tests and the TST showed good fit to the data. Through sensitivity analyses, we showed that our estimates of accuracy and CPTB prevalence are robust to changes to the prior distributions and the assumed dependence structure.

Our results are in agreement with those from pre-existing reports in which the sensitivities of confirmatory tests for CPTB were low (3). We estimated that a single mycobacterial culture test—generally regarded as the most sensitive confirmatory test for CPTB currently available and often the preferred reference standard—fails to detect almost 40% of all children with a positive estimated CPTB status. The numbers of children with a positive estimated CPTB status missed by Xpert (approximately 50%) and microscopy (approximately 77%) are larger. Sensitivities of the TST- and chest radiography–based diagnoses are somewhat higher than the that of the culture test, although the specificities of these tests are much lower.

We also found evidence that the sensitivity of the microbiological tests depends on the bacterial load in respiratory secretions. Our estimates of sensitivity might therefore not be generalizable to ambulatory settings because children with true CPTB who present in outpatient clinics may be expected to be on average less severely diseased (35) and may therefore (on average) have lower bacillary burden in their respiratory secretions. Sensitivity of the microbiological tests in outpatient settings may thus be lower (35). TST sensitivity is strongly dependent on the immune status and thus is lower in HIV-infected and malnourished children. Our results additionally indicated that TST sensitivity varied with the random effect, providing some evidence that—as in adult tuberculosis—TST sensitivity is lower in persons with more severe CPTB disease.

Because of the lack of an accurate diagnostic testing procedure for CPTB, doctors often make treatment decisions under great uncertainty. Reflecting this uncertainty, a definite CPTB diagnosis based on a clinical CPTB case definition as defined in the original study protocol could not be made for 48% of children suspected to have CPTB in the present study. Taking into account this uncertainty of true CPTB status, using our latent class model, we estimated that in our cohort the proportion of children with a negative estimated CPTB status who received CPTB treatment was approximately 46%, whereas the proportion of children with a positive estimated CPTB status who did not receive CPTB treatment was estimated to be 7% (with a wide credible interval). This points to the possibility of a substantial amount of overtreatment and limited (low) undertreatment, which reflects the use a low implicit threshold probability for a decision to treat hospitalized children for CPTB by the treating doctors in a country with a high HIV prevalence. We stress that this low threshold and consequent high level of overtreatment were likely clinically appropriate in the study cohort because of the high uncertainty of true CPTB, the high prevalence of tuberculosis in this geographical area, and the high morbidity and mortality associated with untreated CPTB.

Diagnostic test evaluation in the absence of an accurate reference standard remains a challenging problem. In recognition of this, the US National Institutes of Health convened an expert panel to propose a uniform clinical case definition for PTB. This panel recently issued revised definitions (36). The proposed PTB case definition contains 3 different classes (confirmed tuberculosis, unconfirmed tuberculosis, and unlikely tuberculosis) based on a set of clinical, radiological, and microbiological criteria. Although development of this case definition is clearly an important step forward, the middle category (unconfirmed tuberculosis) prohibits unambiguous evaluation of diagnostic tests in terms of diagnostic test accuracy and estimation of PTB prevalence and of degrees of possible under- and overtreatment. In future work, we intend to compare the estimates of diagnostic test accuracy obtained from using the 3 National Institutes of Health categories to those obtained from our model. We also intend to evaluate our model in children who present with suspected PTB with less severe disease in an ambulatory setting.

Although our latent class analyses have been carefully designed, we acknowledge that our results depend on the assumptions we have made and that it may be difficult to appreciate the validity of such analyses. However, we have made our assumptions explicit in this paper and presented a variety of sensitivity analyses, and we consider this preferable to making assumptions that are both left implicit and known to be untenable, for example, assuming culture is a perfect test. In the absence of a gold-standard test (i.e., a test with perfect accuracy) for CPTB, making essentially unverifiable assumptions is inevitable when quantifying the

accuracy of the diagnostic tests for CPTB. Similarly, studies on the effectiveness of CPTB treatments are complicated by the absence of methods that can distinguish between children who do and do not suffer from active CPTB.

## ACKNOWLEDGEMENTS

Author affiliations: Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada (Samuel G. Schumacher, Nandini Dendukuri, Lawrence Joseph, Madhukar Pai); Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands (Maarten van Smeden); Division of Medical Microbiology, Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa (Mark P. Nicol); National Health Laboratory Service, Groote Schuur Hospital, Cape Town, South Africa (Mark P. Nicol); McGill International TB Centre, McGill University Health Centre, Montreal, Quebec, Canada (Madhukar Pai); Department of Paediatrics and Child Health, and Medical Research Council Unit on Child & Adolescent Health, University of Cape Town, Cape Town, South Africa (Heather J. Zar); and Red Cross War Memorial Children's Hospital, Cape Town, South Africa (Heather J. Zar).

The first 2 authors contributed equally to this work.

This study was made possible by funding from the Canadian Institutes of Health Research (grant 89857). H.Z. was supported by the National Institutes of Health and the Medical Research Council of South Africa (grant R01HD058971).

Conflict of interest: none declared.

## REFERENCES

1. Jenkins HE, Tolman AW, Yuen CM, et al. Incidence of multidrug-resistant tuberculosis disease in children: systematic review and global estimates. *Lancet*. 2014; 383(9928):1572–1579.
2. World Health Organization. *Global Tuberculosis Control: WHO Report 2010*. World Health Organization; 2010. http://apps.who.int/iris/bitstream/10665/44425/1/9789241564069_eng.pdf.
3. Swaminathan S, Rekha B. Pediatric tuberculosis: global overview and challenges. *Clin Infect Dis*. 2010;50(suppl 3): S184–S194.
4. Zar HJ, Connell TG, Nicol M. Diagnosis of pulmonary tuberculosis in children: new advances. *Expert Rev Anti Infect Ther*. 2010;8(3):277–288.
5. 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–S208.
6. 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–S215.
7. Nelson LJ, Wells CD. Global epidemiology of childhood tuberculosis. *Int J Tuberc Lung Dis*. 2004;8(5):636–647.
8. Perez-Velez CM, Marais BJ. Tuberculosis in children. *N Engl J Med*. 2012;367(4):348–361.
9. 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(11):819–824.
10. Detjen AK, DiNardo AR, Leyden J, et al. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. *Lancet Respir Med*. 2015;3(6):451–461.
11. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282(11):1061–1066.
12. Rutjes AWS, Reitsma JB, Coomarasamy A, et al. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess*. 2007;11(50):iii–ix-51.
13. Reitsma JB, Rutjes AW, Khan KS, et al. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol*. 2009;62(8): 797–806.
14. Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol*. 2013;66(10): 1093–1104.
15. Hesseling AC, Schaaf HS, Gie RP, et al. A critical review of diagnostic approaches used in the diagnosis of childhood tuberculosis. *Int J Tuberc Lung Dis*. 2002;6(12):1038–1045.
16. Hatherill M, Hanslo M, Hawkridge T, et al. Structured approaches for the screening and diagnosis of childhood tuberculosis in a high prevalence region of South Africa. *Bull World Health Organ*. 2010;88(4):312–320.
17. van Smeden M, Naaktgeboren CA, Reitsma JB, et al. Latent class models in diagnostic studies when there is no reference standard—a systematic review. *Am J Epidemiol*. 2014; 179(4):423–431.
18. Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: a review of latent class methods. *Stat Med*. 2014;33(24):4141–4169.
19. Tuyisenge L, Ndimubanzi CP, Ndayisaba G, et al. Evaluation of latent class analysis and decision thresholds to guide the diagnosis of pediatric tuberculosis in a Rwandan reference hospital. *Pediatr Infect Dis J*. 2010;29(2):e11–e18.
20. Zar HJ, Workman L, Isaacs W, et al. Rapid molecular diagnosis of pulmonary tuberculosis in children using nasopharyngeal specimens. *Clin Infect Dis*. 2012;55(8):1088–1095.
21. de Onis M, Onyango AW, Borghi E, et al. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ*. 2007;85(9):660–667.
22. Howard WL, Klopfenstein MD, Steininger WJ, et al. The loss of tuberculin sensitivity in certain patients with active pulmonary tuberculosis. *CHEST*. 1970;57(6):530–534.
23. Holden M, Dubin MR, Diamond PH. Frequency of negative intermediate-strength tuberculin sensitivity in patients with active tuberculosis. *N Engl J Med*. 1971;285(27):1506–1509.
24. Battershill JH. Cutaneous testing in the elderly patient with tuberculosis. *CHEST*. 1980;77(2):188–189.
25. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*. 1996;52(3):797–810.

26. Hadgu A, Qu Y. A biomedical application of latent class models with random effects. *J R Stat Soc Ser C Appl Stat*. 1998;47(4):603–616.
27. Plummer M. *JAGS Version 3.1.0 User Manual*. Lyon, France: International Agency for Research on Cancer; 2011.
28. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2003. http://www.R-project.org.
29. Gelman A, Carlin JB, Stern HS, et al. *Bayesian Data Analysis*. London, UK: Chapmann & Hall/CRC; 2003.
30. Jasra A, Holmes CC, Stephens DA. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat Sci*. 2005;20(1):50–67.
31. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol*. 1995;141(3): 263–272.
32. Gustafson P. On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables. *Stat Sci*. 2005;20(2): 111–140.
33. Gustafson P. The utility of prior information and stratification for parameter estimation with two screening tests but no gold standard. *Stat Med*. 2005;24(8):1203–1217.
34. Park T, Casella G. The Bayesian Lasso. *J Am Stat Assoc*. 2008;103(482):681–686.
35. Zar HJ, Workman L, Isaacs W, et al. Rapid diagnosis of pulmonary tuberculosis in African children in a primary care setting by use of Xpert MTB/RIF on respiratory specimens: a prospective study. *Lancet Glob Health*. 2013;1(2):e97–e104.
36. Graham SM, Cuevas LE, Jean-Philippe P, et al. Clinical case definitions for classification of intrathoracic tuberculosis in children: an update. *Clin Infect Dis*. 2015;61(suppl 3): S179–S187.