CLINICAL AND VACCINE IMMUNOLOGY, Apr. 2010, p. 596–601 Vol. 17, No. 4
1556-6811/10/$12.00 doi:10.1128/CVI.00320-09

Assessment of Imprecision in Gamma Interferon Release Assays for the Detection of Exposure to *Mycobacterium tuberculosis*▽

Tamara Tuuminen,[1,2] Esko Tavast,[1]* Riitta Väisänen,[2] Jaakko-Juhani Himberg,[3] and Ilkka Seppälä[2]

Haartman Institute, University of Helsinki, Department of Bacteriology and Immunology, PL21, Fl-00014 Helsinki, Finland[1]; Division of Clinical Microbiology, Helsinki University Hospital, HUSLAB, Haartmaninkatu 3, P.O. Box 400, Fl-00029 HUS, Helsinki, Finland[2]; and Helsinki University Hospital, HUSLAB, Administration Department, Paciuksenkatu 29, P.O. Box 720, Fl-00029 HUS, Helsinki, Finland[3]

Received 6 August 2009/Returned for modification 3 November 2009/Accepted 8 January 2010

New gamma interferon (IFN-γ) release assays (IGRAs) to detect an exposure to *Mycobacterium tuberculosis* have recently been launched. The majority of the studies in temperate-climate countries agree that these methods have superior specificity and equal or even superior sensitivity over tuberculin skin tests (TSTs) in the diagnosis of latent tuberculosis (TB) infection (LTBI). However, reproducibility data of IGRAs are virtually missing. We assessed within-run, between-run, and total imprecision of two commercial IGRAs by testing samples from subjects with a stable state of TB infection or treated pulmonary TB, a sample from a healthy volunteer, and internal quality control samples. We calculated coefficients of variance (CV%s) to describe assays variability and compared the obtained results to the reported CV%s for other commercial immunodiagnostic methods. We illustrate an example of assay variability near the cutoff zone to demonstrate the necessity of a gray zone. Due to the strict adherence to the standard operation procedures (SOP) adopted in our laboratory, the total imprecision of enzyme-linked immunospot (ELISPOT)- and enzyme immunoassay (EIA)-based IGRAs was at a maximum CV% of 37.8% for the samples with moderate and high reactivities. Imprecision of testing samples with very low reactivity levels or nonreactive samples may, however, exceed 100%. In conclusion, despite multiple steps of the method performance, the analytical imprecision of IGRAs, which in our study design included also between-lot variability and had a component of normal biological variation, was well in accordance with the reported imprecisions of other manual immunodiagnostic tests. The recognition of the variability around the cutoff point advocates the use of a gray zone to avoid ambiguous result interpretations.

Evaluation of immunometric tests for infectious diseases abides by the same rules as methods for clinical chemistry and is based on the same general principles (12). Evaluation of analytical performance includes, among other parameters, reproducibility data. As a prerequisite for CE mark registration to get a license to market the products for *in vitro* diagnostic use, manufacturers should provide reproducibility characteristics as a part of the overall performance data.

Two new kits, namely, the T-SPOT.*TB* (Oxford Immunotec, Oxford, United Kingdom) (8) and QuantiFERON-TB Gold In-Tube (Cellestis Limited, Carnegie, Victoria, Australia) (www.cellestis.com/IRM/content/aust/qtfproducts_tbgoldintube_techinfo-perfparameters.html) kits, have been recently launched. The kits utilize the ability of sensitized $CD8^+$ and $CD4^+$ T lymphocytes to release gamma interferon (IFN-γ) when stimulated with synthetic peptides specific to *Mycobacterium tuberculosis* and detect exposure to *Mycobacterium tuberculosis*. While the first method measures the frequency of reactive lymphocytes in the peripheral blood mononuclear cell (PBMC) fraction, the latter measures the concentration of released IFN-γ into supernatants. These methods are collectively called IFN-γ release assays (IGRA). Although these were launched for diagnostics, clinically relevant information on assay reproducibility was available from only one test series (www.cellestis.com/IRM/content/aust/qtfproducts_tbgoldintube_techinfo-perfparameters.html) as of May 2009. From the literature search, we have found only a few publications that are related to this topic (4, 9, 14), whereas test sensitivity and specificity have been extensively tested and reviewed in recent meta-analyses (7, 10). The importance of the reproducibility parameter is emphasized by the demand to assess immunological conversions and reversions, in other words, a significant decrease in immunological responses that exceeds the total imprecision of the method. The clinical phenomenon of immunological reversion was reported to associate with, e.g., successful chemotherapy (2). A spontaneous reversion, which is a phenomenon that is not yet well understood, may be, indeed, a very important observation meaning pathogen clearance. However, the decrease in the response should be well documented and should clearly exceed the method's total imprecision. Immunological conversion may mean a rise in the reactivity that is above the variation of technical and biological noise. Because the data on reproducibility are scarce, we assumed that ethical considerations may have constituted the major obstacle. Indeed, it may be difficult to obtain an ethical permission to collect blood samples consecutively from tuberculosis (TB) patients who may need urgent treatment.

Both ethical considerations and unstable sample material

* Corresponding author. Mailing address: Haartman Institute, University of Helsinki, Department of Bacteriology and Immunology, PL21, Fl-00014 Helsinki, Finland. Phone: 358456750818. Fax: 358-9-19126382. E-mail: esko.tavast@helsinki.fi.

▽ Published ahead of print on 24 February 2010.

make imprecision study of IGRAs difficult. In fact, IGRAs were the first diagnostic methods to exploit cell-mediated immunity and utilize the *ex vivo* activity of vital lymphocytes. For example, another recently introduced CE mark-registered test to evaluate the effect of immunosuppression on the function of lymphocytes (ImmuKnow; Cylex, Columbia, MD) provides only repeatability data (results obtained in different laboratories from the same venipuncture). The information on the between-run imprecision was not yet available in the kit instructions as of January 2009. The shortage of information reflects obvious practical problems in obtaining samples from critically ill subjects to study between-run imprecision as required.

Imprecision data represent a very important parameter when the cutoff point and the width of the gray zone should be considered. Surprisingly, the interpretation guides for the results in both IGRA kit inserts do not discuss the topic of analytical uncertainty, i.e., the variation of a positive response around the cutoff point. Based on our pilot results, we have suggested earlier the use of a gray zone (11). This concept has been introduced recently in only one method (8), albeit without reference to the method imprecision.

The aims of this study were (i) to provide an assessment of the total imprecisions of both IGRAs, (ii) to assess between-run imprecisions of both methods by using internal quality control (QC) samples, and (iii) to demonstrate an example from our daily routine for the need of a cautious interpretation of the result falling on a single cutoff point (per manufacturers' instructions).

MATERIALS AND METHODS

Samples. (i) For the conventional assessment of the total imprecision, we obtained samples from three people. Sample 1 was from a male donor 63 years of age who was diagnosed with vertebral TB at 2.5 years of age. The diagnosis was confirmed by an X ray of his vertebra. He did not receive any antituberculosis therapy then or later. The treatment at that time consisted of rest and abundant feeding. This subject was almost asymptomatic throughout his later life, and at the time of the investigation, his status was stable. For the between-run imprecision data, he attended our laboratory for venipuncture nine times, with an interval average of 1 week. During one episode of flu, his venipuncture was postponed. Sample 2 was from a 45-year-old female subject who was born outside Finland and had pulmonary TB 17 years ago. She received a complete course of combination antituberculosis treatment. At the time of investigation, she was asymptomatic. Sample 3 was from a 53-year-old healthy female subject who was vaccinated but who had no history of exposure to TB. The two female subjects were laboratory personnel; their venipunctures were taken regularly at 2- to 8-day intervals and were assayed like any other routine clinical specimen. The blood samples from all participants were taken for the Ly-TbSpot analysis (11) in 10-ml CTP tubes (BD Diagnostics, Helsinki, Finland) and concurrently in the three special tubes of the QuantiFERON-TB Gold In-Tube kit with lyophilized *M. tuberculosis* antigens to perform the B-TbIFNg test (see below). No permission from an Ethical Committee for this study was requested, because all subjects in the study participated voluntarily as health care professionals.

(ii) To assess between-run imprecision without either the interference of biological variation or venipuncture, we used internal quality control samples that were prepared for our routine practice. It is of note that these samples were used with different lots of reagents; therefore, the component of between-lot variability in these experiments cannot be excluded. For the between-run variation of the enzyme immunoassay (EIA) method in the B-TbIFNg analysis, we used an artificially prepared sample that mimics a low-positive sample. This control was prepared by stimulation of the whole blood with phytohemagglutinin (PHA). The supernatant was obtained by centrifugation and adjusted to the level of approximately 0.5 IU/ml. Therefore, it contained naturally produced gamma interferon (IFN-γ). This control was then aliquoted, stored in ampoules at −20°C, and analyzed with each run. The between-run imprecision in the Ly-TbSpot method was assessed using a sample from a male laboratory doctor who was vaccinated. To avoid false-positive reactions, this sample was cryopreserved in CTL reagents (Cellular Technology Ltd., Shaker Heights, OH) (13), aliquoted, and stored in liquid nitrogen before use. According to the manufacturer, this cryopreservation provides on the average 90% cell viability. In our preliminary experiments with trypan blue staining, we confirmed the manufacturer's claims.

(iii) One sample to demonstrate the necessity of cautious interpretation around the cutoff zone was picked up from our routine laboratory practice. This was a sample from a 76-year-old female who attended Helsinki University Eye Disease Hospital for impaired vision in her right eye. She presented with chorioretinitis of uncertain origin; other infection sources of her chorioretinitis were excluded. A Ly-TbSpot test to study exposure to *M. tuberculosis* was requested because she had pneumonia of uncertain etiology in her thirties and her recent chest X ray showed lesions compatible with earlier TB.

IGRAs. IGRAs were performed according to the standard operation procedures (SOP) adopted for both methods in our laboratory. For each test, we developed an internal quality control sample, i.e., the preparation resembling the actual clinical sample that was divided into frozen aliquots and assayed regularly. No external quality control surveillance is yet established for IGRAs.

The Ly-TbSpot assay is a modified version of the commercial T-SPOT.*TB* assay. The complete procedure with modifications was described earlier (11). Briefly, the modifications include the following. (i) Results are expressed as a number of reactive spots/10^6 lymphocytes. The lymphocyte count from isolated PBMC preparation is calculated with an automated hematologic analyzer (Advia 60; Bayer, Germany) for cell quantification and purity assessment. (ii) An additional positive control, i.e., purified protein derivative (PPD; Statens Serum Institut, Copenhagen, Denmark), is used. (iii) A provisional gray zone for the results falling between 25 and 55 spots/10^6 lymphocytes is adopted based on our pilot imprecision test results that showed the average variation around the cutoff as 40%. The lower limit of the gray zone was calculated as 6 reactive cells (per the manufacturer) multiplied by 4 and adjusted to 10^6 lymphocytes, assuming 95% purity of lymphocyte fraction from the PBMCs. On the contrary, the upper limit was calculated based on a high coefficient of variance (CV%) of 40% and low purity of the lymphocyte fraction at 65% from the PBMCs. For each test, we plated 250,000 cells/well. The purity of PBMCs was analyzed for each sample separately before plating using the hematological analyzer.

The B-TbIFNg test is a modified version of the QuantiFERON-TB Gold In-Tube assay. The major modification of this method was the replacement of the original EIA for IFN-γ (Cellestis Limited, Carnegie, Victoria, Australia) with the PeliKine Compact human EIA (Sanquin, Amsterdam, the Netherlands). The latter gave a steeper calibration curve and ensured more-accurate result interpretation in the cutoff zone (11). In contrast to the original procedure, also with this test, we adopted a gray zone policy on the basis of our pilot imprecision data. We interpreted all results showing the reactivity between 0.35 and 0.50 IU/ml as borderline reactive. Sample 2 presented a very high reactivity that was outside the linearity of the calibration curve. To ensure an adequate EIA reading, the supernatants obtained after centrifugation of incubated whole blood with the specific antigens were diluted 1:30 before every analysis. The dilution coefficient was then taken into account when reporting the final B-TbIFNg results.

Data recording. For the spot counts in the Ly-TbSpot method, we used automated evaluation with predefined settings (EliSpot software, version 4.0; AID GmbH, Strasburg, Germany) enabling elimination of operator-dependent and operator-caused variability (5). B-TbIFNg results were calculated from the calibration curve performed with each run. The optical densities were read with iEMS Reader MF (Labsystems, Helsinki, Finland) at 405 nm.

To demonstrate practical problems of the sample interpretation and the importance of the recognition of the variability around the cutoff in the Ly-TbSpot method, we selected a representative sample from our routine and recorded the optical images (see Fig. 1).

Experimental design and data analysis. (i) Calculation of the total imprecision. Within-run imprecisions of the Ly-TbSpot and B-TbIFNg methods were calculated conventionally as coefficients of variance (CV%s) between replicated analyses performed from the same venipuncture. The pure component of between-run imprecision was difficult to assess due to the low viability of lymphocytes. We considered that performing multiple venipuctures in one day for the same person would be unethical. Therefore, we designed the imprecision study for samples obtained at separate venipunctures with intervals ranging from 2 to 8 days. This design allowed assessment of combined between-run imprecision and the impact of the variability related to venipuncture, a preanalytical step. This design might have introduced a component of biological variation. However, because all volunteers were in a stable clinical state and the knowledge that the effector memory T (T_{EM}) cells involved in this analysis are long-lived, we assumed that the biological variability component had only a marginal effect. The

TABLE 1. Within-run imprecisions of B-TbIFNg and Ly-TbSpot

Sample	Method (no. of replicates)	Detected IFN-γ response[a]		CV%
		Mean (range)	SD	
1	B-TbIFNg (10)	2.7 (2.48–2.86) IU/ml	0.11 IU/ml	4.4
	Ly-TbSpot with AgA (8)	227 (165–290) SFU/10^6 lymph	40 SFU/10^6 lymph	17.5
	Ly-TbSpot with AgB (8)	310 (267–358) SFU/10^6 lymph	29 SFU/10^6 lymph	9.4
2	B-TbIFNg (6)	97 (93–102) IU/ml	2.95 SFU/10^6 lymph	3.0
	Ly-TbSpot with AgA (6)	901 (866–931) SFU/10^6 lymph	28 SFU/10^6 lymph	3.1
	Ly-TbSpot with AgB (6)	540 (455–668) SFU/10^6 lymph	73 SFU/10^6 lymph	13.5

[a] SFU/10^6 lymph, spot-forming units per million lymphocytes.

imprecisions were calculated as the CV%s for antigens A (AgA) and B (AgB) separately for Ly-TbSpot and for B-TbIFNg. The total imprecisions for the tests were calculated as described previously (6).

(ii) Calculation of between-run imprecision from quality control samples. We collected retrospectively results obtained by testing internal quality control samples and calculated CV%s. This design avoided imprecision caused by biological variations and venipuncture but introduced additional components in the thawing of cryopreserved cells and between-lot imprecision.

RESULTS

Assessment of the total imprecision of the Ly-TbSpot and B-TbIFNg assays. The within-run imprecisions of the two assays were assessed for the two levels of reactivities, as shown in Table 1. The total imprecisions of both assays, evaluating the components of between-run imprecision, venipuncture, and short-term biological variation, are presented in Table 2. The CV%s of the analysis of the healthy volunteer (sample 3) were far beyond 100%; however, these results have practically no clinical relevance and represented the variability of background noise. It is of note that occasionally this sample produced some reactivity (maximum, 5 spots/10^6 lymphocytes) under stimulation with *M. tuberculosis*-specific peptides. These responses were, however, far below the set cutoff.

Assessment of between-run imprecision of the Ly-TbSpot and B-TbIFNg assays using quality control samples. The between-run imprecision, including a component of the between-lot imprecision with quality control samples, is shown in Table 3. As shown, the QC sample for the B-TbIFNg analysis produced a CV% of only 23% when it was analyzed in 89 runs. This result provided data practically only in the EIA part of the analysis, but the variation introduced by venipuncture and sample incubation was not included. Thus, this result may underestimate the real variation of analysis of low-positive samples. The variations of the QC samples for the Ly-TbSpot analysis were higher, but this analysis included also the components of cell thawing, washing, and calculations.

Demonstration of the necessity of the gray zone approach. An image of the Ly-TbSpot analysis from a patient sample with borderline reactivities in the Ly-TbSpot method is shown in Fig. 1. As is evident, the blank wells had no reactive spots and there were only a couple of reactive cells that recognized antigen A (presumably early secretory antigen target-6 [ESAT-6]). These reactivities were clearly below the cutoff (less than 6 spots/well, per the manufacturer's instructions). However, a disparity in reactivities to antigen B (presumably culture filtrate protein-10 [CFP-10], another antigen specific for *Mycobacterium tuberculosis*) was observed. Interestingly, the number of reactive lymphocytes in one well was three (below the cutoff of the manufacturer), but there were six spots in a replicate well (at the cutoff that should be interpreted as reactive). In other words, a single analysis and a strict adherence to the manufacturer's interpretation recommendations might have resulted in a conflicting interpretation of the this sample as being reactive or nonreactive depending on the result of a single well that might occur by chance.

DISCUSSION

Here, we demonstrate an example of the total magnitude of imprecision of the IGRA methods. As far as we are aware, this is the first report to evaluate IGRA imprecision according to recommendations adopted for clinical chemistry. The total imprecision for the Ly-TbSpot and B-TbIFNg assays reached 37.8%. At moderately high and high reactivities, the obtained variability was surprisingly low, taking into account the multi-

TABLE 2. Total imprecisions of B-TbIFNg and Ly-TbSpot

Sample	Method (no. of replicates)	Data from extended between-run imprecision[a]			Total imprecision CV%
		Detected IFN-γ response		CV%	
		Mean (range)	SD		
1	B-TbIFNg (7)	3.8 (2.6–6.3) IU/ml	1.45 IU/ml	37.7	37.8
	Ly-TbSpot with AgA (9)	204 (153–326) SFU/10^6 lymph	51 SFU/10^6 lymph	24.8	31.5
	Ly-TbSpot with AgB (8)	308 (189–425) SFU/10^6 lymph	77 SFU/10^6 lymph	24.9	26.6
2	B-TbIFNg (6)	135 (95–211) IU/ml	42 IU/ml	31.2	31.2
	Ly-TbSpot with AgA (6)	834 (684–1,045) SFU/10^6 lymph	140 SFU/10^6 lymph	16.8	17.2
	Ly-TbSpot with AgB (6)	486 (385–636) SFU/10^6 lymph	62 SFU/10^6 lymph	12.7	19.7

[a] SFU/10^6 lymph, spot-forming units per million lymphocytes.

TABLE 3. Between-run (including between-lot) imprecisions of the B-TbIFNg and Ly-TbSpot methods with quality control samples

Method (no. of replicates)	Detected IFN-γ response[a]		CV%
	Mean	SD	
B-TbIFNg (89)	0.71 IU/ml	0.16 IU/ml	23
Ly-TbSpot			
Blank (7)[b]	3.5 SFU/10^6 lymph	2.5 SFU/10^6 lymph	73
PPD (7)	940 SFU/10^6 lymph	296 SFU/10^6 lymph	32

[a] SFU/10^6 lymph, spot-forming units per million lymphocytes.
[b] Incubation with culture media only.

ple stages of the analysis. Indirectly, these results imply that no meaningful biological variation occurred when the clinical condition was stable and when no treatment intervention had commenced (sample 1). Indeed, this volunteer was healthy throughout the period of sample collection, with the exception of a flu episode for which the obtaining of the sample was postponed. There was no antigenic boosting during the time of the investigation. This also illustrates an interesting immunological phenomenon of the longevity of immunological memory and very slow attrition of the reactive cell pool (sample 2). One plausible explanation for the acceptably moderate variability of the Ly-TbSpot method is the advantage of normalization of reading outcome, i.e., the calculation of frequencies of reactive spots in relation to the purified lymphocyte fraction (11). A steeper calibration curve adopted for the B-TbIFNg assay (11) is also advantageous and contributed to the low variability.

Using a high number of replicates ($n = 89$), we observed the between-run imprecision, including a component of the between-lot imprecision, of B-TbIFNg at the level around the cutoff of a 23% CV%. It is of note, however, that this design underestimated the variability that comes from the activation and displays only the variability of the EIA-based analyte determination. The between-run imprecision with a cryopreserved quality control sample for the Ly-TbSpot was in the range of an approximate CV% of 30%. This study design, on the contrary, may have overestimated the imprecision introduced by steps that are usually not involved in clinical sample analysis, i.e., cryopreservation and thawing, but these steps were necessary because of the instability of lymphocytes.

Combined, the results show that despite multiple stages of the analysis (each contributing to the total imprecision), the obtained CV%s are acceptable for clinical diagnosis, although they are somewhat higher than the CV%s reported for the majority of serological methods (Table 4). It is self-evident that every step of the assay contributes to the total imprecision. For instance, the calculation of the white blood cell count by using a blood cell analyzer may add to the total imprecision by a CV% as high as 2.5% (acceptable imprecision according to the Advia 60 manual [1]). As a matter of fact, the more robust and automated the test, the lower the CV%s are, as illustrated by the comparison of methods that were randomly selected in our laboratory (Table 4; the data were collected from the kit instructions).

The limitation in our study design was attributed to ethical problems in getting more serial clinical samples for the assessment. The obtained between-run imprecision comprised either (in theory) a component of short-term biological variation or additional steps normally not involved in the analysis. The strength of our approach, however, is that we were able to assess the total imprecision, including the component of venipunctures. We avoided the impact of between-operator imprecision and any subjective interpretation because all readings were performed using the same operator and the same instrument settings.

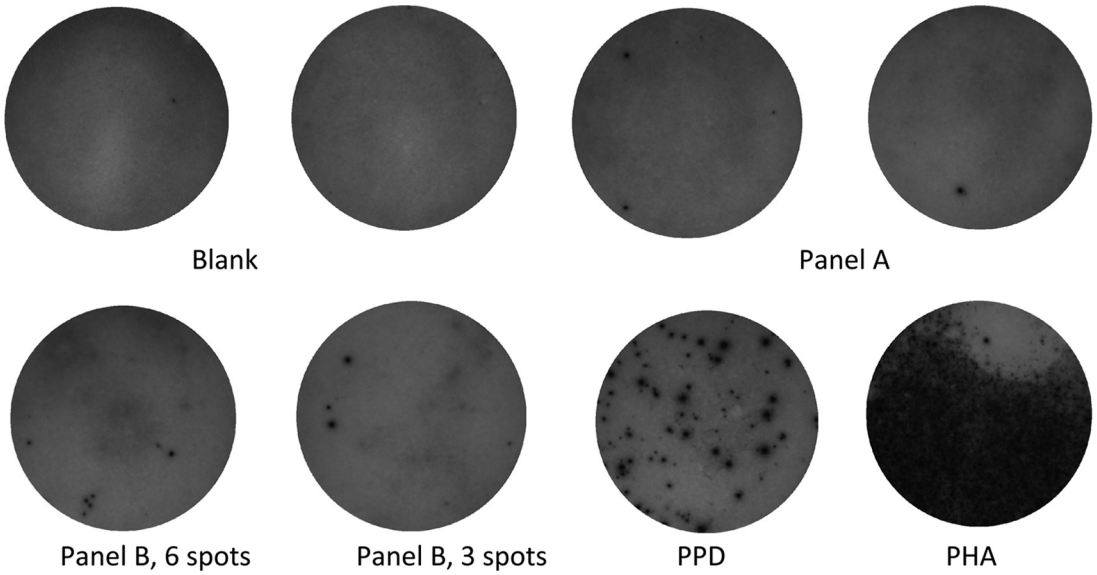

FIG. 1. Illustration of imprecision in the reactive spot counts in the ELISPOT images. The sample was picked up from our routine analysis. (Upper panel) No reactivity in culture media (Blank, negative control of the kit), with one or two reactive spots under stimulations with antigen A (Panel A). (Lower panel) Reactivities of six spots under stimulation with antigen B (positive per manufacturer) and three spots in a replicate well (negative per manufacturer) (Panel B) as well as reactivities to PPD and PHA (positive control of the kit).

TABLE 4. Comparison of the imprecision data from different laboratory methods[a]

Method	Manufacturer	CV% (no. of replicates)		
		Within-run imprecision	Between-run imprecision	Total imprecision[b]
Glucose, automated (Cobas Integra 400 Plus)	Roche (data obtained from HUSLAB clinical chemistry laboratory evaluation)	0.7–2 (36–188)	1.2–2.4 (36–188)	1.7–3.1
Axsym HIV Ab/Ag Combo, automated[c]	Abbott	3.7–7.2	4.6–9.3	NR
TSH receptor autoantibodies, manual method	RSR Limited, Cardiff, UK	4.2–5.5 (21)	8.7–8.8 (25)	NR
Anti-C1q autoantibodies, manual method	Bűhlmann Laboratories Ag, Schönenbuch, Switzerland	4.4–7.1	7.1–14.3	NR
MPO ANCA, manual method	Euro-Diagnostica, Malmö, Sweden	4–12 (8)	4–25 (6)	NR
ImmuKnow manual method	Cylex, Columbia, MD	7.7–15.6 (5)[d]		NR

[a] The data are collected from the manufacturers' kit instructions. Data for the glucose measurement are taken from the internal evaluation of Cobas Integra 400 Plus at HUSLAB, Department of Clinical Chemistry.
[b] NR, not reported.
[c] This study had 10 sites, 10 instruments, and 2 or 3 different lots (total number of replicates, 189).
[d] Nonconventional study design.

In this study design, we made several assumptions. First, we assumed that the responses to antigens A and B are independent and are mediated through different clones of reactive T lymphocytes. Second, we presumed that without any therapeutic intervention that might change the balance between the antigen, antigen-presenting cells, or cytokine milieu, the frequencies of T_{EM} cells are more or less stable. Third, we assumed that the kinetics of attrition of the reactive cell pool corresponding to the treatment 17 years ago is very slow and will not influence the results of our imprecision studies. In fact, biological variation might have had some impact on our results, but we assumed that the frequencies of T_{EM} cells, being the target of this investigation, although possibly fluctuating, were possibly at homeostasis because no new antigenic insult, anti-TB therapeutic intervention, or iatrogenic immunosuppression was recorded for our volunteers during the period of sample collection. We also excluded sample collection during the flu period to avoid nonspecific lymphocyte activation. In other words, the CV%s observed in this study were related to the imprecision caused by multiple stages of the method performance and reflected also normal biological variation. These results corroborated our previous notion (11) that in IGRAs, like in other immunodiagnostic methods, imprecision does exist and should be considered especially at the cutoff point setting.

The importance of recognition of analytical imprecision around the cutoff zone for decision making was further illustrated by the example taken from our routine practice. Indeed, this sample could have been interpreted as a positive or a negative purely by chance. When the method imprecision is known, ambiguous interpretations, especially around the cutoff zone, can be avoided. The results of the illustrated case were reportedly borderline and interpreted as compatible with an immunological scar of an earlier encounter with *M. tuberculosis*. In theory, the gray zone might be ranging also below the cutoff given by the manufacturer. We have no data to evaluate this point.

In the ELISPOT assay, one of the major impacts on the reading outcome is the correct definition of what optical image is considered a spot. Such parameters as size, sharpness, and roundness of the spots are chosen subjectively by the operator when calibrating the spot reader. These problems with instrument calibration do not exist in enzyme-linked immunosorbent assay (ELISA)-based methods, as all ELISA readers measure optical densities with the same physically defined scale. The lack of a standard for the definition of the spot inevitably implies that each laboratory should establish its cutoff point and the gray zone. It also implies that the numerical results obtained from different laboratories may not be comparable. In this study, we used prefixed settings of the ELISPOT reader. Because the definition of the spot is subjective, we may have introduced a systematic error, but this error did not affect the imprecision results.

Imprecision has a decisive influence on the clinical interpretation of the dynamics of IGRA responses. In a recently published study (4), immunological responses in children defined as having LTBI and active TB and in those who were healthy contacts were monitored. Blood samples were collected from all tested subjects at regular time intervals. By plotting the IFN-γ values (from the QuantiFERON) as a function of time, the researchers concluded that they were able to observe an increase of the analyte at day 10 of the treatment. In their interpretation, the increase confirmed the diagnosis while the decrease at later time points suggested success in the curative therapy. These results may have indispensable clinical value; however, the analytical imprecision of the methods was not addressed. The algorithm rules for decision making regarding demonstration of an increase or decrease in magnitude remain unclear.

In another study (14), the assessment of reproducibility in the QuantiFERON-TB assay was performed when blood samples were taken from 14 volunteers from India at days 0, 3, 9, and 12 and analyzed batchwise in two EIA runs with an interval of 1 week. This study design, however, does not provide information on within- or between-run imprecision or the total imprecision. The numerical values with a clinical judgment for each donor were presented, and 2/14 (14%) discordant results (change of "positive" interpretation to "negative") were ob-

served. Upon repetition of the analysis, the discordant results could not be confirmed. It is noteworthy that there were, however, considerable variations in the numerical IFN-γ values throughout all measurement time points. Without awareness of the total imprecision of the method, these discordances may be taken as a biological phenomenon and may be erroneously interpreted as an immunological reversion.

In the most recent relevant publication, a short-term reproducibility study of the QuantiFERON-TB Gold In-Tube assay was presented (3). Only two venipunctures were taken, at an interval of three days. The authors considered the method to be robust and reproducible and concluded that considerable variability was due to intraindividual variations. However, an alternative explanation that many discordant results (5 out of 27) could have been due to the method itself and not due to the human biology was not presented. In our opinion, the immune system is at homeostasis at least within the three consequent days if not challenged, and the observed variability was most likely related to the variability of the method.

In conclusion, although IGRAs have made a great breakthrough in our current arsenal of diagnostics of LTBI, nearly none of the published studies have thoroughly assessed their results through the prism of assay variability. We advocate more studies on imprecision and an acceptance of a concept of the gray zone to avoid interpretation ambiguities.

ACKNOWLEDGMENTS

We thank for financial support in the evaluation of new IGRA methods the following private Finnish foundations: Finnish Lung Health Association (Filha Ry), Pulmonary Association Heli, and the Tuberculosis Association of the University of Tampere.

Tamara Tuuminen has participated in the Biomerieux-sponsored seminar "Latent Tuberculosis Forum." The other authors declare that they have no financial relationship with a commercial entity that has an interest in the subject of this paper.

REFERENCES

1. **Bayer HealthCare.** 2003. Advia 60 manual. Bayer HealthCare, Tarrytown, NY.
2. **Carrara, S., D. Vincenti, N. Petrosillo, M. Amicosante, E. Girardi, and D. Goletti.** 2004. Use of T cell-based assay for monitoring efficacy of antituberculosis therapy. Clin. Infect. Dis. **38:**754–756.
3. **Detjen, A. K., L. Loebenberg, H. M. S. Greval, K. Stanley, A. Gutschmidt, C. Kruger, N. Du Plessis, M. Kidd, N. Beyrs, G. Walzi, and A. C. Hesseling.** 2009. Short-term reproducibility of a commercial interferon gamma release assay. Clin. Vaccine Immunol. **16:**1170–1175.
4. **Herrmann, J.-L., M. Belloy, R. Porcher, N. Simonney, R. Aboutaam, M. Lebourgeois, J. Gaudelus, L. De Losangeles, K. Chadelat, P. Schneinmann, N. Beydon, B. Fauroux, M. Bingen, M. Terki, D. Barraud, P. Cruaud, C. Offredo, A. Ferroni, P. Berche, D. Moissenet, H. Vuthien, C. Doit, E. Bingen, and P. H. Lagrange.** 2009. Temporal dynamics of interferon gamma responses in children evaluated for tuberculosis. PLoS One **4:**e4130.
5. **Janetzki, S., S. Schaed, N. E. Blachere, L. Ben-Porat, A. N. Houghton, and K. S. Panageas.** 2004. Evaluation of elispot assays: influence of method and operator on variability of results. J. Immunol. Methods **291:**175–183.
6. **Krouwer, J. S., and R. Rabinowitz.** 1984. How to improve estimates of imprecision. Clin. Chem. **30:**290–292.
7. **Menzies, D., M. Pai, and G. Comstock.** 2007. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. Ann. Intern. Med. **14:**340–354.
8. **Oxford Immunotec.** 2009. T-SPOT.*TB* package insert. Oxford Immunotec, Marlborough, MA. http://www.oxfordimmunotec.com/USpageinsert.
9. **Pai, M., and R. O'Brien.** 2007. Serial testing for tuberculosis: can we make sense of T cell assay conversions and reversions. PLoS One **4:**e208.
10. **Pai, M., A. Zwerling, and D. Menzies.** 2008. Systematic review: T-cell-based assays for the diagnostics of latent tuberculosis infection: an update. Ann. Intern. Med. **149:**177–184.
11. **Tavast, E., I. Seppälä, E. Salo, and T. Tuuminen.** 2009. IGRA tests perform similarly to TST but cause no adverse reactions: pediatric experience in Finland. BMC Res. Notes **2:**9.
12. **The TDR Diagnostics Evaluation Expert Panel.** 2006. Evaluation of diagnostic tests for infectious diseases: general principles. Nat. Rev. Microbiol. **4:**S20–S32.
13. **Tuuminen, T., S. Sorva, K. Liippo, T. Vasankari, H. Soini, B. Eriksén-Neuman, A. Miettinen, and I. Seppälä.** 2007. Feasibility aspects of commercial interferon (IFN)-γ based methods for the diagnosis of latent *Mycobacterium tuberculosis* infection in Finland, a country of low incidence and high BCG vaccination coverage. Clin. Microb. Infect. **13:**836–838.
14. **Veerapathran, A., R. Joshi, K. Goswami, S. Dogra, E. E. Modie, M. V. Reddy, S. Kalantri, K. Schwartzman, M. A. Behr, D. Menzies, and M. Pai.** 2008. T-cell assays for tuberculosis infection: deriving cutoffs for conversion using reproducibility data. PLoS One **3:**e1850.