Review Article

wiener klinische wochenschrift
the middle european journal of medicine
Printed in Austria

Wien Klin Wochenschr (2002) 114/13–14: 591–600
© Springer-Verlag 2002

Quality of Lyme disease serology

Lessons from the German Proficiency Testing Program 1999–2001

A preliminary report*

Klaus-Peter Hunfeld[1], **Gerold Stanek**[2], **Eberhard Straube**[3], **Hans-Jochen Hagedorn**[3], **Christoph Schörner**[3], **Fritz Mühlschlegel**[3], and **Volker Brade**[1]

[1] Central Laboratory of the German Proficiency Testing Program for Bacteriological Infection Serology, Institute of Medical Microbiology, University Hospital of Frankfurt/Main, Germany
[2] Clinical Institute of Hygiene and Medical Microbiology, Vienna, Austria
[3] The Bacteriologic Infection Serology Study Group of Germany (BISSGG)

Summary. *Objective:* External quality control surveys are an important tool in regulating the quality of infection serology in general and of borreliosis serology in particular. We report on the results of a Lyme disease proficiency testing program which is regularly organised twice a year by our institutions in close cooperation with the Institute of Standardisation in the Medical Laboratory (INSTAND).

Methods: From 1999 to 2001, between 226 and 337 microbiological laboratories participated in each of the four surveys that have been held so far. In addition, between 23 and 30 laboratories from 13 other European countries also participated in each trial. In each survey two serum samples which had been unambiguously characterised by six reference laboratories to contain or not to contain antibodies against the Lyme disease spirochete were distributed in order to determine the accuracy of the diagnostic methods used in participating laboratories. The laboratories also reported interpretative statements of whether or not the test constellation suggested a possible borrelial infection and if an early or late phase of the specific antibody response was suspected.

Results: Test results were found to be in part highly variable and clearly correlated with the manufacturers and the applied test methodology. It was also clear that IgM tests were more difficult to handle than were IgG tests. ELISA-testing was more reproducible and proved to be more sensitive and specific than IFA and IHA testing. Quantification of test results and reporting of specific immunoblot bands also showed high variability. Moreover, for some assays a high number of false positive and false negative test results were reported by the participants.

Conclusion: In view of our results further standardisation of Lyme disease serology is not just desirable but is urgently needed. Moreover, stronger criteria for the validation of available test kits must be applied.

Key words: *Borrelia burgdorferi*, Lyme disease, serodiagnosis, serological proficiency testing.

Introduction

Symptoms of Lyme disease are known to range from mild, flu-like abnormalities to severe disabling conditions, making differential diagnosis difficult for clinicians and in turn making this human spirochetosis the "new great imitator" of medicine. Since the illness can resemble aseptic meningitis, rheumatoid arthritis, influenza, or other syndromes, clinicians tend to rely on serological testing as the primary indicator for the detection of possible infection with the Lyme disease spirochete [1].

As demonstrated in recent studies, however, some serological tests for Lyme borreliosis perform poorly in the routine laboratory and show considerable deficiencies in their sensitivities and specificities [1–6]. Moreover, significant inter- and intra-laboratory variability of test results exists, as shown by proficiency testing programs in areas where Lyme disease is endemic and in studies dealing with the quality of serodiagnostic testing for Lyme disease in the USA [1]. The law in most European countries does not at present require standardised diagnostic and clinical evaluation of commercially available serological test kits prior to registration. These problems, therefore, continue to contribute to the widely encountered overdiagnosis and misdiagnosis of Lyme borreliosis in the USA and Europe [7, 8]. In view of these studies, further standardisation of the test methods is not just desirable but is urgently needed. An important approach in regulating the quality of infection serology in general, and of Lyme borreliosis serology in particular, is regular

* This study is gratefully dedicated to the memory of Dr. med. Annemarie Groh, Jena, who substantially supported the German proficiency testing program for bacteriologic infection serology.

participation of test laboratories in external quality control measures [9]. It is therefore suggested that laboratories performing serological tests for the detection of anti-*B. burgdorferi* antibodies should be forced to participate in regularly held proficiency testing programs [6, 8]. Ours is the first report evaluating the correct performance of serodiagnostic testing for Lyme disease at the national level in Germany.

Materials and methods

Organisation and structure of the German Lyme Disease Proficiency Testing Program

From August 1999 to March 2001 four Lyme borreliosis serology proficiency testing surveys were held in Germany by the central reference laboratory for bacteriological serodiagnostics of the Institute of Medical Microbiology, University Hospital of Frankfurt/Main in close cooperation with the Institute of Standardisation in the Medical Laboratory (INSTAND) and with the reference laboratories of the Bacteriologic Infection Serology Study Group of Germany (BISSGG, see also Addendum). The organisation and structure of the German Lyme Disease Proficiency Testing Program is summarised in Fig. 1.

Selection of sera

Eight serum samples were obtained from voluntary donors. All subjects were clinically evaluated by experienced physicians. Two serum samples (serum scars) contained specific IgG- and/or IgM-antibodies against *Borrelia burgdorferi* as determined by various commercial test systems, but the donors did not remember a known history of Lyme borreliosis or tick bites. The third serum sample was provided by a patient who remembered a tick bite one year previously but thereafter did not develop clinical symptoms of Lyme disease. The fourth sample was obtained from a patient suffering from clinically diagnosed erythema migrans two weeks after antibiotic therapy was initiated. The remaining four samples were negative for specific antibodies against *Borrelia burgdorferi*. Serological descriptions of samples distributed to the participants of INSTAND Lyme borreliosis proficiency testing program are presented in Table 1.

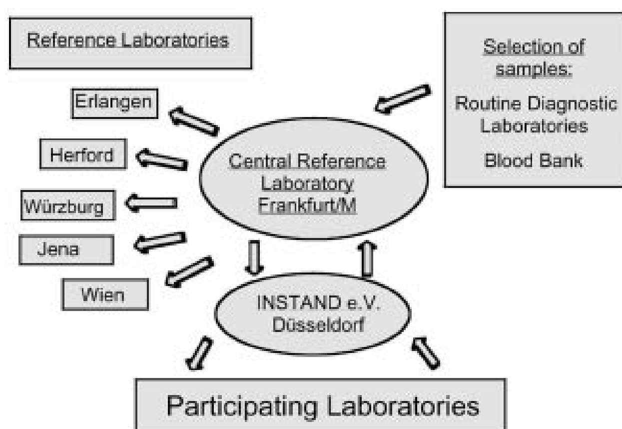

Fig. 1. Organigram of the German proficiency testing program for bacteriologic infection serology

Preparation and shipment of serum samples

Samples were prepared as published recently in more detail [1]. Briefly, approximately 500 ml of whole blood was obtained from each donor. No pool sera were used in the trials. The blood was allowed to clot, and the serum was collected by centrifugation at 500 x g and stored at –20 °C until it was used. Subsequently the samples were thawed, and 500 µl aliquots without preservatives were dispensed in 0.5 ml polypropylene vials (Sarstedt Inc., Germany). Prior to shipment, samples were checked for microbiological sterility and tested for possible reactivity against hepatitis B virus surface antigen and human immunodeficiency virus type 1. Prepared samples were distributed in four shipments (October 1999, March 2000, November 2000, March 2001). In each survey two selected samples were sent to the laboratories participating in the proficiency testing program without providing any clinical information. All samples were shipped in polypropylene boxes and delivered by the German postal service's mail service for receipt within two 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 recently published by the German general council of physicians [9]: 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 specialist laboratories or University laboratories (Fig. 1) with superior expertise on the field of Lyme disease serodiagnostic testing. Each reference laboratory examined the test samples by use of commercially available test kits in addition to well established in-house tests. Qualitative test results were graded as positive, borderline or negative according to the test results of the reference laboratories. The quantitative reference test results for qualitative tests (immunofluorescence assay [IFA], indirect hemagglutination assay [IHA], and enzyme linked immunosorbent assay [ELISA]) were determined for each test by calculating the median from the results obtained for each method by the reference laboratories. For immunoblot testing, the class and number of specific IgG and IgM bands (p100, p58, p43, p41, OspA, OspC, p18, and p17) obtained by use of whole cell and recombinant immunoblots were also reported in order to define reference results for each sample. Thus, all samples were unambiguously characterised with regard to qualitative test results and titres of specific IgM- and IgG-antibodies against *B. burgdorferi* (Table 1).

Study conditions and statistical analysis

Generally, participation in the proficiency testing program is voluntary and so far not enforced by any legal institution in Germany. All laboratories had to register at the Institute for Standardisation in the Medical Laboratory (INSTAND) prior to their participation. No pre-test criteria were established that excluded any laboratories from the survey. In order to guarantee maximum objectivity, all participants were asked to treat the distributed samples as routine and to use their established methods (ELISA, IHA, IFA, and immunoblot) without knowledge of any additional clinical information. Qualitative and quantitative results were reported, together with the methods used, the lot number and manufacturer of the test, and the laboratory machinery utilised. Qualitative immunoblot test re-

Table 1. German Lyme borreliosis proficiency testig program: Characteristics of selected serum samples as determined by the reference laboratories

Sample	IFA			ELISA			Immunoblot				Clinical Infor-mation
	Qual. result (IgM/IgG)	Quant. result (Median) (IgM/IgG)	Quant. result (Range) (IgM/IgG)	ELISA (Poly.) (Qual. result)	IgM (Qual. result)	IgG (Qual. result)	IgM (Qual. result)	IgM (Blot bands)	IgG (Qual. result)	IgG (Blot bands)	
51/1999	IgM: N IgG: N	IgM: 0 IgG: 0	IgM: <20 IgG: 0–20	N	N	N	N	None or p41 only	N	None or p41 only	Blood donor serum
52/1999	IgM: B/P IgG: P	IgM: 64 IgG: 1280	IgM: 40–128 IgG: 640–1280	P	B/P	P	B/P	p41, OspC	P	p100, p58, p43, p41, p39, p30, OspC, p17/p18	Tick bite one year ago
21/2000	IgM: N/B IgG: N	IgM: 0 IgG: 160	IgM: <20 IgG: 80–256	P	N/B	P	N/B	p41, OspA (weak)	P	p100, p58, p43, p41, p39, OspC, p17/p18	Serum scar
22/2000	IgM: N IgG: N	IgM: 0 IgG: 0	IgM: <20 IgG: <20	N	N	N	N	None or p41 only	N	None or p41 only	Blood donor serum
41/2000	IgM: B/P IgG: N	IgM: 20 IgG: 0	IgM: 0–20 IgG: <20	P	B/P	N	B/P	p41, OspC	N	p41, OspA (week)	Erythema migrans
42/2000	IgM: N IgG: N	IgM: 0 IgG: 0	IgM: <20 IgG: <20	N	N	N	N	None or p41 only	N	None or p41 only	Blood donor serum
21/2001	IgM: N IgG: N	IgM: 0 IgG: 0	IgM: <20 IgG: 0–39	N	N	N	N	None or p41 only	N	None or p41 only	Blood donor serum
22/2001	IgM: N IgG: P	IgM: 0 IgG: 40	IgM: <20 IgG: 20–128	P	N	P	N	None or p41 only	P	p100, p41, p39, OspC, OspA, p17/p18	Serum scar

P Positive; *B* borderline; *N* negative.

sults were reported in all surveys. In addition the participating laboratories of the trials in November 2000 and March 2001 reported the number and class of specific IgG and IgM bands (p100, p58, p43, p41, OspA, OspC, p18, and p17) observed in their immunoblots for each of the serum samples. The laboratories also reported interpretative statements of whether or not the test constellation suggested a possible borrelial infection and if an early or late phase of the specific antibody response was suspected. Reports were made in standardised form on defined evaluation sheets by use of a pre-defined code in order to permit statistical analysis after the surveys. Participants were asked to return their reports to INSTAND for further computer-assisted evaluation of results within 10 days of receipt of samples. An overview of the organisation and structure of the proficiency testing program is shown in Fig. 1.

Qualitative results of the participants were accepted if their reported results were congruent with those determined by the reference laboratories (see above). Because of the extreme heterogeneity of the reported quantitative ELISA results, due to the different quantification methods of the test manufacturers, these results were not included in the evaluation detailed below. Quantitative results of classical titre tests (IFA, IHA) were accepted if results of participants for the positive samples were within a range of ± 2 $\log_2$ unit dilutions around the median of the results obtained from the reference laboratories. A certificate of successful participation in the proficiency testing program was forwarded to the laboratories for each parameter, only if their microbiological comments and the qualitative and quantitative test results obtained for *both* samples with their established assay systems met the above criteria.

Results

Participating laboratories

From October 1999 to March 2001, between 229 and 337 microbiological laboratories including hospital labo-

ratories, independent laboratories, physicians office laboratories, and manufacturers of commercially available diagnostic Lyme disease assays participated in each of the four surveys that have been held. Moreover, each time between 23 and 30 laboratories from 13 different European countries (Austria, Bulgaria, Estonia, Finland, France, Great Britain, Hungary, Lichtenstein, The Netherlands, Slovakia, Slovenia, Switzerland, Czech Republic) participated as well (Table 2).

Applied test systems

Tests employed were those used routinely for the serodiagnosis of Lyme disease in the participating laboratories and Fig. 2 gives an overview of the relative frequency of use of the various test systems by the participants during the surveys. It was clear that the laboratories most frequently performed a two-tier protocol, starting with a sensitive ELISA screening and followed by immunoblot confirmation of the results, consistent with the recommendation of CDC and most European scientific expert opinions [11, 12]. For direct immunoglobulin class-specific analysis of samples, IgG and IgM ELISA was used by 60–63% and 75–78% of the participants respectively in the four surveys. An immunoblot confirmatory assay for IgG and IgM antibodies was used by 63–73% of the laboratories. Other serological test methods such as IHA, IFA, and polyvalent ELISA systems were less frequently used (IHA: 7–10%, IFA: 13–26%, polyvalent ELISA: 12–14%).

Accuracy of test results

The percentage of laboratories reporting correct results on the eight serum samples sent out for testing in the four surveys of the German Lyme disease proficiency testing program from 1999 to 2001 are summarised in Table 3. It was clear that IgM tests were more difficult to

Table 2. Number of German and foreign participants in the Lyme disease proficiency testing program surveys from 1999 to 2001

Month/year	No. of participating laboratories		
	German laboratories	Foreign laboratories	Total
10/1999	203	23	226
03/2000	307	27	334
11/2000	222	25	247
03/2001	307	30	337

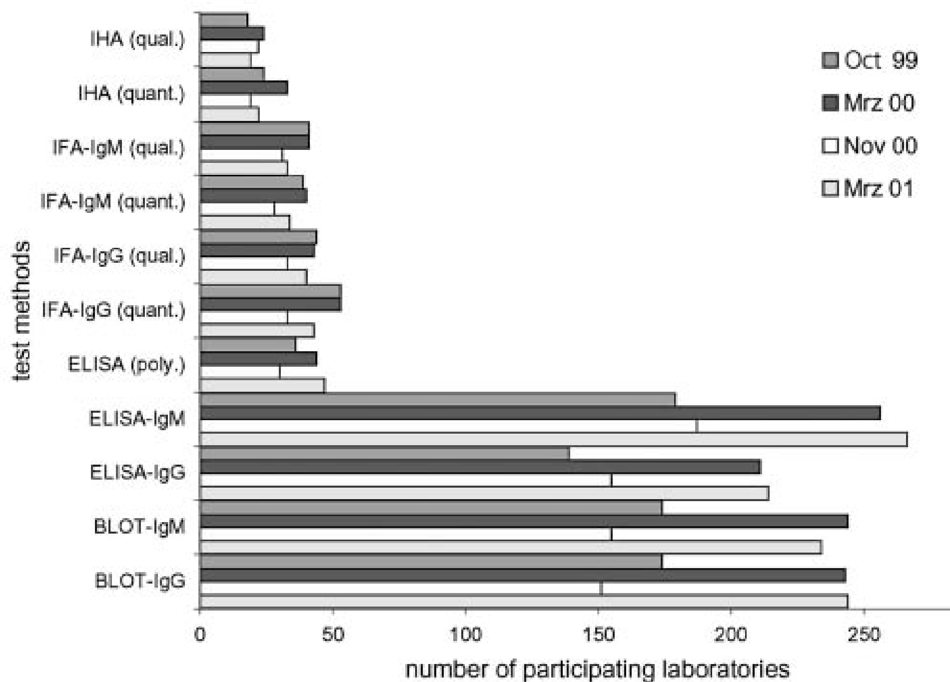

Fig. 2. Relative frequency of test-method. *IHA* Indirect hemaglutination assay; *IFA* immunofluorescence assay; *ELISA* enzyme linked immunosorbent assay; *blot* immunoblot used by the participants of the German Lyme disease proficiency testing program

Table 3. Percentage of laboratories which correctly identified the samples of the German Lyme disease proficiency testing program by use of their established assay systems

Method	Correct test results in %											
	October 1999			March 2000			November 2000			March 2001		
	51/ 1999	52/ 1999	Both samples	21/ 2000	22/ 2000	Both samples	41/ 2000	42/ 2000	Both samples	21/ 2001	22/ 2001	Both samples
IHA qual.	94.5	94	**94**	92	100	**92**	27	95	**23**	100	58	**58**
IHA quant.	78	96	**75**	94	82	**76**	58	53	**21**	86	73	**59**
IgM-IFA qual.	98	67	**66**	80	100	**80**	39	90	**29**	91	91	**85**
IgM-IFA qant.	82	77	**69**	58	85	**57**	42	77	**23**	65	68	**62**
IgG-IFA qual.	86	100	**86**	95	97	**93**	85	94	**82**	82	82	**65**
IgG-IFA quant.	90	83	**77**	89	65	**57**	58	78	**58**	69	81	**65**
ELISA (poly.) qual.	97	100	**97**	100	98	**98**	17	97	**13**	98	80	**79**
IgM-ELISA qual	96	88	**84**	89	100	**88**	88	89	**78**	85	98	**85**
IgG-ELISA qual.	95	99	**95**	97	97	**94**	94	99	**93**	91	85	**79**
IgM-BLOT qual.	93	82	**74**	90	99	**89**	93	87	**81**	91	81	**74**
IgG-BLOT qual.	82	99	**81**	98	90	**89**	76	92	**71**	96	87	**84**
Diagnostic comment	96	100	**96**	87	98	**85**	80	93	**74**	93	87	**81**

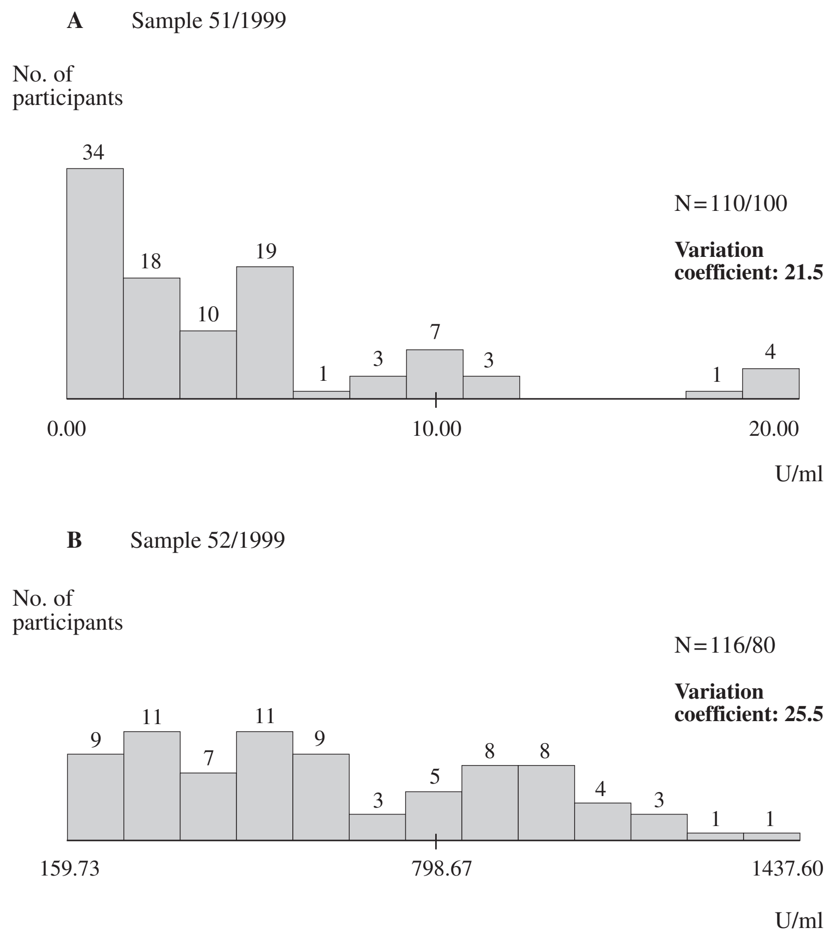

Fig. 3. Variability of quantitative ELISA test results. Representative histogram demonstrating range and relative frequency of quantitative IgG-ELISA test results as reported by 100 of 110 participants for sample 51/1999 (**A**) and by 80 of 116 for sample 52/1999 (**B**) in U/ml using established in-house and commercially manufactured assay systems

handle than were IgG tests, frequently resulting in lower rates of correct test results (Table 3). ELISA-testing was more accurate and proved to be more sensitive and specific than IFA and IHA testing and, consequently, demonstrated superior rates of correct test results. The serum samples showing a predominant IgG response (52/1999, 21/2000, 22/2001) were detected reproducibly by most of the test methods applied. The erythema migrans serum (41/2000) with low titers of IgM antibodies, however, was reliably detected by IgM-ELISA and IgM-immunoblot only, whereas IHA, IFA, and polyvalent ELISA in most cases failed to correctly identify this sample.

Interlaboratory inconsistencies of test results became obvious from the variability of quantitative test results in IFA (Table 4), IHA (data not shown), and ELISA (Fig. 3) in general and the reporting of category and number of specific immunoblot bands (Fig. 4A, B) in particular. Although the calculated median IFA titres of the reference laboratories and those of the participants in the majority of cases varied only for one to two $\log_2$ unit dilutions around the median, the ranges of titres in the group of participants revealed an exorbitant variability of test results (Table 4). Similarly, the quantitative ELISA results demonstrated a very low level of inter-assay standardisation in all the trials, resulting in an extreme heterogeneity of reported quantitative results (Fig. 3) due to methodological differences of commercially manufactured assays and the variable methods of quantification (values of optical density [OD], indices, titres, U/ml) used. Because of that lack of standardisation we decided, therefore, not to include quantitative ELISA results in the final evaluation of the proficiency testing surveys.

Qualitative immunoblot test results were reported by the participants in all surveys that have been performed (Table 3). In addition, the participating laboratories of the trials in November 2000 and March 2001 also reported, in part, the number and category of specific IgG and IgM bands observed in their immunoblots for each of the serum samples (Fig. 4A–D). The immunoblot results of the participating laboratories (Fig. 4A–D) of the surveys in 11/2000 and 03/2001, however, demonstrated that results of individual participants were not fully comparable with regard to the category and number of bands or the combination of bands.

Despite the variability of serological test results, however, the microbiological interpretation reported by the laboratories was more homogeneous than expected: during our surveys most participants (74–96%) reported correct interpretative statements (Table 3) regarding whether or not the test constellation suggested a possible borrelial infection and if an early or late phase of the specific antibody response was suspected.

False positive and false negative results and evaluation of test kit quality

Using a variety of in-house tests and commercially manufactured Lyme disease kits for the serodiagnosis of borreliosis, participants reported an unexpectedly high rate of false positive and false negative test results during the different surveys. For IHA, false negative results were found in 6–73% of reports and false positive results were reported by up to 6% of the participants. For IgM-IFA, the rate of false negative results reached up to 61% and the

Table 4. Analysis of median antibody titres as calculated from the IFA results of reference laboratories in comparison to the median titres calculated from the IFA results of all participants

Sample	Reference laboratories		Participants		
	Antibody titres (median)	Range of titres	Antibody titres (median)	Range of titres	Variation coeficient
51/1999	IgM: 0 IgG: 0	IgM: <20 IgG: 0–20	IgM: 0 IgG: 0	IgM: 0–80 IgG: 0–256	IgM: 10.3 IgG: 12
52/1999	IgM: 64 IgG: 1280	IgM: 40–128 IgG: 640–1280	IgM: 64 IgG:1280	IgM: 0–640 IgG: 20–16384	IgM: 28 IgG: 21.5
21/2000	IgM: 0 IgG: 160	IgM: 0–20 IgG: 80–256	IgM: 10 IgG: 256	IgM: 0–2048 IgG: 40–5120	IgM: n.d. IgG: 22.6
22/2000	IgM: 0 IgG: 0	IgM: <20 IgG: <20	IgM: 0 IgG: 0	IgM: 0–32 IgG: 0–80	IgM: n.d. IgG: n.d.
41/2000	IgM: 20 IgG: 0	IgM: 0–20 IgG: <20	IgM: 20 IgG: 10	IgM: 20–80 IgG: 0–256	IgM: 15.7 IgG: 28.6
42/2000	IgM: 0 IgG: 0	IgM: <20 IgG: <20	IgM: 0 IgG: 0	IgM: 0–19 IgG: 0–256	IgM: 47.1 IgG: n.d.
21/2001	IgM: 0 IgG: 0	IgM: <20 IgG: 0–39	IgM: 0 IgG: 10	IgM: 0–256 IgG: 0–256	IgM: n.d. IgG: n.d.
22/2001	IgM: 0 IgG: 40	IgM: <20 IgG: 20–128	IgM: 0 IgG: 80	IgM: 0–80 IgG: 0–1024	IgM: n.d. IgG: 14.4

n.d. Not determinable.

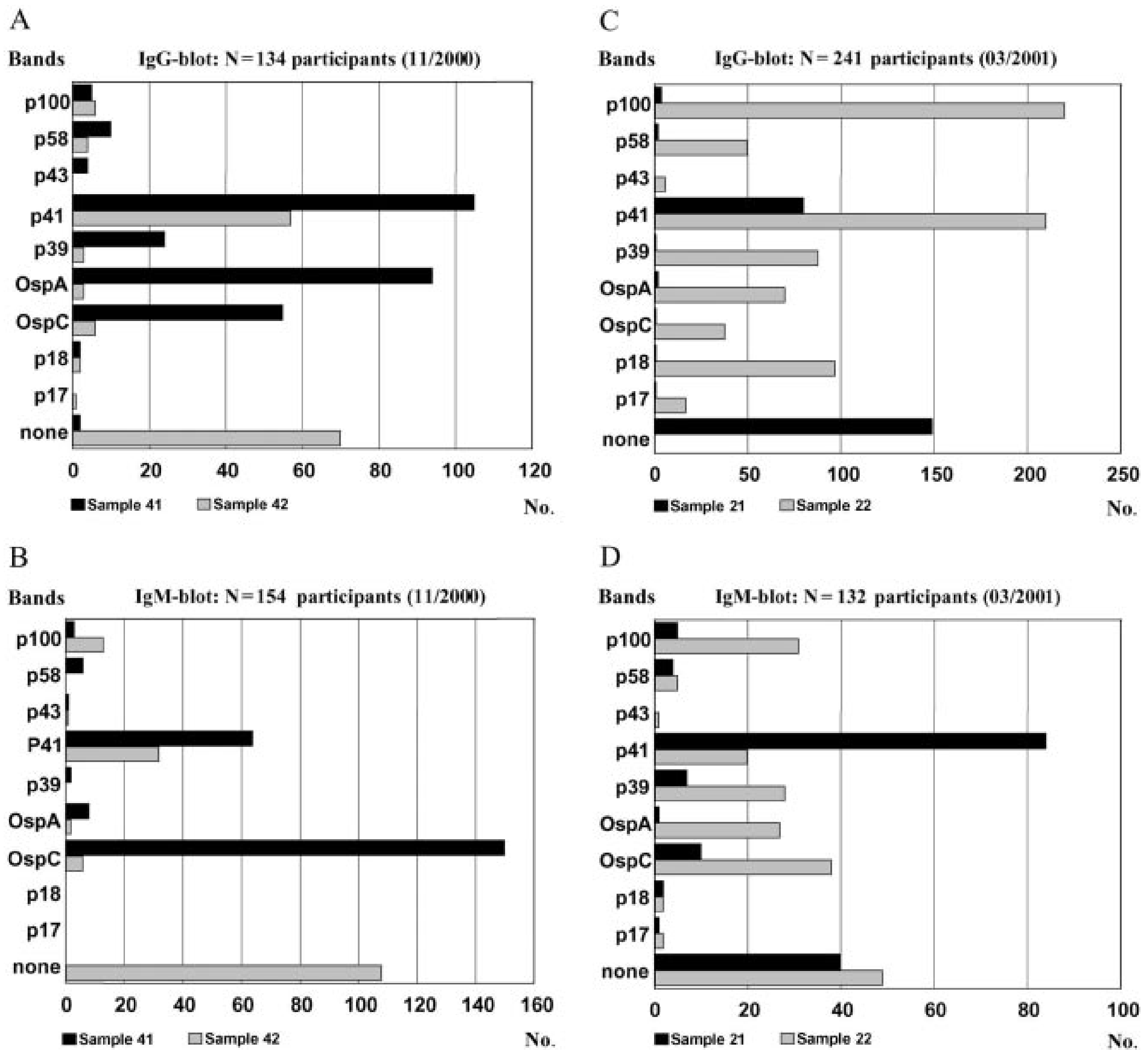

Fig. 4A–D. Relative frequency of number and class of specific bands detected by the participants of the German proficiency testing program using their established immunoblot assays for IgG (**A** and **C**) and IgM (**B** and **D**) for samples 41/2000, 42/2000, 21/2000 and 22/2001. Correct test results as determined for the samples by the reference laboratories are as follows: *Sample 41/2000: IgG:* p41, OspA (weak); *IgM:* p41, OspC. *Sample 42/2000:* no bands or p41 only for *IgG* and *IgM*. *Sample 21/2001:* no bands or p41 only for *IgG* and *IgM*. *Sample 22/2001: IgG:* p100, p41, p39, OspA, OspC, p18/p17; *IgM:* no bands or p41 only

rate of false positive results varied from 2 to 23% during our surveys. For IgG-IFA, both false negative and false positive results were reported by up to 18% of the participants. For polyvalent ELISA, the rate of false negative results ranged from 0 to 83% and was highest in the erythema migrans serum. IgM ELISA and immunoblot tests frequently proved less specific than IgG tests, giving false positive results in up to 15% of the ELISA reports and in 7–19% of the immunoblot reports. False positive results were reported in 1–9% of IgG ELISAs and 4–18% of the IgG immunoblots. As depicted in Table 5, the quality of test results was highly variable in part and also depended on the method used for antibody detection and the manufacturer of the test used in the survey.

Discussion

Several previous interlaboratory studies have compared the performance of serological tests for Lyme borreliosis [1–7]. Many of these studies used a panel of well defined serum samples obtained from clinically ill Lyme disease patients, and from 4 to 45 laboratories participated [1, 5, 6, 13, 14]. Such studies can provide insight into the sensitivity, reliability and reproducibility of serological testing only. However, such investigations cannot give sufficient information on overall performance and relative accuracy of tests at a national level nor provide information on achieving correct serological results with established diagnostic test systems in a large number of partic-

ipating microbiological routine laboratories over longer periods of time. Clinically defined samples from Lyme borreliosis patients are always helpful in the evaluation of the sensitivity and inter-/intra laboratory variability of test results. These sera, however, if sent out without corresponding negative or 'problematic' samples do not at all represent the routine constellation found in the majority of sera sent to microbiological laboratories day by day for diagnosing the presence or absence of possible Lyme disease.

Any study aimed at investigating the quality and overall performance of test systems in the routine microbiological laboratory must be capable of evaluating the broad range of commercially manufactured test systems that are currently used, rather than simply investigating the sensitivity, specificity and reproducibility of experimental serological tests performed in expert laboratories under optimised conditions. Moreover, i. the significance placed on positive, equivocal and negative findings, ii. the weighting of the result obtained by the test systems used in individual laboratories for achieving an overall diagnosis of possible borrelial infection, and iii. the correct classification of the detected antibody response as belonging to possible early or late Lyme disease should all be analysed.

Because it is normally expected that every sample sent out to the participants of a proficiency testing survey will give comparable results in most laboratories, provided that the established diagnostic tests are of comparable diagnostic quality, we decided to use a different approach in our trials investigating the actual situation in the field of Lyme borreliosis serology in a more realistic manner. Thus, for each survey only two samples were selected which were obtained before the trials from routinely diagnosed patients or were harvested from healthy blood donors, but these samples were then used to test several hundred laboratories at the same time. Prior to and/or during each trial these samples had been unambiguously characterised, according to the tentative guidelines of the German chamber of physicians for infection serology proficiency testing [9], with regard to the amount and class of antibody and the number and category of immunoblot bands detected in a variety of test systems by reference laboratories with excellent expertise in the field (Fig. 1). This was done to generate reference test results for each parameter investigated by the participants during the trials. The results obtained by the participants with their established tests were then reported and evaluated together with a diagnostic statement under standardised conditions. As expected, most laboratories relied on ELISA and immunoblot testing of samples. Interestingly, the IFA and IHA median titres of the participating laboratories mostly met the median of titres calculated for the positive samples from the reports of the reference laboratories, but the range of titres reported by the participants showed high inter-laboratory variability (Table 4). The quantitative ELISA results also demonstrated a very low inter-assay standardisation (Fig. 3). Moreover, as already described in previous studies on this topic [1, 2], a high number of both false negative and false positive test results was obvious from our surveys and could be partly correlated with the diagnostic method, the quality of the test kits (Table 5), and the amount of specific antibody present in different sera (Table 4). Qualitative ELISA and immunoblot testing was clearly more reproducible and proved to be more sensitive and specific than IFA and IHA testing, but it was obvious that IgM tests were more difficult to handle than IgG tests and frequently proved less specific (Table 3). Samples containing high amounts of specific antibody were diagnosed more reproducibly in the various assays used by the participants than were samples with low antibody titres.

From the results of our surveys it can be concluded that in routine laboratories the quantity of detected antibody measured in titres or quantitative ELISA results, and similarly the number and category of specific immunoblot bands, can be highly variable for the same sample. In addition, changes in qualitative and quantitative serologic test results may be misleading and can simply result from the use of different assay systems in different laboratories. Consequently, because the exact level of positivity can vary from test to test and the exact level of epidemiological background positivity is not known for many parts of Europe, the specificity and sensitivity of test results and the clinical significance of the findings probably depend most on the quality of the laboratory and the manufacturer of the test kit (Table 5) [14]. Moreover, a more general implementation of diagnostic criteria for the interpretation of immunoblot results [12] may be difficult in the light of the relatively high variability of band patterns found in the same sample by the participants' immunoblot assays (Fig. 4A, B). These finding are important also because it is known that some physicians try to correlate the activity of Lyme disease and the success of subsequent therapy with quantitative serological testing and with qualitative changes in the test results.

However, in contrast to the heterogeneity of serologic test results, most participants (74–96%), irrespective of the test method or the combination of assay systems used, reported correct microbiological comments (Table 3) regarding the question of whether or not the test constellation suggested possible borrelial infection and if an early or late phase of the specific antibody response was suspected. This finding probably results from the application of several test systems on the same sample by most of the participating laboratories and strongly supports the use of assay combinations, i.e. IgG- and IgM-assays as well as screening tests and confirmatory tests for diagnostic testing in infection serology of Lyme disease (Fig. 2).

When considering the results of our proficiency testing surveys it must be borne in mind, however, that four trials held with eight samples investigated cannot definitively determine the general value of Lyme serology or the particular suitability of certain assay systems for routine microbiological laboratory diagnosis. Nevertheless, proficiency testing programs represent the only suitable tool for a general form of external quality control at the national or international level, at the same time providing a more realistic impression of the situation in this difficult field of microbiological diagnostics than do expert studies performed under ideal conditions at the reference laboratory level. Moreover, the in part extreme variability of test results reported by the participants in our surveys is in accordance with previous studies on this topic [1, 8, 13, 14]. The lack of diagnostic quality probably results partly

Table 5. German Lyme serology proficiency testing program: Accuracy of test results of the most frequently used commercially manufactured diagnostic assays

Test	Correct results in % (No. of participants)			
	10/1999 (n=226)	03/2000 (n=334)	11/2000 (n=247)	03/2001 (n=337)
IFA (IgG) qual.	**86 (n=44)**	**93 (n=43)**	**82 (n=33)**	**65 (n=40)**
BAG	100 (n–5)	–	–	75 (n=4)
Bios	60 (n=5)	63 (n=6)	100 (n=4)	50 (n=6)
Euroimmun	75 /n=4)	100 (n=7)	68 (n=3)	60 (n=5)
Gull	80 (n=5)	100 (n=4)	33 (n=3)	–
Mast	100 (n=3)	100 (n=3)	–	33 (n=3)
Viramed	67 (n=3)	100 (n=3)	–	100 (n=3)
IFA (IgM) qual.	**66 (n=41)**	**80 (n=41)**	**29 (n=31)**	**85 (n=33)**
BAG	100 (n=4)	67 (n=3)	60 (n=5)	100 (n=3)
Bios	100 (n=3)	50 (n=6)	0 (n=5)	83 (n=6)
Euroimmun	25 (n=4)	88 (n=8)	33 (n=3)	80 (n=5)
Gull	100 (n=4)	68 (n=3	–	–
Mast	0 (n=3)	100 (n=3)	–	68 (n=3)
Viramed	–	100 (n=3)	–	–
ELISA (IgG)	**95 (n=139)**	**94 (n=211)**	**93 (n=155)**	**79 (n=214)**
Abbott	100 (n=7)	87 (n=13)	100 (n=9)	0 (n=11)
DadeBehring	98 (n=44)	97 (n=65)	100 (n=58)	100 (n=70)
Dako	100 (n=9)	100 (n=10)	100 (n=7)	14 (n=7)
Sorin	100 (n=1)	86 (n=7)	100 (n=4)	100 (n=7)
Mikrogen	67 (n=6)	100 (n=8)	100 (n=13)	85 (n=13)
Virotech	100 (n=10)	100 (n=20)	100 (n=11)	96 (n=26)
Biermann	–	71 (n=7)	100 n=3)	–
Medipan	100 (n=4)	100 (n=5)	100 (n=6)	80 (n=5)
ELISA (IgM)	84 (n=179)	**88 (n=256)**	**78 (n=187)**	**85 (n=266)**
Abbott	83 (n=6)	100 (n=16)	89 (n=9)	100 (n=13)
DadeBehring	97 (n=64)	90 (n=83)	96 (n=71)	100 (n=93)
Dako	100 (n=14)	86 (n=14)	100 (n=12)	82 (n=11)
Sorin	50 (n=2)	100 (n=7)	80 (n=5)	100 (n=6)
Mikrogen	100 (n=10)	100 (n=14)	5.9 (n=17)	76 (n=17)
Virotech	100 (n= 13)	95 (n=20)	33 (n=11)	46 (n=28)
Medipan	100 (n=5)	–	50 (n=6)	100 (n=6)
Immunoblot (IgG)	**81 (n=174)**	**89 (n=243)**	**71 (n=151)**	**84 (n=244)**
Biermann	100 (n=7)	100 (n=11)	100 (n=4)	100 (n=10)
Euroimmun	90 (n=10)	100 (n=25)	82 (n=11)	76 (n=28)
Mikrogen	100 (n=31)	100 (n=50)	84 (n=49)	95 (n=64)
Viramed	31 (n=13)	91 n=23)	22 (n=18)	81 (n=37)
Virotech	59 (n=22)	79 (n=28)	78 (n=18)	81 (n=37)
Immunoblot (IgM)	**74 (n=174)**	**89 (n=244)**	**81 (n=155)**	**74 (n=234)**
Biermann	43 (n=7)	82 (n=11)	100 (n=4)	100 (n=9)
Euroimmun	10 (n=10)	100 (n=24)	70 (n=10)	87 (n=24)
Mikrogen	90 (n=31)	100 (n=49)	82 (n=51)	67 (n=61)
Viramed	86 (n=14)	82 (n=23)	74 (n=19)	89 (n=33)
Virotech	78 (n=20)	100 (n=28)	82 (n=17)	71 (n=34)

from difficulties within the laboratories and also from problems with the widely used but not yet well standardised commercially manufactured serological test kits. To improve the value of Lyme disease serology in the routine microbiological laboratory, a much better inter-assay standardisation of the commercially available test kits is necessary [1, 12, 14]. Moreover, in Europe there is a need for the production of standard sera with high titres of specific antibodies for the uniform calibration of quantitative assays [9]. In addition, repositories of serum specimens covering early and late borrelial infections, other spirochetal infections, and inflammatory disorders that have shown possible cross reactivity in Lyme disease testing as well as normal sera from areas of nonendemicity should be made commercially available to serve as a serum source for a more general standardisation of test kits by the manufacturers, as proposed recently by the Centers of disease control and Prevention (CDC) and the Association

of State and Territorial Public Health Laboratory Directors (ASTPHLD) conference on the Serological Diagnosis of Lyme Disease [15]. Furthermore, manufacturers should be forced by law to undertake a more detailed clinical evaluation of their tests before introducing their product to the market and to make available their peer-reviewed evaluation data to the users of such tests. In view of our results further standardisation of Lyme disease serology is not just desirable but is urgently needed. Moreover, more stringent internationally accepted standardised evaluation criteria must be developed and applied for the validation of available test kits.

Acknowledgments

The authors would like to thank S. Neumann and R. Hilden (INSTAND e.V, Düsseldorf) for excellent technical and statistical support. We also thank Mrs. I. Gutgesell for her excellent technical assistance during the surveys.

References

1. Bakken LL, Callister SM, Wand PJ, Schell RF (1997) Interlaboratory comparison of test results for detection of Lyme disease by 516 participants in the Wisconsin State Laboratory of Hygiene/College of American Pathologists proficiency testing program. J Clin Microbiol 35: 537–543
2. Bakken LL, Case KL, Callister SM, Bourdeau NJ, Schell RF (1992) Performance of 45 laboratories participating in a proficiency testing program for Lyme disease. JAMA 268: 891–895
3. Hedberg CW, Osterholm MT, MacDonald KL, White KE (1987) An interlaboratory study of antibody to *Borrelia burgdorferi*. J Infect Dis 1987 155: 1325–1327
4. Lane RS, Lennette ET, Madigan JE (1990) Interlaboratory comparison of indirect immunoflourescence assays for serodiagnosis of Lyme disease. J Clin Microbiol 28: 1774–1779
5. Schwartz, BS, Goldstein MD, Ribeiro JM, Schulze TL, Shahied SI (1989) Antibody testing in Lyme disease. A comparison of results in four laboratories. JAMA 262: 3431–3434
6. Luger SW, Krauss E (1990) Serologic tests for Lyme disease. Interlaboratory variability. Arch Intern Med 150: 761–763
7. Steere AC, Taylor E, McHugh GL, Logigian EL (1993) The overdiagnosis of Lyme disease. JAMA 269: 1812–1816
8. Hedberg CW, Osterholm MT (1990) Serologic tests for antibody to *Borrelia burgdorferi*. Another Pandora's box for medicine? Arch Intern Med 150: 732–733
9. Hunfeld KP, Brade V (2000) Ringversuche in der bakteriologischen Infektionsserologie – Standortbestimmung und Auswertung des Ringversuchs X/99. Der Mikrobiologe 10: 135–144
10. Magnarelli LA (1989) Quality of Lyme disease tests. JAMA 262: 3464–3465
11. Centers of disease control and Prevention (1995) Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. Morbid Mortal Weekly Rep 44: 590–591
12. Wilske B, Zoeller L, Brade V, Eiffert H, Goebel UB, Stanek G, Pfister HW (2000) Qualitätsstandards in der Mikrobiologisch-infektiologischen Diagnostik: Lyme-Borreliose. Urban & Fischer, München Jena
13. Tammemagi MC, Frank JW, Leblanc M, Artsorb H, Streiner DL (1995) Methodological issues in assessing reproducibility – a comparative study of various indices of reproducibility applied to repeat ELISA serologic tests for Lyme diesease. J Clin Epidemiol 48: 1123–1132
14. Guy EC, Robertson JN, Cimmino M, Gern L, Moosmann Y, Rijpkema SGT, Sambri V, Stanek G (1998) European interlaboratory comparison of Lyme borreliosis serology. Zentbl Bakteriol 287: 241–247
15. Centers of Disease Control and Prevention, Association of State and Territorial Public Health Laboratory Directors (1994) Proceedings of the Second National Conference on Serologic Diagnosis of Lyme disease (Dearborn, MI). Association of State and Territorial Public Health Laboratory Directors, Washington DC

Addendum

Reference laboratories and members of the German Study Group of Bacteriologic Infection Serology:

Prof. Dr. med. Volker Brade, Dr. med. Klaus-Peter Hunfeld, Ringversuchslabor für bakteriologische Infektionsserologie, Institut für Medizinische Mikrobiologie, Universitätsklink der Johann Wolfgang Goethe-Universität Frankfurt, Paul-Ehrlich-Straße 40, D-60596 Frankfurt/Main

Prof. Dr. med. Eberhard Straube, Institut für Medizinische Mikrobiologie, Klinikum der Friedrich-Schiller-Universität Jena (Direktor: Prof. Dr. med. E. Straube), Semmelweisstraße 4, D-07740 Jena

Prof. Dr. med. Hans-Jochen Hagedorn, Medizinale Untersuchungsstelle Herford, Lübbertorwall 18, D-32052 Herford

Prof. Dr. med. Gerold Stanek, Abteilung für Infektionsimmunologie des Klinischen Instituts für Hygiene und Medizinische Mikrobiologie der Medizinischen Fakultät der Universität Wien, Kinderspitalgasse 15, A-1095 Wien

Dr. med. Christoph Schörner, Institut für Klinische Mikrobiologie, Immunologie und Hygiene der Friedrich-Alexander-Universität Erlangen-Nürnberg (Direktor: Prof. Dr. med. M. Röllinghof), Wasserturmstraße 3, D-91054 Erlangen

Dr. med. Fritz Mühlschlegel, Institut für Hygiene und Mikrobiologie der Universität Würzburg (Direktor: Prof. Dr. med. M. Frosch), Josef-Schneider-Straße 2, D-97080 Würzburg

Correspondence: Klaus-Peter Hunfeld, M.D., Institute of Medical Microbiology, University Hospital of Frankfurt, Paul-Ehrlich-Straße 40, D-60596 Frankfurt/Main, Germany,
E-mail: K.Hunfeld@em.uni-frankfurt.de