JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 1997, p. 537–543 Vol. 35, No. 3
0095-1137/97/$04.00+0

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

LORI L. BAKKEN,[1,2]† STEVEN M. CALLISTER,[3] PHILIP J. WAND,[2] AND RONALD F. SCHELL,[2,4]†*

Departments of Continuing and Vocational Education[1] and Medical Microbiology and Immunology,[4] and Wisconsin State Laboratory of Hygiene,[2] University of Wisconsin, Madison, Wisconsin 53706; and Microbiology Research Laboratory, Gundersen Medical Foundation, La Crosse, Wisconsin 54601[3]

Received 12 August 1996/Returned for modification 15 October 1996/Accepted 4 November 1996

In 1991, we reported that 55% of laboratories participating in the Wisconsin Proficiency Testing Program could not accurately identify serum samples from Lyme disease patients containing antibody against *Borrelia burgdorferi*. The purpose of this study was to determine whether the accuracy of Lyme disease test results reported by approximately 500 participants in the Wisconsin State Laboratory of Hygiene/College of American Pathologists Lyme Disease Survey had improved. From 1992 through 1994, 50 serum samples were sent to participants of the survey. Each laboratory received 28 serum samples from individuals with Lyme disease according to the case definition of the Centers for Disease Control and Prevention and 22 serum samples from healthy individuals. Unfortunately, the serodiagnosis of Lyme disease by participants had not improved. The specificity of the Lyme disease assays steadily decreased from approximately 95% to approximately 81% during the 3-year period of the survey. False-positive test results approached 55% with some of the serum samples from healthy donors. A serum sample containing antibody against *Treponema pallidum* was reported as positive by 70% of the participants. In addition, the sensitivity fluctuated between 93 and 75%, depending upon the conjugate used by the laboratories. These results suggest that stronger criteria must be applied for approving and continuing to approve commercially available kits for the serodiagnosis of Lyme disease.

Lyme borreliosis, caused by the spirochete *Borrelia burgdorferi* sensu lato, is an illness that can affect numerous organ systems including the skin, connective tissue, heart, and central nervous system (20). Symptoms range from mild, flu-like abnormalities to severe disabling conditions. Since the illness can resemble aseptic meningitis, rheumatoid arthritis, influenza, or other syndromes, clinicians tend to rely upon serological tests as the primary indicator for detection of infection with *B. burgdorferi* sensu lato. Serological tests for Lyme borreliosis, however, have demonstrated poor sensitivity and specificity (9–12, 14, 18) and continue to make the overdiagnosis of Lyme borreliosis common (19). Therefore, Magnarelli (15) and others (9, 14) have suggested that laboratories performing serological tests for the detection of anti-*B. burgdorferi* antibodies be required to participate in a proficiency testing program.

We showed previously (1) that significant inter- and intralaboratory variations existed among test results obtained from 45 laboratories performing Lyme disease proficiency testing in an area where Lyme disease is endemic. Approximately 4 to 21% of the laboratories failed to identify correctly positive serum samples with titers of 512 or more by using polyvalent serum or immunoglobulin G (IgG) conjugates. With lower levels of anti-*B. burgdorferi* antibody in the serum samples, approximately 55% of participating laboratories did not identify a case-defined serum sample from an individual with Lyme disease according to the case definition of the Centers for Disease Control and Prevention (CDC; CDC case-defined serum sample). In addition, 2 to 7% of laboratories identified as positive serum samples from individuals with no known exposure to *B. burgdorferi*. The false-positivity rate increased to 27% when an IgG conjugate was used. Furthermore, we showed a striking inability of many laboratories to reproduce their results with split samples from the same individual.

In the present study, we sought to determine if Lyme disease test results had improved since our initial report (1). We evaluated the sensitivity and specificity of "screening tests" performed by 516 laboratories for the Wisconsin State Laboratory of Hygiene (WSLH)/College of American Pathologists (CAP) Lyme Disease Proficiency Survey from 1992 through 1994.

This is the first report that evaluated the performance of serodiagnostic testing for Lyme disease at the national level.

MATERIALS AND METHODS

Selection of patients. Twenty-eight serum samples were obtained from patients diagnosed with Lyme borreliosis according to the case definition of Lyme disease established by CDC (6). In addition, a serum sample with Venereal Disease Research Laboratory (VDRL) test positivity and antibodies against *Treponema pallidum* was selected, as was a serum sample containing rheumatoid factor. Furthermore, 20 serum samples were obtained from healthy individuals with no history of Lyme borreliosis. All subjects were evaluated and diagnosed by physicians with expertise in the diagnosis and treatment of Lyme borreliosis. Descriptions of the serum samples distributed to participants of the WSLH/CAP Lyme Disease Proficiency Survey are presented in Table 1.

Serum sample preparation. Approximately 500 ml of whole blood was obtained from each donor. The blood was allowed to clot, and the serum was collected by centrifugation at 50 × *g* and was stored at −20°C until it was used. Subsequently, the serum samples were thawed and 250-μl aliquots, without preservatives, were dispensed into 0.5-ml polypropylene vials (Sarstedt Inc.,

* Corresponding author. Mailing address: Wisconsin State Laboratory of Hygiene, 465 Henry Mall, Madison, WI 53706. Phone: (608) 262-3634. Fax: (608) 265-3451.

† Member of the Diagnostic Immunology Resource Committee (DIRC). Other members of the DIRC who participated in this study are listed in the appendix.

TABLE 1. Serum samples obtained for distribution in the WSLH/CAP Lyme disease proficiency testing survey

Sample identification no.	Type of serum sample[a]	Date of onset (mo/day/yr)	Date of sample collection (mo/day/yr)
92-01	Lyme	4/1/90	10/1/91
92-02	Normal	NA[b]	
92-03	Lyme	10/1/91	12/1/91
92-04	Normal	NA	
92-05	Lyme	6/1/91	9/1/91
92-06	Normal	NA	
92-07	Lyme	8/1/91	12/1/91
92-08	Lyme	8/1/91	8/1/91
92-09	Lyme	6/1/91	10/1/91
92-10	Lyme	6/1/91	8/1/91
92-11	Lyme	6/1/83	7/1/92
92-12	Lyme	5/1/92	7/1/92
92/13	Normal	NA	
92-14	Normal	NA	
92-15	Normal	NA	
92-16	Lyme	7/1/92	8/1/92
92-17	Lyme	7/1/92	10/1/92
92-18	Normal	NA	
92-19	Lyme	8/1/92	10/1/92
92-20	Lyme	7/1/92	8/1/92
93-01[c]	Lyme	5/1/92	11/1/92
93/02[c]	Lyme	5/1/92	11/1/92
93-03	Normal	NA	
93-04	Normal	NA	
93-05	Lyme	9/1/90	3/1/91
93-06	Normal	NA	
93-07	Lyme	10/1/92	3/1/93
93-08	Normal	NA	
93-09[d]	Lyme	7/1/92	7/1/92
93-10[d]	Lyme	7/1/92	8/1/92
93-11	Normal	NA	
93-12	Normal	NA	
93-13	Normal	NA	
93-14	Normal	NA	
93-15	Lyme	7/1/93	9/1/93
94-01	Lyme	8/1/93	8/1/93
94-02[e]	Lyme	8/1/93	9/1/93
94-03	Lyme	7/1/93	7/1/93
94-04	Lyme	8/1/93	7/1/93
94-05	Normal	NA	
94-06	Normal	NA	
94-07	Normal	NA	
94-08	Normal	NA	
94-09	Normal	NA	
94-10	Lyme	7/1/93	7/1/93
94-11	Normal	NA	
94-12	Lyme	7/1/93	9/1/93
94-13[e]	Lyme	8/1/93	8/1/93
94-14	Normal	NA	
94-15	Lyme	8/1/93	8/1/93

[a] Lyme, serum sample from patient with Lyme disease; Normal, serum sample from a healthy subject.

[b] NA, not applicable.

[c] Serum sample 93-01 and 93-02 were replicates of the same specimen.

[d] Serum samples 93-09 and 93-10 were collected 4 weeks apart from one individual.

[e] Serum samples 94-02 and 93-13 were collected 4 weeks apart from one individual.

Princeton, N.J.). Most of the serum samples (all samples except specimen 93-13) were nonreactive for rheumatoid factor, hepatitis B virus surface antigen, and human immunodeficiency virus type 1. The serum samples were also tested for cross-reactivity to *T. pallidum*. Serum sample 93-03 was reactive when tested by the VDRL and the fluorescent treponemal antibody-absorption tests. No changes in titers by an indirect fluorescent-antibody test for detection of anti-*B. burgdorferi* antibody were detected with stored serum samples during the duration of the program. A minimum of 1% of the serum samples from each donor was tested prior to each shipment.

Absorption of serum samples from healthy individuals. Two serum samples were absorbed for demonstration of the absence of cross-reactive or natural antibodies against *B. burgdorferi*. Serum samples 92-13 and 92-15 from healthy individuals were absorbed with *B. burgdorferi* sensu stricto 297. This isolate expresses the major outer surface proteins of *B. burgdorferi* (13) and is representative of the dominant genomic (2) and seroprotective (13) groups of *B. burgdorferi* sensu stricto found in the United States. Briefly, a suspension of low-passage *B. burgdorferi* sensu stricto 297 was thawed, and an aliquot was used to inoculate five centrifuge bottles containing 250 ml of fresh Barbour-Stoenner-Kelly (BSK) medium (4). The cultures were inoculated at 32°C for 3 days, washed three times by centrifugation with phosphate-buffered saline (PBS), pooled with 3 ml of PBS, and used to inoculate 250 ml of serum specimens 92-13 and 92-15. After incubation for 1 h at 21°C the spirochetes were pelleted by centrifugation (10,000 rpm) for 30 min. The sera were removed and were stored at −20°C until used.

Selection of serum samples for shipment to participants. Serum samples obtained from 22 individuals with no history of Lyme borreliosis (including the VDRL test- and rheumatoid factor-reactive specimens) and 28 patients with Lyme disease were randomized for distribution into 10 shipments. Sera from patients with Lyme disease were collected 1 week to 111 months after the onset of symptoms. The average time between the onset of symptoms and the collection date was 7 months. Serum samples 93-01 and 93-02 were replicates of the same specimen. Paired serum samples 93-09 and 93-10 and serum samples 94-02 and 94-13 were collected from two individuals, respectively, and the serum samples in the pairs were drawn at 1-month intervals. Five serum samples were distributed per shipment, for a total of 50 randomized serum samples tested by each participating laboratory.

Testing of serum samples prior to shipment. A random selection of 1% of the aliquots of each serum sample was tested by an in-house enzyme immunoassay for reactivity to *B. burgdorferi* 7 days prior to shipment to the survey participants. Flat-bottom microtiter plates (Nunc) were coated with 100 μl of 2.0 mg of low-passage (<10) *B. burgdorferi* sensu stricto B31 antigen per ml in coating buffer and were incubated overnight at 4°C. The plates were washed with PBS-Tween buffer and were dried for 1 h by incubation at 21°C. The plates were then

TABLE 2. Reference index values for serum samples from healthy donors[a]

Sample identification no.	Mean ± SD index value	
	IgM	IgG
92-02	0.44 ± 0.01	0.60 ± 0.17
92-04	0.60 ± 0.07	0.29 ± 0.01
92-06	0.77 ± 0.05	0.52 ± 0.08
92-13[b]	0.93 ± 0.05	0.23 ± 0.03
92-14	0.22 ± 0.03	0.31 ± 0.04
92-15[b]	0.65 ± 0.03	0.01 ± 0.01
92-18	0.68 ± 0.03	0.22 ± 0.03
93-03[c]	0.19 ± 0.02	1.11 ± 0.04
93-04	0.77 ± 0.05	0.08 ± 0.01
93-06	0.08 ± 0.06	0.25 ± 0.03
93-08	0.38 ± 0.06	0.14 ± 0.05
93-11	0.18 ± 0.01	0.21 ± 0.01
93-12	0.31 ± 0.02	0.57 ± 0.02
93-13[d]	0.02 ± 0.00	0.15 ± 0.01
93-14	0.20 ± 0.01	0.43 ± 0.01
94-05	0.26 ± 0.01	0.17 ± 0.01
94-06	0.97 ± 0.22	0.47 ± 0.03
94-07	0.18 ± 0.01	0.25 ± 0.01
94-08	0.69 ± 0.06	0.47 ± 0.03
94-09	0.50 ± 0.05	0.21 ± 0.01
94-11	0.30 ± 0.03	0.87 ± 0.04
94-14	0.08 ± 0.01	0.18 ± 0.01
Mean	0.43 ± 0.04	0.35 ± 0.03

[a] A total of 22 serum samples were obtained from healthy donors. Values are mean index values for 1% of each serum sample. A positive result had an index value of >1.1. Equivocal results were reported for values of between 0.9 and 1.1, and negative results were reported for those with index values of <0.9.

[b] Serum samples 92-13 and 92-15 were treated with *B. burgdorferi* sensu stricto 297.

[c] Serum samples 93-03 were reactive when tested by the VDRL and FTA-ABS tests.

[d] Serum samples 93-13 was positive for rheumatoid factor.

TABLE 3. Reference index values for sera from patients with Lyme disease[a]

Sample identification no.	Mean ± SD index value	
	IgM	IgG
92-01	1.22 ± 0.03	4.66 ± 0.11
92-03	2.18 ± 0.11	0.95 ± 0.06
92-05	1.33 ± 0.01	3.28 ± 0.06
92-07	0.74 ± 0.02	2.91 ± 0.12
92-08	1.35 ± 0.12	4.49 ± 0.04
92-09	1.35 ± 0.15	4.87 ± 0.29
92-10	1.07 ± 0.04	1.80 ± 0.09
92-11	1.01 ± 0.03	4.73 ± 0.19
92-12	4.16 ± 0.08	0.56 ± 0.07
92-16	2.83 ± 0.12	2.71 ± 0.17
92-17	1.87 ± 0.21	0.59 ± 0.06
92-19	2.08 ± 0.24	0.88 ± 0.13
92-20	1.93 ± 0.13	0.74 ± 0.06
93-01[b]	2.64 ± 0.21	0.52 ± 0.04
93-02[b]	2.61 ± 0.24	0.51 ± 0.04
93-05	0.43 ± 0.05	4.16 ± 0.32
93-07	1.81 ± 0.15	4.04 ± 0.18
93-09[c]	6.68 ± 0.21	1.78 ± 0.07
93-10[c]	3.85 ± 0.38	0.99 ± 0.09
93-15	1.26 ± 0.07	1.77 ± 0.07
94-01	2.16 ± 0.24	2.43 ± 0.12
94-02[c]	1.13 ± 0.09	1.42 ± 0.06
94-03	1.93 ± 0.18	1.07 ± 0.09
94-04	0.26 ± 0.02	0.68 ± 0.04
94-10	2.62 ± 0.22	1.59 ± 0.11
94-12	0.87 ± 0.15	1.57 ± 0.17
94-13[c]	3.38 ± 0.45	1.40 ± 0.14
94-15	2.84 ± 0.32	1.05 ± 0.11
Mean	2.06 ± 0.15	2.08 ± 0.11

[a] A total of 28 serum samples were obtained from patients with Lyme disease. Values are mean index values for 1% of each serum sample. A positive result had an index value of >1.1. Equivocal results were reported for values of between 0.9 and 1.1, and negative results were reported for those with index values of <0.9.

[b] Serum samples 93-01 and 93-02 were replicates of the same serum sample.

[c] Serum samples 93-09 and 93-10 along with serum samples 94-02 and 94-13 were drawn 4 weeks apart from two individuals, respectively.

placed in a desiccator jar under vacuum and were stored at 4°C for 24 h before being sealed with a desiccant in aluminum packs (Riley and Geehr, Evanston, Ill.) and stored at −70°C until they were used.

Microtiter plates containing *B. burgdorferi* sensu stricto B31 antigen were overlaid with a 100-fold dilution of each serum sample, a calibrator, or a reference serum sample contained in PBS with 1% bovine serum albumin. The microtiter plates were incubated for 1 h at 25°C in a humid chamber and were then washed three times with PBS containing 0.05% Tween 20 (Sigma Chemical Co., St. Louis, Mo.). This step was followed by incubation with antibody to human IgM or IgG labeled with horseradish peroxidase (Organon-Teknika-Cappel, Malvern, Pa.) and subsequent incubation with *o*-phenylenediamine phosphate (Sigma). The reaction was stopped by the addition of 0.1 ml of 3 M NaOH, and the absorbances were determined immediately at a wavelength of 490 nm. A mean index value was obtained for each of the test samples by dividing the absorbance value of the sample by the absorbance of the positive control. For this assay, a positive result had an index value of greater than 1.1. Equivocal results were reported for values between 0.9 and 1.1, and negative results were those with index values of less than 0.9. All serum samples were tested five times by this method prior to shipment.

Shipments. Serum samples were sent to participants in styrofoam boxes containing two refrigeration packs (Mid-Lands Chemical, Inc., Omaha, Nebr.). Serum samples were delivered by the U.S. Postal Service's priority mail service for receipt within 2 days. Samples shipped outside the United States were sent by Federal Express.

Participating laboratories. No criterion was established for excluding laboratories from participating in the Lyme Disease Proficiency Survey. Test protocols were not consistent among the participants. Some laboratories tested only for IgM, while others used only IgG-specific assays. Other laboratories used both assays. Furthermore, protocols for testing varied (e.g., some laboratories performed an IgM test only if a polyvalent assay was positive). An average of 388 (76%) laboratories voluntarily tested serum samples with polyvalent conjugates with each shipment of the serum samples. Seventy-eight (15%) laboratories tested the specimens by IgG assays, and 50 (9%) laboratories used IgM assays. The percentage of laboratories performing polyvalent, IgM, or IgG tests varied slightly (<5%) over the 3-year study period. In 1992, 84% of the polyvalent tests performed by participants were enzyme immunoassays, while 67 and 48% of the laboratories used enzyme immunoassays to perform IgM- and IgG-specific tests, respectively. These percentages changed in 1993. Ninety-three percent of the laboratories using polyvalent tests performed enzyme immunoassays. In addition, 80 and 70% of the IgM- and IgG-specific tests were performed by enzyme immunoassays. The percentages of laboratories that used enzyme assays in 1994 were similar to those reported in 1993.

Most of the participants were located in areas of the United States where Lyme disease is endemic. Twelve laboratories were located in countries outside the United States.

Data collection. An average of 516 laboratories reported results from 1992 through 1994. During this period, a laboratory may have chosen not to report results on a sample or set of samples, thus causing variability in the number of tests performed. Laboratories were instructed to treat and process the serum samples used for proficiency testing as routine samples and to perform their established methods for determining the serodiagnosis of Lyme disease. Results were requested within 10 days of receiving the shipment. Quantitative (numerical values) and qualitative (test interpretations) results were collected. Other procedural data such as specimen treatment, source of conjugate, reagents, controls, reporting schemes, incubation time, and temperatures were also recorded.

Statistical analysis. A *t* test was used to determine if there was a significant difference between the reference index values for the sera from individuals with Lyme disease and those from healthy individuals.

RESULTS

Reference sera used for proficiency testing. Table 2 presents the mean anti-*B. burgdorferi* IgM and IgG index values obtained for 1% of each of the serum samples from healthy individuals (including two cross-reactive serum samples; samples 93-03 and 93-13). The serum samples were evaluated by an enzyme immunoassay before being distributed to participants of the Lyme Disease Proficiency Survey. Only serum sample 93-03 had a significant level of immunoglobulin, specifically IgG, against *B. burgdorferi*. All normal and cross-reactive sera had IgG index values of ≤1.11, while index values with the IgM conjugate were ≤0.97. The overall mean anti-*B. burgdorferi* IgM and IgG index values were 0.43 ± 0.04 and 0.35 ± 0.03, respectively.

By contrast, the overall mean index values for anti-*B. burgdorferi* IgM and IgG antibodies with CDC case-defined sera were 2.06 ± 0.16 and 2.08 ± 0.11, respectively (Table 3). Serum samples 92-03, 92-12, 92-17, 92-19, 92-20, 93-01, 93-02, and 93-10 had elevated levels of anti-*B. burgdorferi* IgM but relatively low levels of IgG when they were compared by the in-house reference enzyme immunoassay. Low levels of anti-*B. burgdorferi* IgM and elevated levels of anti-*B. burgdorferi* IgG were detected in serum samples 92-07, 93-05, and 94-12. All other serum samples except sample 94-04 had elevated anti-*B. burgdorferi* IgM and IgG antibodies. Sample 94-04 had very low

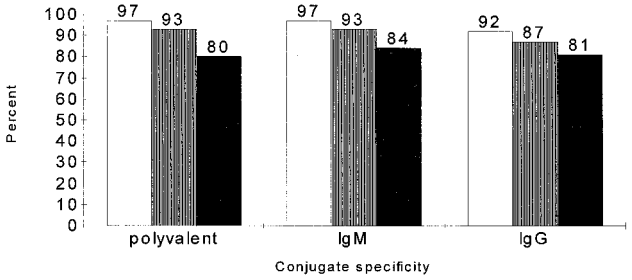

FIG. 1. Specificity of polyvalent, IgM, and IgG conjugate-specific tests for identification of serum samples from healthy donors for 1992 (n = 516), 1993 (n = 515), and 1994 (n = 496), as indicated by the bars from left to right, respectively.

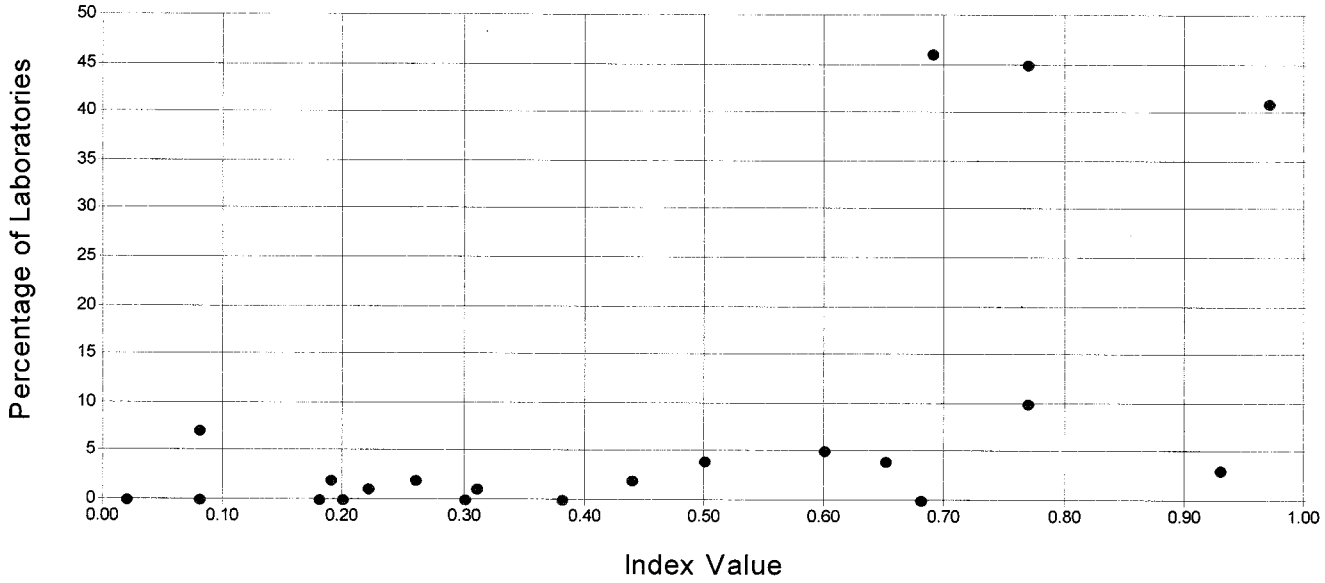

FIG. 2. Percentage of laboratories (n = 78) reporting positive IgM test results for each of the 22 serum samples from healthy donors with various index values. There were two serum samples with an index value of 0.77.

levels of anti-*B. burgdorferi* IgM and IgG antibodies. The patient was diagnosed with lumbar neuropathy and arthralgia. The IgG and IgM index values for case sera (Table 3) were significantly different ($P < 0.05$) than those for sera from donors without Lyme disease (Table 2).

Specificity. Three hundred eighty-eight laboratories reported results of assays in which a polyvalent conjugate was used to test 22 serum samples from healthy individuals. The specificity was 97% in 1992 and decreased to 93 and 80% in 1993 and 1994, respectively (Fig. 1). Similar results were obtained when participants used IgM or IgG conjugate-specific tests. The specificity of the IgM conjugate-specific tests reported by 78 laboratories decreased from 97% in 1992 to 84% in 1994. Likewise, the specificity of the IgG conjugate-specific tests used by 50 laboratories decreased from 92 to 81% during the survey period (Fig. 1).

Figure 2 presents the percentage of laboratories (n = 78) reporting positive IgM test results for each of the 22 serum samples from healthy donors that had various index values. The percentage of laboratories reporting false-positive results rapidly increased, with serum samples from healthy donors having an index value of 0.69 or more. A total of 46, 45, 10, and 41% of laboratories reported false-positive test results when serum samples from healthy donors had index values of 0.69, 0.77, 0.77, and 0.97, respectively. Furthermore, 3 to 7% of laboratories reported false-positive test results when serum samples from healthy donors had index values of 0.08, 0.50, 0.60, 0.65, and 0.93.

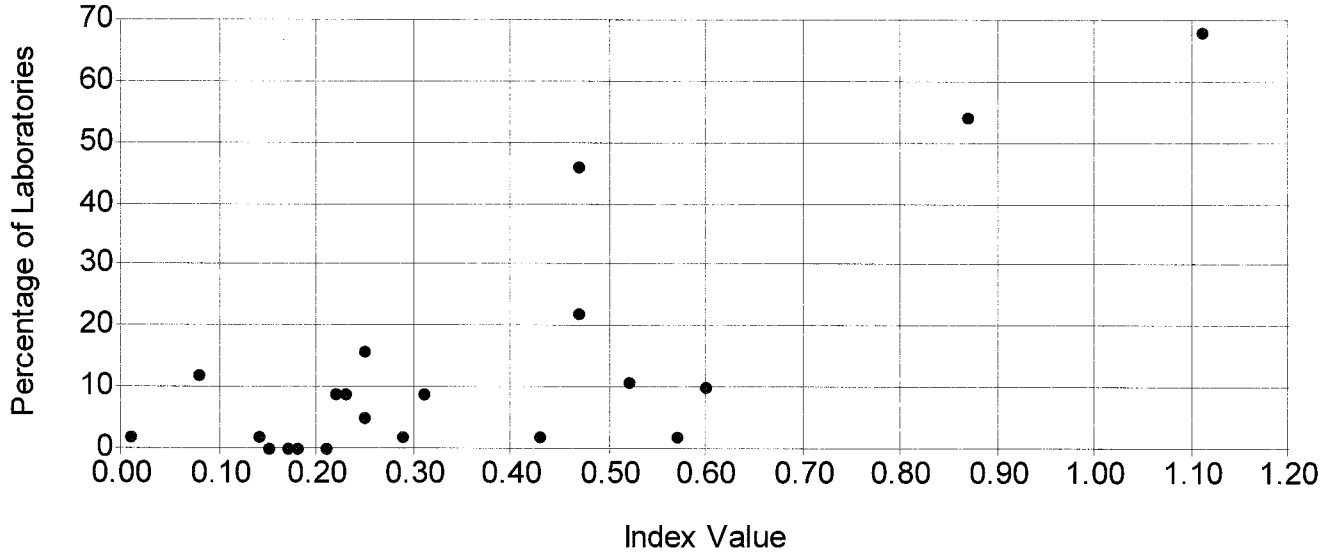

FIG. 3. Percentage of laboratories (n = 50) reporting positive IgG test results for each of the 22 serum samples from healthy donors with various index values. There were two serum samples with an index value of 0.25.

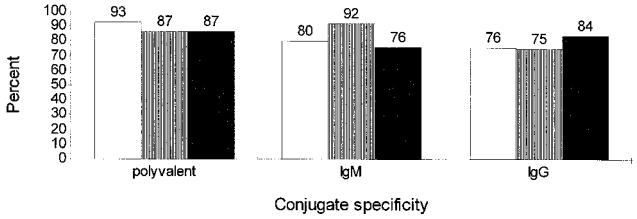

FIG. 4. Sensitivity of polyvalent, IgM, and IgG conjugate-specific tests for identification of CDC case-defined serum samples for 1992 (n = 516), 1993 (n = 515), and 1994 (n = 496), as indicated by the bars from left to right, respectively.

A similar pattern of false-positive test results occurred with laboratories performing IgG conjugate-specific tests (Fig. 3). A total of 12, 16, 22, 46, 11, 10, and 54% of laboratories reported false-positive results with serum samples from healthy donors with index values of 0.08, 0.25, 0.47, 0.47, 0.52, 0.60, and 0.87, respectively. Serum sample 93-03, which was obtained from a subject who was VDRL and FTA-ABS test positive, had an index value of 1.11. Sixty-eight percent of laboratories reported this serum sample to be positive for antibodies against *B. burgdorferi* (Fig. 3). In addition, laboratories that used polyvalent-specific tests reported false-positive test results at rates similar to those for the IgM and IgG conjugate-specific tests (data not shown).

Sensitivity. Twenty-eight CDC case-defined Lyme disease serum samples were tested by 388, 78, and 50 laboratories by polyvalent, IgM, or IgG conjugate-specific tests, respectively. The sensitivity varied from 87 to 93% when polyvalent conjugates were used for the years 1992, 1993, and 1994 (Fig. 4). When IgM conjugate-specific tests were evaluated, the sensitivity was 80% in 1992, peaked at 92% in 1993, and then declined to 76% in 1994. The sensitivity of the IgG conjugate-specific tests was 84% or less during the 3-year duration of the survey (Fig. 4).

Figure 5 illustrates the percentage of laboratories that reported positive IgM test results with each of the 28 CDC case-defined serum samples that had different index values. Ninety percent or more of the laboratories identified all serum samples except sample 94-10 with an index value of 1.50 or more by IgM conjugate-specific tests; serum sample 94-10 had an IgM index value of 2.62. By contrast, 8 of 12 of the serum samples with an index value of 1.50 or less were identified as positive by less than 90% of the laboratories. Furthermore, five of six of these serum samples with an IgM index value of 1.10 or less were reported to be positive for antibodies to *B. burgdorferi* by only 3 to 61% of the participants.

When IgG conjugate-specific tests were used (Fig. 6), more than 90% of the laboratories identified all serum samples except samples 93-09 and 94-03 as positive with an index value of 1.10 or more; serum samples 93-09 and 94-03 had index values of 1.78 and 1.07, respectively. Nine serum samples, however, with index values of 0.99 or less were identified by 75% of the laboratories or less. Specifically, serum samples 92-03, 92-12, 92-17, 92-19, 92-20, 93-01, 93-02, 93-10, and 94-04 were reported to be positive by 74, 48, 37, 57, 28, 45, 49, 49, and 36% of the laboratories, respectively. Similar test results were obtained when polyvalent conjugate tests were used (data not shown).

DISCUSSION

The serodiagnosis of Lyme disease has been plagued with problems of sensitivity and specificity (9–12, 14, 18) since the development of the first serologic test for the detection of antibodies against *B. burgdorferi* (17). Concomitantly, investigators (9, 14, 15) and medical organizations (5) increasingly called for the development of a national proficiency testing program as one means for improving the quality of test results. Initially, WSLH and CAP introduced Lyme disease proficiency programs for laboratories and manufacturers to evaluate the

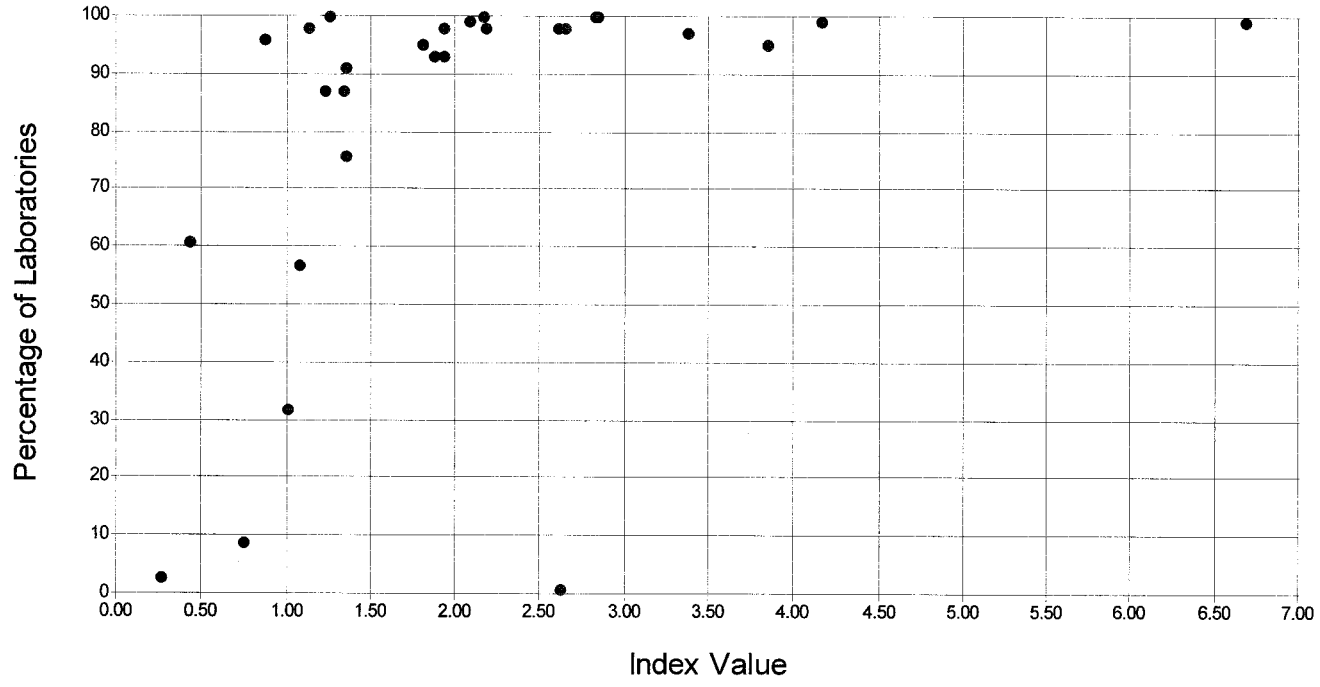

FIG. 5. Percentage of laboratories (n = 50) reporting positive IgM test results for each of the 28 CDC case-defined Lyme disease serum samples with various index values. There were two serum samples with an index value of 1.35.

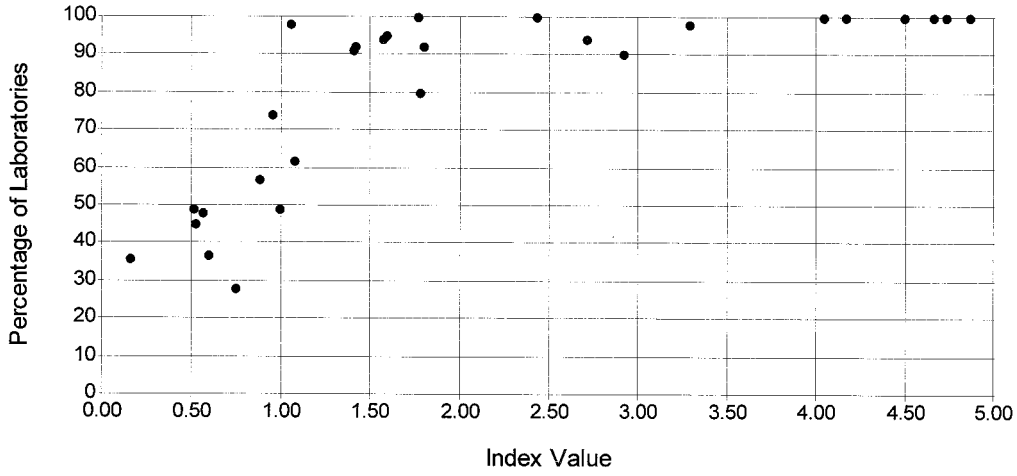

FIG. 6. Percentage of laboratories ($n = 78$) reporting positive IgG test results for each of the 28 CDC case-defined Lyme disease serum samples with various index values.

quality of their Lyme disease assays. These programs, specifically the WSLH Lyme Disease Proficiency Survey (1), affirmed that the serodiagnosis of Lyme disease needed improvement. Fourteen to 21% of laboratories failed to identify correctly a positive serum sample with high levels of antibodies against *B. burgdorferi* by using polyvalent or IgG conjugates. When serum samples contained lesser amounts of anti-*B. burgdorferi* antibody, accuracy varied by approximately 55%. Most disconcerting was the inability of many laboratories to reproduce their results. In addition, none of the commercial or home-brewed assays demonstrated a clear superiority.

Those results (1), however, were obtained from laboratories participating in WSLH's regional (Wisconsin) Lyme disease proficiency program. Furthermore, those results were obtained during a period of rapid commercialization of serodiagnostic kits, the use of a less stringent criterion for determination of case-defined serum, development of multiple home-brewed assays, poorer understanding of the antibody response against *B. burgdorferi*, and increased testing of patients without indications of Lyme borreliosis. Subsequently, the WSLH and CAP Lyme disease proficiency testing programs merged to offer a comprehensive national program to test approximately 500 laboratories. The cosponsored program now evaluates laboratories during a period of enhanced governmental regulation of commercially available kits, renewed awareness about the sensitivity and specificity of testing methods, and increased focus on quality assurance (5).

Unfortunately, our results demonstrated that the serodiagnosis of Lyme disease has not improved. We showed that the specificity of the Lyme disease assays steadily decreased from approximately 95% to approximately 81% during the 3 years of the survey. False-positive test results approached 55% with some of the serum samples from healthy donors. A serum sample containing antibody against *T. pallidum* was reported to be positive by 70% of the participants. In addition, the sensitivity fluctuated between 93 and 75%, depending upon the conjugates used by laboratories for the detection of anti-*B. burgdorferi* antibodies. In general, the sensitivity of the Lyme disease assays decreased during the duration of the survey. These results are discouraging and suggest that even more stringent criteria must be applied by the Food and Drug Administration for approving or continuing to approve commercially available kits for the serodiagnosis of Lyme disease.

The extent of the problem can be illustrated with the results obtained with some of the serum samples from healthy donors that had relatively high IgM index values. As the reference index value increased, the percentage of false-positive test results increased. Approximately 45% of laboratories reported false-positive test results when the IgM index values for three proficiency test serum samples were between 0.69 and 1.10. In addition, false-positive results increased for IgG tests when the index values were between 0.40 and 0.60. Ten to 54% of the laboratories reported false-positive IgG test results for five of seven serum samples within this range of index values. Fifty-four and 68% of laboratories reported false-positive IgG tests for two serum samples, respectively, that had index values of between 0.87 and 1.11. The false-reactivity of the IgM and IgG tests can only be explained for 1 of the 22 serum samples. This serum sample (sample 93-03) was reactive when tested by the VDRL and FTA-ABS tests. It is well known that infection with *T. pallidum* can be detected with Lyme disease assay systems (3, 16).

We also observed a steady decline in the sensitivity of the tests from 1992 through 1994. We found that the sensitivity of the IgM and IgG tests markedly declined when the index values for the serum samples were 1.50 or less. The decline in sensitivity was even greater with the IgG and IgM tests when the index values for the serum samples were 1.10 or less (1.10 was the cutoff value of our reference immunoassay). For example, nine serum samples with an index value of 1.10 or less were identified by only 74% or less (range, 36 to 74%) of the laboratories that used the IgG tests. Similarly, five of six serum samples with index values of 1.10 or less were reported to be positive by 61% or less (range, 3 to 61%) of the laboratories that used the IgM conjugates. Collectively, these results affirm our previous finding (1) that laboratories have the greatest difficulty in identifying serum samples with lower levels of *B. burgdorferi* antibody.

Our results also demonstrated that the performance of laboratories performing tests for Lyme disease can change drastically with the selection of serum samples. In 1995, serum samples with high index values for IgM and IgG antibodies were shipped to survey participants. Concurrently, serum samples from healthy donors were selected for the absence of cross-reactivity. The sensitivity and specificity of Lyme disease tests performed by survey participants increased to 95%. These serum samples clearly did not reflect the spectrum of patients with Lyme disease or their test results.

Under current proficiency testing regulations (7), 90% consensus is required to score each serum sample. If the Lyme disease survey were scored under these regulations, only samples with high levels of antibody would be evaluated. This gives the laboratory a false sense of security since the most difficult samples are not scored and laboratories are given full credit for those samples. As indicated by our results, the greater challenge of test performance is with samples containing low levels of antibody. Although *Clinical Laboratory Improvement Amendments of 1988* regulations require that proficiency testing programs provide a wide range of challenges, the scoring mechanism does not allow low levels of antibody to be appropriately evaluated when reference methods or standards do not exist. Thus, laboratories evaluate their own performance based on scientific evidence provided with the case history and scientific literature.

Although Lyme disease proficiency testing is not regulated, our results demonstrated that grading must be flexible to allow laboratories to be tested with samples that represent the spectrum of samples from patients with Lyme disease. Although the sensitivity and specificity will vary from challenge to challenge, it will reflect the current testing methods until a "gold standard" that can be used to compare or eliminate test methods that are inferior is developed. Again, commercially available tests appear to be excellent if high levels of anti-*B. burgdorferi* antibodies are present. They fail when the level of antibody to *B. burgdorferi* is low. A lack of sensitivity may not be a problem when the prevalence of Lyme disease is low; however, in areas where Lyme disease is endemic, this may prohibit the early detection of the disease. Furthermore, these tests will not be useful as screening tests unless their sensitivity is improved and standardized assays with high specificities are available to confirm the results.

Recently, CDC along with the Association of State and Territorial Public Health Laboratory Directors have recommended that sera submitted for serology be tested by an enzyme immunoassay or indirect fluorescent-antibody assay and that borderline or positive samples be tested by Western blotting (immunoblotting) (5, 8). Our data indicated that the sensitivity and specificity of the currently used tests for Lyme disease are not adequate to meet the two-tier test approach being recommended. Ideally, a screening test should have a high degree of sensitivity (>95%). The current methodologies need to be improved to adequately screen serum samples for confirmatory testing.

In conclusion, our results suggest that stronger measures need to be taken by the Food and Drug Administration to control the quality of commercially available Lyme disease assay kits. One solution is to force all currently used or approved assay systems through an evaluation with sera obtained from culture-positive patients. An arbitrary sensitivity of 90% could be selected. The specificity could also be set at 95%. Commercially available tests that did not reach these performance levels would be removed from the market. Although this is a drastic solution, its time has come. This measure would lessen the overdiagnosis of Lyme disease (19) and prevent the use of costly laboratory solutions, like Western immunoblotting, to make a serodiagnosis of Lyme disease.

APPENDIX

The other members of the Diagnostic Immunology Resource Committee are as follows: Michael J. Borowitz, Department of Pathology, Johns Hopkins Medical Institution, Baltimore, Md.; John S. Coon, Department of Pathology, Rush-Presbyterian Medical College, Chicago, Ill.; Robert Hamilton, Johns Hopkins Asthma and Allergy, Baltimore, Md.; Sandra Larsen, Treponemal Pathogenesis and Immunology Branch, Centers for Disease Control, Atlanta, Ga.; Helene Paxton, Maryland Medical Laboratories, Baltimore, Md.; Rebecca L. Johnson, Department of Pathology, Berkshire Medical Center, Pittsfield, Mass.; David J. Bylund, Box 2701, Rancho Santa Fe, Calif.; Charles A. Horwitz, Department of Pathology, Abbott Northwestern Hospital, Minneapolis, Minn.; David F. Keren, Department of Pathology, Warde Medical Laboratory, Ann Arbor, Mich.; Russell H. Tomar, Departments of Pathology and Laboratory Medicine, University of Wisconsin Hospital and Clinics, Madison, Wis.; Felice Miller Soifer, Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, N.Y.; Henry A. Homburger, Clinical Immunology Laboratory, Mayo Clinic, Rochester, Minn.; and Robert M. Nakamura, Department of Pathology, Scripps Clinic and Research Foundation, La Jolla, Calif.

ACKNOWLEDGMENTS

We thank the staff of the Wisconsin State Laboratory of Hygiene proficiency testing section, particularly Pamela Rogers and Rosemary Graesslin, for assistance in preparing survey samples and providing data for this study. We also thank Kay L. Case for supplying sera.

REFERENCES

1. **Bakken, L. L., K. L. Case, S. M. Callister, N. J. Bourdeau, and R. F. Schell.** 1992. Performance of 45 laboratories participating in a proficiency testing program for Lyme disease serology. JAMA **268:**891–895.
2. **Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J. C. Piffaretti, M. Assous, and P. A. D. Grimont.** 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. Int. J. Syst. Bacteriol. **42:**378–383.
3. **Bruckbauer, H. R., V. Preac-Mursic, R. Fuchs, and B. Wilske.** 1992. Cross-reactive proteins of *Borrelia burgdorferi*. Eur. J. Clin. Microbiol. Infect. Dis. **11:**224–232.
4. **Callister, S. M., K. L. Case, W. A. Agger, R. F. Schell, R. C. Johnson, and J. L. E. Ellingson.** 1990. Effects of bovine serum albumin on the ability of Barbour-Stoenner-Kelly medium to detect *Borrelia burgdorferi*. J. Clin. Microbiol. **28:**363–365.
5. **Centers for Disease Control.** 1990. Case definitions for public health surveillance. Morbid. Mortal. Weekly Rep. **39:**19–21.
6. **Centers for 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.
7. **Clinical Laboratory Improvement Amendments of 1988.** 1988. An Act to amend the Public Health Service Act to revise the Authority for the Regulation of Clinical Laboratories. Government Printing Office, Washington, D.C.
8. **Dressler, F., J. A. Whalen, B. N. Reinhardt, and A. C. Steere.** 1993. Western blotting in the serodiagnosis of Lyme disease. J. Infect. Dis. **167:**392–400.
9. **Hedberg, C. W., and M. T. Osterholm.** 1990. Serologic tests for antibody to *Borrelia burgdorferi*: another Pandora's box for medicine? Arch. Intern. Med. **150:**732–733.
10. **Hedberg, C. W., M. T. Osterholm, K. L. MacDonald, and K. E. White.** 1987. An interlaboratory study of antibody to *Borrelia burgdorferi*. J. Infect. Dis. **155:**1325–1327.
11. **Jones, J. M.** 1991. Serodiagnosis of Lyme disease. Ann. Intern. Med. **114:** 1064.
12. **Lane, R. S., E. T. Lennette, and J. E. Madigan.** 1990. Interlaboratory and intralaboratory comparison of indirect immunofluorescence assays for serodiagnosis of Lyme disease. J. Clin. Microbiol. **28:**1774–1779.
13. **Lovrich, S. D., S. M. Callister, L. C. L. Lim, B. K. Duchateau, and R. F. Schell.** 1994. Seroprotective groups of Lyme borreliosis spirochetes from North America and Europe. J. Infect. Dis. **170:**115–121.
14. **Luger, S. W., and E. Krauss.** 1990. Serologic tests for Lyme disease: interlaboratory variability. Arch. Intern. Med. **150:**761–816.
15. **Magnarelli, L. A.** 1989. Quality of Lyme disease tests. JAMA **262:**3464–3465.
16. **Magnarelli, L. A., J. F. Anderson, and R. C. Johnson.** 1987. Cross-reactivity in serological tests for Lyme disease and other spirochetal infections. J. Infect. Dis. **156:**183–188.
17. **Russell, H., J. S. Sampson, G. P. Schmid, H. W. Wilkinson, and B. Plikaytis.** 1984. Enzyme-linked immunosorbent assay and indirect immunofluorescence assay for Lyme disease. J. Infect. Dis. **149:**465–470.
18. **Schwartz, B. S., M. D. Goldstein, J. M. C. Ribeiro, T. L. Schulze, and S. I. Shahied.** 1989. Antibody testing in Lyme disease: a comparison of results in four laboratories. JAMA **262:**3431–3434.
19. **Steere, A. C., R. L. Grodzicki, A. N. Kornblatt, J. E. Craft, A. G. Barbour, W. Burgdorfer, G. P. Schmid, E. Johnson, and S. E. Malawista.** 1983. The spirochetal etiology of Lyme disease. N. Engl. J. Med. **308:**733–740.
20. **Steere, A. C., E. Taylor, G. L. McHugh, and E. L. Logigian.** 1993. The overdiagnosis of Lyme disease. JAMA **269:**1812–1816.