INT J TUBERC LUNG DIS 12(8):895–902

Improving the estimation of tuberculosis infection prevalence using T-cell-based assay and mixture models

M. Pai,*† N. Dendukuri,*‡ L. Wang,§ R. Joshi,¶ S. Kalantri,¶ H. L. Rieder#

*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, †Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, Montreal, ‡Technology Assessment Unit, McGill University Health Center, Montreal, Quebec, §Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada; ¶Mahatma Gandhi Institute of Medical Sciences, Sevagram, Maharashtra, India; #International Union Against Tuberculosis and Lung Disease, Paris, France

SUMMARY

BACKGROUND: The prevalence of latent tuberculosis infection (LTBI) is traditionally estimated using the tuberculin skin test (TST). Highly specific blood-based interferon-gamma release assays (IGRAs) are now available and could enhance the estimation of LTBI prevalence in combination with model-based methods.

DESIGN: We compared conventional and model-based methods for estimating LTBI prevalence among 719 Indian health care workers who underwent both TST and QuantiFERON-TB Gold In-Tube (QFT-G). In addition to using standard cut-off points on TST and QFT-G, Bayesian mixture model analyses were performed with: 1) continuous TST data and 2) categorical data using both TST and QFT-G results in a latent class analysis (LCA), accounting for prior information on sensitivity and specificity.

RESULTS: Estimates of LTBI prevalence varied from 33.8% to 60.7%, depending on the method used. The mixture model based on TST alone estimated the prevalence at 36.5% (95%CI 28.5–47.0). When results from both tests were combined using LCA, the prevalence was 45.4% (95%CI 39.5–51.1). The LCA provided additional results on the sensitivity, specificity and predictive values of joint results.

CONCLUSION: The availability of novel, specific IGRAs and development of methods such as mixture analyses allow a more realistic and informative approach to prevalence estimation.

KEY WORDS: tuberculosis; prevalence; tuberculin skin test; interferon-gamma release assay; mixture model; latent class analysis

NEARLY A THIRD of the world's population is estimated to be infected with *Mycobacterium tuberculosis*.[1] In populations such as health care workers in developing countries, the prevalence of latent tuberculosis infection (LTBI) has been estimated to be about 50%.[2,3] Such prevalence estimates are used to quantify the extent of tuberculosis (TB) transmission, ascertain time trends and evaluate control programmes.[4,5] However, given the lack of a gold standard test for LTBI, there is no guarantee that prevalence estimates are accurate.

LTBI prevalence is traditionally estimated using the tuberculin skin test (TST). Although the TST is useful in clinical practice, it has several limitations, including variable specificity attributable to cross-reactivity with bacille Calmette-Guérin (BCG) vaccination and infection with non-tuberculous mycobacteria.[6,7]

For the first time, an alternative to the TST has emerged in the form of T-cell based interferon-gamma (IFN-γ) release assays (IGRAs).[8,9] Two commercial IGRAs are available—QuantiFERON-TB Gold In-Tube (QFT-G)® (Cellestis Ltd, Carnegie, VIC, Australia) and T-SPOT.*TB*® (Oxford Immunotec, Oxford, UK). Although the specificity of IGRAs is definitely higher than TST, their sensitivity is probably comparable to TST.[8–10] Lack of a gold standard for LTBI makes it difficult to estimate the accuracy of both TST and IGRAs. There is thus uncertainty around LTBI prevalence estimates, especially as both tests are imperfect, and little is known about the validity of IGRA cut-offs.[9–12]

In addition to the inherent limitations of the TST, there are limitations with the approaches used to convert TST data into prevalence estimates. Although the TST provides continuous data (induration in mm),[13] the prevalence of LTBI is usually estimated by dichotomising the results using cut-offs such as $\geqslant 5$, $\geqslant 10$ and $\geqslant 15$ mm, depending on risk.[14] This approach amounts to assuming the test characteristics to be 100% sensitive and specific. Furthermore, a cut-off approach underutilises the available data. Both

M. Pai and N. Dendukuri contributed equally to this study.

Correspondence to: Madhukar Pai, Department of Epidemiology, Biostatistics & Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Canada H3A 1A2. Tel: (+1) 514 398 5422. Fax: (+1) 514 398 4503. e-mail: madhukar.pai@mcgill.ca

Article submitted 3 January 2008. Final version accepted 21 March 2008.

commercial IGRAs use cut-offs for LTBI diagnosis and they, too, underutilise the continuous data on T-cell IFN-γ response.

Recognising these limitations, a few studies have used modelling approaches, called mixture models, to estimate prevalence using TST data.[15–17] In the infectious diseases literature, there is growing interest in another type of mixture model, called a latent class model, for analysing the results of multiple dichotomised tests.[18] Such models have also been applied to TB data.[19] Latent class analysis (LCA) is based on the notion that the observed results of various imperfect tests for the same disease are influenced by a common, underlying latent (unobserved) variable, the true disease status. Increasing the number of imperfect tests increases our knowledge of the latent disease status, analogous to a large dark room becoming more illuminated with every additional light turned on.[18]

In this study, we use the results from a previously established cohort, illustrate the application and interpretation of two mixture models and compare them with traditional approaches to estimating LTBI prevalence.

METHODS

Study design

In 2004, we established a cohort of health care workers at the Mahatma Gandhi Institute of Medical Sciences (MGIMS), a rural medical school in India.[20] Between January and May 2004, 719 health care workers (median age 22 years, 62% women) underwent TST and IGRA testing after providing written informed consent. Approval for this study was obtained from the ethics committee of the MGIMS. This cohort was comprised of 352 (49%) medical students and nursing students, 73 (10%) interns and residents, 160 (22%) nurses, 12 (2%) attending physicians/faculty, and 122 (17%) orderlies and laboratory workers. About 71% of the cohort had BCG vaccine scars.

Tuberculin skin test

TST was performed using 1 tuberculin unit (TU) of purified protein derivative (PPD) RT23 (Statens Serum Institut, Copenhagen, Denmark), the standard dosage used in India[21] and the dosage originally recommended by the World Health Organization (WHO).[13] One TU of PPD was administered intradermally by a certified technician and the induration was read after 48–72 h using a blinded caliper.

QFT-G assay

The QFT-G assay was performed as per the manufacturer's recommendations. IFN-γ values (international units [IU] per ml) for TB-specific antigens and mitogen were corrected for background by subtracting the value obtained for the respective negative control. Valid QFT-G results were obtained in all subjects and no indeterminate results were noted. Because the QFT-G enzyme-linked immunosorbent assay (ELISA) cannot accurately resolve the IFN-γ values when they exceed 10 IU/ml, values larger than 10 IU/ml were treated as 10 IU/ml in all the analyses.

Methods for estimation of LTBI prevalence

1) Cut-off point methods using dichotomised TST and QFT-G results

For TST, we used the standard 5 mm, 10 mm and 15 mm cut-off points.[14] For QFT-G, we used the cut-off point of IFN-γ $\geqslant$ 0.35 IU/ml, as recommended by the manufacturer.[22,23] We calculated 95% confidence intervals (CIs) for each prevalence estimate using the method based on the normal approximation.

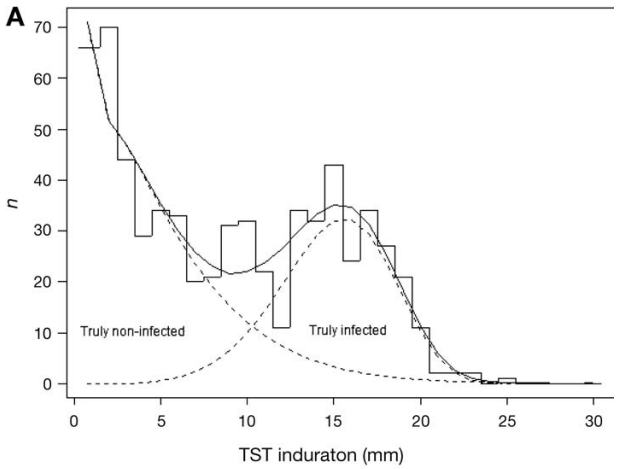
2) Mixture models

We implemented two different mixture models: 1) a mixture model for continuous TST data; and 2) a latent class model using the joint dichotomised results of TST and QFT-G tests. We chose not to fit a mixture model for the continuous QFT-G data, as the statistical probability distribution of the continuous IFN-γ data did not appear to be one of the standard distributions that were dealt with by the available software programme.[24]

There are some aspects that are common to both models. Both models assumed that while the observed data arise from two groups, i.e., truly infected and truly not infected, the group membership variable is unobserved (latent). Thus, under these models, the group of patients with a high test value, e.g., a tuberculin induration of 14 mm or a QFT-G result of IFN-γ 0.45 IU/ml, would not automatically be all classified as positive. Instead, they would be treated as a mixture of truly infected and non-infected individuals.

In Figure 1 we illustrate how the observed data are assumed to be split into infected and non-infected groups under the two models. In Figure 1A, the dashed lines indicate the distribution of TST among the infected and non-infected groups. The goal of this mixture model is to estimate the parameters of each distribution. In Figure 1B, we see how each cell in the cross-tabulation between TST and QFT-G can be broken up into infected and non-infected persons. The goal of this latent class model is to estimate the proportion of infected and non-infected patients in each cell. These proportions can be expressed in terms of the sensitivity and specificity of each test, and the prevalence.

The other common feature of both models is that they were estimated using a Bayesian approach (reviewed elsewhere).[18,24–27] The Bayesian approach requires that each unknown parameter in the model has a prior distribution (Table 1 shows the priors used for both tests). For example, based on the LTBI literature, we can reasonably say that the sensitivity of the TST lies in the range of 75–90% (Table 1).[6,8–10] This information can be summarised as a statistical probability distribution, as illustrated in Figure 2A. If no prior



B

Observed data		
	QFT-G +	QFT-G –
TST +	226	62
TST –	72	359

Truly infected		
	QFT-G +	QFT-G –
TST +	y11	y10
TST –	y01	y00

Truly non-infected		
	QFT-G +	QFT-G –
TST +	226-y11	62-y10
TST –	72-y01	359-y00

Figure 1 **A.** Mixture model for continuous TST data: distributions of TST reactions in LTBI positive and negative groups (n = 719). The X-axis displays the TST induration in mm and the Y-axis displays the frequency. The dashed lines show the estimated probability density of TST results for the cross-reactors and infected groups. The solid line is a smoothed density plot approximating the histogram of the observed frequency distribution data. **B.** Latent class model for dichotomous TST and QFT-G data: cross-tabulation of results in the truly infected and non-infected groups. The number of truly infected individuals in each cell of the cross-tabulation is denoted by y11, y10, y01 and y00. These numbers are unobserved (or latent). TST = tuberculin skin test; QFT-G = QuantiFERON-TB Gold In-Tube; LTBI = latent tuberculosis infection.

information is available, or if we prefer that our results are not influenced by prior information, we may choose to use a 'non-informative' prior distribution. For example, in both types of models discussed below we used a non-informative prior distribution for the prevalence of LTBI, allowing for equal weight of all values from 0% to 100% (Figure 2A).

1) MIXTURE MODEL USING CONTINUOUS TST RESULTS: We fit this model to the TST data, using *R*-statistical programmes developed for the International Union Against Tuberculosis and Lung Disease (The Union).[24] The unknown parameters in this model are the percentage of patients in the truly infected and non-infected groups, and the parameters of the distribution (e.g., mean and variance) of TST results within each group. The software package requires the user to select the statistical probability distribution of TST values within the infected and non-infected groups (details are provided in the Appendix). This programme automatically uses non-informative prior distributions for all parameters. Whereas, in theory, mixture models can be fitted to continuous data from multiple tests, the programme we used was able to fit models for results from a single test only. Mixture models with TST data have been successfully used in many settings, even in populations where TB infection rates were low (i.e., a large proportion of zero TST values).[15]

Table 1 Prior information on sensitivity and specificity of tuberculin skin test and QuantiFERON-TB Gold In-Tube tests*

Parameter	Prior distribution (95%CrI)
TST sensitivity	75–90
TST specificity	70–90
QFT-G sensitivity	75–90
QFT-G specificity	95–100

* Prior estimates were derived from previous systematic reviews and meta-analyses.[6–10]
CrI = credibility interval; TST = tuberculin skin test; QFT-G = QuantiFERON-TB Gold In-Tube assay.

2) LCA USING DICHOTOMISED TST AND QFT-G RESULTS: We used the cut-offs of 10 mm for TST and 0.35 IU/ml for QFT-G to define the dichotomous tests.

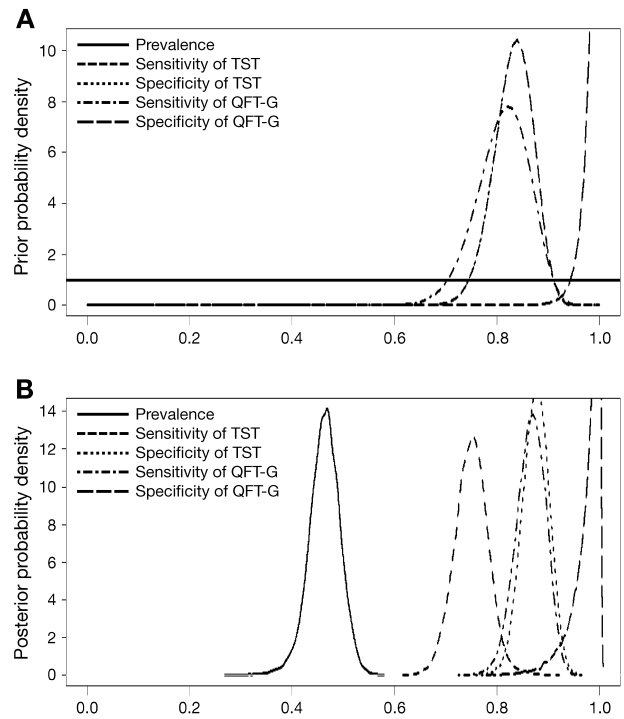


Figure 2 Prior and posterior distributions in LCA. **A.** Prior probability distributions for LTBI prevalence, TST and QFT-G accuracy (based on previous literature). These prior distributions reflect the relative importance of values between 0 and 1 for prevalence, sensitivity and specificity before LCA. **B.** Posterior distributions after LCA. These posterior distributions reflect the effect of updating the prior distributions with the observed data. For example, the distribution of the prevalence (solid black line in **A**) has changed from being uniform across the (0,1) range prior to LCA, to a more peaked distribution about 45.4% (solid black line in **B**). LCA = latent class analysis; LTBI = latent tuberculosis infection; TST = tuberculin skin test; QFT-G = QuantiFERON-TB Gold In-Tube.

For the QFT-G assay, we used the standard cut-off provided by the manufacturer. For TST, we used the 10 mm cut-off based on the original study, where this cut-off had the best agreement with QFT-G and was also associated with known risk factors for LTBI.[20] The LCA was implemented using Bayes Latent Class Models (BLCM), a user-friendly statistical programme available from the website of one of the authors.[28] This is the only method for which we discuss how results of both TST and QFT-G tests can be used simultaneously to estimate disease prevalence.

The unknown parameters in this model were the prevalence, and the sensitivity and specificity of the two tests. For this model, prior information was needed on a minimum of two parameters.[27] We used the prior information on the sensitivity and specificity parameters listed in Table 1 (technical details are presented in the Appendix). Although our primary focus was the prevalence of LTBI, the LCA model also provided estimates of the sensitivity and specificity of the tests, and the positive predictive value for each combination of test results, along with 95% credible intervals (CrI).*

RESULTS

Cut-off point methods using dichotomised TST and QFT-G results

Valid TST and QFT-G results were both available for a total of 719 health care workers. Table 2 shows the estimates of LTBI prevalence, obtained by using cut-off point based analyses of TST and QFT-G data. With the TST, the prevalence estimate was 60.7% with a low TST cut-off of 5 mm, and 23.2% with a high cut-off of 15 mm. With a 10 mm cut-off, the LTBI prevalence estimate was 41.4%. With QFT-G, the manufacturer's cut-off resulted in a prevalence estimate of 40.1%.

Mixture models

Mixture model using continuous TST results

The output of the mixture model based on continuous TST results is shown in Figure 1A. The dashed lines show the overlapping TST density plots among the truly infected and not infected groups. The solid line is a smoothed density plot of all observed TST results. The estimate of the prevalence of LTBI from this model was 36.5%. This is essentially the percentage of individuals whose TST values fall under the density plot on the right. By default, the statistical programme assumes that there are no false-negatives, i.e., all subjects with a 0 mm induration (10.4%) are automatically classified as truly non-infected. We can therefore estimate the percentage of cross-reactors as 100% − 36.5% − 10.4% = 53.1%. The median of the TST values among the infected group was 15.1 (95%CrI 14.1–15.9), while among the cross-reactors it was 4.03 (95%CrI 3.11–4.89).

* CrIs are the Bayesian analogue of CIs.

Table 2 Estimates of LTBI prevalence from the different methods

Method used to estimate prevalence	LTBI prevalence %	95%CI or CrI %
Cut-off point based analysis of TST data		
TST (⩾5 mm cut-off point)	60.7	57.1–64.2
TST (⩾10 mm cut-off point)	41.4	37.7–44.9
TST (⩾15 mm cut-off point)	23.2	20.1–26.3
Cut-off point based analysis of QFT-G data		
QFT-G (IFN-γ ⩾ 0.35 IU/ml, manufacturer's cut-off point)	40.1	36.6–43.7
Mixture analysis of continuous TST data		
Mixture model of TST (assuming Weibull distributions for both infected and cross-reacting subgroups)	36.5	28.5–47.0
LCA of TST and QFT-G data		
LCA (using prior information on TST and QFT-G)	45.4	40.1–49.7

LTBI = latent tuberculosis infection; CI = confidence interval; CrI = credible interval; TST = tuberculin skin test; QFT-G = QuantiFERON-TB Gold In-Tube assay; LCA = latent class analysis.

Two other useful plots from this model are shown in Figure 3. In Figure 3A, we have a plot of the relation between the probability of infection and induration. The probability of infection increases from 40% at 10 mm induration to 92% at 19 mm induration. Figure 3B is a receiver operating characteristic (ROC) plot of sensitivity vs. 1-specificity for each possible TST cut-off point. The plot shows that the optimal combination of sensitivity and specificity of 92% was obtained at 10 mm induration.

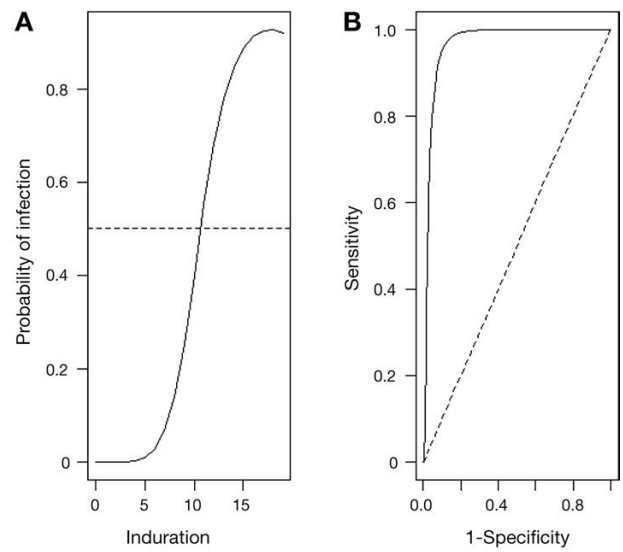


Figure 3 **A.** Plot of the probability of tuberculosis infection at each induration as estimated by the selected mixture model for continuous TST data. The dashed line indicates the probability of infection of 50%. **B.** Receiver operating characteristic (ROC) curve from the same model plotting the sensitivity corresponding to 1-specificity across the induration scale. The dashed line denotes the 45 degree line corresponding to a test with sensitivity = specificity = 50%. Note: optimal combination of sensitivity and specificity of 92% was obtained at 10 mm induration. TST = tuberculin skin test.

LCA using dichotomised TST and QFT-G results

The cross-tabulation of the TST and QFT-G results on which the LCA model was based was: TST+/QFT-G+, 226; TST+/QFT-G–, 62; TST–/QFT-G+, 72; TST–/QFT-G–, 359. Based on the LCA model, the prevalence estimate was 45.4% (Table 2). A plot of the posterior density for the prevalence is shown in Figure 2B. This figure shows that the distribution of the prevalence has changed from being uniform across the (0,1) range prior to using the data, to a more peaked distribution about 45.4%. Using this distribution we were also able to determine a 95%CrI for the prevalence. Based on the CrI, there is a 95% probability that the prevalence of LTBI lies between 40.1% and 49.7%.

In addition to the prevalence, the LCA also provided estimates of the sensitivity and specificity of both tests, and the correlation between the tests within classes defined by infection status (Table 3 and Figure 2B). The estimate of the sensitivity of TST was lower than its prior distribution, while the sensitivity of QFT-G was higher. The median specificity of TST increased closer to 87%. We calculated predictive values based on the prevalence, sensitivity and specificity. For example, an individual testing positive by both TST and QFT-G is estimated to have a 99% probability of having LTBI, as compared to a 2% probability if both tests are negative (Table 3). An individual testing TST-positive and QFT-G-negative is estimated to have a 46% probability of having LTBI, as compared to an 85% probability for an individual testing TST-negative and QFT-G-positive.

DISCUSSION

Prevalence and annual risk of LTBI is often used to determine the extent of TB transmission and TB risk trends over time.[4,5,29,30] However, because there is no gold standard for LTBI, estimation of prevalence relies on cut-off point based analyses of the TST.[4,29,30] The TST has limitations, and there are limitations with the approaches used to dichotomise TST data. For example, to account for the frequently recognised deficiency in specificity with a cut-off of ⩾10 mm induration, tuberculin surveys have used methods such as mirror image, or other cut-offs, similarly correcting the loss of sensitivity by the gain in specificity to estimate LTBI prevalence and annual risk of infection.[4,29–31] However, these methods effectively reduce to a cut-off point analysis.

Table 3 Results on positive predictive values, sensitivity and specificity from latent class analysis model

	Posterior distribution	
Variable	Median %	95%CrI
P (LTBI+|TST+, QFT-G+)	99.2	99.0–100.0
P (LTBI+|TST+, QFT-G−)	46.0	29.0–65.0
P (LTBI+|TST−, QFT-G+)	85.0	69.0–94.0
P (LTBI+|TST−, QFT-G−)	2.0	1.0–4.0
Sensitivity of TST	79.5	74.9–84.4
Specificity of TST	87.4	82.3–91.8
Sensitivity of QFT-G	89.9	86.1–93.7
Specificity of QFT-G	97.4	94.2–98.9

CrI = credible interval; LTBI = latent tuberculosis infection; TST = tuberculin skin test; QFT-G = QuantiFERON-TB Gold In-Tube assay.

The availability of model-based techniques offers a more realistic approach to prevalence estimation that accounts for the imperfect nature of the test, and allows simultaneous analysis of multiple imperfect tests. These models also provide estimates of sensitivity, specificity and predictive values. IGRAs are also substantially more specific than the TST.[10] Incorporation of IGRAs therefore offers yet another option for improving the estimation of LTBI prevalence, especially in settings where BCG affects TST specificity.[7] However, because IGRAs are not perfect, they cannot be used as a standard to calibrate TST.

In this analysis, we used TST and QFT-G results from a large cohort of health care workers to compare various approaches for estimating prevalence. Although cut-off methods are easy to use, the choice of the cut-off is subjective and different cut-offs provide different prevalence estimates. Furthermore, cut-off approaches do not provide any additional statistics such as sensitivity, specificity or predictive values. The two mixture models required carefully considered assumptions, but were fairly straightforward to apply given the availability of software. The LCA model required prior knowledge about the accuracy of the individual tests, and these were derived from systematic reviews.[6,8–10] Both mixture models provide several other statistics in addition to prevalence.

Our results showed that estimates of prevalence varied widely, depending on the method. This suggests that prevalence estimates from different surveys may produce heterogeneous results, at least in part because of the methods and tests used. The cut-off based methods all provided prevalence estimates of around 40%. Based on TST results alone, both model-based results gave similar estimates of the prevalence of around 36.5%; when results from both tests were combined using LCA, the estimated prevalence was 45.4%. Estimates of TST sensitivity and specificity at 10 mm induration from the two models were also different—sensitivity was 92% based on the continuous mixture model compared to 79.5% based on LCA, while specificity was 92% based on the continuous mixture model compared to 89.9% based on LCA. The difference in the results was in part because the latter model took into account the observed results of both tests, as well as prior information that the QFT-G specificity was higher than that of TST at 10 mm induration.

The LCA provided predictive values that may be helpful when both TST and QFT-G results are available. An individual positive by both tests had a 50 times higher likelihood of having LTBI than an individual negative by both tests. The model also suggests

that an individual with a TST-negative/QFT-G-positive discordant result had a high likelihood (85% probability) of having LTBI, and this could be driven by the higher specificity of QFT-G. Thus, estimates from LCA could be useful in clinical decision making.

The choice of a particular model will be guided by the type of data available and whether model assumptions are satisfied. Both models have their advantages and disadvantages. The mixture model for continuous data has the advantage of using all of the collected information on the continuous test results. On the other hand the user needs to make a careful choice of the probability distribution which, if mis-specified, could bias the prevalence estimate. Moreover, while we can incorporate prior information on the parameters of these probability distributions or the distribution of the prevalence, we cannot incorporate prior information on test sensitivity and specificity.

The advantage of LCA is that it allows us to account for prior information on prevalence, sensitivities and specificities. However, this approach is based on dichotomous test results that do not use all the information from continuous test results. This is a limitation. It involves fewer assumptions about the probability distribution of the data and can be more easily extended to multiple tests. Both types of models are sensitive to choice of prior information. This is particularly the case when the number of tests available is small. With increasing numbers of tests, the observed data begin to dominate any prior information. Future studies should evaluate if LCA with three tests (QFT-G, T-SPOT.*TB* and TST) will improve the estimation of prevalence. In general, both types of models can be extended to the case of multiple tests, to the case when there are more than two latent classes,[32] and to incorporate covariates that may affect prevalence.[33]

In conclusion, we have shown that traditional cut-off point methods, although easy to implement, have several limitations. On the other hand, statistical models incorporating more than one test, while providing more informative and useful results, are sensitive to assumptions and require software and expertise. We were limited by the available software in our ability to apply the continuous mixture model to QFT-G results and to the joint TST and QFT-G results. While it is theoretically feasible to build mixture models that can handle multiple continuous test results, such models are very difficult to implement in practice. In particular, the problems we encountered were: 1) the poorly understood frequency distribution of IFN-γ—a highly skewed distribution with a large proportion of zero values and a long tail of positive values; and 2) large IFN-γ values are not precisely measured by the QFT-G ELISA—thus the right tail of the distribution is poorly resolved. We are currently pursuing methodological approaches that will allow us to use non-parametric continuous data distributions in LCA models. Lastly, there is a need for population-based surveys using IGRAs.[12] IGRAs may enable researchers to revisit and revise some of the risk and rate estimates traditionally used in TB epidemiology,[12] and enable better monitoring of TB trends.[5]

Acknowledgements

This work was supported in part by a grant from the Canadian Institutes of Health Research (CIHR-MOP-81362). MP is a recipient of a CIHR New Investigator Career Award. CIHR had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

The *R* programmes for mixture model analyses of TST results can be downloaded from The Union Tuberculosis Department website, http://www.tbrieder.org/. BLCM software for latent class analysis can be downloaded from http://www.medicine.mcgill.ca/epidemiology/dendukuri/index.html. The development of this software was supported by the United Nations Children's Fund/United Nations Development Programme/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), Geneva.

References

1 Dye C, Scheele S, Dolin P, Pathania V, Raviglione M C. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 1999; 282: 677–686.
2 Joshi R, Reingold A L, Menzies D, Pai M. Tuberculosis among health care workers in low- and middle-income countries: a systematic review. PLoS Med 2006; 3: e494.
3 Menzies D, Joshi R, Pai M. Risk of tuberculosis infection and disease associated with work in health care settings. Int J Tuberc Lung Dis 2007; 11: 593–605.
4 Arnadottir T, Rieder H L, Trebucq A, Waaler H T. Guidelines for conducting tuberculin skin test surveys in high prevalence countries. Tubercle Lung Dis 1996; 77 (Suppl 1): 1–19.
5 Dye C, Bassili A, Bierrenbach A, et al. Measuring tuberculosis burden, trends, and the impact of control programmes. Lancet Infect Dis 2008; 8: 233–243.
6 Menzies R I. Tuberculin skin testing. In: Reichman L B, Hershfield E S, eds. Tuberculosis: a comprehensive international approach. New York, NY, USA: Marcel Dekker, 2000: pp 279–322.
7 Farhat M, Greenaway C, Pai M, Menzies D. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? Int J Tuberc Lung Dis 2006; 10: 1192–1204.
8 Pai M, Riley L W, Colford J M Jr. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis 2004; 4: 761–776.
9 Pai M, Kalantri S, Dheda K. New tools and emerging technologies for the diagnosis of tuberculosis: Part 1. Latent tuberculosis. Expert Rev Mol Diagn 2006; 6: 413–422.
10 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: 340–354.
11 Pai M, Kalantri S, Menzies D. Discordance between tuberculin skin test and interferon-gamma assays. Int J Tuberc Lung Dis 2006; 10: 942–943.
12 Pai M, Dheda K, Cunningham J, Scano F, O'Brien R. T-cell assays for the diagnosis of latent tuberculosis infection: moving the research agenda forward. Lancet Infect Dis 2007; 7: 428–438.
13 Deck F, Guld J. The WHO tuberculin test. Bull Int Union Tuberc 1964; 34 (1): 53–70.
14 American Thoracic Society. Targeted tuberculin testing and

treatment of latent tuberculosis infection. Am J Respir Crit Care Med 2000; 161 (Suppl): S221–S247.
15 Neuenschwander B E, Zwahlen M, Kim S J, Engel R R, Rieder H L. Trends in the prevalence of infection with *Mycobacterium tuberculosis* in Korea from 1965 to 1995: an analysis of seven surveys by mixture models. Int J Tuberc Lung Dis 2000; 4: 719–729.
16 Neuenschwander B E, Zwahlen M, Kim S J, Lee E G, Rieder H L. Determination of the prevalence of infection with *Mycobacterium tuberculosis* among persons vaccinated with bacillus Calmette-Guerin in South Korea. Am J Epidemiol 2002; 155: 654–663.
17 Trebucq A, Guerin N, Ali Ismael H, Bernatas J J, Sevre J P, Rieder H L. Prevalence and trends of infection with *Mycobacterium tuberculosis* in Djibouti, testing an alternative method. Int J Tuberc Lung Dis 2005; 9: 1097–1104.
18 Hadgu A, Dendukuri N, Hilden J. Evaluation of nucleic acid amplification tests in the absence of a perfect gold-standard test: a review of the statistical and epidemiologic issues. Epidemiology 2005; 16: 604–612.
19 Scott A N, Joseph L, Belisle P, Behr M A, Schwartzman K. Bayesian modelling of tuberculosis clustering from DNA fingerprint data. Stat Med 2008; 27: 140–156.
20 Pai M, Gokhale K, Joshi R, et al. *Mycobacterium tuberculosis* infection in health care workers in rural India: comparison of a whole-blood, interferon-γ assay with tuberculin skin testing. JAMA 2005; 293: 2746–2755.
21 Chadha V K, Jagannatha P S, Vaidyanathan P S, Jagota P. PPD RT23 for tuberculin surveys in India. Int J Tuberc Lung Dis 2003; 7: 172–179.
22 Mori T, Sakatani M, Yamagishi F, et al. Specific detection of tuberculosis infection: an interferon-gamma-based assay using new antigens. Am J Respir Crit Care Med 2004; 170: 59–64.
23 Mazurek G H, Jereb J, Lobue P, Iademarco M F, Metchock B, Vernon A. Guidelines for using the QuantiFERON-TB Gold test for detecting *Mycobacterium tuberculosis* infection, United States. MMWR Recomm Rep 2005; 54: 49–55.
24 Neuenschwander B E. Bayesian mixture analysis for tuberculin induration data. Paris, France: International Union Against Tuberculosis and Lung Disease, 2007. http://www.tbrieder.org Accessed May 2008.
25 Goodman S N. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med 1999; 130: 1005–1013.
26 Goodman S N. Introduction to Bayesian methods. I: measuring the strength of evidence. Clin Trials 2005; 2: 282–290; discussion 301–304, 64–78.
27 Joseph L, Gyorkos T W, 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: 263–272.
28 Dendukuri N. WinBUGS programs for modeling results of two conditionally dependent, non-gold standard diagnostic tests using latent class analysis. Montreal, Canada: McGill University, 2007. http://www.medicine.mcgill.ca/epidemiology/dendukuri/stat%20software.htm Accessed May 2008.
29 Rieder H. Annual risk of infection with *Mycobacterium tuberculosis*. Eur Respir J 2005; 25: 181–185.
30 Rieder H L. Methodological issues in the estimation of the tuberculosis problem from tuberculin surveys. Tubercle Lung Dis 1995; 76: 114–121.
31 World Health Organization. Generic guidelines for the estimation of the annual risk of tuberculosis infection. New Delhi, India: WHO, SEARO, 2006.
32 Dendukuri N, Hadgu A, Wang L. Modeling conditional dependence between diagnostic tests due to multiple latent variables: a hierarchical latent class model. In: Proceedings of the Joint Statistical Meetings of the American Statistical Association, 2005. Minneapolis, MN, USA: ASA, 2005.
33 Hadgu A, Qu Y. A biomedical application of latent class models with random effects. Applied Statistics 1998; 47: 603–616.

APPENDIX

Mixture model for continuous TST results

The analysis was carried out using a library for the *R*-statistical package developed by B Neuenschwander for the The Union (http://www.tbrieder.org/). The software is freely available along with a manual.

Mixture analysis provides a framework for analysing data arising from different subgroups. It is generally not known to which subgroup an individual belongs (i.e., group membership is unknown). However, the number of subgroups is usually known. Moreover, the type of distribution for the subgroups can be approximated by some well-known distribution (e.g., the normal or lognormal distribution). If the observed data meet these assumptions, estimation of mixture models is feasible.

The programme requires users to specify the statistical probability distribution of TST induration results among infected and non-infected patients. Three probability distributions are allowed by the software programme: normal, lognormal or Weibull. The normal distribution is symmetric, the lognormal is always skewed to the right, and the Weibull distribution is very flexible and can be symmetric or skewed in either direction depending on its shape and scale parameters. Based on the histogram of the observed data, we felt a probability distribution skewed to the right was suitable among the cross-reactors and a symmetric distribution was suitable for the infected subjects. We selected a Weibull distribution for TST scores in both groups. This was also supported by a statistical criterion reported by the software programme, the log-likelihood, which attained its highest value for this model (data not shown).

LCA of dichotomised TST and QFT-G results

LCA is based on the notion that the observed results of various imperfect tests for the same disease are influenced by a common, underlying latent (unobserved) variable, the true disease status. Increasing the number of imperfect tests increases our knowledge of the latent disease status. One medical application of LCA is the evaluation of diagnostic tests in the absence of a gold standard. For example, if one has several tests for detecting the presence/absence of a disease, but no comparison 'gold standard' that indicates disease status with certainty, LCA can be used to provide estimates of diagnostic accuracy (sensitivity, specificity, predictive value, etc.) of the different tests.

LCA was performed using the Bayes Latent Class Models [BLCM] software (freely available with accompanying manual and files at: http://www.medicine.mcgill.ca/epidemiology/dendukuri/index.html). BLCM is a programme that was developed to estimate diagnostic test properties and population disease prevalence in the context of simultaneous use of multiple possibly correlated diagnostic tests. It uses a Bayesian

approach that allows substantive prior information on the prevalence, sensitivities and specificities to be incorporated in the analysis.

Dichotomous TST and QFT-G test results were used in the model. The latent class model for two diagnostic tests is 'not identifiable', i.e., we have fewer degrees of freedom than parameters to estimate. The number of degrees of freedom is given by the number of possible combinations of test results minus 1. With two dichotomous tests we have four possible combinations of test results and therefore 3 degrees of freedom. The parameters that are to be estimated are the prevalence of LTBI, and the sensitivity and specificity of each test, i.e., 5 parameters. Informative prior distributions are required on a minimum of $5 - 3 = 2$ parameters. We had reasonable prior information on the range of values of the sensitivity and specificity of each test (Table 1). These ranges were entered as the limits of the 95% prior CrI for each parameter. The programme converts this information into the posterior distributions illustrated in Figure 2. Alternatively, we could have selected a distribution allowing for equal weight for all values within the ranges given in Table 1.

In addition to providing results on the estimated prevalence of LTBI, the LCA model also provided estimates of the sensitivity and specificity of the tests, and the positive predictive value for each combination of test results, along with 95%CrIs. CrIs are the Bayesian analogue of CIs.

RESUMEN

CONTEXTE : On étudie traditionnellement la prévalence de l'infection tuberculeuse latente (LTBI) par le test cutané tuberculinique (TCT). Des tests sanguins hautement spécifiques de libération d'interféron-gamma (IGRA) sont actuellement disponibles et pourraient renforcer l'estimation de la prévalence de la LTBI en combinaison avec des méthodes basées sur des modèles.

SCHÉMA : Nous avons comparé des méthodes conventionnelles et des méthodes basées sur des modèles pour l'estimation de la prévalence de la LTBI chez 719 travailleurs des soins de santé indiens qui ont subi à la fois le TCT et le QuantiFERON-TB Gold In-Tube (QFT-G). En plus de l'utilisation de limites standard de positivité pour les TCT et le QFT-G, des analyses des modèles mixtes Bayesiens ont été pratiquées sur 1) des données continues de TCT et 2) des données catégorielles utilisant à la fois les résultats du TCT et du QFT-G dans une analyse de classe latente (LCA) tenant compte de préalables concernant la sensibilité et la spécificité.

RÉSULTATS : Selon la méthode utilisée, les estimations de la prévalence de la LTBI ont varié de 33,8% à 60,7%. Le modèle mixte basé sur le seul TCT a estimé la prévalence à 36,5% (IC95% 28,5–47,0). Lorsque l'on combine les résultats des deux tests en utilisant une LCA, la prévalence est de 45,4% (IC95% 39,5–51,1). La LCA a fourni des résultats complémentaires sur la sensibilité, la spécificité et les valeurs prédictives des résultats conjoints.

CONCLUSION : La disponibilité des IGRA novateurs et spécifiques et l'élaboration de méthodes telles que les analyses de mixture permettent une approche plus réaliste et plus informative de l'estimation de la prévalence.

RÉSUMÉ

MARCO DE REFERENCIA : Tradicionalmente la prevalencia de infección tuberculosa latente (LTBI) se calcula utilizando la prueba cutánea de la tuberculina (TCT). En la actualidad, se cuenta con pruebas sanguíneas extremadamente específicas basadas en la liberación de interferón gama (IGRA), las cuales podrían reforzar la estimación de la prevalencia, en asociación con métodos de modelización.

MÉTODOS : Se compararon métodos convencionales con métodos de modelización a fin de calcular la prevalencia de LTBI en 719 profesionales de salud hindúes, a quienes se practicó la TCT y la prueba del interferón-γ (QuantiFERON-TB Gold In-Tube). Además de aplicar los límites usuales de significación en ambas pruebas, se realizaron análisis bayesianos de modelos mixtos con : 1) datos continuos y 2) datos categóricos usando los resultados de ambos tipos de pruebas en un análisis de clase latente (LCA) que consideró distribuciones previas de sensibilidad y especificidad.

RESULTADOS : La estimación de la prevalencia de LTBI osciló entre 33,8% y 60,7% en función del método utilizado. Con el modelo mixto basado exclusivamente en la TCT se calculó una prevalencia de 36,5% (IC95% 28,5–47,0). Los resultados de ambas pruebas combinadas usando el LCA aportaron una prevalencia de 45,4% (IC95% 39,5–51,1). Este análisis procuró resultados complementarios de sensibilidad, especificidad y factores de predicción de los resultados combinados.

CONCLUSIÓN : Las nuevas IGRA y la concepción de métodos como los análisis mixtos ofrecen un enfoque más realista e informativo de la estimación de la prevalencia.