

Intra- and interlaboratory reproducibility of an ELISA serological test for Lyme disease

MARTIN C TAMMEMAGI BSc DVM MSc, JOHN W FRANK BSc MD CCFP MSc FRCPC,
MICHAEL LEBLANC BSc MMath PhD, HARVEY ARTSOB PhD

MC TAMMEMAGI, JW FRANK, M LEBLANC, H ARTSOB. Intra- and interlaboratory reproducibility of an ELISA serological test for Lyme disease. *Can J Infect Dis* 1995;6(2):90-95.

OBJECTIVES: Lyme disease has been increasingly diagnosed throughout North America since the late 1970s. The clinical diagnosis and epidemiological monitoring of Lyme disease are aided by serological testing for the etiological agent, *Borrelia burgdorferi*. Numerous authorities have questioned the reproducibility of these serological tests. This study assessed the intra- and interlaboratory reproducibility of an ELISA used to aid in the diagnosis of Lyme disease.

METHODS: Twenty-seven sera from cases and noncases were tested by three laboratories. Two of the laboratories repeated the tests once. These testings were part of the 1991 quality control assessment of provincial laboratories carried out by the Laboratory Centre for Disease Control (LCDC), Ottawa.

RESULTS: The mean weighted kappa statistics were 0.87 for interlaboratory comparisons and 0.89 for intralaboratory comparisons.

CONCLUSIONS: Overall, the ELISA assessed in this study demonstrated good to excellent intra- and interlaboratory reproducibility in the LCDC 1991 quality control assessment when the data were assessed in the categorical scale using the weighted kappa statistic. Generalization of these findings to clinical laboratory settings must be done with caution.

Key Words: ELISA, Lyme borreliosis, Lyme disease, Reliability, Reproducibility, Serology

Reproductibilité d'un test sérologique ELISA pour le dépistage de la maladie de Lyme entre les laboratoires et à l'intérieur des laboratoires

OBJECTIFS : La maladie de Lyme est de plus en plus diagnostiquée partout en Amérique du Nord depuis la fin des années 1970. Le diagnostic clinique et la surveillance épidémiologique de la maladie de Lyme sont facilités par des épreuves sérologiques pour le dépistage de l'agent étiologique, *Borrelia burgdorferi*. De nombreuses autorités ont mis en doute la reproductibilité de ces épreuves sérologiques. Cette étude a évalué la reproductibilité d'un test ELISA utilisé dans le diagnostic de la maladie de Lyme, entre différents laboratoires et à l'intérieur des laboratoires.

MÉTHODES : Vingt-sept échantillons de sérum provenant de cas et de non-cas ont été testés par trois laboratoires. Deux des laboratoires ont répété les tests une fois. Ces épreuves faisaient partie d'une évaluation du contrôle de la qualité de 1991 des laboratoires provinciaux menée par le Laboratoire de lutte contre la maladie d'Ottawa.

voir page suivante

Graduate Department of Community Health, and Department of Preventative Medicine and Biostatistics, University of Toronto, Toronto, and Laboratory Centre for Disease Control, Ottawa, Ontario

Correspondence and reprints: Dr M Tammemagi, Graduate Department of Community Health, Fourth Floor, McMurrich Building, University of Toronto, Toronto, Ontario M5S 1A1. Telephone (416) 978-2058, Fax (416) 978-1883, e-mail mtammem@pmb.med.utoronto.ca

Received for publication June 16, 1994. Accepted September 21, 1994

RÉSULTATS : Les données sur les moyennes kappa pondérées étaient de 0,87 pour les comparaisons entre les laboratoires et de 0,89 pour les comparaisons à l'intérieur des laboratoires.

CONCLUSIONS : De façon globale, les épreuves ELISA évaluées dans cette étude ont démontré une reproductibilité de bonne à excellente entre les laboratoires et dans les laboratoires, dans le cadre de l'évaluation du contrôle de la qualité de 1991 du Laboratoire de lutte contre la maladie, lorsque les données ont été évaluées dans une échelle de catégories utilisant les moyennes kappa pondérées. La généralisation de ces résultats au contexte des laboratoires cliniques ne peut se faire sans exercer une certaine prudence.

LYME DISEASE IS A MULTISYSTEMIC TICK-BORNE BACTERIAL INFECTION of humans and animals caused by the spirochete *Borrelia burgdorferi*. Lyme disease has spread rapidly across the United States and has increased dramatically in incidence since its first recognition in the mid-1970s (1). Clinical diagnosis and public health surveillance are aided by the detection of antibodies against *B burgdorferi*. The accuracy and reproducibility of available serological tests have not been demonstrated convincingly and have been questioned by many authorities (2-10).

It has been recommended by a consensus conference on Lyme disease that all laboratories in Canada use a commercial ELISA for initial testing of patients' sera for antibodies to *B burgdorferi* (11). Since 1990, an annual national quality control assessment (QCA) of provincial diagnostic laboratories that test for Lyme disease has been carried out by the Laboratory Centre for Disease Control (LCDC) in Ottawa. Some Canadian laboratories employ the Whittaker Lyme Stat (WLS) ELISA Kit (Whittaker Bioproducts, Maryland) to measure immunoglobulin (Ig) G and IgM antibodies in human serum. This paper reports the results of a study of the inter- and intralaboratory reproducibility of the WLS ELISA used by three laboratories in the LCDC 1991 QCA.

MATERIALS AND METHODS

In May 1991, the LCDC conducted a QCA of eight provincial laboratories that perform tests to detect antibodies to *B burgdorferi*. Three provincial laboratories used the WLS ELISA, while the other laboratories used alternative tests. The results of only two provincial laboratories are reported here, because one set of sera was spoiled. The sera of 27 individuals were included in this study. The sera were diluted 1 in 10 parts of phosphate buffered saline, and each laboratory was provided 0.24 mL aliquots. Sera were shipped frozen on dry ice and the LCDC maintained frozen aliquots for comparative testing. All laboratories tested the samples blind to the putative 'gold standard' status of the sera except for the LCDC, and all laboratories tested the sera blind to the outcomes from the other laboratories. Laboratory 6 and the LCDC used kits from the same lot for their repeat testing, but the kit lots differed between laboratory 6 and the LCDC. The lot number for laboratory 2 was not available. Due to a scarcity of material, Western blots were not performed by these laboratories.

Sera in the sample came from both infected and uninfected individuals, and represented as close to bona fide 'gold standard' cases and noncases as could practically be obtained. The 'gold standard' status was based on a detailed clinical and diagnostic work-up (including Western blots) by

experts in the field. Sixteen sera were considered positive. These patients had been exposed in endemic areas, had clinical manifestations characteristic of Lyme disease, and experts (LCDC and/ or Centers for Disease Control and Prevention, Fort Collins, Colorado) felt reasonably certain about their status. In many positive cases the pathognomonic sign, erythema chronicum migrans, was observed or the etiological agent was detected bacteriologically. Patient background and source of sera are presented in Table 1. All 11 negative sera came from individuals living in nonendemic areas and were considered by expert opinion to represent individuals free from Lyme disease. Five negative and 12 positive sera were obtained from the Centers for Disease Control and Prevention at Fort Collins. The serum from one of these patients was consistently nonreactive on serological testing despite the fact that *B burgdorferi* had been isolated from the patient. Thus, the maximum of serological reactors expected was 15. Five of the 11 negative sera came from healthy patients and six came from patients who were thought to have a higher likelihood of being false reactors (9,12-14): three from syphilis patients, two from patients who were antinuclear antibody-positive and one from a patient who was rheumatoid-positive.

The WLS test has its colorimetric outcome measured optically. The diagnostic significance of the optical density or absorbance is not interpreted directly, but rather the reading is converted into a predictive index value (PIV), which is a standardized score that is then interpreted as a positive, equivocal or negative result. The absorbance readings for a set of results derived from one test kit are transformed into PIVs (diagnostic scores) by use of a calibration (regression) line obtained by measuring the absorbance of three standards provided in each test kit and fixing those absorbance readings to predetermined index values called labelled index values. Although the labelled index values may vary slightly among kits, they are set so that the PIVs among kits in different laboratories and at different times are always interpreted in the same way. A PIV below 0.80 is considered seronegative; equal to or greater than 1.00 is considered positive; and in between these values is considered an equivocal result.

The 27 sera were tested twice by the LCDC, once by laboratory 2, and twice by laboratory 6. This allowed the assessment of interlaboratory reproducibility to be made among three laboratories, and intralaboratory reproducibility in two different laboratories. Intra- and interlaboratory reproducibility were assessed by computation of the weighted kappa statistic (k_w) (15-17) for the data in the categorical scale, that is, through use of the diagnostic classifications of the PIVs. An unweighted kappa statistic for multiple raters and multiple

TABLE 1
The panel of sera used in the LCDC 1991 quality control assessment

Serum number	Status	History supplied with serum	Received from
1	+	Lyme disease patient from New York	CDC
2	-	Normal	CDC
3	-	Normal	CDC
4	-	Normal	CDC
5	-	Normal	CDC
6	-	Normal	CDC
7	+	Lyme disease patient from New York	CDC
8	+	Lyme disease patient from Illinois	CDC
9	+	Lyme disease patient from Wisconsin	CDC
10	+*	Lyme disease patient from Germany via Wisconsin	CDC
11	+	Lyme disease patient from Wisconsin	CDC
12	+	Lyme disease patient from Illinois	CDC
13	+	Lyme disease patient from Connecticut	CDC
14	+	Lyme disease patient from Connecticut	CDC
15	+	Lyme disease patient from Connecticut	CDC
16	+	Lyme disease patient from Massachusetts	CDC
17	+	Lyme disease/syphilis patient from New York	CDC
18	-	Syphilis patient	LCDC
19	-	Syphilis patient	LCDC
20	+	Lyme disease via Quebec laboratory [†]	LSPQ
21	-	Syphilis patient	LCDC
22	+	Lyme disease via Quebec laboratory [†]	LSPQ
23	+	Lyme disease patient via Saskatchewan laboratory [†]	SPHL
24	-	Antinuclear antibody-positive	Ottawa Civic Hospital
25	-	Antinuclear antibody-positive	Ottawa Civic Hospital
26	-	Rheumatoid positive	Ottawa General Hospital
27	+	Lyme disease patient from Switzerland	Virion Diagnostics

CDC Centers for Disease Control and Prevention, Fort Collins, Colorado; LCDC Laboratory Centre for Disease Control, Ottawa; LSPQ Laboratoire de Santé Publique du Québec, Ste Anne de Bellevue; SPHL Saskatchewan Public Health Laboratory, Regina. All Lyme disease sera from CDC were from clinically characterized Lyme disease patients; Normal is serum from healthy donors in nonendemic areas; *Bacteriologically confirmed Lyme disease but negative in serological tests; Patient 17 clinically fit Lyme disease but serological tests were more characteristic of tertiary or latent syphilis; [†]Patients had histories of travel to recognized Lyme-endemic areas

categories (16) was computed for interlaboratory comparisons. Only one set of results was included from each laboratory in computation of the latter statistic. From the two sets of results provided by the LCDC and laboratory 6, the least reproducible set was used, so that the estimate of the overall kappa would tend to be conservative. The sensitivity and specificity were calculated via the method described by Poynard *et al* (18) and Simel *et al* (19) for tests with trichotomously categorized results. Essentially, this method excludes intermediates from the calculation. Sensitivity is the ratio of true positives:(true positives + false negatives). Specificity is the ratio of true negatives:(true negatives + false positives).

RESULTS

Test results and the diagnostic interpretation for each patient are presented in Table 2. The sensitivity for the different sets of results ranged from 0.88 to 0.94 (mean \pm SD 0.91 \pm 0.04) and the specificity ranged from 0.82 to 1.0 (0.96 \pm 0.08). In two sets of results there were no equivocal results. The remaining three sets of results had only one equivocal result each, and the patient was different in each set of results.

Ten different pair-wise comparisons could be made from the five sets of results: eight interlaboratory and two intralaboratory. The weighted kappa statistics for the different comparisons are reported in Table 3. The weighted kappa statistics for the 10 comparisons range from 0.74 to 0.96. The intralaboratory weighted kappa statistics were 0.96 and 0.81 (mean 0.89). The weighted kappa statistics for the eight interlaboratory comparisons range from 0.74 to 0.93 (mean \pm SD 0.87 \pm 0.07). The overall unweighted kappa for multiple raters and multiple categories for the three laboratories was 0.80 (95% CL 0.70, 0.91).

DISCUSSION

Past studies of Lyme serological tests have concluded that these tests lacked reproducibility (8-10), but questionable statistical methods and study designs were used. For example, some studies used measures of trend or association, not concordance, to appraise agreement, such as Spearman's rank correlation coefficient (9) or Wilcoxon's signed rank test (10). Wilcoxon's rank test measures for a difference in the means between two sets of results, but it is possible for two sets of results to have very discrepant results and yet have similar

TABLE 2
Results and interpretation of results of the Whittaker Lyme Stat ELISA for the 27 sera in the LCDC 1991 quality control assessment

Serum number (diagnosis)	Predicted index value and interpretation				
	LCDC first test	LCDC second test	Laboratory 2	Laboratory 6 first test	Laboratory 6 second test
1 (+)	0.88, E	1.40, R	0.75, N	1.64, R	0.39, N
2 (-)	-0.60, N	-0.66, N	0.06, N	-0.36, N	-0.70, N
3 (-)	-0.81, N	-0.82, N	0.00, N	-0.91, N	-1.28, N
4 (-)	-0.62, N	-0.68, N	0.08, N	-0.81, N	-0.97, N
5 (-)	-0.75, N	-0.93, N	0.00, N	-0.83, N	-1.25, N
6 (-)	-0.63, N	-0.65, N	0.16, N	-0.36, N	-0.87, N
7 (+)	5.64, R	10.00, R	4.41, R	9.15, R	7.86, R
8 (+)	2.71, R	3.60, R	2.04, R	4.77, R	5.21, R
9 (+)	5.00, R	6.98, R	3.54, R	7.18, R	8.23, R
10 (+)	-0.67, N	-0.15, N	0.08, N	0.06, N	0.26, N
11 (+)	2.77, R	3.21, R	2.73, R	2.55, R	2.26, R
12 (+)	1.33, R	1.60, R	0.83, E	2.55, R	2.04, R
13 (+)	7.57, R	13.22, R	5.26, R	12.19, R	9.61, R
14 (+)	4.82, R	7.43, R	2.96, R	6.19, R	5.52, R
15 (+)	5.12, R	8.72, R	3.49, R	7.21, R	6.00, R
16 (+)	9.65, R	16.88, R	6.17, R	14.69, R	12.67, R
17 (+)	3.03, R	5.59, R	1.92, R	5.08, R	5.54, R
18 (-)	0.22, N	0.51, N	0.56, N	1.13, R	0.65, N
19 (-)	0.03, N	0.56, N	0.67, N	1.68, R	0.89, E
20 (+)	1.79, R	2.74, R	1.20, R	3.38, R	2.45, R
21 (-)	-0.46, N	-0.53, N	0.00, N	-0.74, N	-0.99, N
22 (+)	7.67, R	15.75, R	3.96, R	14.31, R	10.89, R
23 (+)	2.83, R	4.31, R	1.94, R	6.87, R	5.12, R
24 (-)	-0.70, N	-0.61, N	0.00, N	-0.91, N	-1.22, N
25 (-)	-0.87, N	-0.83, N	0.00, N	-1.08, N	-1.44, N
26 (-)	-0.65, N	-0.68, N	0.03, N	-0.62, N	-0.90, N
27 (+)	6.76, R	13.53, R	4.22, R	13.73, R	11.17, R

E Equivocal; *LCDC* Laboratory Centre for Disease Control; *N* Negative; *R* Reactive or positive

TABLE 3
Weighted kappa statistics for the different comparisons made for the repeat Whittaker Lyme Stat ELISA in the LCDC 1991 quality control assessment

	LCDC second test	Laboratory 2	Laboratory 6 first test	Laboratory 6 second test
LCDC first test	0.963 (0.589, 1.0)	0.926 (0.562, 1.0)	0.811 (0.434, 1.0)	0.925 (0.559, 1.0)
LCDC second test		0.889 (0.519, 1.0)	0.847 (0.458, 1.0)	0.888 (0.515, 1.0)
Laboratory 2			0.741 (0.371, 1.0)	0.926 (0.562, 1.0)
Laboratory 6 first test				0.811 (0.434, 1.0)

k_w weighted kappa (95% CL); *LCDC* Laboratory Centre for Disease Control

means. Biased samples (9) or restricted samples (8) were used in some reproducibility studies. Also, some studies failed to deal with intermediate or equivocal results (8,9), grouping these results together with either positive results (8) or negative results (9). Some studies were not conducted blind (8).

For clinical purposes the PIVs are interpreted in the ordinal scale as negative, equivocal or positive. The weighted kappa statistic is considered by many authorities to be the measure of reproducibility of choice for such ordinal data (16,17). It not only measures concordance but also takes into consideration chance agreement, and it weights the degree of disagree-

ment. A kappa statistic of 1 reflects perfect agreement, of 0 indicates random agreement, and of less than 0 suggests disagreement. The range of values encountered in this study are generally believed to reflect good to excellent agreement (16,20).

As the test results, the PIVs, naturally occur in the continuous form, why was the analysis of reproducibility not carried out with the data in the continuous scale, via methods such as the intraclass correlation coefficient (21-23) or Bland and Altman's limits of agreement (24-26)? Measures of reproducibility with the data in the continuous form may overemphasize

lack of reproducibility at extremes where the exact result does not matter diagnostically and may underemphasize reproducibility around important diagnostic interpretive cut-points. Because medical decision-making is based on the interpretation of the PIV and not the particular value itself, ordinal assessment of reproducibility is appropriate here.

The overall unweighted kappa for multiple raters and multiple categories for the three laboratories demonstrates excellent reproducibility ($k_w=0.80$). The narrower 95% CLs (0.70, 0.91) compared with the pair-wise weighted kappa is in large part due to an increase in sample size. The interlaboratory reproducibility (mean $k_w=0.89$) did not appear to be appreciably worse than the intralaboratory reproducibility (mean $k_w=0.87$) as might have been expected, but it is unlikely that this study had the power to demonstrate a difference if it existed. Also, lack of intralaboratory blindness to the first set of results does not appear to be a biasing factor because results are remarkably uniform across and within laboratories. Interlot reproducibility as reflected in comparisons of results between the LCDC and laboratory 6 appears high (mean $k_w=0.87$).

It is somewhat reassuring that the number of equivocal results was small (range 0 to 3.7%), because these results often present more of a diagnostic conundrum than clearcut positive or negative cases. No consistent pattern could be detected indicating whether certain individuals were more likely to have equivocal results or equivocal outcomes occurred more in true positives than true negatives, or vice versa. Given the sample size, this study lacks the power to detect such patterns if they exist.

CONCLUSIONS

As measured by the weighted kappa statistic, the WLS ELISA demonstrated good to excellent inter- and intralaboratory reproducibility in the LCDC 1991 QCA. Ideally, a test should be tested in a sample that is similar to the population in which it will be used. In the case of rare diseases, such as Lyme disease, this is impractical because an extraordinarily large sample size is required to obtain an adequate number of true positive cases. The LCDC 1991 QCA is a much more heterogeneous sample, with many more positive cases, than is encountered in the homogeneous population tested in public diagnostic laboratories in which most sera test negative. The variance of the study sample results are at least four times greater than those observed in the Ontario Provincial Laboratory (27). Increased heterogeneity inflates the kappa statistic and leads to an overestimate of reproducibility. When marginal proportions are unbalanced, as would occur for results obtained in most North American laboratories, the maximum achievable magnitude of kappa drops to well below 1 (28,29).

Users of the WLS test may therefore feel fairly confident that it is relatively reproducible by the standards applied to most laboratory tests. Laboratory factors that can improve and maintain the reproducibility of Lyme serological tests are discussed by Duffey and Salugsugan (30). However, as others have pointed out (31,32), this high reliability does not translate into high validity. In low prevalence situations the positive predictive value of the test may still be quite low, and

this has led to the practice of confirming positive ELISAs with Western Blot tests, which are believed to be more specific. Deficiencies of Western Blot tests as confirmatory tests for Lyme disease have recently been pointed out (33). It is therefore important to emphasize that the diagnosis of Lyme disease is primarily based on appropriate clinical presentation, with laboratory tests providing backup (33).

ACKNOWLEDGEMENTS: This research was supported in part by the National Health Research and Development Program through a National Health Fellowship to Dr Martin Tammemagi.

REFERENCES

1. Steere AC, Malawista SE, Snyderman DR, et al. Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. *Arthritis Rheum* 1977;20:7-17.
2. Osterholm MT, Forfang JC, White KE, Kuritsky JN. Lyme disease in Minnesota: epidemiology and serologic findings. *Yale J Biol Med* 1984;57:677-83.
3. Wilkinson HW, Russell H, Sampson JS. Caveats on using nonstandardized serologic tests for Lyme disease. *J Clin Microbiol* 1985;21:291. (Lett)
4. Shrestha M, Grodzicki RL, Steere AC. Diagnosing early Lyme disease. *Am J Med* 1985;78:235.
5. Craft JE, Grodzicki RL, Steere AC. Antibody response in Lyme disease: evaluation of tests. *J Infect Dis* 1984;149:789-95.
6. Grodzicki RL, Steere AC. Comparison of immunoblotting and indirect enzyme-linked immunosorbent assay using different antigen preparations for diagnosing early Lyme disease. *J Infect Dis* 1988;157:790-7.
7. Hansen K, Hindersson P, Pedersen NS. Measurement of antibodies to the *Borrelia burgdorferi* flagellum improves serodiagnosis in Lyme disease. *J Clin Microbiol* 1988;26:338-46.
8. Luger S, Krause E. Serologic tests for Lyme disease: inter-laboratory variability. *Arch Intern Med* 1990;150:761-3.
9. Schwarz BS, Goldstein MD, Ribeiro JMC, Schulze TL, Shahied SI. Antibody testing in Lyme disease. A comparison of results in four laboratories. *JAMA* 1989;262:3431-4.
10. Hedberg CW, Osterholm MT, MacDonald KL, White KE. An inter-laboratory study of antibody to *Borrelia burgdorferi*. *J Infect Dis* 1987;156:1325-7.
11. Consensus conference on Lyme disease. *Can J Infect Dis* 1991;2:49-54.
12. Magnarelli LA, Anderson JF, Johnson RC. Cross-reactivity in serological tests for Lyme disease and other spirochetal infections. *J Infect Dis* 1987;156:183-8.
13. Magnarelli LA, Anderson JF. Enzyme-linked immunosorbent assays for the detection of class-specific immunoglobulins to *Borrelia burgdorferi*. *Am J Epidemiol* 1988;127:818-25.
14. Golightly MG, Thomas JA, Viciano AL. The laboratory diagnosis of Lyme borreliosis. *Lab Med* 1990;21:299-304.
15. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;70:213-20.
16. Fleiss JL. *Statistical Methods for Rates and Proportions*, 2nd edn. New York: John Wiley & Sons, 1982.
17. Kramer MS, Feinstein AR. The biostatistics of concordance. *Clin Pharmacol Ther* 1981;29:111-23.
18. Poynard T, Chaput JC, Etienne JP. Relations between effectiveness of a diagnostic test, prevalence of the disease, and percentages of uninterpretable results – an example in the diagnosis of jaundice. *Med Decis Making* 1982;2:287-97.
19. Simel DL, Feussner JR, Delong ER, Matchar DB. Intermediate,

- indeterminate, and uninterpretable diagnostic results. *Med Decis Making* 1987;7:107-14.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
 21. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966;19:3-11.
 22. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York: John Wiley & Sons, 1986.
 23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
 24. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 1990;20:337-40.
 25. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983;32:307-17.
 26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-10.
 27. Tammemagi MC. "A Study of Agreement Statistics for Lyme Disease Serology." Masters Thesis, University of Toronto, 1992.
 28. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
 29. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551-8.
 30. Duffey PS, Salugsugan J. Serodiagnosis of Lyme borreliosis. *Clin Microbiol Newsletter* 1993;15:81-6.
 31. Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology, the Essentials*, 2nd edn. Baltimore: Williams & Wilkins, 1988.
 32. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston: Little, Brown, 1987.
 33. Artsob A. Western Blot as a confirmatory test for Lyme disease. *Can J Infect Dis* 1993;4:115-6.
-
-



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

