Estimating diagnostic accuracy of tests for latent tuberculosis infection without a gold standard among healthcare workers

E Girardi (girardi@inmi.it)[1], C Angeletti[1], V Puro[1], R Sorrentino[2], N Magnavita[3], D Vincenti[1], S Carrara[1], O Butera[1], A M Ciufoli[4], S Squarcione[4], G Ippolito[1], D Goletti[1,5]

1. Department of Epidemiology and Preclinical Research, Istituto Nazionale per le Malattie Infettive Lazzaro Spallanzani, IRCCS (National Institute for Infectious Diseases "Lazzaro Spallanzani"), Rome, Italy
2. San Camillo-Forlanini Hospital, Rome, Italy
3. Institute of Occcupational Medicine, Università Cattolica del Sacro Cuore, (Catholic University), Rome, Italy
4. Office of the Hospital Director, Istituto Nazionale per le Malattie Infettive Lazzaro Spallanzani, IRCCS (National Institute for Infectious Diseases "Lazzaro Spallanzani"), Rome, Italy
5. Clinical Department, Istituto Nazionale per le Malattie Infettive Lazzaro Spallanzani, IRCCS (National Institute for Infectious Diseases "Lazzaro Spallanzani"), Rome, Italy

This article was published on 29 October 2009.
Citation style for this article: Girardi E, Angeletti C, Puro V, Sorrentino R, Magnavita N, Vincenti D, Carrara S, Butera O, Ciufoli AM, Squarcione S, Ippolito G, Goletti D. Estimating diagnostic accuracy of tests for latent tuberculosis infection without a gold standard among healthcare workers. Euro Surveill. 2009;14(43):pii=19373. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19373

The evaluation of diagnostic accuracy of new *in vitro* diagnostic assays for tuberculosis infection has been hampered by the lack of a standard reference test. The aim of this study was to compare sensitivity and specificity of interferon gamma assays for latent tuberculosis infection by assessing the association of test results with tuberculosis occupational exposure and by using latent class analysis. We analysed data from 115 healthcare workers on whom tuberculin skin test (TST) and the following *in vitro* tests were performed: in-house ELISPOT for RD1 proteins, T.SPOT-TB and Quantiferon-TB Gold. Results of all tests were associated with increased occupational risk of exposure to *Mycobacterium tuberculosis*, but only TST was associated with *Bacillus* Calmette-Guérin (BCG) vaccination. Sensitivity/specificity (95% confidence intervals) estimated by a latent class model were: 99.9%/64.2% (53.0-74.1) for TST, 95.3% (61.8-99.6)/87.5% (78.0-93.2) for in-house ELISPOT, 96.7% (69.3-99.7)/85.6% (75.3-92.0) for T.SPOT-TB, and 76.3% (55.9-89.1)/93.6% (85.4-97.3) for Quantiferon. The estimated specificity of *in vitro* assays was higher than that of TST also among individuals who were not BCG-vaccinated. In conclusion, when used in healthcare workers, *in vitro* assays may provide a significant increase of specificity for tuberculosis infection compared to TST, even among non vaccinated individuals, at the cost of some sensitivity.

Introduction

Identification and treatment of individuals with latent tuberculosis infection is an important component of tuberculosis elimination strategies in low incidence countries, and may contribute to the global tuberculosis control efforts [1-4]. In this context, healthcare workers represent an important target population for latent tuberculosis infection screening programmes [5]. The effectiveness of these programmes, however, has been limited by the fact that the standard tool used to diagnose latent tuberculosis infection, the tuberculin skin test (TST), has a limited diagnostic accuracy, mainly because it relies on the use of protein purified derivative (PPD), which is a mixture of antigens shared by many pathogenic and non-pathogenic mycobacteria, including *Bacillus* Calmette-Guérin (BCG) strains used for vaccination [6].

Recently, new immunologic tests have been introduced for diagnosing tuberculosis infection [7,8]. These tests, often referred to as interferon gamma release assays (IGRAs) are based on the detection of *in vitro* response to proteins encoded by genes located within the region of difference 1 (RD1) of *M. tuberculosis* genome, the early secreted antigenic target 6 protein (ESAT-6) and the culture filtrate protein 10 (CFP-10), that are not shared with BCG strains or most environmental mycobacteria [9,10]. Two of these tests have been made commercially available. Both measure interferon gamma released *in vitro* in response to RD1-encoded antigens, although they use different antigen preparations (overlapping peptides spanning the entire length of these proteins) and different assay formats (ELISA and ELISPOT) [11,12]. Recent guidelines recommend that these tests be used instead of [1,2] or in addition to [13] TST.

A number of studies have evaluated IGRA, in comparison to TST, as a tool for screening latent tuberculosis infection among healthcare workers [14-19]. To our knowledge, however, no study has compared different IGRAs in this population group.

The lack of a gold standard for the diagnosis of latent tuberculosis infection has hampered the assessment of the diagnostic accuracy of IGRAs. Different strategies have been used so far to address this issue, including the evaluation of the proportion of positive tests among individuals with active tuberculosis (as a proxy for sensitivity), and of the proportion of negative tests among individuals at low risk for tuberculosis infection (as a proxy for specificity) [1,2,7]. Another approach that has been proposed for the validation of IGRAs is based on the assessment of the association of test results with risk factors for tuberculosis infection

[11,20]. Finally, latent class analysis, a statistical method which has been proposed for the assessment of diagnostic tests in the absence of a gold standard, could be used in this context [21]. In the frequentist statistical approach used in the present study, this analysis requires availability of results from at least three different diagnostic tests on the same individual, and it is based on the concept that different tests for the same disease are influenced by a common latent variable, the disease status, which cannot be measured directly [21-23].

Healthcare workers remain at risk for tuberculosis infection also in countries with low tuberculosis incidence [24]. However, especially in countries such as Italy where until recently BCG vaccination has been widely used in healthcare workers, surveillance of tuberculosis infection has been hampered by the low specificity of TST. In the present paper, we analysed data on healthcare workers in Italy who were tested by TST and by three *in vitro* interferon gamma tests, an in-house ELISPOT assay based on RD1 proteins [25], a commercial ELISPOT assay and a commercial whole blood ELISA using RD1 peptides. To validate the use of these tests in this population group, we assessed their association with occupational tuberculosis risk and estimated their sensitivity and specificity by using a latent class analysis.

Methods

Study design and participants

We conducted a cross-sectional study in 2004-2005 at two tertiary care hospitals in Rome, Italy, which include wards that routinely treat pulmonary tuberculosis patients. Healthcare workers at these institutions who had had a routine periodic health check in 2004 or 2005 were considered for inclusion, if they had a positive TST result in the 12 months, or a negative TST result in the three months before we did the *in vitro* tests. There was no formal calculation of the sample size prior to the study. No incentive was offered for participation. The study was approved by the ethics committees at participating institutions and study participants gave written informed consent.

For each individual enrolled in the study, the following data were abstracted from personal charts: age, sex, place of birth, job category, ward or service of present and past employment, BCG vaccination, household tuberculosis contacts. Ward or service of employment were classified either as high risk if more than one patient with tuberculosis was cared for per year, or as low risk if that was not the case.

Diagnostic assays

The TST was administered by trained nurses at participating institutions by the Mantoux procedure using 5 IU of PPD (Chiron). Results were read after 48 to 72 hours. For the purpose of the present analysis an induration of at least 10 mm was scored as a positive response [1,2].

The in-house ELISPOT assay based on ESAT-6 and CFP-10 proteins (Lionex) was performed as previously described [25], and results were scored positive if the average number of spot-forming cells (SFCs) in cultures stimulated with these antigens was at least three-fold higher than the average number of SFCs in the control. Interferon gamma values are presented as number of SFCs per million PBMC, after subtraction of the appropriate control according to the described criteria.

The commercial ELISPOT assay used was the T-SPOT.TB (Oxford Immunotec) and it was performed as previously described [11]. Responses were scored positive if the test wells contained a mean of at least six spot-forming cells more than the mean of the negative control wells, and if this number was at least twice the mean of the negative control wells.

The commercial ELISA assay was the enhanced 'in-tube' version of QuantiFERON-TB Gold (QFT-G, Cellestis Limited).This assay is based on peptides spanning the entire sequences of ESAT-6 and CFP-10 as well as another peptide representing a portion of the TB7.7 antigen [12]. It involves two stages: incubation of whole blood with the antigens, and measurement of interferon gamma production in harvested plasma by ELISA. As recommended by the manufacturer, the cut-off value for a positive test was 0.35 interferon gamma IU/ml.

All blood test were performed on the same blood sample. For 47 individuals (45.3%), the blood sample was taken on the day the TST was performed, while for the remaining individuals, it was taken eight to 365 days after the TST. ELISA and ELISpot were performed at the study site, and all assays met quality control standards.

Statistical methods

Standard univariable methods were used to describe the association between participant characteristics and results of diagnostic assays.

The association of test results with risk factors for tuberculosis infection was studied by fitting four multivariable logistic regression models, one for each diagnostic test, with the same covariates, and results were shown as odds ratios (OR) with the associated 95% confidence intervals (CI). Risk factors introduced in the models were age (as a continuous variable), sex and all variables that were significant in the univariable anlaysis for at least one diagnostic test. Whether the association with each risk factor varied by type of diagnostic assay was assessed by testing the hypothesis of homogeneity of the relative odds ratios. The test was performed using seemingly unrelated regression that takes into account the correlation between diagnostic test results of the same participant.

To estimate sensitivity and specificity of different diagnostic tests we performed a latent class analysis [21-23,26] a family of statistical models based on the concept of 'latent variable', that can simply be thought as an unobservable random variables. LCA is appropriate to study situations in which categorical responses are observed on n subjects and these responses are dependent by a categorical unobservable characteristic of the subject. Briefly, parameters of interest were estimated by modelling the relations between an unobservable (latent) and observable variables. In this respect, the observed results of the diagnostic tests are considered as a measure, prone to error, of an unobservable dichotomous latent variable, the true disease status. From these imperfect measures we can estimate a 'consensus' gold standard used, in turn, to evaluate sensitivity and specificity of the tests as well as the prevalence of the disease [22].

Let us assume that D represents the unknown disease status for each subject (1 for diseased and 0 for not diseased) and θ_d (d=0,1) its probability. Moreover let t_j be the observed result of our jth test (j=0,…,p) that can take on the values 0, negative,

or 1, positive. If we denote with π_{jd} the conditional probability of a positive response at the jth test given D=d, the parameters of interest for our study, i.e. the sensitivity and specificity of each test, are π_{j1} and $1-\pi_{j0}$, respectively. Each subject i (i=1,.....,n) will have a vector of observed responses, $T_i=(t_1,\ldots,t_p)$, and the marginal probability of T_i that follows a multivariate Bernoulli distribution is given by

$$Pr(\mathbf{T}_i) = \sum_{d=0}^{1} \theta_d Pr(T_i \mid D = d) \quad (1).$$

Assuming for each subject the independence between responses to the p tests, given the true disease status, equation (1) can be written as:

$$Pr(\mathbf{T}_i) = \sum_{d=0}^{1} \theta_d \prod_{j=1}^{p} \pi_{jd}^{t_j}(1-\pi_{jd})^{1-t_j} \quad (2).$$

Both θ_d and π_{jd} were modelled on a log odds, or logit, scale and we could also account for the effect of covariates using the usual approach of logistic model. The equations describing prevalence and conditional probabilities of positive response were as follows:

$$Logit(\theta_d \mid \mathbf{v}) = \alpha \quad (3) \text{ and}$$

$$Logit(\pi_{jd} \mid \mathbf{x}) = \gamma_j + \lambda_j \eta_d + \mathbf{x}'\boldsymbol{\beta} \quad (4),$$

where:

1. x was a vector of covariates for the ith subject, with their relatives vectors of parameters β;
2. η_d was the (random) effect, common for all tests, exerted by the unknown true disease status;
3. λ_j were the factor loadings that allow the effect of ηd to differ between tests and
4. γ_j represented the (fixed) effect of each test on conditional probability [22,26].

In order to make a latent class model estimable, the number p of diagnostic tests used on the same study sample must provide at least as many degrees of freedom as the number of parameters to be estimated, in other words the condition $(2^p-1)\geq(2p+1)$ has to be satisfied and this imply that at least three tests are requested for our study. Prevalence as well as sensitivity and specificity were modeled as logit (log odds). We included BCG as a covariate in the model for sensitivity and specificity. The fit of the model without covariates was assessed by using the Pearson's chi-squared statistic (the sum of squared difference between observed and expected frequencies over the expected). Nested models were compared using the log-likelihood ratio (LR) test [27-29].

The significance of the difference in accuracy between pairs of diagnostic assays was evaluated by using Wald test for fixed coefficients of the latent class model.

In traditional latent class analysis, it is assumed that the results of each individual for a given disease status are independent (the so-called conditional independence) or, in other words, that the observed associations between tests are explained only by the latent variable. In our study this condition could not be satisfied, regarding the similarities in technological characteristics of assays. To verify whether a lack of conditional independence between tests could have influenced our estimates, we introduced in the equation (4) an additional subject-specific random variable z with Gaussian distribution to take into account the correlation between the assays that was not due to the disease status [27,29]. The results from the traditional latent class analysis were then compared with those from the model with random effect using the Akaike Information Criterion (AIC) and Pearson's statistic.

Statistical analyses were performed with Stata, Release 9 (Stata Corp). The programme "gllamm" in Stata [30] and "randomLCA" package for R [31] were used to fit latent class analysis models.

Results

Study population

Included in the present analysis were 115 healthcare workers. Of these, 39 (33.9%) were currently employed in wards in which the risk of being exposed to tuberculosis was high (such as wards for infectious diseases and respiratory diseases), and 76 (66.1%) were employed in hospital services in which the risk of exposure to tuberculosis was low (such as paediatrics, internal medicine and hospital epidemiology). Of those currently employed in low-risk services, seven had worked in services with high exposure risk in the past. The median age of the participants was 41 years and the majority were female. BCG vaccination was documented for 43 participants (37.4%).

Association of results in the four diagnostic assays with participants characteristics

Overall 61 individuals (53.0%) were TST-positive, 40 (38.4%) were positive by in-house ELISPOT, 42 (36,5%) by T-SPOT.TB and 29 (25,2%) by QFT-G . The results of the different diagnostic assays by participant characteristics are shown in Table 1. A higher proportion of positive tests was observed among those who had at one point been employed in high-risk services, compared to those employed only in other hospital services. This difference was statistically significant for all tests except for the QFT-G test. In addition, older study participants were more likely to be positive in all tests. A positive result in the TST only was associated with a previous BCG vaccination. Physicians had the lowest prevalence of positive results in all tests, but this difference was significant for QFT-G only. Surprisingly, the prevalence of positive results in the three *in vitro* assays was not elevated among those reporting household tuberculosis contact, and differences were not statistically significant.

As shown in Table 2, 40 individuals (34.8%) were negative in all the four tests, while 75 (65.2%) individuals were positive in at least one test. Of those 75, 22 (19.1%) were positive in all the four tests. Nineteen individuals (16.5%) were positive only in the TST.

In a multivariable analysis (Table 3), having worked in high-risk tuberculosis services increased the probability of a positive result for all diagnostic tests (homogeneity test: p=0.52), although the effect was significant only for the T-SPOT.TB and the in-house ELISPOT. Sex was not significantly associated with the probability of a positive result and the odds ratios were not significantly different among diagnostic tests (p=0.41). Older individuals, however, had a significantly higher probability of a positive result for all tests. The effect of BCG vaccination was not homogeneous among diagnostic tests (p=0.001) and significant only for the TST, with a higher odds ratio for a positive result for BCG-vaccinated compared to not vaccinated subjects. Physicians were at a lower risk of a positive result compared to nurse assistants; this result was significant for TST and QFT-G.

Table 1

Results of diagnostic tests for tuberculosis infection by characteristics of healthcare workers in Rome, Italy (n=115)

Characteristic (no.)	Tuberculin skin test no. of positives (%)	In-house RD1 ELISPOT no. of positives (%)	T-SPOT.TB no. of positives (%)	QuantiFERON TB Gold no. of positives (%)
Ward/service				
Low TB risk (69)	30 (44)	17 (25)	18 (26)	16 (23)
High TB risk* (46)	31 (67) †	23 (50) †	24 (52) †	13 (28)
Sex				
Male (48)	22 (46)	17 (35)	19 (40)	11 (29)
Female(67)	35 (52)	23 (34)	23 (34)	18 (27)
Place of birth				
EU (110)	57 (53)	38 (35)	40 (37)	26 (24)
Non-EU (5)	3 (60)	1 (20)	1 (20)	2 (40)
BCG vaccination				
No (72)	30 (42)	26 (36)	24 (33)	22 (31)
Yes (43)	31 (72) †	14 (32)	18 (42)	7 (16)
Household TB contact				
No (102)	53 (52)	37 (36)	40 (39)	27 (27)
Yes (13)	8 (62)	3 (23)	2 (15)	2 (15)
Job category				
Physician (18)	6 (33)	4 (22)	6(33)	1 (5.6)
Nurses (67)	40 (60)	24 (36)	23 (34)	16 (24)
Nurse assistant (30)	15 (50)	12 (40)	13 (43)	12 (40) †
Age (years)				
≤41 (59)	41 (36)	11 (19)	12 (20)	8 (14)
>41 (56)	40 (71) †	29 (52) †	30 (54) †	21 (38) †

BCG: Bacillus Calmette-Guérin; EU: European Union; TB: tuberculosis.
* currently or in the past
† p<0,05

Table 2

Response patterns to four different diagnostic tests for tuberculosis infection observed among healthcare workers in Rome, Italy, and predicted by a latent class analysis model with and without a random effect (n=115)

Response pattern				Observed		Predicted LCA	Predicted LCA with random effect
Tuberculin Skin test	In-house RD1 ELISPOT	T-SPOT.TB	QuantiFERON TB Gold	No.	%	No.	No.
-	-	-	-	40	34.8	37.8	39.9
+	+	+	+	22	19.1	21.8	21.9
+	-	-	-	19	16.5	21.1	19.4
+	+	+	-	7	6.1	7.3	7.1
+	-	+	-	7	6.1	3.9	4.7
-	+	-	-	5	4.3	5.4	4.6
-	-	-	+	4	3.5	2.6	2.1
+	+	-	-	3	2.6	3.2	3.9
-	-	+	-	3	2.6	6.4	5.3
-	+	+	-	2	1.7	0.9	1.0
+	+	-	+	1	0.9	1.0	0.9
+	-	+	+	1	0.9	1.3	1.3
+	-	-	+	1	0.9	1.5	1.8
-	+	+	+	0	0.0	0.1	0.2
-	+	-	+	0	0.0	0.4	0.3
-	-	+	+	0	0.0	0.4	0.5

LCA: latent class analysis.

Estimation of the accuracy of the assays by latent class analysis

The tuberculosis infection prevalence in the population estimated in the latent class analysis model was 26.9% (95% CI: 18.1% to 35.7%). The predicted frequencies for the patterns of response to the four tests (Table 2) showed a good fit with the observed data (Pearson's statistic p-value=0.25).

In the latent class analysis (Table 4), TST had the highest estimated sensitivity but a very low specificity. The two ELISPOT-based tests, the in-house ELISPOT and the T-SPOT.TB, both had a sensitivity close to that of the TST, while their estimated specificity was still high. QFT-G had a very high estimated specificity, although its sensitivity was lower than that of the other three tests. When

Table 3

Multivariable odds ratios (95% confidence intervals) of a positive result for selected risk factors by diagnostic test among healthcare workers in Rome, Italy (n=115)

	Diagnostic test assumed as outcome variable				
	Tuberculin Skin test	In-house RD1 ELISPOT	T-SPOT.TB	QuantiFERON TB Gold	
	MOR# (95% CI)	MOR# (95% CI)	MOR# (95% CI)	MOR# (95% CI)	p*
Ward/service					
Low TB risk	1.00	1.00	1.00	1.00	
High TB risk	2.48 (0.97-6.35)	3.88 (1.52-9.91)	3.10 (1.28-7.48)	1.68 (0.63-4.49)	0.519
p**		0.472	0.681	0.491	
BCG Vaccination					
No	1.00	1.00	1.00	1.00	
Yes	4.32 (1.56-11.95)	0.62 (0.23-1.67)	1.49 (0.58-3.81)	0.41 (0.14-1.23)	0.001
p**		0.001	0.060	<0.001	
Gender					
Male	1.00	1.00	1.00	1.00	
Female	2.13 (0.73-6.21)	1.23 (0.46-3.26)	1.28 (0.50-3.26)	0.82 (0.29-2.31)	0.413
p**		0.449	0.401	0.107	
Age (per five years increase)	1.86 (1.39-2.48)	1.69 (1.29-2.22)	1.56 (1.21-2.02)	1.50 (1.16-1.95)	0.485
p**		0.599	0.231	0.215	
Job category					
Physician	0.20 (0.04-0.92)	0.25 (0.05-1.23)	0.39 (0.09-1.63)	0.07 (0.01-0.70)	0.480
p**		0.758	0.393	0.377	
Nurses	1.64 (0.49-5.51)	1.21 (0.38-3.87)	0.67 (0.22-2.04)	0.63 (0.21-1.91)	0.211
p**		0.721	0.159	0.156	
Nurse assistant	1.00	1.00	1.00	1.00	

BCG: Bacillus Calmette-Guérin; CI: Confidence Interval; MOR: multivariable odds ratio. TB: tuberculosis.
Adjusted for all the variables in the table by fitting a logistic regression model.
* p-value for the hypothesis of no difference among OR, obtained by fitting a seemingly unrelated regression model.
**p-value for the hypothesis of no difference to the OR for tuberculin skin test, obtained by fitting a seemingly unrelated regression model.

Table 4

Specificity and sensitivity of four diagnostic assays for tuberculosis infection estimated among 115 healthcare workers in Rome, Italy by a latent class analysis model

	Specificity [%]			Sensitivity [%]		
	Estimate	95% confidence interval		Estimate	95% confidence interval	
Tuberculin skin test	64.2	53.0	74.1	99.9	NC	NC
In-house RD1 ELISPOT	87.5	78.0	93.2	95.3	61.8	99.6
T-SPOT.TB	85.6	75.3	92.0	96.7	69.3	99.7
QuantiFERON TB Gold	93.6	85.4	97.3	76.3	55.9	89.1

NC: not computable.

the tests were compared in pairs to evaluate differences in their diagnostic accuracy, statistically significant differences were recorded for the comparison between TST and the other three tests (p=0.003, p=0.005 and p<0.001, respectively, for the comparison with in-house ELISPOT, T-SPOT.TB and QFT-G), while the difference between the T-SPOT.TB and QFT-G was of borderline statistical significance (p=0.057).

To explore the impact of BCG vaccination on the diagnostic accuracy of the TST, we also fitted a latent class analysis models solely for those subjects who had not been vaccinated against BCG. In this analysis, the estimated prevalence of tuberculosis infection was 26.3%. As shown in Table 5, the sensitivity of the TST was similar to that estimated for the entire population. In contrast, an increased specificity was estimated for TST among not BCG-vaccinated subjects (79.1%), although it remained lower than that estimated for the *in vitro* assays. The estimated accuracy of IGRAs did not vary markedly in this analysis, except for QFT-G sensitivity which increased from 76.3 to 94.8.

Finally, we compared the traditional latent class analysis model to a model with a subject-specific random effect in order to assess whether the removal of conditional independence assumption among tests had an impact on the results. The estimate of tuberculosis infection prevalence in the latter model was 25.0%, and the predicted frequencies for the patterns of response to the four tests were similar to the former model with a slight worsening of the AIC (476.97 and 477.77 in the latent class analysis and the model with subject-specific random effect, respectively), and an equally slight improvement in Pearson's statistic (p=0.267). The estimates of diagnostic accuracy were remarkably similar in the two models (Table 6).

Discussion

We compared the results obtained in the TST and three *in vitro* assays for tuberculosis infection in healthcare workers. We found that positive results in all four assays were associated with increased occupational risk of exposure to *M. tuberculosis*, but only the TST was correlated with BCG vaccination. Taking advantage of the fact that the results of four different assays for tuberculosis infection were available for the same groups of individuals, we provided an estimate of the diagnostic accuracy of these assays by using a latent class analysis model. In this analysis, the *in vitro* tests were found to be more specific for tuberculosis infection than the TST, even among non-vaccinated individuals, at the cost of some sensitivity. Moreover, our data suggest that ELISPOT-based tests may differ in accuracy from the ELISA-based test.

Previous studies conducted among healthcare workers in countries with low and high tuberculosis incidence [14-17] have shown an association between QFT-G results and occupational exposure to patients with active tuberculosis. Our results are consistent with these findings and show an even stronger association with occupational exposure for ELISPOT-based assays, although no statistically significant differences were recorded when association coefficients for the four different tests were compared. Moreover, as in previous studies [32,33], we found that TST results were associated with previous vaccination, while this was not the case for *in vitro* assays.

We also used latent class analysis to estimate and compare the sensitivity and specificity of different tests for tuberculosis infection. Latent class analysis allows addressing a major issue in the evaluation of diagnostic tests, i.e. the estimation of diagnostic accuracy when a gold standard test is not available, and for this reason it has been used in different infectious conditions in which a

Table 5

Comparison of specificity and sensitivity of four diagnostic assays for tuberculosis infection estimated among 72 not BCG-vaccinated healthcare workers by a latent class analysis model

	Specificity %			Sensitivity %		
	Estimate	95% confidence interval		Estimate	95% confidence interval	
Tuberculin skin test	79.1	65.9	88.1	100.0	N.C.	N.C.
In-house RD1 ELISPOT	84.6	72.2	92.1	94.4	65.8	99.3
T-SPOT.TB	90.4	78.4	96.1	100.0	N.C.	N.C.
QuantiFERON TB Gold	92.3	81.3	97.1	94.8	63.1	99.5

NC: not computable.

Table 6

Comparison of specificity and sensitivity of four diagnostic assays for tuberculosis infection estimated among 115 healthcare workers by a latent class analysis model with and without a subject-specific random effect

	Specificity %		Sensitivity %	
	LCA	LCA with random effect	LCA	LCA with random effect
Tuberculin skin test	64.2	64.4	99.9	100.0
In-house RD1 ELISPOT	87.5	88.5	95.3	97.5
T-SPOT.TB	85.6	86.9	96.7	98.8
QuantiFERON TB Gold	93.6	94.3	76.3	81.4

LCA: latent class analysis.

definitive demonstration of the infecting organism was not feasible [22].

As reported in a recently published systematic review, the sensitivity of IGRAs for tuberculosis infection has previously been estimated in a number of studies by calculating the proportion of positive patients among those diagnosed with culture-proven tuberculosis [32]. The sensitivity in these studies ranged from 55% to 93% for QFT-G with a pooled estimate of 78% for the first version of the QFT-G or 70% for the in tube version of this assay, and from 83 to 100% for T-Spot.TB with a pooled estimate of 90%. In the studies in which both IGRAs were performed on the same group of patients, the positivity rate tended to be higher for the ELISPOT assay. Our estimates of the sensitivity of interferon gamma tests for latent infection, obtained by latent class analysis, were above 95% for ELISPOT-based assays and 76.3% for the ELISA assay, thus consistent with those obtained from patients with active tuberculosis. Nevertheless, the TST had the highest estimated sensitivity (99.9%) in our study, which is in contrast to the results of studies on patients with active tuberculosis, most of which reported a higher sensitivity for interferon gamma assays compared to the TST [34]. However, there is evidence that estimates of sensitivity of TST for active infection may differ from that for latent infection: On average 10 to 25% of patients with active TB do not respond to the TST, and reactivity may be restored after initiation of treatment in most of the patients who were initially negative [35]. In contrast, sensitivity estimates derived from studies on healthy individuals may exceed 95% [36]. Moreover, some studies conducted to assess the accuracy of diagnosis of latent tuberculosis infection suggest that the sensitivity of interferon gamma tests may indeed be somewhat lower than or equal to that of the TST [33,37,38]. On the other hand, in a recent study carried out among healthcare workers in India, in which a Bayesian latent class analysis was used to compare accuracy of QFT-G and TST, Pai *et al.* estimated that the QFT-G had an higher sensitivity than the TST (89.9% and 79.5 %, respectively) [39]. The results reported by Pai *et al*. are not directly comparable to those of the present study since a different statistical approach was used to construct the latent class model and results from only two different tests were available for each subject. Moreover, the subjects in the two studies were enrolled in countries with very different tuberculosis incidence.

In this study, specificity was estimated to be consistently higher for IGRAs compared to the TST. This finding was not unexpected since these *in vitro* assays are based on antigens that, differently from the PPD antigens used in the TST, are present almost exclusively in bacteria of the *M. tuberculosis* complex. Previous studies included in the aforementioned systematic review [34] have shown that, among individuals at low risk for tuberculosis infection, QFT-G is negative in 92-98% of cases (estimated pooled specificity 99% and 96% in BCG-vaccinated and non-vaccinated individuals, respectively), and T-SPOT.TB in 85-100% of cases (estimated pooled specificity 93%). These figures are consistent with specificity values estimated for IGRAs in our study. Moreover, there is indirect evidence that these tests have higher specificity for latent tuberculosis infection than the TST. It has in fact been shown that, when used in contact tracing studies, these tests yield a better correlation to the degree of exposure to tuberculosis cases than the TST, and that their results are not influenced by the BCG vaccination status [32,33,37]. The specificity of the TST estimated in our study was quite low. It has been shown that large variations in the specificity of the TST can be observed when the test is applied to different populations [38], and in our study, the high prevalence of previous BCG vaccination among healthcare workers may be one cause of low specificity. However, TST specificity was estimated to be low also among non-vaccinated healthcare workers. A similar finding has been reported for healthcare workers in the United States, and it has been attributed to infection with non-tuberculous mycobacteria [40]. In contrast, a higher value for the specificity of the TST (87.4%) resulted from the application of a Bayesian latent class model in spite of the fact that 71% of subjects were BCG-vaccinated [39].

The statistical model we used also allowed an overall comparison of diagnostic accuracy of the tests analysed. We found that the diagnostic accuracy of the TST was significantly different from that of blood tests. This finding is not surprising if it is considered, in addition to the higher specificity of the antigens used, that the *in vitro* tests avoid a series of operational problems that may affect the accuracy of the TST, including variability in the intradermal injection of the antigen and in the reading of the response [8].

When the three *in vitro* tests were compared, we found a difference of borderline significance between QFT-G and T-SPOT.TB. The reasons for this difference are unclear. One may speculate that the ELISPOT technique, thanks to the ability to detect single cells that secrete interferon gamma in response to specific stimuli, may provide a higher sensitivity at the cost of some specificity. The cut-off value used to define positivity could also account for differences in sensitivity and specificity, at least in part. In fact, a study in which the commercial T-SPOT.TB and ELISA were used, has shown that the differences in diagnostic accuracy between the two tests become negligible when new cut-off points are used that have been optimised on the same population [41].

Before drawing firm conclusions, it is important to appreciate the limitations of the statistical method we used [21,22]. Latent class analysis assumes the existence of a 'true disease status' which influences the results of diagnostic tests, and this mathematically defined entity does not necessarily have a clear clinical or biological sense. There is consistent evidence that the TST predicts the development of active tuberculosis [6]. Thus the presence of latent tuberculosis infection, as identified by a positive TST, is associated with an increased risk of active disease. It remains to be determined if the same meaning could be attributed to the random variable identified as 'latent tuberculosis infection' in the present analysis.

Another drawback of the traditional version of latent class analysis is the assumption of conditional independence, i.e. the absence of correlation among test results given the disease status. This is often unrealistic in practice due to similarities among tests. However, following the approach proposed by Qu *et al.* [27] to relax this assumption, we used an additional random effect, with which it is possible to model all the non-observable factors at the subject level that could introduce correlation between test results. The estimates of diagnostic accuracy for the model with subject-specific random effect were very similar to those obtained in the traditional latent class analysis, and the measures of goodness of fit were comparable in the two models as well.

Other limitations of the present study need to be mentioned. First, all the individuals included were healthy adults, and thus our results should not be generalised for different populations, in particular for children or immunocompromised individuals in whom a significant proportion of indeterminate results may be observed,

in particular when using ELISA-based assays [40]. Similarly, the diagnostic accuracy estimated for latent tuberculosis infection is not necessarily similar to that obtained when using these tests to diagnose active tuberculosis infection. Second, tuberculin skin tests have been administered and read by different trained nurses, and thus inter-reader variability in interpreting the results should be expected. Third, the confidence intervals around our estimates of association coefficients and of sensitivity and specificity were rather wide because of the limited size of the population studied. Nevertheless, we were able to demonstrate statistically significant differences in the diagnostic accuracy of the different tests used.

Longitudinal studies comparing the ability of the TST to predict the risk of active tuberculosis with that of interferon gamma assays would be needed to establish the usefulness of the new tests for tuberculosis infection. Preliminary data suggest that positive IGRAs results may indeed be associated with the risk of active tuberculosis [42]. However, these studies will be difficult to perform in populations such as healthcare workers. In this context, the present study provides further evidence on the advantages in terms of specificity, and on the potential loss of sensitivity for latent tuberculosis infection of blood tests in comparison to the TST. Moreover, it provides comparative estimates of diagnostic accuracy of different blood tests and thus may contribute to choosing the strategies for diagnosing tuberculosis infection among heath careworkers. In particular, our results may suggest the use of IGRAs, either alone or as confirmatory tests in TST-positive individuals, in a population with a high prevalence of previous BCG vaccination. These choices, however, will also need to take other considerations into account, including the economical and operational aspect, and the stability of test results over time [43].

Acknowledgements

The authors thanks S Carrara and F Bizzoni for their help with laboratory tests, L Pischedda, S Lodi, S Console, A Moretti, M Paparatti and M Sorcelli for their help with participants' data collection, and C Nisii for editing the manuscript.

References

1. National Tuberculosis Controllers Association. Centers for Disease Control and Prevention (CDC). Guidelines for the investigation of contacts of persons with infectious tuberculosis. Recommendations from the National Tuberculosis Controllers Association and CDC. MMWR Recomm Rep. 2005;54(RR-15):1-47.
2. Guidelines for Using the QuantiFERON®-TB Gold Test for Detecting Mycobacterium tuberculosis Infection, United States. MMWR Recomm Rep. 2005;54(RR-15):49-55.
3. Broekmans JF, Migliori GB, Rieder HL, Lees J, Ruutu P, Loddenkemper R, et al. European framework for tuberculosis control and elimination in countries with a low incidence. Recommendations of the World Health Organization (WHO), International Union Against Tuberculosis and Lung Disease (IUATLD) and Royal Netherlands Tuberculosis Association (KNCV) Working Group. Eur Respir J. 2002;19(4):765-75.
4. Dye C, Watt CJ, Bleed DM, Hosseini SM, Raviglione MC. Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. JAMA. 2005;293(22):2767-75.
5. Taylor Z, Nolan CM, Blumberg HM, American Thoracic Society; Centers for Disease Control and Prevention; Infectious Diseases Society of America. Controlling tuberculosis in the United States: recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. MMWR Recomm Rep. 2005;54(No. RR-12):1- 81.
6. Huebner RE, Schein MF, Bass JB Jr. The tuberculin skin test. Clin Infect Dis. 1993;17(6):968-75.
7. Pai M, Riley LW, Colford JM Jr. Interferon-gamma assays in the immunodiagnosis of tuberculosis: a systematic review. Lancet Infect Dis. 2004;4(12):761-76.
8. Richeldi L. An update on the diagnosis of tuberculosis infection. Am J Respir Crit Care Med. 2006; 174(7):736-42.
9. Sørensen AL, Nagai S, Houen G, Andersen P, Andersen AB. Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. Infect Immun. 1995; 63(5):1710-7.
10. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science. 1999;284(5419):1520-3.
11. Lalvani A, Pathan AA, Durkan H, Wilkinson KA, Whelan A, Deeks JJ, et al. Enhanced contact tracing and spatial tracking of Mycobacterium tuberculosis infection by enumeration of antigen-specific T cells. Lancet. 2001;357(9273):2017-21.
12. Mori T, Sakatani M, Yamagishi F, Takashima T, Kawabe Y, Nagao K, et al. Specific detection of tuberculosis infection: an interferon-gamma-based assay using new antigens. Am J Respir Crit Care Med. 2004;170(1):59-64.
13. National Collaborating Centre for Chronic Conditions. Tuberculosis: clinical diagnosis and management of tuberculosis, and measures for its prevention and control. London: Royal College of Physicians; 2006.
14. Pai M, Gokhale K, Joshi R, Dogra S, Kalantri S, Mendiratta DK, et al. Mycobacterium tuberculosis infection in health care workers in rural India: comparison of a whole-blood interferon gamma assay with tuberculin skin testing. JAMA. 2005;293(22):2746-55.
15. Drobniewski F, Balabanova Y, Zakamova E, Nikolayevskyy V, Fedorin I. Rates of latent tuberculosis in health care staff in Russia. PLoS Med. 2007; 4(2): e55.
16. Stebler A, Iseli P, Mühlemann K, Bodmer T. Whole-blood interferon-gamma release assay for baseline tuberculosis screening of healthcare workers at a Swiss university hospital. Infect Control Hosp Epidemiol. 2008; 29(7):681–3.
17. Mirtskhulava V, Kempker R, Shields KL, Leonard MK, Tsertsvadze T, del Rio C, et al. Prevalence and risk factors for latent tuberculosis infection among health care workers in Georgia. Int J Tuberc Lung Dis. 2008;12(5):513-9.
18. Carvalho AC, Crotti N, Crippa M, Baschè R, De Iaco G, Signorini S, et al. QuantiFERON-TB Gold test for healthcare workers. J Hosp Infect. 2008 ;69(1):91-2.
19. Pollock NR, Campos-Neto A, Kashino S, Napolitano D, Behar SM, Shin D, et al. Discordant QuantiFERON-TB Gold test results among US healthcare workers with increased risk of latent tuberculosis infection: a problem or solution? Infect Control Hosp Epidemiol. 2008 ;29(9):878-86.
20. Kunst H, Khan KS. New tests for the diagnosis of latent tuberculosis infection. Ann Intern Med. 2007;147(9):672-3.
21. Pepe MS. The statistical evaluation of medical test for classification and prediction. Oxford University Press. New York. 2003.
22. 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(5):604-12.
23. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. Stat Methods Med Res. 1998;7(4):354-70.
24. Seidler A, Nienhaus A, Diel R. Review of epidemiological studies on the occupational risk of tuberculosis in low-incidence areas. Respiration. 2005;72(4):431-46.
25. Goletti D,Carrara S, Vincenti D, Saltini C, Rizzi EB, Schininà V, et al. Accuracy of an immune diagnostic assay based on RD1 selected epitopes for active tuberculosis in a clinical setting: a pilot study. Clin Microbiol Infect. 2006;12(6):544-50.
26. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: Multilevel, longitudinal and structural equation models. Florida: Chapman & Hall/ CRC. 2004.
27. 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.
28. Alvord WG, Drummond JE, Arthur LO, Biggar RJ, Goedert JJ, Levine PH, et al. A method for predicting individual HIV infection status in the absence of clinical information. AIDS Res Hum Retroviruses. 1988;4(4):295-304.
29. Goetghebeur E, Liinev J, Boelaert M, Van der Stuyft P. Diagnostic test analyses in search of their gold standard: latent class analyses with random effects. Stat Methods Med Res. 2000;9(3):231-48.
30. Rabe-Hesketh S, Skrondal A. Multilevel and Longitudinal Modeling using stata. College Station, TX: Stata Press. 2005.
31. The R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Austria. 2009.
32. Zellweger JP, Zellweger A, Ansermet S, de Senarclens B, Wrighton-Smith P. Contact tracing using a new T-cell-based test: better correlation with tuberculosis exposure than the tuberculin skin test. Int J Tuberc Lung Dis. 2005;9(11):1242-7.
33. Brock I, Weldingh K, Lillebaek T, Follmann F, Andersen P. Comparison of tuberculin skin test and new specific blood test in tuberculosis contacts. Am J Respir Crit Care Med. 2004;170(1):65-9.
34. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med. 2008;149(3):177-84.

35. Rooney JJ Jr, Crocco JA, Kramer S, Lyons HA. Further observations on tuberculin reactions in active tuberculosis. Am J Med. 1976;60(4):517-22.

36. Berkel GM, Cobelens FG, de Vries G, Draayer-Jansen IW, Borgdorff MW. Tuberculin skin test: estimation of positive and negative predictive values from routine data. Int J Tuberc Lung Dis. 2005;9(3):310-6.

37. Shams H, Weis SE, Klucar P, Lalvani A, Moonan PK, Pogoda JM, et al. Enzyme-linked immunospot and tuberculin skin testing to detect latent tuberculosis infection. Am J Respir Crit Care Med. 2005;172(9):1161-8.

38. Rieder HL. Epidemiologic basis of tuberculosis control. International Union Against Tuberculosis and Lung Disease. Paris. 1999.

39. Pai M, Dendukuri N, Wang L, Joshi R, Kalantri S, Rieder HL. Improving the estimation of tuberculosis infection prevalence using T-cell-based assay and mixture models. Int J Tuberc Lung Dis. 2008;12(8):895-902.

40. von Reyn CF, Horsburgh CR, Olivier KN, Barnes PF, Waddell R, Warren C, et al. Skin test reactions to Mycobacterium tuberculosis purified protein derivative and Mycobacterium avium sensitin among health care workers and medical students in the United States. Int J Tuberc Lung Dis. 2001;5(12):1122-8.

41. Lee JY, Choi HJ, Park IN, Hong SB, Oh YM, Lim CM, et al. Comparison of two commercial interferon gamma assays for diagnosing Mycobacterium tuberculosis infection. Eur Respir J. 2006;28(1):24–30.

42. Bakir M, Millington KA, Soysal A, Deeks JJ, Efee S, Aslan Y, et al. Prognostic value of a T-cell-based, interferon-gamma biomarker in children with tuberculosis contact. Ann Intern Med. 2008;149(11):777-87.

43. Pai M, Joshi R, Dogra S, Mendiratta DK, Narang P, Kalantri S, et al. Serial testing of health care workers for tuberculosis using interferon-gamma assay. Am J Respir Crit Care Med. 2006;174(3):349-55.