



Multicenter Clinical Evaluation of Modified Two-Tiered Testing Algorithms for Lyme Disease Using Zeus Scientific Commercial Assays

Maroun M. Sfeir,^a Jennifer K. Meece,^b  Elitza S. Theel,^c Dane Granger,^c Thomas R. Fritsche,^b Allen C. Steere,^d  John A. Branda^d

^aUConn Health, Farmington, Connecticut, USA

^bMarshfield Clinic Health System, Marshfield, Wisconsin, USA

^cDivision of Clinical Microbiology, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA

^dMassachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

Maroun M. Sfeir and Jennifer K. Meece contributed equally to this article. Author order was determined by drawing straws.

ABSTRACT Modified two-tiered testing (MTTT) algorithms for Lyme disease (LD), which involve the sequential use of orthogonal enzyme immunoassays (EIAs) without immunoblotting, are acceptable alternatives to standard two-tiered testing (STTT; EIA followed by immunoblots) provided the EIAs have been FDA-cleared for this intended use. We evaluated four Zeus Scientific LD EIAs used in two distinct MTTT algorithms for FDA review. MTTT 1 used a VlsE1/pepC10 polyvalent EIA followed by a whole-cell sonicate (WCS) polyvalent EIA. MTTT 2 used the same first-tier EIA followed by separate IgM and IgG WCS EIAs. In a retrospective phase, we compared each MTTT algorithm to STTT using archived samples from LD patients or control subjects. In a prospective phase, we used the same algorithms to analyze consecutive excess samples submitted for routine LD serology to three clinical laboratories. For the retrospective phase, MTTTs 1 and 2 were more sensitive (56% and 74%) than STTT (41%; $P \leq 0.03$) among 61 patients with acute erythema migrans (EM). In LD patients with neuroborreliosis, carditis, or arthritis ($n = 75$), sensitivity was comparable between algorithms (96 to 100%; $P = 1.0$). Among 190 control subjects without past LD, all algorithms were highly and comparably specific ($\geq 99\%$, $P = 0.48$). For the prospective phase, ($n = 2,932$), positive percent-agreement (PPA), negative percent-agreement (NPA), and overall agreement of MTTT 1 with STTT were 93%, 97.7% and 97.4% (kappa 0.80). MTTT 2 yielded higher PPA (98%) but lower NPA (96.1%) and overall agreement (96.2%, kappa 0.74; all $P < 0.05$). Compared with STTT, both MTTT algorithms provided increased sensitivity in EM patients, comparable sensitivity in later disease and non-inferior specificity.

KEYWORDS *Borrelia*, *Borrelia burgdorferi*, Lyme disease, diagnostics, immunodiagnostics, serology

Lyme disease is a multisystem infection caused by the tick-transmitted spirochete *Borrelia burgdorferi sensu lato*. Serologic testing is the mainstay laboratory diagnostic approach, and two-tiered testing is required to optimize specificity (1).

Standard two-tiered testing (STTT) for Lyme disease involves a first-tier enzyme immunoassay (EIA) followed (if the first-tier test is reactive) by IgM- and IgG-specific immunoblots interpreted according to specified criteria (2). Recently, several studies have demonstrated the clinical validity, advantages and limitations of an alternative approach termed modified two-tiered testing (MTTT) (3–5). Importantly, we and others have shown that MTTT for Lyme disease often is more sensitive in early infection compared with STTT, and specificity is non-inferior (5–13).

Editor Patricia J. Simner, Johns Hopkins

Copyright © 2022 American Society for Microbiology. All Rights Reserved.

Address correspondence to John A. Branda, jbranda@partners.org.

The authors declare a conflict of interest. This study was funded by Zeus Scientific. J.A.B. has received research support for other studies from Pfizer, bioMerieux, Immunetics, Alere, DiaSorin, and the Bay Area Lyme Foundation (BALF) and has consulted for DiaSorin and Roche Diagnostics. E.S.T. has consulted for Euroimmune Inc. and Serimmune Scientific. T.R.F. and J.E.M. have received research support for other studies from bioMerieux, Immunetics and DiaSorin. TRF has consulted for DiaSorin. All other authors: No conflicts.

Received 20 December 2021

Returned for modification 22 January 2022

Accepted 28 March 2022

Published 14 April 2022

In basic MTTT algorithms, two orthogonal polyvalent EIAs (IgM/IgG) are applied sequentially, without the use of immunoblots. In this context, “orthogonal” tests are those that are different enough in their antigen targets or test principles that applying them sequentially in a two-tiered algorithm significantly improves specificity compared with the individual tests alone. Other MTTT algorithms use immunoglobulin class-specific EIAs in one tier of the two-tiered algorithm, allowing separate determination of IgM and IgG reactivity (4).

The U.S. Centers for Disease Control and Prevention (CDC) recently updated its Lyme disease serologic testing recommendations, which now advise that MTTT is an acceptable alternative to STTT if the first- and second-tier assays employed have been cleared by the FDA with an indication for use in a MTTT algorithm (14).

Here, we report the findings of a multicenter clinical evaluation for FDA review of four commercial Lyme disease EIAs and their use in 2 distinct MTTT algorithms.

MATERIALS AND METHODS

The study was approved by the Institutional Review Board at each institution and consisted of a retrospective and a prospective phase. In the retrospective phase, two MTTT algorithms were evaluated in comparison with STTT using archived serum samples obtained from well-characterized Lyme disease patients and control subjects. In the prospective phase, consecutive residual samples submitted to three clinical laboratories for routine Lyme disease serology were analyzed using the same MTTT algorithms and the results were compared with those of the STTT.

The retrospective phase included 280 archived serum samples obtained from the CDC Serum Repository for Lyme Disease Diagnostic Test Development and Evaluation (15). This panel includes samples from patients with common Lyme disease manifestations, control subjects with “look-alike” illnesses, and healthy control subjects. The CDC panel was supplemented with a convenience set of 96 archived samples contributed by one of the authors (A.C.S.) to improve the sample size of true cases in some categories. Table 1 provides a detailed description of archived samples used in the study (total $n = 376$).

In the prospective phase, consecutive excess serum samples submitted for routine Lyme disease serologic testing to three clinical laboratories were collected between July 2018 and August 2018 (total $n = 2,932$). The laboratories were in Rochester, MN (Mayo Clinic; 1,042 samples contributed); Marshfield, WI (Marshfield Clinic Research Institute; 990 samples contributed); and Boston, MA (Massachusetts General Hospital; 900 samples contributed). Samples were tested as described below.

Serologic testing. In the retrospective phase, all 376 archived samples were assayed using the following four Zeus commercial Lyme EIAs:

1. *Borrelia* VlsE1/pepC10 IgG/IgM Test System;
2. *Borrelia burgdorferi* IgG/IgM Test System (polyvalent whole-cell sonicate [WCS] EIA);
3. *Borrelia burgdorferi* IgM Test System (monovalent IgM WCS EIA); and
4. *Borrelia burgdorferi* IgG Test System (monovalent IgG WCS EIA).

The sensitivity and specificity of two distinct MTTT algorithms involving different assay combinations (described below) were then determined by *post hoc* data analysis. First-tier-reactive samples were subsequently analyzed using MarDx *B. burgdorferi* IgM and IgG Marblot tests (STTT), except for the 20 samples from control subjects with a remote history of Lyme disease, which were analyzed using ViraMed Biotech *Borrelia* B31 IgG and IgM ViraStripe tests because the MarDx blots had been discontinued. Assays were performed by Zeus Scientific employees who were blinded to clinical category for the CDC panel until all retrospective testing was complete. ViraStripe assays were performed in the laboratory of A.C.S. Immunoblots were interpreted according to CDC criteria (2), except that the “1-month rule” was not applied (exclusion of IgM immunoblot results when duration of symptoms at the time of sample collection exceeds 1 month, a rule that is not part of MTTT interpretive criteria). Non-application of the 1-month rule allows comparison of STTT and MTTT performance without the potential for the rule itself to affect (reduce) STTT sensitivity compared with MTTT sensitivity, which can occur (7).

In the prospective phase, EIAs were performed on site at the Minnesota and Massachusetts centers; for the Wisconsin site, samples were shipped frozen to Zeus Scientific for testing. All samples were assayed using the same first-tier test, the VlsE1/pepC10 IgG/IgM EIA, followed (only when reactive) by second-tier tests according to the STTT and MTTT algorithms described below, which were followed in parallel. Immunoblotting for STTT was performed by Zeus Scientific employees after frozen transport of first-tier-reactive samples. Immunoblots were interpreted according to CDC criteria, except that the “1-month” rule was not applied.

Two-tiered testing algorithms. Two distinct MTTT algorithms were evaluated in comparison to one STTT algorithm (Fig. 1):

- MTTT 1: VlsE1/pepC10 IgG/IgM EIA followed by polyvalent WCS EIA.
- MTTT 2: VlsE1/pepC10 IgG/IgM EIA followed by monovalent IgM and IgG EIAs.
- STTT: VlsE1/pepC10 IgG/IgM EIA followed by IgM and IgG immunoblots.

Statistical analysis. Differences between proportions were considered statistically significant if the two-tailed P value was < 0.05 as determined using McNemar’s test. The modified Wald method was used to calculate 95% confidence intervals (95% CIs). Agreement between two-tiered algorithms was measured using the

TABLE 1 Archived samples used in the retrospective phase of this study^a

Clinical category	No. samples	Sample description
Lyme disease	166	
EM	91	
EM, acute- and convalescent- phases	30 each phase, total 60	Paired samples from patients with solitary or multiple EM. Most patients were culture or PCR positive for <i>B. burgdorferi</i> . Time between development of an EM rash and collection of acute-phase sera was <30 days in all cases. Samples obtained from CDC repository (15).
Acute EM with hematogenous dissemination	31	Samples from patients with solitary EM plus a positive blood PCR for <i>B. burgdorferi</i> ($n = 23$) or multiple EM ($n = 2$) or multiple EM plus a positive blood PCR test ($n = 6$). Time between development of an EM rash and collection of acute-phase sera was <30 days in all cases. Samples were obtained from one of the authors (A.C.S.).
Lyme neuroborreliosis or carditis	25	Samples from patients with cranial nerve palsy, lymphocytic meningitis, radiculopathy, or heart block. Samples were obtained from CDC repository ($n = 10$) (15) and one of the authors (A.C.S., $n = 15$). Time between symptom onset and sample collection was ≤ 30 days in most cases; in other cases, the IgG antibody response was sufficiently expanded to meet IgG immunoblot criteria for a positive result (15).
Lyme arthritis	50	Samples were obtained from CDC repository ($n = 20$) (15) and one of the authors (A.C.S., $n = 30$). Time between symptom onset and sample collection was >30 days in most cases; in all cases, regardless of symptom duration, the IgG antibody response was sufficiently expanded to meet IgG immunoblot criteria for a positive result (15).
Control subjects	210	
With a remote past history of LD and another current illness	20	Current illnesses included psoriatic arthritis ($n = 4$), fibromyalgia (2), rheumatoid arthritis (2), degenerative arthritis (2), arthralgia (2), inflammatory bowel arthritis (1), spondyloarthropathy (1), polymyalgia rheumatica (1), fatigue (1), psychiatric illness (1), multiple sclerosis (1), peripheral neuropathy (1), and osteoarthritis (1). Samples obtained from one of the authors (A.C.S.).
With "look-alike" illnesses but no past history of LD	90	Subjects had fibromyalgia ($n = 15$), multiple sclerosis (15), mononucleosis (15), periodontitis (15), rheumatoid arthritis (15), or syphilis (15). Samples obtained from CDC repository (15).
Healthy and no past history of LD	100	
Residing in region where LD is endemic	50	All resided in New York. Obtained from CDC repository (15).
Residing in region where LD is non-endemic	50	All resided in Texas. Obtained from CDC repository (15).

^aNo., number; EM, erythema migrans; LD, Lyme disease.

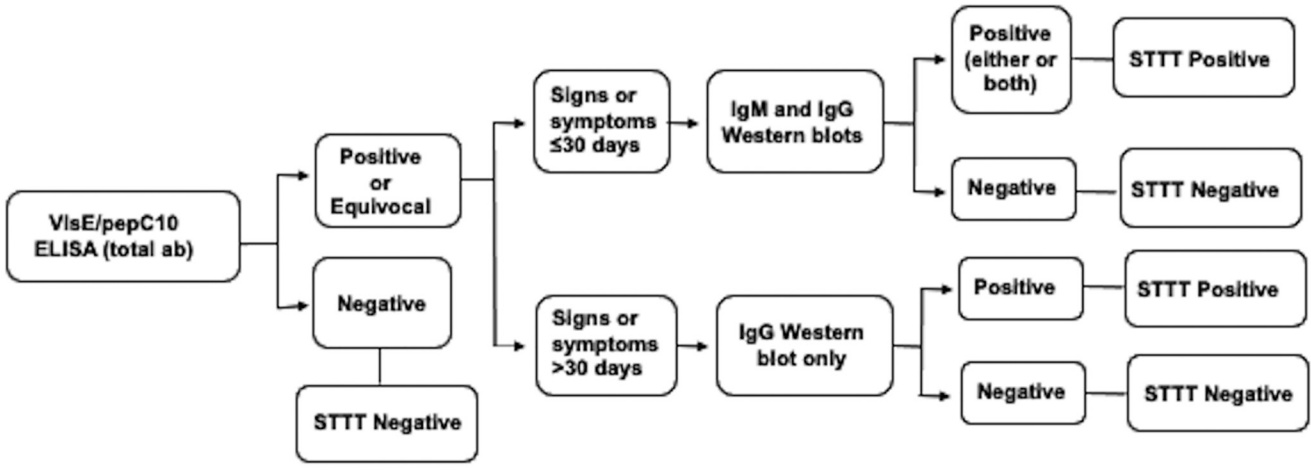
kappa statistic, with the level of agreement classified based on published standards: 0.81 to 1.00 (almost perfect), 0.61 to 0.80 (substantial), 0.41 to 0.60 (moderate), 0.21 to 0.40 (fair), and 0 to 0.20 (slight) (16). All analyses were performed using publicly available software (GraphPad QuickCalcs).

RESULTS

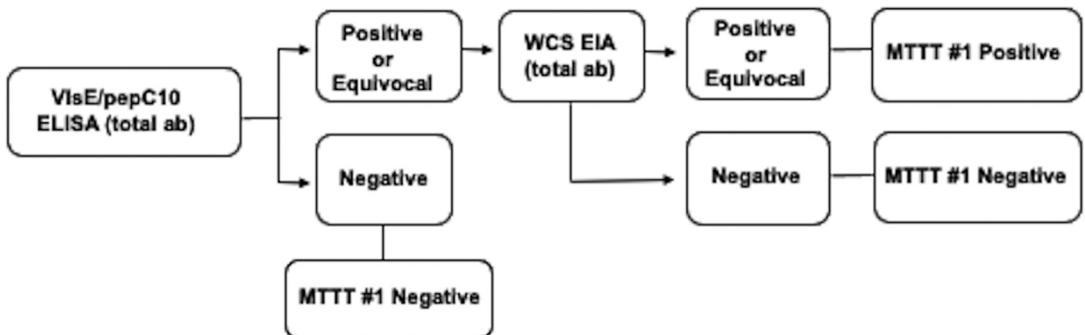
Retrospective phase. Using the panel of archived samples, MTTT algorithms 1 and 2 produced significantly higher specificity compared with each individual EIA alone, showing that the tests have sufficient orthogonality for use together (Tables 2 and 3). Moreover, when these tests were used sequentially, both MTTT algorithms were significantly more sensitive in patients with EM ($N = 61$) compared to STTT (56% and 74% versus 41%, $P = 0.03$ and 0.0001 , respectively) (Table 4). MTTT 1 was less sensitive in this cohort than MTTT 2 (56% versus 74%, $P = 0.003$). The difference was greatest in the subset of patients with EM and hematogenous dissemination. In this group, MTTT 2 was 71% sensitive, whereas MTTT 1 was 39% sensitive ($P = 0.004$).

Nearly all samples collected from patients with non-cutaneous manifestations of Lyme disease (early neuroborreliosis, carditis, or arthritis) were positive using any of

Standard Two-tiered Testing (STTT) Algorithm



Modified Two-tiered Testing (MTTT) Algorithm #1



Modified Two-tiered Testing (MTTT) Algorithm #2

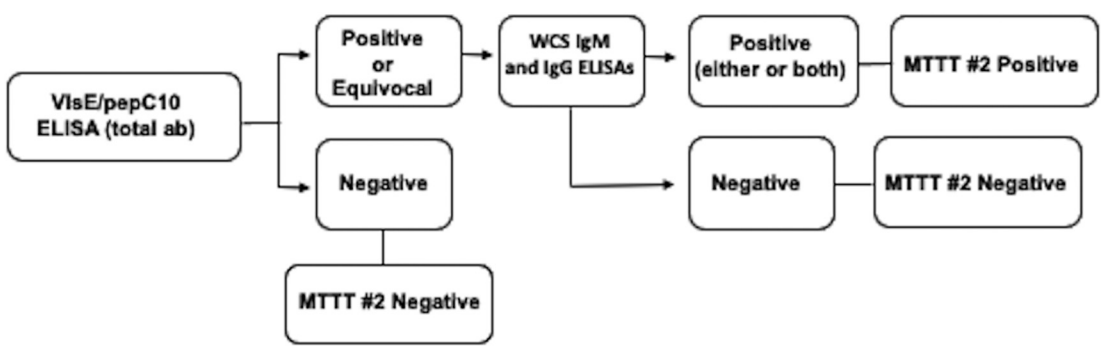


FIG 1 Flowcharts illustrating the two-tiered serologic testing algorithms evaluated in this study.

TABLE 2 Specificity of individual EIAs used in this study^a

Control subject category	No. in cohort	No. of negative results (specificity [%])				
		VlsE1/pepC10 IgG/IgM EIA	WCS IgG/IgM EIA	WCS IgM EIA	WCS IgG EIA	WCS IgM or IgG
"Look alike" illnesses	90	83 (92)	79 (88)	78 (87)	80 (89)	70 (78)
Healthy						
From area of endemicity	50	50 (100)	50 (100)	46 (92)	50 (100)	46 (92)
From area of non-endemicity	50	47 (94)	48 (96)	45 (90)	50 (100)	45 (90)
All control subjects	190	180 (95)	177 (93)	169 (89)	180 (95)	161 (85)

^aEIA, enzyme immunoassay; WCS, whole-cell sonicate.

the MTTT or STTT algorithms; there were no significant differences in sensitivity (Table 4). Minor differences in sensitivity between MTTT and STTT algorithms were also not significant using convalescent-phase samples from patients with recent EM (83 to 90% versus 77%, $P = 0.13$ to 0.48).

Among archived samples collected from control subjects with other illnesses and no history of Lyme disease ($n = 90$), negative results were obtained in 88/90 (98% specificity) using either MTTT 1 or MTTT 2, compared with 90/90 cases (100% specificity) using the STTT algorithm ($P = 0.48$; Table 5). The same two-control subject sera were falsely positive in both MTTT algorithms; each of the two subjects had acute infectious mononucleosis.

Using samples collected from healthy subjects living in New York, an area of endemicity for Lyme disease ($n = 50$), or Texas, an area of non-endemicity ($n = 50$), 100% specificity was obtained using any of the STTT or MTTT algorithms. When results obtained using samples from all control subjects with no history of Lyme disease were considered in aggregate ($n = 190$), each MTTT algorithm was 99% specific (188/190 negative), whereas the STTT algorithm was 100% specific ($P = 0.48$).

Among subjects with a remote history of Lyme disease and another current illness ($n = 20$), STTT was negative more frequently (9/20, 45%) than either MTTT 1 (6/20, 30%; $P = 0.25$) or MTTT 2 (7/20, 35%; $P = 0.48$).

Prospective phase. Across the three clinical sites, 2,932 consecutive samples were collected prospectively. In most cases, the investigators did not have access to clinical information, and classification of subjects into case and control groups was not possible. Thus, clinical sensitivity and specificity of the two-tiered algorithms could not be determined in this phase. Instead, MTTT algorithms were evaluated by assessing agreement with the reference method, STTT.

STTT was positive in 179/2,932 samples (6%) and negative in 2,753/2,932 samples (94%). Positive percent agreement (PPA), negative percent agreement (NPA), and overall agreement determinations for MTTTs 1 and 2 are reported in Table 6. MTTT 1 had higher overall agreement with STTT compared to MTTT 2 (97.4%, kappa 0.80 versus 96.2%, kappa 0.74; $P = 0.00001$), although both kappa values fell into the category of "substantial agreement" (16). MTTT 2 was more balanced in its performance, with PPA and NPA of >95% (98% and 96.1%, respectively) whereas MTTT 1 met that standard only for NPA (97.7%) and not for PPA (93%).

Because both STTT and MTTT 2 use separate IgM- and IgG-specific tests in the second tier, we also assessed agreement with immunoglobulin class-specific reactivity

TABLE 3 Specificity of individual EIAs compared with MTTT algorithms^a

MTTT algorithm	First-tier test		Second-tier test(s)		MTTT algorithm specificity (%; 95% CI)
	Specificity (%)	P^b	Specificity (%)	P^b	
1 (VlsE1/pepC10 IgG/IgM EIA f/b WCS IgG/IgM EIA)	180/190 (95)	0.01	177/190 (93)	0.003	188/190 (99, 96–100)
2 (VlsE1/pepC10 IgG/IgM EIA f/b WCS IgM and IgG EIAs)	180/190 (95)	0.01	161/190 (85)	0.0001	188/190 (99, 96–100)

^aMTTT, modified two-tiered testing; CI, confidence interval; f/b, followed by; WCS, whole-cell sonicate; EIA, enzyme immunoassay.

^b P values refer to the comparison between the individual EIA and the relevant MTTT algorithm.

TABLE 4 Sensitivity of two-tiered testing algorithms for Lyme disease^a

Algorithm	Test 1	Test 2	No. positive results (sensitivity [%]; 95% CI)					
			Paired samples from CDC repository		Acute EM with hematogenous dissemination (n = 31)	All acute EM combined (n = 61)	Neuritis or carditis (n = 25)	Lyme arthritis (n = 50)
			EM acute phase (n = 30)	EM conv. phase (n = 30)				
STTT	VisE1/pepC10 IgG/IgM EIA	IgM and IgG immunoblots	15 (50; 33–67)	23 (77; 59–88)	10 (32, 18–50)	25 (41, 30–54)	24 ^b (96, 79–100)	50 (100, 91–100)
MTTT 1	VisE1/pepC10 IgG/IgM EIA	WCS EIA (IgG/IgM)	22 (73, 55–86), <i>P</i> = 0.02 ^c	25 (83, 66–93), <i>P</i> = 0.48	12 (39, 24–56), <i>P</i> = 0.68	34 (56, 43–68), <i>P</i> = 0.03	25 (100, 84–100), <i>P</i> = 1.0	50 (100, 91–100), <i>P</i> = 1.0
MTTT 2	VisE1/pepC10 IgG/IgM EIA	WCS IgM and IgG EIAs	23 (77, 59–88), <i>P</i> = 0.01	27 (90, 74–97), <i>P</i> = 0.13	22 (71, 53–84), <i>P</i> = 0.002	45 (74, 61–83), <i>P</i> = 0.0001	25 (100, 84–100), <i>P</i> = 1.0	50 (100, 91–100), <i>P</i> = 1.0

^aCI, confidence interval; EM, erythema migrans; conv., convalescent; STTT, standard two-tiered testing; MTTT, modified two-tiered testing algorithm; EIA, enzyme immunoassay; WCS, whole-cell sonicate.

^bOne sample was negative by IgM immunoblot (strong 23-kd band, absent 39-kd band, weak 41-kd band), IgG immunoblot (only 41-kd and 45-kd bands present).

^cAll *P* values refer to comparison with standard two-tiered testing (VisE1/pepC10 polyvalent EIA followed by IgM and IgG immunoblots).

TABLE 5 Specificity of two-tiered testing algorithms for Lyme disease^a

Algorithm	Test 1	Test 2	No. negative (specificity [%]; 95% CI)			Previous history of LD and another current illness (n = 20)
			With "look-alike" illnesses (n = 90)	Healthy	Control subjects	
			From area of endemicity (n = 50)	From area of non-endemicity (n = 50)	No past history of LD, combined (n = 190)	
STTT	VisE1/pepC10 IgG/IgM EIA	IgM and IgG immunoblots	90 (100, 95–100)	50 (100, 91–100)	190 (100, 98–100)	9 (45, 26–66)
MTTT 1	VisE1/pepC10 IgG/IgM EIA	WCS EIA (IgG/IgM)	88 (98, 92–100), P = 0.48 ^b	50 (100, 91–100), P = 1.0	188 (99, 96–100), P = 0.48	6 (30, 14–52), P = 0.25
MTTT 2	VisE1/pepC10 IgG/IgM EIA	WCS IgM and IgG EIAs	88 (98, 92–100), P = 0.48	50 (100, 91–100), P = 1.0	188 (99, 96–100), P = 0.48	7 (35, 18–57), P = 0.48

^aCI, confidence interval; STTT, standard two-tiered testing; MTTT modified two-tiered testing algorithm; EIA, enzyme immunoassay; WCS, whole-cell sonicate.

^bAll P values refer to the comparison with standard two-tiered testing (VisE1/pepC10 polyvalent EIA followed by IgM and IgG immunoblots).

TABLE 6 Prospective evaluation of modified two-tiered testing algorithms for Lyme disease^a

Test method	Reference method	Agreement			
		PPA (95% CI), no. results (test/reference)	NPA (95% CI), no. results (test/reference)	Overall (95% CI), no. results ^b	κ (range)
MTTT 1 (VlsE1/pepC10 IgG/IgM EIA f/b WCS IgG/IgM EIA)	STTT (VlsE1/pepC10 IgG/IgM EIA f/b IgM and IgG immunoblots)	93 (89–96), 167/179	97.7 (97.1–98.2), 2,690/2,753	97.4 (96.8–98.0), 2,857/2,932	0.80 (0.76–0.85)
MTTT 2 (VlsE1/pepC10 IgG/IgM EIA f/b WCS IgM and IgG EIAs)	STTT (VlsE1/pepC10 IgG/IgM EIA f/b IgM and IgG immunoblots)	98 (95–100), 176/179; $P = 0.008^c$	96.1 (95.3–96.8), 2,646/2,753; $P < 0.000001^c$	96.2 (95.5–96.9), 2,822/2,932; $P = 0.00001^c$	0.74 (0.70–0.79)
MTTT 2 IgM (VlsE1/pepC10 IgG/IgM EIA f/b WCS IgM EIA)	STTT IgM (VlsE1/pepC10 IgG/IgM EIA f/b IgM immunoblot)	96 (90–99), 101/105	95.5 (94.7–96.3), 2,701/2,827	95.6 (94.8–96.2), 2,802/2,932	0.59 (0.53–0.65)
MTTT 2 IgG (VlsE1/pepC10 IgG/IgM EIA f/b WCS IgG EIA)	STTT IgG (VlsE1/pepC10 IgG/IgM EIA f/b IgG immunoblot)	92 (86–96), 115/125; $P = 0.27^d$	97.3 (96.6–97.8), 2,730/2,807; $P = 0.0006^d$	97.0 (96.4–97.6), 2,845/2,932; $P = 0.004^d$	0.71 (0.65–0.77)

^aCI, confidence interval; PPA, positive percent agreement; NPA, negative percent agreement; MTTT, modified two-tiered testing; EIA, enzyme immunoassay; f/b, followed by; WCS, whole-cell sonicate; STTT, standard two-tiered testing.

^bCombined positive and negative results in agreement, divided by total number of prospectively collected samples ($n = 2,932$).

^c P values refer to the Comparison between MTTT 1 and MTTT 2.

^d P values refer to the Comparison between MTTT 2 IgM and MTTT 2 IgG.

(Table 6). MTTT 2-IgM (VlsE1/pepC10 polyvalent EIA followed by WCS IgM EIA only) was in moderate agreement with IgM-only STTT based on the kappa statistic (95.6%, kappa 0.59), and MTTT 2-IgG (VlsE1/pepC10 polyvalent EIA followed by WCS IgG EIA only) was in substantial agreement with IgG-only STTT (97.0%, kappa 0.71).

DISCUSSION

In this study, we evaluated two distinct MTTT algorithms, a polyvalent VlsE1/pepC10 EIA followed by a polyvalent WCS EIA (MTTT 1) and the same VlsE1/pepC10 EIA followed by separate monovalent IgM and IgG WCS EIAs (MTTT 2). Both MTTT algorithms were significantly more sensitive in patients with acute EM compared with STTT. The difference was most pronounced with MTTT 2; using this approach, the number of EM cases detected nearly doubled in comparison with STTT (74% versus 41%; $P = 0.0001$). However, by convalescence, after treatment with antibiotic therapy, the differences between the two-tiered algorithms had narrowed and were no longer significantly different. Moreover, sensitivity in patients with later manifestations of the infection (Lyme neuroborreliosis, carditis, or arthritis) was equal or non-inferior between the MTTT and STTT algorithms.

In healthy subjects or those with “look-alike” (non-Lyme) illnesses but no history of Lyme disease, both MTTT algorithms were significantly more specific than any of the individual assays alone, indicating a useful degree of orthogonality between the first- and second-tier assays. However, each MTTT algorithm produced two false-positive results among control subjects with heterophile-reactive mononucleosis that STTT did not, although the difference in overall specificity between the two approaches was nonsignificant (99% versus 100%; $P = 0.48$). Such cross-reactivity is well-described in WCS Lyme disease EIAs (17). Because heterophile antibody cross-reactivity affected all EIAs used in this study except for the IgG-specific WCS EIA, the MTTT algorithm component assays were susceptible to the same errors and were not orthogonal in this respect.

In the multicenter prospective phase of this study, 2,932 consecutive samples submitted to three clinical testing centers were analyzed according to MTTTs 1 and 2 and STTT. Using STTT as the reference method, the PPA of MTTT 1 was lower than that of MTTT 2 (93% versus 98%, $P = 0.008$). This finding may reflect the lower sensitivity of MTTT 1 and STTT compared with MTTT 2 among patients with early LD, demonstrated in the retrospective phase of this study.

Findings from the prospective phase demonstrated high degrees of agreement in negative results between the MTTT algorithms and STTT. The NPA of MTTT 1 was 97.7% and that of MTTT 2 was 96.1%. Although small, the difference in NPA between

the MTTT algorithms was highly significant ($P < 0.000001$). The difference may be attributable in part to false-positive results with MTTT 2 that correctly resulted as negative with MTTT 1. However, this may not be the major factor because MTTT 1 had equal specificity in the retrospective phase among control subjects with no history of LD (99%) and numerically, but nonsignificantly, lower specificity (30% versus 35%, $P = 1.0$) among control subjects with a history of LD and another current illness. Instead, much of this difference may be attributable to the greater sensitivity of MTTT 2 for patients with EM (especially patients with EM and hematogenous dissemination), compared with MTTT 1 or STTT, as demonstrated in the retrospective phase. In a recent survey of large U.S. commercial laboratories performed by the CDC, it was estimated that 86% of true Lyme disease patients among the tested population had localized disease (EM) rather than disseminated disease (18), and we assume that most true cases in our tested population also had EM.

Among 47 samples in the prospective study that were negative using STTT and MTTT 1, but positive using MTTT 2, IgM reactivity, but not IgG reactivity, was found in 22 (47%) using isotype-specific WCS EIAs (data not shown), which is consistent with expected findings in EM cases. Thus, while MTTT 2 had lower NPA than MTTT 1, this may partly reflect its higher sensitivity in patients with acute EM (i.e., false-negative results with MTTT 1 and STTT in some early cases). The difference in early disease sensitivity between the MTTT algorithms is counterintuitive and perplexing, considering that they employ the same first-tier test, and the second-tier tests are all made using WCS. However, the second-tier tests are still distinct assays with individually determined interpretive cut-points and slightly different chemistry.

The significant difference in NPA between the MTTT algorithms also had a pronounced effect on overall agreement measurements in the prospective phase, a calculation that factors in both PPA and NPA and gives them equal weight. MTTT 1 had higher overall agreement with STTT than MTTT 2 (97.4% versus 96.2%; $P = 0.00001$), with the driving factor in the difference being the relatively large number of discrepancies with STTT-negative samples using MTTT 2.

Based on this analysis, we conclude that one MTTT algorithm did not appreciably outperform the other. MTTT 2 is more sensitive in patients with acute EM but may produce more false-positive results than MTTT 1. From a performance standpoint, MTTT 1 may be preferable in low-prevalence areas and during the off-season, when specificity is paramount. MTTT 2 may be preferable during peak season in areas of endemicity, due to its high sensitivity among patients with early LD. Beyond performance, however, there are several other considerations. One disadvantage of MTTT 1 is that a positive result does not give information about immunoglobulin (Ig) class specificity (IgM or IgG), which can be useful in correlating serologic test results with clinical features and disease timeline. MTTT 2 does provide this information, and our findings demonstrate that Ig class-specific results obtained using MTTT 2, whether positive or negative, correlated well with those obtained using STTT. On the other hand, MTTT 1 involves only two tests, both of which are polyvalent (IgM/IgG) tests. This reduces cost, streamlines testing and training, and simplifies result reporting and interpretation compared with MTTT 2.

The use of MTTT algorithms is not ideal in every clinical context. MTTT strategies provide only a categorical result (positive or negative for *B. burgdorferi* antibodies). This is sufficient for routine cases, as when the patient has objective signs compatible with common manifestation of Lyme disease and there is no history of Lyme disease (3, 4). More complex cases may be better evaluated using second-tier assays that provide detailed information about the spirochetal targets of the humoral immune response, such as immunoblots or multiplexed assays (3, 4).

Particularly problematic are patients with a history of previous LD and a current clinical picture with a large differential diagnosis. Although the antibody response declines after spirochetal killing with antimicrobial therapy, the response often persists at a low level. In this study, the STTT algorithm was still positive in 55% of patients who had a history of previous LD and another current illness, compared with 65 to

70% of patients using the MTTT algorithms. Thus, positive antibody responses with either method cannot clearly distinguish between active and past infection. However, immunoblotting or multiplexed serologic assays may provide clues not available in MTTT approaches. For example, when considering whether a patient with a history of early LD now has a late manifestation of the disease or another illness, it can be helpful to know whether the patient has an immunoblot with a response more suggestive of early or late LD. However, the distinction between past and current infection is not always possible even with immunoblotting.

Our study has several limitations. In the prospective phase, our research team often did not have access to medical information about the patients whose serum samples were included. This prevented an unbiased, systematic analysis of discrepancies between the MTTT algorithms and STTT, which would have better informed us about which approach (MTTT or STTT) had most likely provided the correct result based on the clinical diagnosis. In addition, the prospective phase involved sample collection in three distinct geographical areas, but all were within the continental U.S. Thus, we do not know how the MTTT algorithms would perform in diagnosing Lyme disease acquired elsewhere or in distinguishing Lyme disease from that caused by non-*Burgdorferi* species of Lyme-related borreliae. An additional limitation is the immunoblot used in most cases for the second tier of the STTT algorithm. The MarDx blot employs a *B. burgdorferi* WCS which is known to contain spirochetal proteins that are nonspecific for that infection. That may lead to false-positive results, especially with the IgM blot. Newer techniques for blot manufacture employ recombinant or purified *Borrelia* proteins, making them easier to interpret and potentially providing better performance (19, 20). Thus, the reference standard used here (STTT with MarDX blots) may be suboptimal and reactive in some true-negative cases.

In summary, this multicenter study demonstrated the clinical validity of two distinct MTTT algorithms for the serodiagnosis of Lyme disease, both of which employ commercial EIAs. The results of this performance evaluation have been reviewed by FDA through the premarket notification [510(k)] pathway. In July 2019, the four Zeus EIAs evaluated here, which had previously been FDA-cleared only for marketing as first-tier tests for use in STTT algorithms with immunoblotting, were cleared by FDA for marketing with the additional indication of use in MTTT algorithms 1 and 2. This was the first instance of Lyme disease serologic tests being labeled with an indication for use in MTTT algorithms. The availability of FDA-cleared commercial test kits for use in MTTT algorithms removes a major impediment to widespread adoption of this approach in the United States.

ACKNOWLEDGMENTS

We are grateful to Vedrana Eleta, Dina Ryan, Kathleen Cichonski, Danielle Accardi, Julius Torres, Sarah Ma, and Christine Keselica for technical assistance.

This study was funded by Zeus Scientific. J.A.B. has received research support for other studies from Pfizer, bioMérieux, Immunetics, Alere, DiaSorin, and the Bay Area Lyme Foundation (BALF), and has consulted for DiaSorin and Roche Diagnostics. E.S.T. has consulted for Euroimmune Inc. and Serimmune Scientific. T.R.F. and J.E.M. have received research support for other studies from bioMérieux, Immunetics, and DiaSorin. T.R.F. has consulted for DiaSorin. The other authors have no conflicts of interest.

REFERENCES

1. Moore A, Nelson C, Molins C, Mead P, Schriefer M. 2016. Current guidelines, common clinical pitfalls, and future directions for laboratory diagnosis of Lyme disease, United States. *Emerg Infect Dis* 22:1169–1177. <https://doi.org/10.3201/eid2207.151694>.
2. CDC. 1995. Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. *MMWR Morb Mortal Wkly Rep* 44:590–591.
3. Branda JA, Body BA, Boyle J, Branson BM, Dattwyler RJ, Fikrig E, Gerald NJ, Gomes-Solecki M, Kintrup M, Ledizet M, Levin AE, Lewinski M, Liotta LA, Marques A, Mead PS, Mongodin EF, Pillai S, Rao P, Robinson WH, Roth KM, Schriefer ME, Slezak T, Snyder J, Steere AC, Witkowski J, Wong SJ, Schutzer SE. 2018. Advances in serodiagnostic testing for Lyme disease are at hand. *Clin Infect Dis* 66:1133–1139. <https://doi.org/10.1093/cid/cix943>.
4. Branda JA, Steere AC. 2021. Laboratory diagnosis of Lyme borreliosis. *Clin Microbiol Rev* 34:e00018-19. <https://doi.org/10.1128/CMR.00018-19>.
5. Pegalajar-Jurado A, Schriefer ME, Welch RJ, Couturier MR, MacKenzie T, Clark RJ, Ashton LV, Delorey MJ, Molins CR. 2018. Evaluation of modified two-tiered testing algorithms for Lyme disease laboratory diagnosis using

- well-characterized serum samples. *J Clin Microbiol* 56:e01943-17. <https://doi.org/10.1128/JCM.01943-17>.
6. Wormser GP, Schriefer M, Agüero-Rosenfeld ME, Levin A, Steere AC, Nadelman RB, Nowakowski J, Marques A, Johnson BJ, Dumler JS. 2013. Single-tier testing with the C6 peptide ELISA kit compared with two-tier testing for Lyme disease. *Diagn Microbiol Infect Dis* 75:9–15. <https://doi.org/10.1016/j.diagmicrobio.2012.09.003>.
 7. Branda JA, Linskey K, Kim YA, Steere AC, Ferraro MJ. 2011. Two-tiered antibody testing for Lyme disease with use of 2 enzyme immunoassays, a whole-cell sonicate enzyme immunoassay followed by a VlsE C6 peptide enzyme immunoassay. *Clin Infect Dis* 53:541–547. <https://doi.org/10.1093/cid/cir464>.
 8. Molins CR, Delorey MJ, Replogle A, Sexton C, Schriefer ME. 2017. Evaluation of bioMérieux's dissociated VIDAS Lyme IgM II (LYM) and IgG II (LYG) as a first-tier diagnostic assay for Lyme disease. *J Clin Microbiol* 55: 1698–1706. <https://doi.org/10.1128/JCM.02407-16>.
 9. Branda JA, Strle K, Nigrovic LE, Lantos PM, Lepore TJ, Damle NS, Ferraro MJ, Steere AC. 2017. Evaluation of modified 2-tiered serodiagnostic testing algorithms for early Lyme disease. *Clin Infect Dis* 64:1074–1080. <https://doi.org/10.1093/cid/cix043>.
 10. Molins CR, Delorey MJ, Sexton C, Schriefer ME. 2016. Lyme borreliosis serology: performance of several commonly used laboratory diagnostic tests and a large resource panel of well-characterized patient samples. *J Clin Microbiol* 54:2726–2734. <https://doi.org/10.1128/JCM.00874-16>.
 11. Davis IRC, McNeil SA, Allen W, MacKinnon-Cameron D, Lindsay LR, Bernat K, Dibernardo A, LeBlanc JJ, Hatchette TF. 2020. Performance of a modified two-tiered testing enzyme immunoassay algorithm for serologic diagnosis of Lyme Disease in Nova Scotia. *J Clin Microbiol* 58:e01841-19. <https://doi.org/10.1128/JCM.01841-19>.
 12. Baarsma ME, Schellekens J, Meijer BC, Brandenburg AH, Souilljee T, Hoffuis A, Hovius JW, van Dam AP. 2020. Diagnostic parameters of modified two-tier testing in European patients with early Lyme disease. *Eur J Clin Microbiol Infect Dis* 39:2143–2152. <https://doi.org/10.1007/s10096-020-03946-0>.
 13. Branda JA, Lemieux JE, Blair L, Ahmed AA, Hong DK, Bercovici S, Blauwkamp TA, Hollemon D, Ho C, Strle K, Damle NS, Lepore TJ, Pollock NR. 2021. Detection of *Borrelia burgdorferi* cell-free dna in human plasma samples for improved diagnosis of early Lyme borreliosis. *Clin Infect Dis* 73:e2355–e2361. <https://doi.org/10.1093/cid/ciaa858>.
 14. Mead P, Petersen J, Hinckley A. 2019. Updated CDC recommendation for serologic diagnosis of Lyme disease. *MMWR Morb Mortal Wkly Rep* 68: 703. <https://doi.org/10.15585/mmwr.mm6832a4>.
 15. Molins CR, Sexton C, Young JW, Ashton LV, Pappert R, Beard CB, Schriefer ME. 2014. Collection and characterization of samples for establishment of a serum repository for Lyme disease diagnostic test development and evaluation. *J Clin Microbiol* 52:3755–3762. <https://doi.org/10.1128/JCM.01409-14>.
 16. Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174. <https://doi.org/10.2307/2529310>.
 17. Fawcett PT, Gibney KM, Rose CD, Dubbs SB, Doughty RA. 1992. Frequency and specificity of antibodies that crossreact with *Borrelia burgdorferi* antigens. *J Rheumatol* 19:582–587.
 18. Hinckley AF, Connally NP, Meek JI, Johnson BJ, Kemperman MM, Feldman KA, White JL, Mead PS. 2014. Lyme disease testing by large commercial laboratories in the United States. *Clin Infect Dis* 59:676–681. <https://doi.org/10.1093/cid/ciu397>.
 19. Binnicker MJ, Jespersen DJ, Harring JA, Rollins LO, Bryant SC, Beito EM. 2008. Evaluation of two commercial systems for automated processing, reading, and interpretation of Lyme borreliosis Western blots. *J Clin Microbiol* 46:2216–2221. <https://doi.org/10.1128/JCM.00200-08>.
 20. Kadkhoda K, Gretchen A. 2018. Higher sensitivity of the *recom* line *Borrelia* IgG Immunoblot Kit than of the standard Lyme IgG Immunoblot Kit according to CDC testing criteria. *J Clin Microbiol* 56:e00527-18. <https://doi.org/10.1128/JCM.00527-18>.