Within-Subject Variability of *Mycobacterium tuberculosis*-Specific Gamma Interferon Responses in German Health Care Workers[▽]

Felix C. Ringshausen,[1]* Albert Nienhaus,[1] José Torres Costa,[2] Heiko Knoop,[3] Stephan Schlösser,[4]
Gerhard Schultze-Werninghaus,[3] and Gernot Rohde[5]

University Hospital Hamburg–Eppendorf, Institute for Health Services Research in Dermatology and Nursing, Hamburg, Germany[1];
Faculty of Medicine, University of Porto, Alameda Professor Hernâni Monteiro, Porto, Portugal[2]; University Hospital Bergmannsheil,
Department of Pneumology, Allergology, and Sleep Medicine, Ruhr-University Bochum, Bochum, Germany[3];
University Hospital Bergmannsheil, Department of Occupational Medicine, Ruhr-University Bochum,
Bochum, Germany[4]; and Maastricht University Medical Centre, Department of
Respiratory Medicine, Maastricht, Netherlands[5]

Gamma interferon (IFN-γ) release assays (IGRAs) are used increasingly for the periodic tuberculosis (TB) screening of health care workers (HCWs), although data regarding the reproducibility and interpretation of serial testing results in countries with a low incidence of TB are scarce. The present study evaluated and compared the within-subject variability of dichotomous and continuous results of two commercial IGRAs, the QuantiFERON-TB Gold In-Tube (QFT) and the T-SPOT (T-SPOT), in German HCWs during a 4-week period. Thirty-five immunocompetent HCWs with low or medium TB screening risk and without known recent TB exposure or tuberculin skin test application were tested repeatedly with both IGRAs at weekly intervals. A total of 158 valid results were obtained for each IGRA. Changes of about ±70% (QFT) and ±60% (T-SPOT) from the mean IFN-γ response accounted for 95% of the within-subject variability. However, according to the manufacturers' cutoffs, inconsistent results were observed more frequently for the QFT (28.6%; four conversions, six reversions) than for the T-SPOT (8.6%; three reversions; *P* < 0.001). The overall agreement between the IGRAs was good. Regression toward the means accounted for a significant decline in mean IFN-γ responses of about 25% between successive visits for both IGRAs. Although both assays were highly reliable and reproducible, we observed substantial within-subject variability and regression toward the means during a 4-week period, which should be considered when interpreting serial testing results in comparable populations and settings. Our data support the use of borderline zones for the interpretation of serial IGRA results and the retesting of borderline positive results before offering preventive chemotherapy.

Health care workers (HCWs) are at increased risk of occupational exposure to tuberculosis (TB) (18, 30). Hence, the periodical screening of HCWs for latent tuberculosis infection (LTBI) according occupational safety and health (OSH) measures and preventive treatment are essential components of TB infection control programs in countries with a low incidence of TB (4, 35).

For decades, the tuberculin skin test (TST) has been the tool of choice for the diagnosis of LTBI among HCWs (14). However, the TST has well-known limitations (12, 17). In contrast, the commercially available gamma interferon (IFN-γ) release assays (IGRAs) QuantiFERON-TB Gold In-Tube (QFT; Cellestis, Carnegie, Australia) and T-SPOT.TB (T-SPOT; Oxford Immunotec, Abingdon, United Kingdom) possess distinct advantages for use in serial testing and have emerged as attractive alternatives for the TB screening of HCWs (15). These one-visit, *ex vivo* blood tests avoid sensitization, the boosting of immune response, cross-reactivity with most nontuberculous mycobacteria, or loss to follow-up, as frequently observed with

the TST (12, 17). Recently, systematic reviews have shown that IGRAs have excellent specificity for the diagnosis of LTBI even in Bacillus Calmette-Guérin (BCG)-vaccinated populations (8, 23). Furthermore, they correlate with occupational risk factors of TB exposure in low-incidence settings (37).

Although data regarding the reproducibility of IGRA results and their performance in serial testing are limited, the use of IGRAs is rapidly expanding. Two recently published systematic reviews on the within-subject variability and performance in the serial testing of HCWs identified only 4 and 10 studies, respectively, which met the predefined eligibility criteria (33, 37). Studies on reproducibility and within-subject variability were predominantly from high-incidence settings (India and South Africa) and included only small numbers of subjects (67 in total) (6, 32, 34). In contrast, serial testing studies were mostly from countries with a low to intermediate TB incidence (5, 13, 20, 24–26, 36). Across the board, these studies indicate that IGRA conversions and reversions occur at variable rates due to significant variability and that the manufacturers' simplistic dichotomous negative-to-positive cutoffs may not be appropriate for the interpretation of repeat IGRA results. Accordingly, the most recent guidelines on the use of IGRAs for the diagnosis of TB infection are more cautious regarding their use in serial testing than certain earlier guidelines (4, 7, 11, 15, 16).

Various borderline zones around the manufacturers' thresh-

---

* Corresponding author. Mailing address: Institute for Health Services Research in Dermatology and Nursing, University Hospital Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. Phone: 49 40 7410 55 428. Fax: 49 40 7410 55 348. E-mail: f.ringshausen@uke.de.
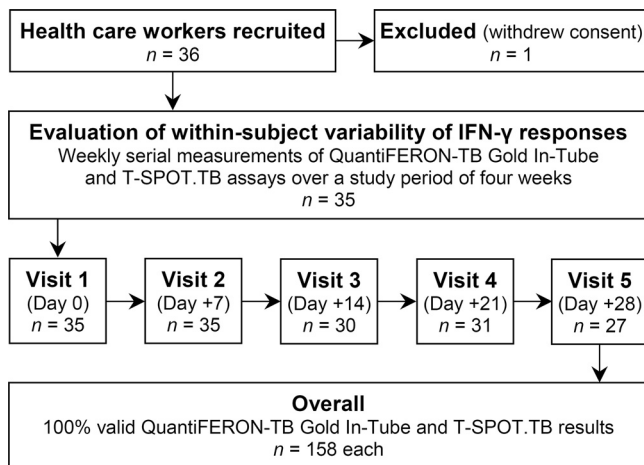
FIG. 1. Study flow chart.

olds and definitions of conversions and reversions have been proposed to improve their interpretation in serial testing (15, 20, 21, 24, 32, 34). Apparently, the performance of IGRAs depends on the population and epidemiologic setting in which these tests are applied (9). Thus, it is not surprising that consensus on the definition and prognosis of IGRA conversions, reversions, and borderline zones still is lacking (22). To date, only a few studies systematically determined different components of technical and biological IGRA variability (6, 34). A single study assessed within-subject variability in a head-to-head comparison between both commercial IGRAs (32), but no peer-reviewed study reporting data on the within-subject variability of IFN-γ responses from a low-incidence setting has been published yet.

The objective of the present study was to evaluate and compare the reproducibility and within-subject variability of dichotomous and continuous measures of two commercial IGRAs, the QFT and the T-SPOT, in German HCWs during a 4-week study period.

MATERIALS AND METHODS

Study design and subjects. The present study was conducted at the University Hospital Bergmannsheil in Bochum, Germany, between July 2010 and February 2011. All IGRAs were performed at the Pulmonary Research Laboratory by a team with profound experience in IGRA research (26, 27). Eligible HCWs were recruited prospectively by an occupational physician. Occupational health records were consulted to identify HCWs with prior positive IGRA or TST results so that at least 10 subjects with positive baseline results for each IGRA could be included. Hence, the frequency of reported positive IGRA results does not reflect the prevalence of LTBI among HCWs at our institution, which previously has been shown to be around 9% (27). Subjects were interviewed and repeatedly tested with the QFT and the T-SPOT assay at weekly intervals during a 4-week period (Fig. 1). Inclusion criteria were an age of 18 years and above, current engagement in health care work, valid baseline QFT and T-SPOT results, and written and informed consent. Exclusion criteria were recruitment from a setting with evidence of ongoing TB transmission, current active TB or treatment for LTBI, unprotected TB exposure, TST application less than 6 and 3 months prior to or during the study, respectively, absence from more than two and/or one of the first two study visits, and withdrawal of consent. All HCWs with medium TB screening risk or persistently positive IGRA results were subjected to further follow-up according to German occupational safety and health legislation.

Diagnostic methods. The standardized interview and questionnaire covered individual and occupational risk factors for recent TB infection, LTBI reactivation, and false-negative or -positive TB immune responses (symptoms suggestive for TB, previous TB and antituberculous treatment, past and recent exposure,

previous TST results, BCG vaccination, comorbidity, immunosuppression, profession, affiliation, and workplace). BCG status was confirmed via medical records or vaccination scars. TB screening risk was classified according to the Centers for Disease Control and Prevention (4).

The QFT, an enzyme-linked immunosorbent assay (ELISA), measures IFN-γ secreted by T-cell lymphocytes after stimulation with the highly *Mycobacterium tuberculosis*-specific antigens ESAT-6, CFP-10, and TB7.7. The T-SPOT employs the enzyme-linked immunospot technique (ELISPOT) and measures IFN-γ secreting spot-forming T cells per well and 2.5×10^5 peripheral blood mononuclear cells using ESAT-6 and CFP-10. Both IGRAs were manually performed and repeated by the same technician in strict adherence to the manufacturers' instructions (3, 19). Blood was collected in the provided QFT test tubes and lithium heparin tubes for the T-SPOT and immediately processed. Blood sampling was performed at the same time of day for each subject to minimize diurnal variation. Serum samples for the QFT were stored at 4°C, and IFN-γ ELISAs were performed within 2 weeks of blood sampling (median, 2; interquartile range, 1 to 8 days). All samples from one subject were tested with kits from the same batch. All T-SPOT spot counts were read manually (to a maximum spot count of 100) by the same two independent technicians using a stereomicroscope (Zeiss, Göttingen, Germany). In cases of disagreement, a technician from Oxford Immunotec performed the decisive spot count. The technicians were blinded to subject details and QFT results.

All assays met the manufacturers' quality standards; in particular, all mitogen (positive) controls showed strong responses of >4.5 IU/ml and >20 SFCs, respectively. Additional quality assurance was implemented. Within-test variability was assessed via test-retest reproducibility on the same set of aliquoted specimens for both assays and by ELISA reader reliability through repeated determination of optical density values of the same QFT plate for 1 h (450-nm filter, 620-nm reference filter; Zenyth 340R ELISA reader; Anthos, Krefeld, Germany). All experiments assessing test-retest reliability were carried out on the same microtiter plate.

The results were considered positive if the QFT IFN-γ response of TB antigen minus the negative control was ≥0.35 IU/ml (and ≥25% of the negative control value), and T-SPOT results were considered positive if the number of SFCs was ≥6 SFCs for either ESAT-6 or CFP-10 after subtracting the spot count of the negative control, respectively. If the negative control had ≥6 SFCs, ESAT-6 or CFP-10 had to exhibit a spot count of at least twice the spot count of the negative control. Conversion was defined as baseline IFN-γ concentrations of <0.35 IU/ml followed by a result of ≥0.35 IU/ml and baseline spot counts of <6 SFCs followed by a result of ≥6 SFCs, respectively. Reversion was defined as the opposite progression. As previously proposed (21, 29, 32), uncertainty zones of 0.2 to 0.7 IU/ml (QFT) and 4 to 8 SFCs (T-SPOT) and increases from <0.35 to >0.7 IU/ml (QFT) and from <4 to >8 SFCs (T-SPOT) were used for borderline zone analysis and as definitions for conversions (and reversions), respectively.

Statistical analysis. Categorical data were compared using Pearson's chi-squared or Fisher's exact test, where appropriate. Normal distributions in continuous variables were determined using the Kolmogorov-Smirnov test. Kappa values were calculated to assess agreement between dichotomous measures (28). Within-subject variability of IFN-γ responses was assessed using coefficient of variation and linear mixed-effect model analyses. For the CV analysis, mean IFN-γ responses were calculated for each subject during the 4-week study period. Within-subject variability was calculated by determining the mean standard deviations (SDs) for all subjects. Assuming that 95% of subjects are located within the range of 2 SDs, expressing this SD as a percentage of the mean IFN-γ response allowed the calculation of the 95% range of within-subject variability. Continuous IFN-γ responses were log transformed to normalize their differences. Linear mixed-effect model analysis was performed using the restricted maximum-likelihood method with repeated measures, an autoregressive covariance structure, and random subject intercepts to adjust for patterned relationships across time (10). In addition, they were used to assess the association of covariates and factors with log IFN-γ responses. Reliability was assessed using intraclass correlation coefficients, which were computed using unconditional intercept-only random effect models. The Akaike information criterion was used to assess the model fit. Significance was assessed using P values computed for t statistics corresponding to parameter estimates or from Wald statistics, respectively. Data analysis was performed using SPSS, version 11.5 (SPSS Inc., Chicago, IL). $P < 0.05$ was considered significant.

The study was approved by the ethics committee of the Ruhr-University Bochum, Germany. All participants gave their written and informed consent.

TABLE 1. Characteristics of the study population

Variable	No. (%)
Subjects, total	35 (100)
Age	
20–29 yr	5 (14.3)
30–39 yr	8 (22.9)
40–49 yr	16 (45.7)
≥50 yr	6 (17.1)
Sex	
Male	7 (20.0)
Female	28 (80.0)
Foreign country of birth	6 (17.1)
Birth in high-TB-incidence country[a]	2 (5.7)
Personal history of TB[b]	1 (2.9)
TB screening risk classification	
Low	17 (48.6)
Medium	18 (51.4)
Duration of employment in health care	
1–9 yr	7 (20.0)
10–19 yr	11 (31.4)
20–29 yr	8 (22.9)
≥30 yr	9 (25.7)
Prior tuberculin skin test[c]	30 (85.7)
Results of prior tuberculin skin test	
Positive	14 (46.7)
Negative	16 (53.3)
BCG vaccination	
Yes	20 (57.1)
No	15 (42.9)
Health care profession	
Nursing or physician	14 (40.0)
Other	21 (60.0)

[a] Countries with an annual TB incidence of >50 per 100,000 population (according to the World Health Organization) included Estonia ($n = 1$) and Kazakhstan ($n = 1$).
[b] The year of personal history of TB was 1948.
[c] The prior tuberculin skin test had been administered a means of 11 ± 6.1 (range, 3 to 21) years ago. In the majority of subjects (24/30; 80.0%) the qualitative multipuncture method was used.

RESULTS

Study population. Thirty-five subjects constituted the final study population (Fig. 1). The characteristics of the study population are shown in Table 1. The mean age was 42 ± 10.5 (range, 20 to 62) years. The mean duration of employment in health care was 20 ± 11.4 (range, 1 to 40) years. None of the study participants was HIV positive or reported symptoms suggestive of TB disease, immunosuppressive medication, comorbidity, concomitant illnesses, or unprotected TB exposure during the study period. The participants were recruited from a variety of different in- and outpatient departments (accident and emergency, anesthesiology, endocrinology, gastroenterology, gynecology, laboratory, patient services, physical therapy, pneumology, radiology, and surgery) and professions (administration, cleaning staff, endoscopy assistant, nursing, physical therapist, physician, and technical assistant).

Within-subject variability. A total of 158 valid results from 35 subjects were available for analysis of within-subject variability for each IGRA (Fig. 1). Five serial measurements at weekly intervals were obtained for a total of 19 subjects, four for 15 subjects, and three for 1 subject.

IFN-γ responses varied considerably within subjects during the 4-week study period (Fig. 2). According to the manufacturers' predefined negative-to-positive cutoffs, the QFT showed four unstable conversions (4/23; 17.4%) and six definite reversions (6/12; 50.0%), while the T-SPOT showed zero
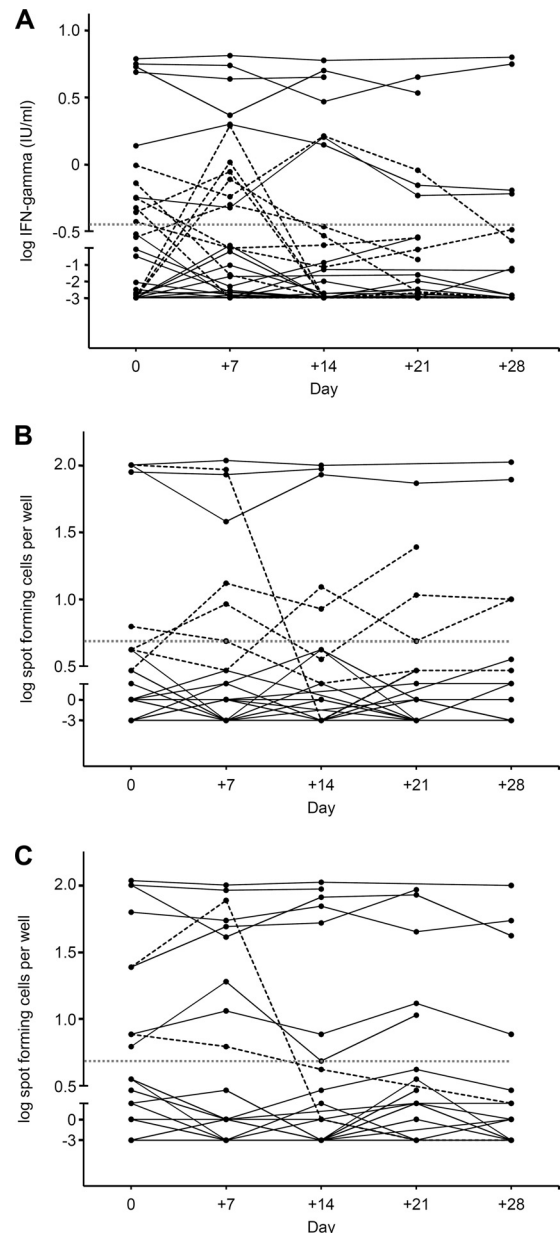

FIG. 2. Within-subject variability of log IFN-γ responses during five successive study visits for QuantiFERON-TB Gold In-Tube (A), T-SPOT.TB panel A (B) (ESAT-6), and T-SPOT.TB panel B (CFP-10) (C). The dotted gray lines indicate the manufacturers' diagnostic assay thresholds. The dashed lines represent individuals crossing this threshold.

TABLE 2. Consistency of IGRA during the 4-week study period stratified by baseline IGRA result

Baseline IGRA result	Total no. of subjects	Overall trend	
		Consistent	Inconsistent
		n (%)	n (%)
QFT (IU/ml)			
Total	35	25 (71.4)	10 (28.6)
<0.2	21	18 (85.7)	3[a] (14.3)
0.2–<0.35[b]	2	1 (50.0)	1[a] (50.0)
0.35–0.7[b]	5	1 (20.0)	4[c] (80.0)
>0.7–<3.0	3	1 (33.3)	2[c] (66.7)
≥3.0	4	4 (100)	0
T-SPOT (SFC)[d]			
Total	35	32 (91.4)	3 (8.6)
0–3	24	24 (100)	0
4–5[b]	1	1 (100)	0
6–8[b]	2	1 (50.0)	1[c] (50.0)
9–29	2	1 (50.0)	1[c] (50.0)
≥30	6	5 (83.3)	1[c] (16.7)

[a] All conversions were unstable (increase across the diagnostic threshold of 0.35 IU/ml from a baseline IFN-γ concentration of <0.35 IU/ml and subsequent permanent decrease to below 0.35 IU/ml for the QFT).

[b] The ranges from 0.2 to 0.7 IU/ml and 4 to 8 SFC represent proposed borderline zones for the interpretation of repeated QFT and T-SPOT results, respectively.

[c] All reversions were definite (permanent decrease below 0.35 IU/ml from a baseline IFN-γ concentration of ≥0.35 IU/ml for the QFT, and permanent decrease below 6 SFC from an SFC count of ≥6 SFC for the T-SPOT).

[d] Maximum spot count either from the ESAT-6 or the CFP-10 panel (minus the negative control spot count), whichever was higher.

FIG. 3. Mean log IFN-γ responses at each study visit for Quanti-FERON-TB Gold In-Tube (●), T-SPOT.TB panel A (□) (ESAT-6), and T-SPOT.TB panel B (CFP-10) (◆). Error bars depict the standard errors of the means from each assay at each point in time.

conversions and three definite reversions (3/10; 30.0%). Overall, inconsistent results were more frequently observed for the QFT (10/35; 28.6%) than for the T-SPOT (3/35; 8.6%; $P < 0.001$). Nineteen and 25 subjects were persistently QFT and T-SPOT negative, while six and seven subjects remained QFT and T-SPOT positive during the study period, respectively (Table 2). All of the six persistently QFT-positive subjects and six out of seven persistently T-SPOT-positive subjects had been classified as having a medium TB screening risk.

The maximum range in the magnitude of IFN-γ response within a subject was 2.58 IU/ml for the QFT assay and 99 SFCs for the T-SPOT assay. Seven and five subjects showed ranges of >1.0 IU/ml and >10 SFCs, respectively. In contrast, five and four subjects showed absolutely no variability and remained stationary at values of 0.00 IU/ml and 0 SFC, respectively. We estimated variance components from a linear mixed-effect model and found significant within-subject variation of IFN-γ
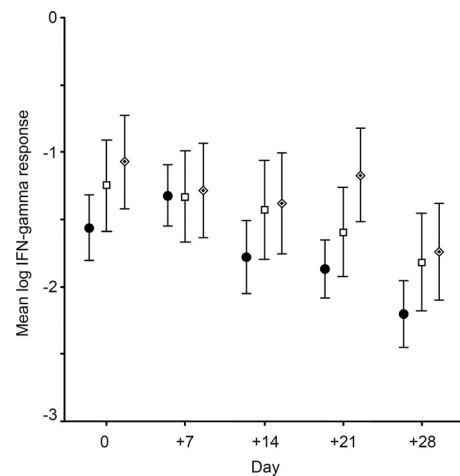
responses at each study visit for both IGRAs (all $P \leq 0.01$), except for the QFT assay at the fourth visit ($P = 0.064$). Coefficient of variation and linear mixed-effect model analyses demonstrated that 95% of the within-subject variability of IFN-γ responses during the 4-week study period occurred within 67 and 61% of the mean IFN-γ response for the QFT and the T-SPOT assay, respectively. Hence, changes from mean responses of more than roughly ±70% (QFT) and ±60% (T-SPOT) are unlikely to represent nonspecific variation and may indicate a true change in TB infection status.

Moreover, we observed a substantial decline in the positivity rate (Table 3) and correspondingly in mean IFN-γ responses for both IGRAs over time (Fig. 3). The average change in IFN-γ responses between successive study visits was estimated by controlling for the respective baseline IFN-γ response. We found a significant decrease in the mean IFN-γ responses of 26% (95% confidence interval [CI], 12.1 to 39.3%; $P < 0.001$) and 24% (95% CI, 9.3 to 39.1%; $P < 0.001$) for the QFT and T-SPOT, respectively.

Using a simple linear mixed-effect model, we identified age as being significantly associated with the magnitude, and hence the absolute variability, of successive IFN-γ responses for both IGRAs ($P < 0.05$), while no relation with BCG vaccination and foreign origin was observed ($P \geq 0.40$). More complex linear

TABLE 3. Concordance between the two commercial IGRAs stratified by study visit

Visit no. (study day)	No. of subjects per visit	No. (%) of positive QFT results	No. (%) of positive T-SPOT results	No. (%) of discordant IGRA results	Agreement between QFT and T-SPOT results	
					Raw (%)	Kappa (P)
1 (0)	35	12 (34.3)	10 (28.6)	8 (22.9)	77.1	0.47 (0.005)
2 (+7)	35	12 (34.3)	10 (28.6)	6 (17.1)	82.9	0.60 (<0.001)
3 (+14)	30	7 (23.3)	7 (23.3)	2 (6.7)	93.3	0.81 (<0.001)
4 (+21)	31	5 (16.1)	5 (16.1)	2 (6.5)	93.5	0.76 (<0.001)
5 (+28)	27	4 (14.8)	4 (14.8)	2 (7.4)	92.6	0.71 (<0.001)
Total	158	40 (25.3)	36 (22.8)	20 (12.7)	87.3	0.65 (<0.001)

TABLE 4. Characteristics of subjects with inconsistent IGRA results over the 4-week study period

Subject no.	Age (yr)	Sex[c]	TB risk	IGRA	Result from visit (study day)[d]:					Overall trend[a]
					1 (0)	2 (+7)	3 (+14)	4 (+21)	5 (+28)	
7	41	M	Medium	QFT	Positive (0.45)	Positive (1.05)	Negative (0.00)		Negative (0.00)	Inconsistent
				T-SPOT	Negative (1/0)	Negative (0/0)	Negative (0/0)		Negative (4/1)	Negative
8	42	F	Low	QFT	Positive (0.49)[b]	Negative (0.00)	Negative (0.00)	Negative (0.00)	Negative (0.00)	Inconsistent
				T-SPOT	Negative (0/0)	Negative (1/0)	Negative (0/0)	Negative (0/0)	Negative (0/0)	Negative
12	34	F	Low	QFT	Negative (0.00)	Positive (2.58)	Negative (0.00)	Negative (0.01)	Negative (0.00)	Inconsistent
				T-SPOT	Negative (0/0)	Negative (0/0)	Negative (1/1)	Negative (0/0)	Negative (0/0)	Negative
13	48	F	Low	QFT	Negative (0.00)	Positive (1.09)	Negative (0.00)	Negative (0.00)		Inconsistent
				T-SPOT	Negative (1/1)	Negative (1/0)	Negative (1/0)	Negative (0/2)		Negative
14	39	F	Low	QFT	Negative (0.00)	Positive (0.82)		Negative (0.02)	Negative (0.00)	Inconsistent
				T-SPOT	Negative (0/3)	Negative (1/1)		Negative (2/2)	Negative (2/2)	Negative
17	48	F	Low	QFT	Positive (0.59)[b]	Negative (0.20)	Negative (0.11)	Negative (0.19)	Negative (0.31)	Inconsistent
				T-SPOT	Negative (0/3)	Negative (1/1)	Negative (0/3)	Negative (0/5)	Negative (0/3)	Negative
19	20	M	Low	QFT	Positive (0.77)	Negative (0.08)	Negative (0.00)	Negative (0.00)	Negative (0.00)	Inconsistent
				T-SPOT	Negative (0/0)	Negative (0/0)	Negative (0/0)	Negative (0/0)	Negative (0/0)	Negative
22	47	F	Medium	QFT	Negative (0.28)	Negative (0.00)	Negative (0.10)		Negative (0.09)	Negative
				T-SPOT	Positive (1/10)	Positive (0/8)	Negative (0/5)		Negative (0/2)	Inconsistent
24	49	F	Medium	QFT	Positive (0.37)[b]	Negative(0.20)	Negative (0.21)	Negative (0.26)		Inconsistent
				T-SPOT	Negative (0/2)	Negative (0/3)	Negative (5/0)	Negative (0/3)		Negative
25	41	M	Medium	QFT	Positive (0.60)	Positive (0.49)	Positive (1.63)	Positive (0.62)	Positive (0.64)	Positive
				T-SPOT	Positive (100/30)	Positive (100/100)	Negative (0/1)	Negative (3/0)	Negative (3/0)	Inconsistent
26	48	F	Low	QFT	Negative (0.26)	Positive (0.52)[b]	Negative (0.33)	Negative (0.14)		Inconsistent
				T-SPOT	Positive (0/8)	Positive (2/24)	Positive (0/6)	Positive (3/14)		Positive
27	44	F	Medium	QFT	Positive (1.04)	Positive (0.61)	Positive (1.66)	Positive (0.96)	Negative (0.24)[b]	Inconsistent
				T-SPOT	Positive (5/10)	Positive (12/15)	Positive (4/10)	Positive (14/17)	Positive (13/10)	Positive
32	55	F	Medium	QFT	Negative (0.19)	Negative (0.03)	Negative (0.12)	Negative (0.26)		Negative
				T-SPOT	Positive (8/4)[b]	Positive (6/0)	Negative (2/0)	Negative (3/1)		Inconsistent

[a] Conversions were defined as a baseline IFN-γ concentration of <0.35 IU/ml, followed by an increase across the threshold of ≥0.35 IU/ml. Reversions were defined as the opposite progression.

[b] IFN-γ responses within proposed borderline zones from 0.2 to 0.7 IU/ml and 4 to 8 SFCs which showed changes in results from negative to positive or vice versa according to the manufacturers' thresholds.

[c] M, male; F, female.

[d] Values in parentheses represent IFN-γ responses in IU/ml and SFCs (ESAT-6 panel/CFP-10 panel) for the QFT and the T-SPOT assay, respectively.

mixed-effect models were constructed. The baseline IFN-γ response ($P < 0.001$), a positive prior TST result ($P < 0.01$), and classification as medium TB screening risk ($P < 0.05$) remained the only variables which showed significant associations with the magnitude of IFN-γ responses for both IGRAs (if the aforementioned variables were excluded from analysis), even after adjustment for covariates and factors such as age, BCG vaccination, and foreign origin. However, clinically relevant variability, i.e., changes of dichotomous IGRA results, was less frequently observed in subjects with clearly negative or highly positive results (Table 2).

Concordance between the two commercial IGRAs and agreement with prior TST results. While the proportion of discordant IGRA results decreased, the proportion of concordantly negative IGRA results increased. Overall, the agreement between both IGRAs was substantial and improved over time (Table 3). The overall agreement of QFT and T-SPOT results with prior TST results was fair (kappa, 0.32; $P < 0.001$) and moderate (kappa, 0.47; $P < 0.001$), respectively. Notably, all five subjects that remained concordantly positive in both IGRAs during the complete study period had had prior positive TST results, 14 ± 6.4 years ago on average, as did five out of six persistently QFT-positive and all seven persistently T-SPOT-positive subjects.

Borderline zone analysis. The characteristics of subjects with inconsistent IGRA results during the study period are shown in Table 4. When applying the proposed borderline

zones of 0.2 to 0.7 IU/ml (QFT) and 4 to 8 SFCs/well (T-SPOT) to those IFN-γ responses that showed a change in results from negative to positive or vice versa, the QFT conversion rate decreased from 17.4% (4/23) to 13.0% (3/23) of subjects. Accordingly, the reversion rate dropped from 50.0% (6/12) to 16.7% (2/12) for the QFT and from 30% (3/10) to 20% (2/10) for the T-SPOT (Table 4).

Quality assessment. The assessment of test-retest reproducibility (and ELISA reader reliability for the QFT) proved that IGRA results were highly reliable and reproducible for both dichotomous and continuous measures (Table 5). There were no changes from negative to positive or vice versa. However, the assessment of T-SPOT test-retest reliability was limited to those assays that showed borderline results. There were no indeterminate results.

DISCUSSION

The present study confirms that the within-subject variability of IFN-γ responses occurs with both IGRAs among HCWs in a setting of low TB incidence. Depending on the respective assay, variable rates of conversions and reversions may complicate the interpretation of repeated results in the short term. We found a significant decline in qualitative and quantitative IGRA results over time with both assays, which may be explained by regression toward the means. Accordingly, both

TABLE 5. Quality assessment of both commercial IGRAs

IGRA	Component of within-test variability	Description	No. of subjects (total no. of measures)	Intraclass correlation coefficient of continuous measures (95% CI)
QFT[a]	Test-retest reproducibility[b]	Duplicate 1 versus duplicate 2	35 (70)	0.996 (0.992–0.998)
	ELISA reader reliability	Repeated readings of optical density at 0, 5, 15, 30, and 60 min	35 (175)	0.998 (0.997–0.999)
T-SPOT	Test-retest reproducibility[b]	Duplicate 1 versus duplicate 2	8 (16)	0.993 (0.967–0.999)

[a] The correlation coefficient (r) of the standard curve, which was calculated from the mean absorbance values of the internal standards, was >0.999 for each QFT ELISA run.
[b] There was perfect agreement of all dichotomous measures (raw, 100%; kappa coefficient, 1.000).

effects need to be considered when interpreting repeat IGRA results in comparable populations and settings.

Data on the reproducibility and within-subject variability of IFN-γ responses from well-designed studies is rare. Notably, most of the available data relates to high-incidence countries (6, 32, 34), and so far only a few preliminary results on IGRA repeatability and reproducibility among HCWs in the United States have been published (2). In line with results from previous studies, we found changes of ±60 to 70% from the mean IFN-γ responses for both IGRAs to be responsible for 95% of the within-subject variability. A recent South African study tested 26 subjects with both IGRAs at weekly intervals during a 3-week period (a total of 88 results per IGRA) (32). Changes of ±80% for the QFT and ±3 SFCs for the T-SPOT assay were found to be within normal limits. The overall frequency of discordant results between both IGRAs of 9% (8/88 subjects) was comparable to our rate of 13%. Inconsistent results occurred more often with the T-SPOT (six episodes) than with the QFT (one episode) and were seen only in subjects from the medium and high TB risk group. In contrast, we found significantly more inconsistent QFT results (Table 2), with 7 of 13 subjects (54%) having inconsistent results and low TB risk (Table 4). Another study from South Africa evaluated day-to-day variability of QFT results in 27 HCWs during a 3-day period (6). Although the QFT proved to be highly reproducible if rigorously handled and no changes in dichotomous results occurred, considerable within-subject variability was found in the magnitude of IFN-γ responses. Veerapathran and colleagues serially tested 14 Indian HCWs with the QFT assay on four occasions during 12 days (total $n = 56$) (34). This study reported a high test-retest and within-subject reproducibility of dichotomous results but only moderate reproducibility with continuous measures. However, inconsistent results were noted in only two subjects, both with IFN-γ responses around the cutoff. Interestingly, this study observed a nonsignificant decline in mean IFN-γ responses of 30% between successive visits. All three studies mentioned above have in common that they were conducted in high-incidence countries, had high overall rates of IGRA positivity of 40 to 57%, had large individual maximum ranges of absolute QFT IFN-γ responses of 8.41 to 11.11 IU/ml, and included only a few subjects with results close to the cutoff.

Our study has several strengths and limitations. Although the number of participating subjects is small, it analyzed more than 150 serial measurements for each IGRA in a head-to-head comparison. Our study population is homogenous, and TB exposure during the study period is unlikely. We performed additional quality assessment and showed that both IGRAs were highly reproducible and reliable. Different potential sources of technical and biological variability, which may contribute to the total variability of IFN-γ responses in clinical practice (6), were systematically eliminated. However, it was not the objective of our study to assess different potential sources of variability, and we were unable to completely avoid some sources of variability, e.g., intraoperator or interplate variability. Moreover, no conclusions about the safety of the borderline zone analysis can be made, as we did not include a longitudinal follow-up beyond the 4-week study period.

As highlighted in a recent systematic review, the reversion of IGRA serial testing results is common among HCWs and more frequently observed than conversion (37). Our data suggest that this observation is explained by regression toward the means, a ubiquitous and inevitable, though often missed, statistical phenomenon that occurs whenever repeated measurements on the same subject are observed with random error, i.e., nonsystematic variation around a true mean (1). Its recognition is critical if subjects are categorized into groups according to baseline measurements and if treatment effects or disease incidence rates are estimated based on repeated surrogate measurements. As the effect of regression toward the means increases with measurement variability, it can be reduced by increasing the number of independent baseline measurements, which gives better estimates of the means and within-subject variability. However, this may not be feasible in clinical practice but can be approached by repeating IGRA results within certain borderline zones. In our study, the use of a borderline zone predominantly reduced the frequency of QFT reversions. The extent of within-subject variability that we determined approximates the proposed borderline zones. Three recent serial testing studies among German and Portuguese HCWs demonstrated that the implementation of a QFT borderline zone of 0.2 to 0.7 IU/ml may safely reduce the rate of inconsistent results (26, 29, 31). For the T-SPOT, evidence for the safety of a specific borderline zone is sparse. While the U.S. Food and Drug Administration recommends a borderline zone of 5 to 7 SFCs, van Zyl-Smit et al. suggested a borderline zone of 4 to 8 SFCs (32). However, the follow-up in this study was only 6 months.

Our findings have important implications for the use of IGRAs in the serial testing of HCWs in countries with low TB incidence. First, the interpretation of serial testing results requires knowledge of quantitative results and clinical data (TB risk, prior IGRA, and/or TST results). Second, the reporting of

IGRA characteristics and quality assessment should become a required standard to appraise the robustness of reported results. Third, the use of borderline zones appears to be reasonable and feasible for clinical practice. However, the appropriateness and safety of definitions of conversions, reversions, and borderline zones need to be derived from large prospective cohort studies in different settings. Fourth, with respect to individual TB risk factors, subjects with borderline positive results should be retested before being offered preventive chemotherapy.

However, it is important to note that our findings are applicable only to comparable settings. Considering the controlled study setting and the fact that the variability of individual measurements may be greater than the variability of means, the total variability of serial IGRA results observed in clinical practice may even exceed our estimates.

In conclusion, we observed substantial within-subject variability and regression toward the means of IFN-γ responses with both IGRAs during a 4-week period, which should be considered when interpreting serial testing results from HCWs. Our data suggest that the use of the T-SPOT assay and of a borderline zone for the interpretation of the QFT result in fewer inconsistent results in the serial testing of HCWs. Taking into account the lack of data on the prognosis of IGRA results in periodical occupational screening in settings with a low TB incidence rate, HCWs with borderline positive results should be retested rather than offered preventive chemotherapy.

REFERENCES

1. **Barnett, A. G., J. C. van der Pols, and A. J. Dobson.** 2005. Regression to the mean: what it is and how to deal with it. Int. J. Epidemiol. **34:**215–220.
2. **Belknap, R., et al.** 2009. Diagnosis of latent tuberculosis infection in U.S. health care workers: reproducibility, repeatability and 6 month follow-up with interferon-gamma release assays (IGRAs). Am. J. Respir. Crit. Care Med. **179:**A4101.
3. **Cellestis.** 2007. QuantiFERON®-TB Gold In-Tube package insert. Cellestis, Melbourne, Australia. http://www.cellestis.com/IRM/Company/ShowPage .aspx?CPID=1255.
4. **Centers for Disease Control and Prevention.** 2005. Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care settings, 2005. MMWR Recomm. Rep. **54**(RR-17):1–141.
5. **Chee, C. B., et al.** 2009. Use of a T cell interferon-gamma release assay to evaluate tuberculosis risk in newly qualified physicians in Singapore health-care institutions. Infect. Control Hosp. Epidemiol. **30:**870–875.
6. **Detjen, A. K., et al.** 2009. Short-term reproducibility of a commercial interferon gamma release assay. Clin. Vaccine Immunol. **16:**1170–1175.
7. **Diel, R., et al.** 2007. Recommendations for background studies in tuberculosis. Pneumologie **61:**440–455.
8. **Diel, R., et al.** 2011. Interferon-gamma release assays for the diagnosis of latent Mycobacterium tuberculosis infection: a systematic review and meta-analysis. Eur. Respir. J. **37:**88–99.
9. **Diel, R., R. Loddenkemper, and A. Nienhaus.** 2010. Evidence-based comparison of commercial interferon-gamma release assays for detecting active TB: a metaanalysis. Chest **137:**952–968.
10. **Diggle, P. J.** 1988. An approach to the analysis of repeated measurements. Biometrics **44:**959–971.
11. **European Centre for Disease Prevention and Control.** 2011. Use of interferon-gamma release assays in support of TB diagnosis. European Centre for Disease Prevention and Control, Stockholm, Sweden.
12. **Farhat, M., C. Greenaway, M. Pai, and D. Menzies.** 2006. False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria? Int. J. Tuberc. Lung Dis. **10:**1192–1204.
13. **Lee, K., et al.** 2009. Annual incidence of latent tuberculosis infection among newly employed nurses at a tertiary care university hospital. Infect. Control Hosp. Epidemiol. **30:**1218–1222.
14. **Mack, U., et al.** 2009. LTBI: latent tuberculosis infection or lasting immune responses to M. tuberculosis? A TBNET consensus statement. Eur. Respir. J. **33:**956–973.
15. **Mazurek, G. H., et al.** 2005. Guidelines for using the QuantiFERON-TB Gold test for detecting Mycobacterium tuberculosis infection, United States. MMWR Recomm. Rep. **54:**49–55.
16. **Mazurek, M., et al.** 2010. Updated guidelines for using interferon gamma release assays to detect Mycobacterium tuberculosis infection-United States, 2010. MMWR Recomm. Rep. **59:**1–25.
17. **Menzies, D.** 1999. Interpretation of repeated tuberculin tests. Boosting, conversion, and reversion. Am. J. Respir. Crit. Care Med. **159:**15–21.
18. **Menzies, D., R. Joshi, and M. Pai.** 2007. Risk of tuberculosis infection and disease associated with work in health care settings. Int. J. Tuberc. Lung Dis. **11:**593–605.
19. **Oxford Immunotec.** 2009. T-SPOT.*TB* package insert (PI-TB8-IVD-UK-V5). Oxford Immunotec, Abingdon, United Kingdom. http://www .oxfordimmunotec.com/8-UK.
20. **Pai, M., et al.** 2006. Serial testing of health care workers for tuberculosis using interferon-gamma assay. Am. J. Respir. Crit. Care Med. **174:**349–355.
21. **Pai, M., et al.** 2009. T-cell assay conversions and reversions among household contacts of tuberculosis patients in rural India. Int. J. Tuberc. Lung Dis. **13:**84–92.
22. **Pai, M., and R. O'Brien.** 2007. Serial testing for tuberculosis: can we make sense of T cell assay conversions and reversions? PLoS Med. **4:**e208.
23. **Pai, M., A. Zwerling, and D. Menzies.** 2008. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann. Intern. Med. **149:**177–184.
24. **Perry, S., et al.** 2008. Reproducibility of QuantiFERON-TB gold in-tube assay. Clin. Vaccine Immunol. **15:**425–432.
25. **Pollock, N. R., et al.** 2008. 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. **29:**878–886.
26. **Ringshausen, F. C., et al.** 2010. Predictors of persistently positive Mycobacterium-tuberculosis-specific interferon-gamma responses in the serial testing of health care workers. BMC Infect. Dis. **10:**220.
27. **Ringshausen, F. C., et al.** 2009. In-hospital contact investigation among health care workers after exposure to smear-negative tuberculosis. J. Occup. Med. Toxicol. **4:**11.
28. **Sachs, L.** 2003. Angewandte statistik: anwendung statistischer methoden, 11th ed. Springer, Berlin, Germany.
29. **Schablon, A., et al.** 2010. Serial testing with an interferon-gamma release assay in German healthcare workers. GMS Krankenhhyg. Interdiszip. **5:**Doc05.
30. **Seidler, A., A. Nienhaus, and R. Diel.** 2005. Review of epidemiological studies on the occupational risk of tuberculosis in low-incidence areas. Respiration **72:**431–446.
31. **Torres Costa, J., R. Silva, R. Sa, M. J. Cardoso, and A. Nienhaus.** 2011. Serial testing with the interferon-gamma release assay in Portuguese health-care workers. Int. Arch. Occup. Environ. Health **84:**461–469.
32. **van Zyl-Smit, R. N., et al.** 2009. Within-subject variability and boosting of T-cell interferon-gamma responses after tuberculin skin testing. Am. J. Respir. Crit. Care Med. **180:**49–58.
33. **van Zyl-Smit, R. N., A. Zwerling, K. Dheda, and M. Pai.** 2009. Within-subject variability of interferon-g assay results for tuberculosis and boosting effect of tuberculin skin testing: a systematic review. PLoS One **4:**e8517.
34. **Veerapathran, A., et al.** 2008. T-cell assays for tuberculosis infection: deriving cut-offs for conversions using reproducibility data. PLoS One **3:**e1850.
35. **World Health Organization.** 2009. WHO policy on TB infection control in health-care facilities, congregate settings and households. WHO, Geneva, Switzerland.
36. **Yoshiyama, T., N. Harada, K. Higuchi, Y. Nakajima, and H. Ogata.** 2009. Estimation of incidence of tuberculosis infection in health-care workers using repeated interferon-gamma assays. Epidemiol. Infect. **137:**1691–1698.
37. **Zwerling, A., S. van den Hof, J. Scholten, F. Cobelens, D. Menzies, and M. Pai.** 12 January 2011, posting date. Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. Thorax doi:10.1136/thx.2010.143180.