

Is Serological Testing a Reliable Tool in Laboratory Diagnosis of Syphilis? Meta-Analysis of Eight External Quality Control Surveys Performed by the German Infection Serology Proficiency Testing Program

Iris Müller, Volker Brade, Hans-Jochen Hagedorn, Erich Straube, Christoph Schörner, Matthias Frosch, Harald Hlobil, Gerold Stanek, and Klaus-Peter Hunfeld*

Central Laboratory of the Bacteriologic Infection Serology Study Group of Germany (BISSGG), Institute of Medical Microbiology, University Hospital of Frankfurt, Paul-Ehrlich-Str. 40, D-60596 Frankfurt/Main, Germany

Received 2 July 2005/Returned for modification 24 August 2005/Accepted 27 January 2006

The accuracy of diagnostic tests is critical for successful control of epidemic outbreaks of syphilis. The reliability of syphilis serology in the nonspecialist laboratory has always been questioned, but actual data dealing with this issue are sparse. Here, the results of eight proficiency testing sentinel surveys for diagnostic laboratories in Germany between 2000 and 2003 were analyzed. Screening tests such as *Treponema pallidum* hemagglutination assay (mean accuracy, 91.4% [qualitative], 75.4% [quantitative]), *Treponema pallidum* particle agglutination assay (mean accuracy, 98.1% [qualitative], 82.9% [quantitative]), and enzyme-linked immunosorbent assays (ELISAs) (mean qualitative accuracy, 95%) were more reliable than Venereal Disease Research Laboratory (VDRL) testing (mean accuracy, 89.6% [qualitative], 71.1% [quantitative]), the fluorescent treponemal antibody absorption test (FTA-ABS) (mean accuracy, 88% [qualitative], 65.8% [quantitative]), and immunoblot assays (mean qualitative accuracy, 87.3%). Clearly, immunoglobulin M (IgM) tests were more difficult to manage than IgG tests. False-negative results for samples that have been unambiguously determined to be IgM and anti-lipoid antibody positive accounted for 4.7% of results in the IgM ELISA, 6.9% in the VDRL test, 18.5% in the IgM FTA-ABS, and 23.0% in the IgM immunoblot assay. For negative samples, the mean percentage of false-positive results was 4.1% in the VDRL test, 5.4% in the IgM ELISA, 0.7% in the IgM FTA-ABS, and 1.4% in the IgM immunoblot assay. On average, 18.3% of participants misclassified samples from patients with active syphilis as past infection without indicating the need for further treatment. Moreover, 10.2% of laboratories wrongly reported serological evidence for active infection in samples from patients with past syphilis or in sera from seronegative blood donors. Consequently, the continuous participation of laboratories in proficiency testing and further standardization of tests is strongly recommended to achieve better quality of syphilis serology.

Syphilis caused by the spirochete *Treponema pallidum* is a re-emerging disease that is sexually transmitted and can progress in stages. In the United States, the rate of syphilis increased 9.1% from 2.2 cases per 100,000 population in 2001 to 2.4 cases per 100,000 population in 2002 (5). In Germany, the number of newly reported cases of syphilis increased dramatically, >100%, since 2001 and reached 4.1/100,000 people in 2004 (19). There is rising evidence that the resurgence of syphilis in Germany is partly due to an ongoing epidemic in men with male sexual partners in Hamburg, Berlin, Frankfurt, and Cologne (19). Evidence also exists for an increase of new heterosexual cases of syphilis owing to the commercial sex trade in those parts of Germany that border Eastern Europe (18). As a result, Germany has the highest incidence of syphilis among the western European countries, and the Robert Koch Institute urges a rapid expansion of surveillance and serological screening at epidemic foci, such as larger cities, and in the main core groups of the epidemic (commercial sex workers and male sexual partners) to rapidly identify potential transmitters (19). New molecular

tests for syphilis are unlikely to replace serology in the short term because they are fairly expensive and require sophisticated equipment (14). Antibody detection by nontreponemal tests (anti-lipoid antibody detection) and treponemal tests (anti-*T. pallidum* antibody detection) is still regarded as the mainstay for diagnosing syphilis and for monitoring the success of subsequent antibiotic treatment (2, 4, 9, 14). The accuracy of diagnostic tests is critical for successful control measures of epidemic syphilis outbreaks, including case finding, prompt therapy of infected individuals, and mandatory testing of potential transmitters (2, 4, 9, 14). Thus, promotion and quality control of diagnostic procedures is a relevant public health issue, but peer reviewed publications on that topic are sparse (11, 15, 20, 21, 22). Here, for the first time, the impact of test quality on the laboratory diagnosis of syphilis in Germany is investigated by use of a meta-analysis of external quality control program data obtained between 2000 and 2003 by the The Bacteriologic Infection Serology Study Group of Germany (BISSGG) (12, 13).

MATERIALS AND METHODS

Organization and structure of the German syphilis proficiency testing program. From March 2000 to September 2003, eight syphilis serology proficiency testing surveys (Table 1) were conducted in Germany by the central reference

* Corresponding author. Mailing address: Institute of Medical Microbiology, University Hospital of Frankfurt, Paul-Ehrlich Str. 40, D-60596 Frankfurt/Main, Germany. Phone: 49 69 6301 6441. Fax: 49 69 6301 5767. E-mail: K.Hunfeld@em.uni-frankfurt.de.

TABLE 1. Number of German and foreign participants in the syphilis proficiency testing program surveys conducted between 2000 and 2003

Mo/yr	No. of participating laboratories		
	German	Foreign	Total
3/2000	398	20	418
11/2000	327	18	345
3/2001	395	25	420
9/2001	398	23	421
3/2002	395	28	423
9/2002	350	26	376
3/2003	392	27	419
9/2003	348	26	374

laboratory for bacteriological serodiagnoses at the Institute of Medical Microbiology, University Hospital of Frankfurt/Main, in cooperation with the Institute of Standardization in the Medical Laboratory e.V. (INSTAND e.V.), Düsseldorf, Germany, and with the six reference laboratories of the BISSGG. The organization and structure of the German proficiency testing program for bacteriologic infection serology is summarized elsewhere in more detail (12, 13).

Sera used throughout the German syphilis proficiency testing program, 2000 to 2003. Sixteen serum samples were obtained from voluntary donors after obtaining written informed consent. All subjects were clinically evaluated by experienced physicians. Nine serum samples contained specific antibodies against *T. pallidum*, as determined by various commercial test systems. All antibody-positive donors could recall a known history of a current or past symptomatic syphilis infection, which also had been documented in the medical records of these patients by the treating physicians. Seven samples tested negative for specific antibodies against *T. pallidum* and were used as negative controls. A current or very recent syphilis infection was excluded in these donors by careful physical examination, evaluation of patients' medical histories, and review of the medical records provided by the referring physicians. Table 2 provides a detailed description of the clinical data available for all 16 samples.

Preparation and shipment of serum samples. Samples were prepared as published recently (13) and then stored at -20°C until use. Subsequently, the samples were thawed, and 500- μl aliquots without preservatives were dispensed in 0.5-ml polypropylene tubes (Sarstaedt, Germany). Prior to shipment, samples were checked for microbiological sterility and tested for possible reactivity against hepatitis B and C antigens as well as for human immunodeficiency virus types 1 and 2. Prepared samples were then distributed into eight shipments (March 2000, November 2000, March 2001, September 2001, March 2002, September 2002, March 2003, and September 2003). In each survey, two selected samples were sent to the participants without providing any additional clinical

information. Samples were shipped in polypropylene boxes and delivered by mail service for receipt within 2 days.

Assessment of correct test results by reference laboratories. Assessment of reference test results for each trial was performed according to the provisional guidelines for the performance of proficiency testing surveys in infection serology as proposed to the German general council of physicians (12). Each time, qualitative and quantitative reference test results were determined for each pair of serum samples during the proficiency testing survey by three to six different local specialized laboratories or university laboratories (13) with extensive expertise in the field of serodiagnostic testing for syphilis. Each reference laboratory examined the test samples using commercially available test kits from different vendors. Qualitative test results were graded positive, borderline, or negative according to the model of test results of the reference laboratories. The reference test results for quantitative tests were determined for each test by calculating the median from the results obtained for each method by the reference laboratories. For immunoblot testing, only qualitative test results obtained in accordance with the instructions of the manufacturers of the test kits used by the reference laboratories were reported to define reference results for each sample. By means of the preceding measures, all samples were unambiguously characterized with regard to qualitative test results and the amount of titers of specific immunoglobulin M (IgM) and IgG antibodies against *T. pallidum*. The characteristics of the serum samples applied in the German syphilis proficiency testing program as determined by the six reference laboratories are shown in Table 2.

Study conditions and evaluation of results. To date, participation in proficiency testing programs is not mandatory in any German legal institution. All laboratories were required to register at INSTAND prior to their participation. No pretest criteria were established to exclude any laboratories from the survey. All participants were instructed to treat samples as routine samples and to perform their established serological test methods on the distributed samples blind to additional clinical information to guarantee maximum objectivity. Qualitative and quantitative results had to be reported together with the methods used, the lot number, test manufacturer, and the laboratory machinery utilized. Moreover, the laboratories reported interpretative statements as to whether the test constellation suggested a possible syphilis infection and whether an active or latent infection was suspected. Reports were made in standardized form on defined evaluation sheets by use of a predefined code to permit statistical analysis after the surveys. Only one test result per test method (Venereal Disease Research Laboratory [VDRL] test, *T. pallidum* particle agglutination assay [TPPA], etc.) was reported to INSTAND by each participant. Participants were requested to return their reports to INSTAND for further computer-assisted evaluation of results within 10 days after receipt of samples (13). Qualitative results from participants were accepted as being accurate if their reported test results were congruent with the model as determined by the reference laboratories (see above). Because the quantitative enzyme-linked immunosorbent assay (ELISA) results reported were so heterogeneous, owing to the different quantification methods of the test manufacturers, these results were not included in

TABLE 2. German syphilis proficiency testing program: characteristics of selected serum samples as determined by the six reference laboratories^a

Sample	TPHA	TPPA	ELISA (polyvalent)	VDRL	CFT (cardiolipin)	ELISA (IgG)	ELISA (IgM)	Immunoblot IgG	Immunoblot IgM	FTA-ABS IgM	Clinical information (time of sampling after therapy)
21/2000	P (5,120)	P (10,240)	P	P (16)	P (40)	P	P	P	P	P (80)	Syphilis stage II (3 wk)
22/2000	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor
41/2000	P (2,560)	P (5,120)	P	P (8)	P (40)	P	P	P	P	P (80)	Syphilis stage II (4 mo)
42/2000	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor
21/2001	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor
22/2001	P (2,560)	P (5,120)	P	P (32)	P (80)	P	B/P	P	P	P (160)	Syphilis stage II (8 mo)
41/2001	P (2,560)	P (5,120)	P	B/N (≤ 1)	B/N (≤ 5)	P	N	P	N/B	N (<5)	Syphilis stage II (4 yr)
42/2001	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor
21/2002	P (20,480)	P (40,960)	P	P (128)	P (320)	P	P	P	P	P (160)	Syphilis stage II (1 wk)
22/2002	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor
51/2002	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor
52/2002	P (1,280)	P (1,280)	P	B/N (≤ 1)	N (<5)	P	N	P	N	N (<5)	Syphilis stage I (5 yr)
21/2003	P (10,240)	P (40,960)	P	P (16)	P (40)	P	N/B	P	N/B	P (20)	Syphilis reinfection stage II (6 mo)
22/2003	P (160)	P (160)	P	N (<1)	N (<5)	P	N	P	N	N (<5)	Syphilis stage I (5 yr)
51/2003	P (2,560)	P (5,120)	P	B/P (4)	B/P (20)	P	B/P	P	B/P	P (80)	Syphilis stage I (6 mo)
52/2003	N (<80)	N (<80)	N	N (<1)	N (<5)	N	N	N	N	N (<5)	Healthy blood donor

^a Legend: P, positive; B, borderline; N, negative. Median titers determined by the reference laboratories are given in parentheses.

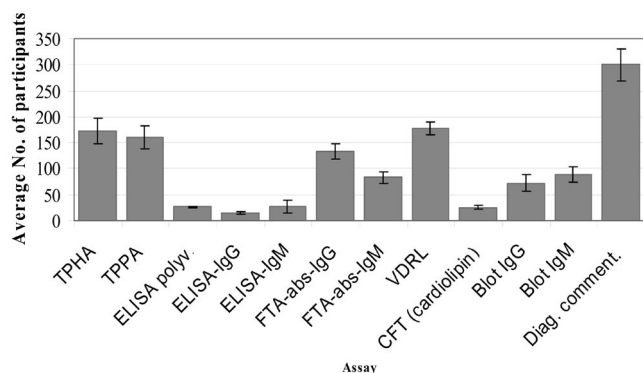


FIG. 1. Number of diagnostic comments and relative frequencies of use of the test methods reported by participants (mean, 400) during the surveys, 2000 to 2003. Blot, immunoblot; polyv., polyvalent; Diag., diagnostic. Bar markers indicate intervals of ± 1 standard deviation around the mean.

the evaluation listed below. Quantitative results of classical titer tests were accepted as being accurate provided results from participants were reported within a range of $\pm 2 \log_2$ unit dilutions around the median of the test results obtained by the reference laboratories. A qualifying certificate was forwarded to successfully participating laboratories for each parameter under the condition that their microbiological commentary and qualitative and quantitative test results for both samples determined with established assay systems met the above-listed criteria (12, 13).

RESULTS

Participating laboratories. From March 2000 to September 2003, between 345 and 423 (mean, 400) microbiological laboratories, including hospital laboratories, independent laboratories, physicians' office laboratories, and manufacturers of commercially available diagnostic syphilis assays, took part in each of the eight syphilis serology proficiency testing surveys (Table 1). On each occasion, between 28 and 18 laboratories from 10 European countries (Austria, Belgium, Czech Republic, Finland, Great Britain, Italy, Lithuania, Lichtenstein, Slovakia, and Switzerland) participated as well.

Application of assay systems. Figure 1 provides an overview of the relative frequencies of use of the various test systems by the participants during the surveys. Classical treponemal tests, such as the *Treponema pallidum* hemagglutination assay (TPHA) and the TPPA, were used more frequently than the more recently introduced diagnostic approaches like class-specific or polyvalent ELISAs and whole-cell or recombinant immunoblots (Fig. 1). As expected, most laboratories relied on stepwise diagnostic protocols, applying a sensitive polyvalent screening test (TPHA, 48%; TPPA, 45%; ELISA, 7%) followed by confirmation of positive results with fluorescent treponemal antibody absorption test (FTA-ABS test; 57%) or immunoblotting (43%). Confirmed cases were subjected to the VDRL test (87%) or cardiolipin complement fixation test (CFT; 13%) to determine the potential activity of the disease, followed by IgM class-specific assays like the FTA-ABS IgM test (42%), IgM immunoblot assay (45%), or IgM ELISA (13%) to test for the presence of specific anti-*T. pallidum* IgM antibodies as an additional marker of active or recent syphilis infection. This diagnostic approach complies with the recommendations of most European scientific expert opinions and

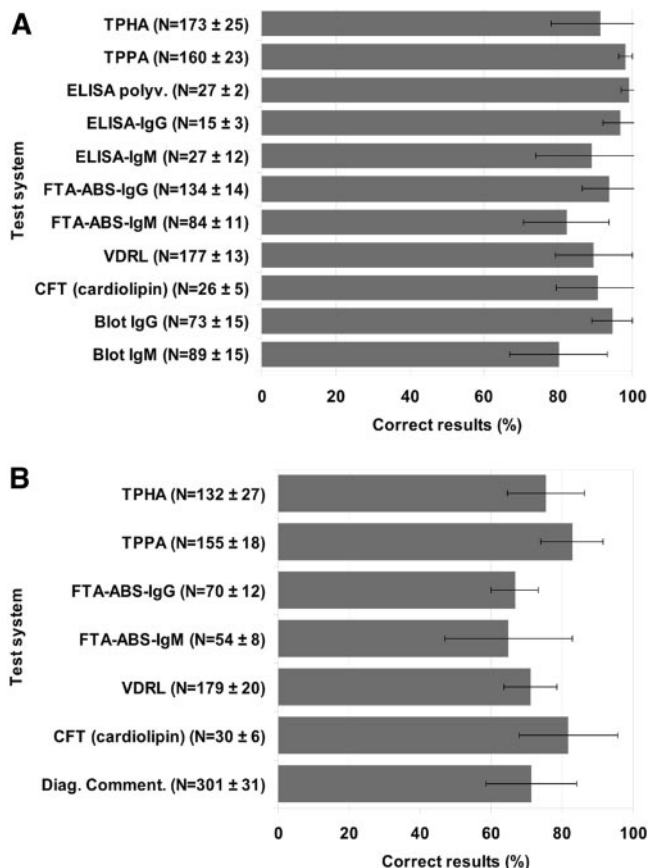


FIG. 2. (A) Average percentage of correct qualitative test results for the given diagnostic methods used throughout the eight proficiency testing trials. Bar markers indicate an interval of ± 1 standard deviation of the mean. (B) Average percentage of correct diagnostic comments and correct quantitative test results for the given diagnostic methods used throughout the eight proficiency testing trials. Bar markers indicate an interval of ± 1 standard deviation of the mean. Blot, immunoblot; polyv., polyvalent.

with the guidelines of the German Society for Microbiology and Hygiene (6, 9).

General findings. Throughout our surveys, the mean accuracy of the reference laboratories was 95% (range, 88 to 100%) for qualitative test results, 90% (range, 82 to 96%) for quantitative test results, and 95% (range, 83 to 100%) for diagnostic comments. The mean percentage of participant laboratories that reported correct results by use of different assays on the 16 serum testing samples sent out in the eight surveys of the German syphilis proficiency testing program from 2000 to 2003 are summarized in Fig. 2. In general, qualitative results were more reliable (range of mean accuracy, 80 to 98%) than quantitative test results (range of mean accuracy, 65 to 83%). Obviously, the test results obtained with the various assays used by the participants were much less reproducible in samples with very low and very high antibody titers than in samples with intermediate amounts of specific antibodies (Tables 2 and 3). From the broad range of quantitative results reported for the same specimen during the individual surveys, it can also be concluded that, in the routine laboratory, the quantity of detected antibody measured in titers (Table 3; Fig. 3) or quanti-

TABLE 3. Analysis of median antibody titers calculated from the VDRL and FTA-ABS IgM test results of reference laboratories in comparison to the median titers calculated from the results of all participating laboratories

Assay	Date (mo/yr)	Reference laboratory result			Participant result						
		Sample no.	Median titer	Range	No. of results	Median titer	Range	Acceptable range	Correct (%)	Total (%)	
VDRL	3/2000	2000/21	16	8–32	197	16	1–128	4–64	98.0	63.6	
		2000/22	0	0–1	181	0	0–16	0–0.9	70.7		
	11/2000	2000/41	8	4–8	158	8	1–64	2–32	98.8	64.8	
		2000/42	0	0–1	143	0	0–2	0–0.9	72.1		
	3/2001	2001/21	0	0	173	0	0–4	0–0.9	79.2	70.8	
		2001/22	32	16–64	188	16	0–128	8–128	96.3		
	9/2001	2001/41	0	0–1	151	0	0–32	0–1	88.1	76.8	
		2001/42	0	0	148	0	0–4	0–0.9	83.1		
	3/2002	2002/21	128	64–512	207	128	0–1,024	32–512	94.2	73.9	
		2002/22	0	0	192	0	0–4	0–0.9	84.4		
	9/2002	2002/51	0	0	168	0	0–<2	0–0.9	92.2	78.6	
		2002/52	0	0–1	168	0	0–64	0–1	82.8		
	3/2003	2003/21	16	8–16	201	16	0–256	4–64	96.0	81.6	
		2003/22	0	0	199	0	0–16	0–0.9	85.9		
	9/2003	2003/51	4	0–4	161	2	0–32	1–16	76.4	59.0	
		2003/52	0	0	161	0	0–<2	0–0.9	82.6		
	FTA-ABS IgM	3/2000	2000/21	80	40–640	65	40	0–5,120	20–320	64.6	50.0
			2000/22	0	0	64	0	0–12	<5	78.1	
11/2000		2000/41	80	80–160	50	10	0–640	20–320	66.0	59.1	
		2000/42	0	0	50	0	0–10	<5	84.0		
3/2001		2001/21	0	0	62	0	0–16	<5	82.2	43.0	
		2001/22	160	40–160	61	160	0–1,280	40–640	52.4		
9/2001		2001/41	0	0	48	0	0–160	<5	85.4	85.4	
		2001/42	0	0	47	0	0–5	<5	93.6		
3/2002		2002/21	160	40–320	61	40	0–2,560	40–640	72.1	59.4	
		2002/22	0	0	60	0	0–12	<5	88.4		
9/2002		2002/51	0	0	47	0	<5	<5	100.0	100.0	
		2002/52	0	0	47	0	<5	<5	100.0		
3/2003		2003/21	20	0–80	55	80	0–320	5–80	56.4	52.7	
		2003/22	0	0	55	0	0–12	<5	96.4		
9/2003		2003/51	80	10–256	43	80	0–2,560	20–320	69.8	69.8	
		2003/52	0	0	43	0	<5	<5	100.0		

tative ELISA units (data not shown) can vary widely for the same sample.

Accuracy of screening test results. Screening tests such as TPHA (qualitative mean accuracy, 91.4%; range, 56.1 to 98.2%; quantitative mean accuracy, 75.4%; range, 55.5 to 95.5%), TPPA (qualitative mean accuracy, 98.1%; range, 93.8 to 100%; quantitative mean accuracy, 82.9%; range, 66.1 to 96%), and polyvalent ELISAs (qualitative mean accuracy, 99.1%; range, 93.2 to 100%) were much more reproducible and proved to be more sensitive and specific than FTA-ABS tests and class-specific ELISAs (Fig. 2a). Clearly, IgM ELISAs (qualitative mean accuracy, 89%; range, 51.6 to 100%) were more difficult to manage than IgG ELISAs (qualitative mean accuracy, 96.7%; range, 86.7 to 100%) and frequently proved less specific (Fig. 2). Although used by only a small number of participants (7%), polyvalent ELISAs turned out to be the most reliable and reproducible test system for the qualitative detection of specific anti-*T. pallidum* antibodies throughout our surveys (Fig. 2A).

Accuracy of anti-lipoid antibody tests and *T. pallidum*-specific IgM test results. The qualitative and quantitative test results obtained by anti-cardiolipin antibody tests and FTA-ABS IgM assays, which are often used to determine possible activity of the infection, demonstrated a very low degree of interassay standardization (Table 4; Fig. 3). The accuracy of

the cardiolipin CFT (qualitative mean accuracy, 90.7%; range, 70 to 100%; quantitative mean accuracy, 81.7%; range, 55.2 to 100%), however, was higher than that of the VDRL test (qualitative mean accuracy, 89.6%; range, 68 to 99%; quantitative mean accuracy, 71.1%; range, 59.0 to 81.6%). With regard to the detection of specific IgM antibodies, qualitative IgM ELISA results (qualitative mean accuracy, 89%; range 51.6 to 100%) were more accurate than FTA-ABS IgM test results (qualitative mean accuracy, 82.3%; range, 64 to 100%) (Fig. 2). Qualitative IgM immunoblot results (Fig. 2a) showed substantial variability throughout our surveys (qualitative mean accuracy, 80.1%; range, 57.8 to 98.9%). Although for the FTA-ABS IgM test (quantitative mean accuracy, 64.9%; range, 43 to 100%) and VDRL test (quantitative mean accuracy, 71.1%; range, 59 to 81.6%) the median titers of the participating laboratories mostly met the median titers calculated for the positive samples from the results of the reference laboratories, the ranges of titers reported by the participants showed high interlaboratory variability, probably owing to methodological difficulties in reading test results correctly and due to the known lack of standardization of the commercially manufactured assays used (Table 3; Fig. 3). If samples with borderline reactivity were excluded from the meta-analysis, for samples that had been unambiguously determined to be IgM and anti-lipoid antibody positive (Table 2), the percentage of false-

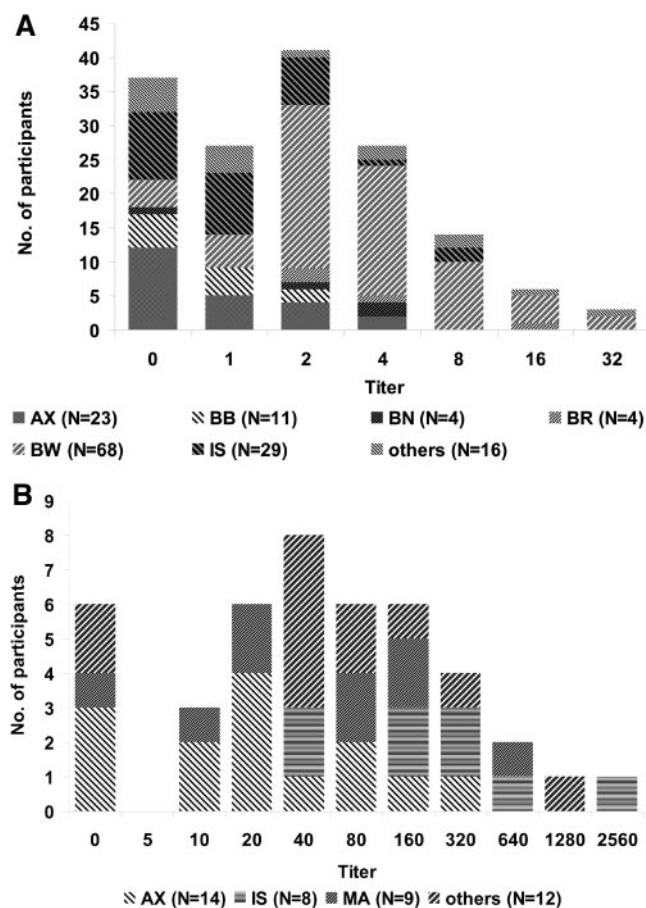


FIG. 3. Representative distribution of quantitative VDRL (A) and FTA-ABS IgM (B) assay titers, as reported by the participants of the proficiency testing trial held in September 2003. Distribution of titers for the positive sample 51/2003 (median VDRL test reference titer, 4; median FTA-ABS IgM test reference titer, 80) clearly demonstrates that test results are dependent on the manufacturer of the assay (for characterization of samples, see Table 3). Distribution of results as obtained by tests from different manufacturers is indicated by different gray scales. AX, bioMerieux; BB, Biokit; BN, Becton-Dickinson; BR, Biorad; BW, Dade Behring; IS, Innogenetics; MA, Mast.

negative results accounted for 6.9% of the VDRL test results, 4.7% of the IgM ELISA and CFT results, 18.5% of the IgM FTA-ABS test results, and 23% of the IgM immunoblot results from the participating laboratories. The mean percentage of false-positive results in clearly negative samples was 2% for the CFT, 4.1% for the VDRL test, 5.4% for the IgM ELISA, 0.7% for the IgM FTA-ABS test, and 1.4% for IgM immunoblotting throughout our studies. Clearly, the number of both false-negative and false-positive test results for anti-lipoid antibody tests and *T. pallidum*-specific IgM tests as encountered in our surveys are correlated with the diagnostic method, the quality of the test kits (Table 4), and the amount of specific antibodies present in different sera (Table 3).

Accuracy of reported diagnostic comments. Although most laboratories adhere to the current guidelines of stepwise serologic testing for syphilis in Germany (9), qualitative and quantitative changes in serologic test results may be misleading and can emerge simply by using different assay systems in different

laboratories (Fig. 3). In addition to these inconsistencies, on average, only 71% of the participants reported correct interpretative statements of test results throughout our surveys. In fact, on average, 18.3% of participants in their diagnostic comments misclassified samples from patients with clinically and serologically defined active syphilis (Table 2) as a past infection without recommending further treatment. Moreover, 10.2% of laboratories incorrectly reported serological evidence for active infection in samples from patients with past syphilis or in sera from seronegative blood donors. This means that, despite the application of a variety of test combinations on the same sample by most of the participating laboratories, a lack of expertise existed regarding whether or not the test constellation suggested possible syphilis and whether an active or past infection was suspected from the results of treponemal and nontreponemal assays.

DISCUSSION

In the scientific literature, the ranges of stage-dependent sensitivity and specificity of diagnostic assays for the serological detection of syphilis have been reported to be 70 to 100% and 97 to 99%, respectively (14). The quality of routine serological diagnosis of syphilis, however, has been questioned by several studies that found significant inter- and intralaboratory variability of test results (11, 12, 15, 20, 21, 22). In the United States, the Food and Drug Administration (FDA) and the Center for Devices and Radiological Health enforce a complex regulatory system for new in vitro diagnostics (8). For assays that represent a substantially new diagnostic approach, independent clinical testing is required in the process of so-called “premarket approval.” Simple test remakes, the so-called “me-too tests,” can be cleared by complying with the 510(k) regulations which substantially require the manufacturer to compare its product against an established device that has already

TABLE 4. German syphilis proficiency testing program 2000 to 2003: accuracy of test results for the most frequently used commercially manufactured VDRL and FTA-ABS IgM tests^a

Assay	Manufacturer	Qualitative testing		Quantitative testing	
		No. of participants	Correct results (%)	No. of participants	Correct results (%)
VDRL	AX	31 (4.3)	89.1 (15.0)	30 (5.2)	73.8 (15.3)
	BN	5 (1.0)	93.8 (16.5)	5 (0.5)	56.7 (24.0)
	IS	32 (6.7)	88.3 (15.9)	29 (7.0)	71.1 (9.9)
	BB	14 (2.7)	81.5 (14.9)	13 (2.5)	72.7 (13.6)
	BW	74 (7.7)	91.7 (6.9)	74 (9.6)	73.6 (5.5)
	ZZ	17 (4.2)	88.1 (13.7)	23 (6.9)	65.8 (17.8)
	Total	177 (13)	89.6 (10.4)	179 (20)	71.1 (7.5)
FTA-ABS IgM	AX	29 (3.9)	83.8 (14.2)	19 (3.5)	70.5 (13.7)
	IS	12 (3.8)	89.7 (9.2)	7 (2.5)	69.8 (18.3)
	BA	5 (1.4)	92.4 (12.6)	4 (1.6)	75.2 (20.7)
	MA	22 (3.3)	79.8 (12.4)	9 (2.1)	48.9 (28.7)
	ZZ	14 (3.8)	74.3 (21.1)	11 (2.6)	58.9 (27.7)
	Total	84 (11)	82.3 (11.5)	54 (8)	64.9 (18.0)

^a Results are means, with standard deviations indicated in parentheses. AX, bioMerieux; BA, BAG; BB, Biokit; BN, Becton-Dickinson; BW, DadeBehring; IS, Innogenetics; MA, Mast; ZZ, other.

been cleared by the FDA (8, 14). In Europe, in general, no independent clinical testing is necessary before placing in vitro diagnostic tests for syphilis on the market. This development resulted after the liberalization of the in vitro diagnostics (IVD) market in Europe, and since the institution of the new European IVD directive in 2000, the law no longer requires extensive, independent, and continuous standardized diagnostic as well as clinical evaluation of commercially available serological test kits for syphilis tests (1). Instead, the IVD directive only enforces quality standards for the production quality and safety of in vitro diagnostic tests in their intended use (1, 17). Consequently, inexpensive test remakes are promoted and increasingly pushed onto the market. Actually, in Germany alone, 42 different companies provide diagnostic tests for syphilis, and not surprisingly, the different methodological approaches of diagnostic tests in themselves may account, in part, for substantial differences with regard to the variable test quality, as noted in our sentinel surveys. In addition, the technically correct application of a test during diagnostic analysis and the individual operator's experience in the evaluation and assessment of test results (e.g., for FTA-ABS and VDRL tests) play a pertinent role in the quality of the findings and their comparability with results obtained by other laboratories (13, 22). Our investigations show that the VDRL test (qualitative mean accuracy, 89.6%; quantitative mean accuracy, 71.1%), the IgM FTA-ABS test (qualitative mean accuracy, 82.3%; quantitative mean accuracy, 64.9%), and IgM immunoblotting (qualitative mean accuracy, 80.1%), in part, perform less reliably than *T. pallidum*-specific screening tests in the routine diagnostic laboratory (Table 4; Fig. 2 and 3). Although our tests were somewhat different in methodology, our proficiency testing survey results do resemble the findings of several preceding international studies demonstrating considerable deficiencies in the quality of syphilis serology in the United States, the United Kingdom, and Taiwan (11, 15, 20, 21). Similarly, a look at the 2004 College of American Pathologists' proficiency testing reports G-A and G-B revealed that qualitative VDRL and RPR testing (range of VDRL test accuracy, 84.3 to 100%; range of RPR test accuracy, 64.2 to 100%) tended to be less reliable than TPPA testing (range of accuracy, 98.1 to 100%) (K. P. Hunfeld, personal communication). According to our study, obviously, the sensitivity and specificity of test results and the significance of the diagnostic findings depend primarily on the expertise of the individual laboratory and the test manufacturer (Tables 3 and 4; Fig. 3). Moreover, changes in test results may be misleading and can result simply from the use of different assay systems or from failure to test follow-up samples in parallel with previously obtained samples from the same patient. This is important because epidemiologists and physicians are known to correlate the disease activity and success of treatment with changes in laboratory tests. Clearly, the level of accuracy for syphilis serology in Germany is higher than that revealed in recent surveys on the quality of Lyme disease or *Chlamydia pneumoniae* serology (13, 16). However, mean accuracy levels below 95% for qualitative tests and below 90% for quantitative tests are unacceptable for diagnosing syphilis whether in screening pregnant woman, blood products, or potentially infected patients. In addition, successful surveillance and control of the current syphilis epidemic in Germany call for better test quality. Furthermore, the fiscal impact of a

flawed test on the health care system is probably largely underestimated. Assuming a prevalence in Germany of 4/100,000 people and ca. 5,000,000 syphilis tests/year, including blood bank testing, a difference of 11% in net sensitivity and of 0.2% in net specificity as calculated for two different IgG test combinations (TPHA screening tests: test 1 sensitivity, 88%; test 1 specificity, 94%; test 2 sensitivity, 95%; test 2 specificity, 99%; FTA-ABS confirmatory assays: test 1 sensitivity, 90%; test 1 specificity, 95%; test 2 sensitivity, 95%; test 2 specificity, 99.9%) would account for 14 false-negative cases, 14,950 false-positive cases, and a total of € 7,249,710 (~\$8,699,652) in excess costs (€ 29, or ~\$35, per test) due to FTA-ABS confirmatory testing ($n = 249,990$ tests). These medical and economic considerations clearly warrant intensified efforts to achieve better quality and standardization in the laboratory diagnosis of syphilis, in general, and in Germany in particular. Guidelines for acceptance and evaluation of new syphilis tests (2) as published by the CDC (Centers for Disease Control and Prevention), the regular participation of diagnostic laboratories in proficiency testing, and the establishment of medical advisory boards for the diagnosis of syphilis represent internationally proven interventions for achieving better test standardization and for regulating the quality of infection serology in general (12, 15, 20, 21, 22). Over the years, the success of such policy options is strongly supported by the results of quality assessment schemes in other countries, including the United Kingdom, Taiwan, and the United States, where the performance of laboratories could be improved, particularly when guidance was provided to poorly performing laboratories (11, 20, 21, 22). In addition, the use of standard preparations can increase accuracy levels $\geq 10\%$ (21, 22). Such interventions were successful in parameters like rheumatoid factor, parvovirus B19 serology, Lyme disease serology, and tick-borne encephalitis ELISA testing (3, 7, 10, 21, 22). To improve the quality of syphilis serology in Germany and possibly in Europe, a network of independent specialist laboratories should deal with the issues of test evaluation, quality promotion, and interassay standardization of commercially available test kits on a more regular basis.

ACKNOWLEDGMENTS

This study was funded by a grant provided by INSTAND e.V., Düsseldorf, Germany.

We thank Jeffrey N. Gibbs for discussing legal aspects of licensing procedures for serologic tests in the United States.

REFERENCES

1. Bundesministerium fuer Gesundheit (BMFG). 2000. Bekanntmachung (AKZ 117-456000-02/2) zur EG-Richtlinie über *in vitro* Diagnostika (98/79/EG). Bundesgesetzblatt 118:12077.
2. Center for Disease Control. 1977. Guidelines for evaluation and acceptance of new syphilis serology tests for routine use. Center for Disease Control, Atlanta, Ga.
3. Centers for Disease Control and Prevention and Association of State and Territorial Public Health Laboratory Directors (ASTPHLD). 1994. Proceedings of the 2nd National Conference on Serologic Diagnosis of Lyme Disease (Dearborn, MI). ASTPHLD, Washington, D.C.
4. Centers for Disease Control and Prevention. 2002. Sexually transmitted diseases treatment guidelines 2002. *Morb. Mortal. Wkly. Rep.* 51:1-80. [Online.] <http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5106a1.htm>. Accessed 1 December 2005.
5. Centers for Disease Control and Prevention. 2003. Primary and secondary syphilis—United States, 2002. *Morb. Mortal. Wkly. Rep.* 52:1117-1120.
6. Egglestone, S. I., and A. J. L. Turner. 2000. Serological diagnosis of syphilis. *Comm. Dis. Public Health* 3:158-162.

7. **Ferguson, M., D. Walker, and B. Cohen.** 1997. Report of a collaborative study to establish the international standard for parvovirus B19 serum IgG. *Biologicals* **25**:283–288.
8. **Gibbs, J. N.** 1998. Regulations and standards. ASRs: FDA issues final rule. *IVD Technol.* [Online.] <http://www.devicelink.com/ivdt/archive/98/01/009.html>. Accessed 12 December 2005.
9. **Hagedorn, H.-J.** 2000. MIQ. Qualitätsstandards in der mikrobiologisch-infektiologischen Diagnostik. Heft 16, Syphilis. Urban & Fischer, München, Germany.
10. **Hofmann, H., F. X. Heinz, and H. Dippe.** 1983. ELISA for IgM and IgG antibodies against tick-borne encephalitis virus: quantification and standardization of results. *Zentralbl. Bakteriol. Mikrobiol. Hyg. 1 Abt. Orig. A* **255**: 448–455.
11. **Hsu, W. S., J. T. Kao, and S. W. Ho.** 2000. Quality assurance in clinical laboratories in Taiwan. *J. Formos. Med. Assoc.* **99**:235–242.
12. **Hunfeld, K.-P., and V. Brade.** 2000. Proficiency testing in bacteriological infection serology—state of the art and results of proficiency testing trial X/99. *Mikrobiologie* **10**:135–144.
13. **Hunfeld, K. P., G. Stanek, E. Straube, H. J. Hagedorn, C. Schoerner, F. Muehlschlegel, and V. Brade.** 2002. Quality of Lyme disease. Lessons from the German Proficiency Testing Program 1999–2001. *Wien. Klin. Wochenschr.* **114/13**:591–600.
14. **Larsen, S. A., B. M. Steiner, and A. H. Rudolph.** 1995. Laboratory diagnosis and interpretation of tests for syphilis. *Clin. Microbiol. Rev.* **8**:1–21.
15. **Neimeister, R. P., R. Teschemacher, I. J. Yankevitch, and J. Cocklin.** 1975. Proficiency testing, trouble shooting and quality control for the RPR test. *Am. J. Med. Technol.* **41**:13–17.
16. **Peeling, R. W., S. P. Wang, J. T. Grayston, F. Blasi, J. Boman, A. Clad, H. Freidank, et al.** 2000. *Chlamydia pneumoniae* serology: interlaboratory variation in microimmunofluorescence assay results. *J. Infect. Dis.* **181**(Suppl. 3):S426–S429.
17. **Place, J. F.** 2004. The coming age of in vitro testing. *IVD Technol.* [Online.] <http://www.devicelink.com/ivdt/archive/00/09/002.html>. Accessed 1 November 2005.
18. **Resl, V., M. Kumpova, L. Cerna, M. Novak, and P. Parazdiora.** 2003. Prevalence of STDs among prostitutes in Czech border areas with Germany in 1997–2001 assessed in project “Jana.” *Sex. Transm. Infect.* **79**:e3. [Online.] <http://www.stijournal.com/cgi/content/full/79/6/e3>. Accessed 1 November 2005.
19. **Robert Koch Institut.** 2005. Syphilis, p. 155–160. *In* Infektionsepidemiologisches Jahrbuch für 2004. Mercedes-Druck, Berlin, Germany.
20. **Snell, J. J., J. V. de Mello, and P. S. Gardner.** 1982. The United Kingdom national microbiological quality assessment scheme. *J. Clin. Pathol.* **35**:82–93.
21. **Taylor, R. N., K. M. Fulford, V. A. Przybyszewsky, and V. Pope.** 1978. Centers for Disease Control diagnostic immunology proficiency testing program results for 1978. *J. Clin. Microbiol.* **8**:388–395.
22. **Taylor, R. N., and K. M. Fulford.** 1981. Assessment of laboratory improvement by the CDC diagnostic immunology proficiency testing program. *J. Clin. Microbiol.* **13**:356–368.