Diagnostic parameters of cellular tests for Lyme borreliosis in Europe (VICTORY study): a case-control study

M E Baarsma, Freek R van de Schoor*, Stefanie A Gauw, Hedwig D Vrijmoeth, Jeanine Ursinus, Nienke Goudriaan, Calin D Popa, Hadewych JM ter Hofstede, Mariska MG Leeflang, Kristin Kremer, Cees C van den Wijngaard†, Bart-Jan Kullberg†, Leo AB Joosten†, Joppe W Hovius†*

Lancet Infect Dis 2022; 22: 1388–96

Published Online June 14, 2022 https://doi.org/10.1016/S1473-3099(22)00205-5

See Comment page 1264

*Joint first authors

†Contributed equally

Center for Experimental and Molecular Medicine (M E Baarsma MD, S A Gauw RN, J Ursinus MD, N Goudriaan MD, Prof J W Hovius MD), **and Department of Epidemiology and Data Science** (M M G Leeflang PhD), **Amsterdam UMC location AMC, Amsterdam Institute for Infection and Immunity, University of Amsterdam, Amsterdam, Netherlands; Department of Internal Medicine, Radboudumc Center for Infectious Diseases and Radboud Institute of Health Sciences** (F R van de Schoor MD, H D Vrijmoeth MD, H J M ter Hofstede MD, Prof B J Kullberg MD, Prof L A B Joosten PhD), **and Department of Rheumatology** (C D Popa MD), **Radboudumc, Nijmegen, Netherlands; Department of Rheumatology, Sint Maartenskliniek, Ubbergen, Netherlands** (C D Popa); **National Institute for Public Health and the Environment, Center for Infectious Disease Control, Bilthoven, Netherlands** (K Kremer PhD, CC van den Wijngaard PhD); **KNCV Tuberculosis Foundation, The Hague, Netherlands** (K Kremer)

Correspondence to: Prof Joppe W Hovius, Center for Experimental and Molecular Medicine, Amsterdam Institute for Infection and Immunology, Amsterdam UMC, University of Amsterdam, Amsterdam, 1105 AZ, Netherlands
victory@amsterdamumc.nl

Summary

Background Cellular tests for Lyme borreliosis might be able to overcome major shortcomings of serological testing, such as its low sensitivity in early stages of infection. Therefore, we aimed to assess the sensitivity and specificity of three cellular tests.

Methods This was a nationwide, prospective, multiple-gate case-control study done in the Netherlands. Patients with physician-confirmed Lyme borreliosis, either early localised or disseminated, were consecutively included as cases at the start of antibiotic treatment. Controls were those without Lyme borreliosis from the general population (healthy controls) and those with potentially cross-reactive conditions (eg, autoimmune disease). We used three cellular tests for Lyme borreliosis (Spirofind Revised, iSpot Lyme, and LTT-MELISA) as index tests, and standard two-tier serological testing (STTT) as a comparator. Clinical data from Lyme borreliosis patients were collected at baseline and at 12 weeks after inclusion, and blood samples were obtained at baseline, 6 weeks, and 12 weeks. Control participants underwent clinical and laboratory assessments at baseline only.

Findings Cases comprised 271 patients with Lyme borreliosis (of whom 245 had early-localised Lyme borreliosis and 26 had disseminated disease) and controls comprised 228 participants without Lyme borreliosis from the general population and 41 participants with potentially cross-reactive conditions. Recruitment occurred between May 14, 2018, and March 16, 2020. The specificity of STTT in healthy controls (216 of 228 samples [94·7%, 95% CI 91·5–97·7]) was higher than that of the cellular tests: Spirofind (140 of 171 [81·9%, 76·1–87·2]), iSpot Lyme (32 of 103 [31·1%, 21·5–40·3]) and LTT-MELISA (100 of 190 [52·6%, 44·9–60·3]). Cellular tests had varying sensitivities: Spirofind (88 of 204 [43·1%, 36·4–50·4]), iSpot Lyme (51 of 94 [54·3%, 44·5–63·7]), and LTT-MELISA (66 of 218 [30·3%, 23·8–36·7]). The Spirofind and iSpot Lyme outperformed STTT for sensitivity, but were similar to the C6-ELISA (C6-ELISA: 135 of 270 [50·0%, 44·5–55·5]; STTT: 76 of 270 [28·1%, 23·0–33·6]).

Interpretation The cellular tests for Lyme borreliosis used in this study have a low specificity compared with serological tests, which leads to a high number of false-positive test results. We conclude that these cellular tests are unfit for clinical use at this stage.

Funding Netherlands Organization for Health Research and Development, AMC Foundation (Amsterdam UMC), and Ministry of Health of the Netherlands.

Copyright © 2022 Elsevier Ltd. All rights reserved.

Introduction

Lyme borreliosis is the most common vector-transmitted disease in Europe and North America and is caused by spirochaetes from the *Borrelia burgdorferi* sensu lato complex. Diagnosis of Lyme borreliosis is based on a combination of clinical evaluation and laboratory testing. The most used diagnostic tool for Lyme borreliosis is two-tiered serological testing. The standard two-tier testing (STTT) algorithm entails an ELISA, followed by an immunoblot in the case of an equivocal or positive ELISA result.[1,2] Alternatively, the first ELISA might be followed by a second ELISA in a modified two-tier testing algorithm.[2,3] The specificity of STTT is excellent when done and interpreted according to current guidelines.[1,2,4,5] Although the sensitivity of serological testing increases steadily among patients with disseminated disease in the weeks to months after the initial infection, it is well known that the sensitivity of STTT is low in the early-localised stage (known as erythema migrans).[4] Full maturation of the B-cell response might take several weeks and can be abrogated by antimicrobial therapy.[6] Another caveat is the persistence of IgG antibodies, which can be detectable in patient serum for decades after clearance of the infection.[6,7] Consequently, health professionals cannot rely on serology to follow disease progression or assess whether the infection has been eliminated.

For these reasons, various researchers have focused on the cellular response of the immune system to infection as a potential marker for Lyme borreliosis. Such cellular tests are based on the memory T-cell-mediated immune response after ex-vivo stimulation of whole blood or

Research in context

Evidence before this study

We searched PubMed using the terms "Borrelia", "borreliosis", "Lyme", "cellular test", "cellular assay", "lymphocyte transformation test", "LTT" and "EliSPOT" for Articles published from inception to May 4, 2021, in any language. All studies assessing diagnostic tests for Lyme borreliosis using cellular immunity were reviewed. The results included one systematic review published in January, 2020. The literature suggests that cellular tests might be able to overcome two major shortcomings of the current diagnostic standard (serology)—namely, the low sensitivity of serology in early disease and its inability to reliably distinguish a current infection from a past infection. Our review of the literature produced contrasting results on the diagnostic parameters of cellular tests for Lyme borreliosis, with sensitivities ranging from 28% to 90% and specificities from 33% to 99%. Many studies had methodological shortcomings, such as unclear case definitions for Lyme borreliosis, the absence of appropriate control groups, or undeclared conflicts of interest. Despite these shortcomings, cellular Lyme borreliosis tests are frequently sought by patients with long-lasting but poorly understood symptoms attributed to Lyme borreliosis.

Added value of this study

Our study is the first to report on the diagnostic parameters of cellular tests in a substantial cohort of patients from Europe with patients confirmed to have Lyme borreliosis who were classified on the basis of recognised guidelines. Controls comprised people without Lyme borreliosis (healthy controls) from the general population and people with potentially cross-reactive conditions. The specificity in healthy controls of the cellular tests varied from 31% to 81%, which was worse than the specificity of standard serological testing (>93%). The cellular assays had varying sensitivities but did not clearly outperform serological assays.

Implications of all the available evidence

These findings suggest that the cellular tests for Lyme borreliosis that have been assessed are unfit for clinical use, as they have not been properly validated. This finding is especially relevant for the two commercially available tests that have been used for patients, and it underscores the need to properly assess novel diagnostic tests for Lyme borreliosis before use. Research on assays with improved diagnostic characteristics is urgently needed.

PBMCs with specific pathogenic antigens. Signal molecules (eg, cytokines) or cell proliferation can serve as a read-out of these tests.[8–10] Previous studies on the diagnostic performance of cellular Lyme borreliosis tests have produced contradicting results, with sensitivities ranging from 28% to 90% and specificities from 33% to 99%.[11–13] Some of the aforementioned studies did not use clear Lyme borreliosis case definitions or used clinical criteria other than those defined by the European Society of Clinical Microbiology and Infectious Diseases, or Infectious Diseases Society of America.[14] Some investigators reported the results of small groups or did not include appropriate control groups.[12] In summary, most studies investigating cellular tests for Lyme borreliosis have had methodological shortcomings.[12]

However, additional research is highly relevant to both clinicians and patients, given that cellular Lyme borreliosis tests are commercially available from various laboratories in Europe and the USA, and they are frequently used for patients with chronic complaints attributed to Lyme borreliosis.[15] In the absence of a comprehensive validation, patients and practitioners cannot rely on the results of such tests to make a diagnosis of Lyme borreliosis or guide clinical management. We argue that more research on these tests is needed, a position that is supported by various health agencies and patient-backed organisations and initiatives.[16,17] In fact, authors of the latest UK National Institute for Health and Care Excellence (NICE)[18] guideline described finding a new test modality for diagnosing Lyme borreliosis at all stages of the infection, such as a cellular test, as a research priority.

We assessed the diagnostic performance of three cellular tests for Lyme borreliosis in an independent study with clearly defined participant categories and case definitions. We included two IFNγ release assays on different platforms and one lymphocyte transformation test (LTT) and compared them with STTT.

Methods

Study design and participants

As described in the published study protocol,[19] we did a prospective multiple-gate case-control study with consecutively included patients with confirmed Lyme borreliosis, healthy (ie, without Lyme borreliosis) controls from the general population, and patients with potentially cross-reactive conditions (CRCs, comprised of patients with autoimmune disease, diagnosed with syphilis, diagnosed with leptospirosis, or positive for cytomegalovirus infection). The inclusion and exclusion criteria were applied as published[19] and as described in the appendix (pp 2–3), as was matching between patients with Lyme borreliosis and healthy controls. All groups were recruited from the Netherlands (where Lyme borreliosis is endemic throughout the country) between May 14, 2018, and March 16, 2020.

See Online for appendix

Briefly, adult patients with acute, physician-confirmed Lyme borreliosis were recruited through the website tekenradar.nl or their treating physician. All eligible patients who consented were included consecutively. Patients were included before or just after the start (≤7 days after initiation) of any antibiotic treatment. Clinical data were collected at baseline and at 12 weeks

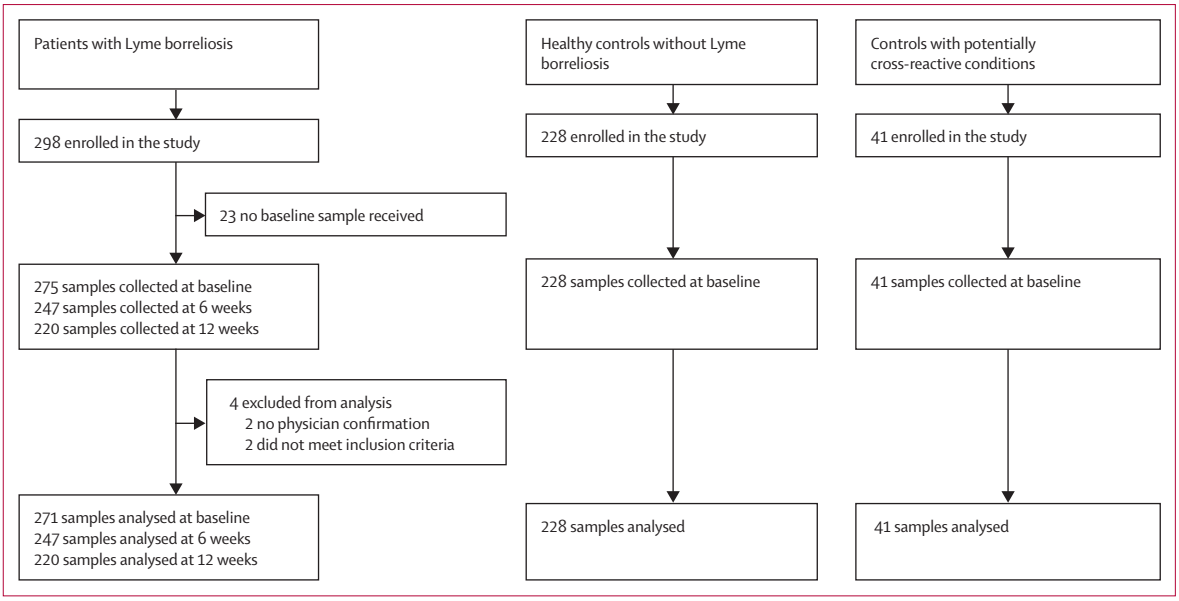

Figure 1: **Study participants**

after inclusion. Blood samples were obtained at baseline, 6 weeks, and 12 weeks. Healthy controls and patients with CRCs underwent clinical and laboratory assessments at baseline only. The study was approved by the Medical Ethics Committee of Noord-Holland and Amsterdam UMC, was registered with the Netherlands Trial Register (NL7732), and was done in accordance with the Declaration of Helsinki, the Dutch Medical Research Involving Human Subjects Act, and institutional guidelines. Written informed consent was obtained from all participants before inclusion. The study was designed and done in consultation with two patient representatives, who were also involved in the interpretation of the results and providing input for the manuscript.

Procedures

We obtained participants' medical history, information on concurrent medication, and history of tick bites. Medical history from healthy controls and controls with CRCs was obtained through an interview with an investigator. A validated online self-report questionnaire[20] was used to obtain this information from patients with Lyme borreliosis. Similarly, healthy controls and controls with CRCs were interviewed by investigators about any tick bites in their past, and patients with Lyme borreliosis self-reported tick bites during the last 5 years. In addition, patients with Lyme borreliosis provided an extensive history of their Lyme borreliosis symptoms and antibiotic treatment.

Laboratory measurements were done as published previously.[19] We utilised three cellular tests for Lyme borreliosis: the Spirofind Revised (Oxford Immunotec, Oxford, UK), the commercially available Lyme iSpot (Autoimmun Diagnostika, Strassberg, Germany), and LTT-MELISA (InVitaLab, Neuss, Germany). Spirofind was done at Amsterdam UMC and Radboudumc and was interpreted as prescribed by the manufacturer. Samples for the iSpot and LTT were transferred to the facilities of their respective manufacturers in Germany by overnight courier to be assessed there. Transfer of samples in this way matches the real-world situation for iSpot or LTT for most patients, wherein blood samples are sent to a German laboratory via overnight mail. Results from the LTT and iSpot were analysed as prescribed by the manufacturer. For all tests, each sample was given a unique alphanumerical code and clinicians doing the tests were masked to the donor's status.

As a comparator, we used STTT. All serum samples were first tested by C6-ELISA (Oxford Immunotec, Oxford, UK). All sera with a positive or equivocal C6-ELISA result were then tested with a recomLine *Borrelia* IgM and IgG immunoblot (Mikrogen, Neuried, Germany) as a confirmatory assay. More details on the execution of serological tests and overall test interpretation are given in the appendix (p 3).

The results of the cellular tests and the C6-ELISA were reported in a primary and an alternate interpretation. For the iSpot, LTT, and C6-ELISA, equivocal results were classified as positive in the primary interpretation and classified as negative in the alternate interpretation. The primary interpretation was defined because we hypothesised that physicians would generally be inclined to treat patients suspected of Lyme borreliosis with equivocal test results. For the Spirofind, which does not have any equivocal results by design, the per-protocol interpretation included only samples that were processed

strictly according to the manufacturer's protocol and the all-samples interpretation included all samples that had an interpretable result, even those with a protocol deviation (eg, those that exceeded the permitted time from blood draw to start of incubation).

Statistical analysis

The 95% CIs of sensitivity and specificity were calculated using bootstrapped Clopper-Pearson interval. Diagnostic parameters of the various cellular and serological tests were compared within participant groups using an exact McNemar test. An exact McNemar test was also used for the comparison of valid versus non-valid tests (ie, all interpretable results *vs* tests that were done but did not yield an interpretable result). For independent groups, proportions were compared using Fisher's exact test and means were compared using Student's *t* test. The area under the curve for the antigens of cellular tests was assessed using a receiver operating characteristics (ROC) analysis, and the accompanying 95% CI were bootstrapped. Missing test results or other datapoints were not imputed and were excluded from analyses. Analyses were done in SPSS (version 26), except for bootstrapping the 95% CIs of the ROCs, which was done in R (pROC 1.18.0), and the second figure, which was rendered in GraphPad (version 9).

Role of the funding source

The Netherlands Organization for Health Research and Development was involved in the design of this study and had reviewed the grant application but was not involved in data collection, data analysis, data interpretation, or the writing of the report. The AMC Foundation and the Ministry of Health of the Netherlands were not involved in any of the aforementioned aspects of the study.

Results

The Lyme borreliosis cohort comprised 298 patients, of whom 275 provided a blood sample at baseline for use in the primary outcome (figure 1). Those who did not provide a blood sample at baseline (23 patients) were excluded from analyses, together with two patients who did not have physician confirmation of their Lyme borreliosis diagnosis and two who retrospectively did not meet the inclusion criteria. Therefore, the final cohort for analyses consisted of 271 patients with Lyme borreliosis, who all provided a sample at baseline and the majority of whom also provided samples at 6 weeks (247 patients, 91%) and 12 weeks (220 patients, 81%) after inclusion. The Lyme borreliosis cohort comprised 245 patients (90%) with physician-confirmed erythema migrans and 26 patients (10%) with disseminated manifestations (appendix p 4). The majority of patients with erythema migrans (187 [76%] of 245) provided their baseline sample in the first 3 weeks after the onset of erythema migrans.

	Patients with Lyme borreliosis (N=271)	Healthy controls (N=228)	p value
Mean age (SD; range), years	53 (14; 18–82)	54 (12; 18–81)	0·43*
Sex			
Male	116 (43%)	104 (46%)	0·59†
Female	155 (57%)	124 (54%)	0·59†
Tick bite incidence in area of residence‡			
Low	92 (34%)	91 (40%)	0·27†
Medium	88 (32%)	74 (32%)	0·27†
High	91 (34%)	63 (28%)	0·27†
Erythema migrans incidence in area of residence‡			
Low	82 (30%)	84 (37%)	0·30†
Medium	93 (34%)	70 (31%)	0·30†
High	96 (35%)	74 (32%)	0·30†
Tick bite in medical history§¶	112 (41%)	91 (40%)	0·78†
Lyme borreliosis in medical history	20 (7%)	20 (9%)	0·62†
Concomitant diagnoses§‖			
0	115 (43%)	76 (33%)	0·071†
1–2	127 (47%)	118 (52%)	0·071†
≥3	28 (10%)	34 (15%)	0·071†
Immunosuppressive medication**			
No	238 (88%)	194 (85%)	..
Mild	29 (11%)	34 (15%)	0·22†,††
High	4 (1%)	0 (0)	0·13†,††

Data are in n (%) unless specified otherwise. *Student's *t* test. †Fisher's Exact test. ‡Low, medium, and high should be read as the first, second, and third tertile of the incidence of tick bites or erythema migrans by postcode of residence. §Missing one patient with Lyme borreliosis. ¶Tick bites were recorded over the past 5 years in the Lyme borreliosis group and over the entire lifespan for healthy controls. ‖Concomitant diagnoses include all diagnoses recorded; for patients with Lyme borreliosis, these were self-reported via an adapted version of the Treatment Inventory of Costs in Patients questionnaires and, for healthy controls, a medical history was taken by the investigators. **The mild category of immunosuppressive medication includes (a combination of) non-steroid anti-inflammatory drugs, inhalation and topical corticosteroids, antihistamines, and miscellaneous (mesalazine, colchicine, allopurinol, and budesonide); high category group includes disease-modifying antirheumatic drugs, high-dose corticosteroids (eg, prednisone [>7·5 mg per day]), and biologicals (alirocumab); use of highly immunosuppressive medication was an exclusion criterion for healthy controls. ††Comparison with no immunosuppressive medication.

Table 1: **Participant characteristics at baseline**

As controls, we recruited 228 healthy controls and 41 controls with CRCs. There were no significant differences in baseline characteristics between patients with Lyme borreliosis and healthy controls (table 1). Baseline characteristics for controls with CRCs are shown in the appendix (p 5).

Patients with Lyme borreliosis and healthy controls provided 966 blood samples. Flowcharts indicating the number of samples processed for each test are given in the appendix (pp 11–16).

As a consequence of mimicking the real-world situation for iSpot or LTT-MELISA, which is often dependent on overnight shipment to the manufacturer's laboratory, we were unable to ship samples arriving on days that did not

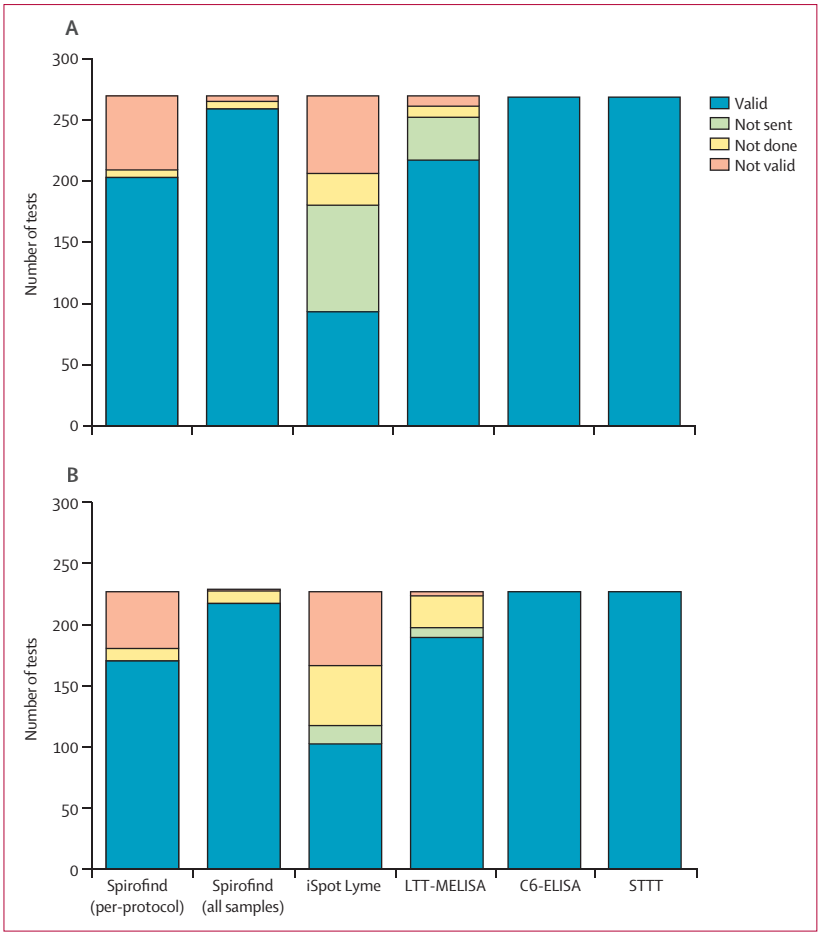

Figure 2: **Distribution of valid versus not-valid results per cellular test**
(A) Tests in patients with Lyme borreliosis at baseline. (B) Tests in participants without Lyme borreliosis (healthy controls).

	Sensitivity at baseline (all patients with Lyme borreliosis)		Specificity at baseline (participants without Lyme borreliosis [healthy controls])	
	True positive, n; N	Sensitivity (95% CI), %	True negative, n; N	Specificity (95% CI), %
Spirofind (per-protocol)	88; 204	43·1% (36·4–50·4)	140; 171	81·9% (76·1–87·2)
Spirofind (all samples)	116; 260	44·6% (38·7–50·6)	174; 216	80·6% (75·3–85·5)
iSpot Lyme (primary)	51; 94	54·3% (44·5–63·7)	32; 103	31·1% (21·5–40·3)
iSpot Lyme (alternate)	11; 94	11·7% (5·5–18·6)	79; 103	76·7% (67·3–84·5)
LTT-MELISA (primary)	66; 218	30·3% (23·8–36·7)	100; 190	52·6% (44·9–60·3)
LTT-MELISA (alternate)	42; 218	19·3% (14·1–25·0)	130; 190	68·4% (61·2–75·0)
C6-ELISA (primary)	135; 270	50·0% (44·5–55·5)	212; 228	93·0% (89·2–96·4)
C6-ELISA (alternate)	126; 270	46·7% (41·1–52·3)	214; 228	93·9% (90·5–97·1)
Standard two-tier testing (STTT C6-ELISA and immunoblot)	76; 270	28·1% (23·0–33·6)	216; 228	94·7% (91·5–97·7)

This table omits tests results that were missing or non-valid for any reason. Details on missing tests results are given in the appendix (p 6). The primary interpretation classified equivocal results as positive, and the alternate interpretation classified equivocal results as negative.

Table 2: **Diagnostic parameters of all tests at baseline**

permit next-day processing. The proportion of tests that were not valid, of the total number of tests done for patients with Lyme borreliosis and healthy controls was significantly higher for iSpot (273 [42·8%] of 638 samples) than for any other test (LTT-MELISA: 23 [2·8%] of 832, p<0·0001; Spirofind [per-protocol interpretation]: 202 [21·7%] of 930, p<0·0001; and STTT: 0 [0%] of 270, p<0·0001). The iSpot's high proportion of not-valid tests did not seem to be solely explained by delays in transit, because we found that 247 (43·9%) of 563 iSpot tests without shipping delay were also deemed not valid by the manufacturer or were not able to analysed. Spirofind and STTT were done in-house, of which 727 (75·3%) of 966 samples could be evaluated with Spirofind and 965 (99·9%) of 966 samples could be evaluated with STTT. Detailed information on missing test results is provided in figure 2 and the appendix (pp 6–7).

Diagnostic parameters of all tests for patients with Lyme borreliosis and healthy controls at baseline are depicted in table 2; data on the reactivity of the tests at 6 weeks and 12 weeks after inclusion are provided in table 3.

iSpot had the highest sensitivity at baseline (54·3% [95% CI 44·5–63·7]), but this dropped to 11·7% (5·5–18·6) when interpreting equivocal results as negative. Of all interpretable iSpot tests from patients with Lyme borreliosis and healthy controls, 160 (43·8%) of 365 samples were reported as equivocal by the manufacturer. The sensitivity of LTT-MELISA was 30·3% (23·8–36·7) in the primary interpretation, dropping to 19·3% (14·1–25·0) when equivocal results were deemed negative. Spirofind's per-protocol interpretation and its all-samples interpretation yielded similar results, 43·1% (36·4–50·4) and 44·6% (38·7–50·6), respectively. STTT had the lowest sensitivity of all tests (28·1% [23·0–33·6]). STTT's low sensitivity was primarily due to a substantial proportion of positive or equivocal C6-ELISA results not confirmed by immunoblot. The C6-ELISA itself had a sensitivity of 46·7% (44·5–55·5) in its alternate interpretation or 50·0% (41·1–52·3) in its primary interpretation. We found that the primary and per-protocol interpretations of the iSpot and Spirofind were comparable to each other (p=0·42) in terms of sensitivity, whereas both outperformed LTT (iSpot *vs* LTT: p=0·0064; Spirofind *vs* LTT: p=0·0065) and STTT (iSpot *vs* STTT: p=0·0094; Spirofind *vs* STTT: p=0·0012). STTT was comparable to LTT-MELISA (LTT *vs* STTT: p=0·91).

Serological tests had the highest specificity, with STTT (94·7% [91·5–97·7]) differing only slightly from C6-ELISA as a standalone test (primary interpretation: 93·0% [89·2–96·4]; alternate interpretation: 93·9% [90·5–97·1]). Spirofind had the highest specificity among the cellular tests (all samples: 80·6% [75·3–85·5]; per-protocol: 81·9% [76·1–87·2]). As with sensitivity, iSpot and LTT had major variations in their specificity, which were dependent on the interpretation of equivocal results. The specificity of iSpot ranged from 31·1% (21·5–40·3) in the primary interpretation to 76·7% (67·3–84·5) in the alternate

	Reactivity at 6 weeks after baseline (all patients with Lyme borreliosis, N=247)		Reactivity at 12 weeks after baseline (all patients with Lyme borreliosis, N=220)		Cumulative reactivity (baseline, 6 weeks after baseline, and 12 weeks after baseline; all patients with Lyme borreliosis, N=271)	
	Positive tests, n; N	Reactivity (95% CI), %	Positive tests, n; N	Reactivity (95% CI), %	Positive tests, n; N	Reactivity (95% CI), %
Spirofind (per-protocol)	64; 184	34·8% (27·8–41·7)	76; 168	45·2% (38·0–51·9)	145; 260	55·8% (50·2–61·6)
Spirofind (all samples)	82; 233	35·2% (29·2–42·0)	89; 202	44·1% (36·8–50·9)	157; 269	58·4% (52·4–64·6)
iSpot Lyme (primary)	53; 95	55·8% (44·9–66·7)	41; 73	56·2% (45·6–67·6)	107; 176	60·8% (53·5–67·4)
iSpot Lyme (alternate)	12; 95	12·6% (6·8–19·0)	9; 73	12·3% (5·3–20·5)	32; 176	18·2% (12·8–23·3)
LTT-MELISA (primary)	101; 213	47·4% (40·4–54·2)	91; 187	48·7% (41·7–55·3)	180; 263	68·4% (61·4–75·3)
LTT-MELISA (alternate)	69; 213	32·4% (26·2–39·2)	62; 187	33·2% (26·4–39·5)	129; 263	49·0% (43·2–55·4)
C6-ELISA (primary)	102; 247	41·3% (35·9–46·6)	77; 220	35·0% (28·7–41·3)	142; 271	52·4% (46·9–58·0)
C6-ELISA (alternate)	95; 247	38·5% (32·9–43·7)	71; 220	32·3% (26·5–38·5)	131; 271	48·3% (42·8–54·1)
Standard two-tier testing (C6 and immunoblot)	55; 247	22·3% (17·9–26·9)	43; 220	19·5% (14·8–24·9)	81; 271	29·9% (24·6–35·4)

This table omits tests results that were missing or non-valid for any reason. Details on missing tests results are given in the appendix (p 7). For the cumulative reactivity, a positive result at any timepoint was interpreted as a positive cumulative result. Missing data were considered negative, unless no datapoints were available, in which case the cumulative sensitivity could not be calculated. The primary interpretation classified equivocal results as positive, and the alternate interpretation classified equivocal results as negative.

Table 3: **Reactivity of tests in patients with Lyme borreliosis at follow-up timepoints**

interpretation, whereas that of LTT varied from 52·6% (44·9–60·3) in the primary interpretation to 68·4% (61·2–75·0) in alternate interpretation. Looking only at each test's primary interpretation (iSpot and LTT-MELISA) or per-protocol interpretation (Spirofind), we found that all comparisons between the specificity of cellular tests were significant. Spirofind's specificity exceeds that of LTT (p<0·0001), while both outperformed iSpot in terms of specificity (Spirofind *vs* iSpot: p<0·0001; LTT *vs* iSpot: p=0·0008). We also found that STTT's specificity exceeded those of all cellular tests in their respective primary and per-protocol interpretations (STTT *vs* Spirofind: p=0·0001; STTT *vs* iSpot and STTT *vs* LTT: p<0·0001).

We did not observe any meaningful changes to each cellular test's reactivity in patients with Lyme borreliosis at follow-up after participants had finished their antibiotic treatment (table 3). An exception was LTT-MELISA, which showed an increase in reactivity from 30·3% to 46·4% from baseline to 6 weeks follow-up. The reactivity of the C6-ELISA and STTT decreased at follow-up. C6-ELISA reactivity at baseline (primary interpretation) was 50·0% (95% CI 44·5–55·5), decreasing to 41·3% (35·9–46·6) at 6 weeks and 35·0% (28·7–41·3) at 12 weeks. The comparisons between cellular tests at follow-up time points were not statistically tested because of the low specificity of these tests and accompanying considerable chance of a false-positive result at follow-up.

Predefined sensitivity analyses on the tests' diagnostic parameters showed no significant differences in the sensitivity of the cellular tests between patients with erythema migrans and patients with disseminated Lyme borreliosis, or in the specificity of the cellular tests in healthy controls (with or without previous potential exposure to *Borrelia*; appendix p 8). The sensitivity of both serological test methods in patients with disseminated Lyme borreliosis was significantly higher than in patients with erythema migrans (appendix p 8).

The diagnostic parameters of hypothetical combinations of the C6-ELISA and cellular tests, and that of both commercially available cellular tests, are given in the appendix (p 9), showing no added value of combining serological and cellular tests using the current assays. ROC analyses did produce any usable cut-off optimisation strategies (appendix pp 17–19).

The assays' specificity in controls with CRCs is given in the appendix p 10, showing that Spirofind (primary interpretation: 94·4% [86·2–100·0]), STTT (100% [91.4–100·0]), and the C6-ELISA (92·7% [83·0–100·0]) all performed comparably. In the primary interpretation, the specificity of LTT-MELISA in controls with CRCs (57·9% [42·1–72·9]) was lower than the aforementioned assays (LTT *vs* STTT: p<0·0001; LTTT *vs* C6-ELISA: p=0·0098; LTT *vs* Spirofind: p=0·0018), as was that of iSpot (27·3% [10·5–44·6]; iSpot *vs* C6-ELISA and Spirofind: p=0·0002; iSpot *vs* STTT: p<0·0001). LTT-MELISA and iSpot were comparable in controls with CRCs (appendix p 10).

Discussion

In this prospective study, three cellular tests for Lyme borreliosis had a lower specificity than conventional serological testing. Of the three cellular tests, Spirofind maintained the highest specificity in healthy controls (per-protocol analysis: 81·9%), but this was still lower than that of STTT (94·7%) and the C6-ELISA (primary interpretation: 93·0%), which had specificities that were

in line with findings from earlier prospective studies.[4,21] Depending on the interpretation, the iSpot or LTT-MELISA tests had specificities as low as 31·1% (for iSpot) and 52·6% (for LTT). These findings render the cellular tests unfit for clinical use at present. A similar lack of clinical suitability has been shown for other cellular tests.[22]

The sensitivity of cellular tests varied substantially between tests and interpretations. The primary interpretations of Spirofind and iSpot outperformed STTT in their in terms of sensitivity at baseline but did not exceed the sensitivity of the C6-ELISA. The sensitivity of LTT-MELISA at baseline was similar to that of STTT but lower than that of all other assays.

In theory, cellular tests could be complementary to serological assays since the combination of both test methods would provide a more complete picture of the host's immune response to *Borrelia* infection. Unfortunately, the current diagnostic parameters make such combined testing unusable for clinical practice. Any combination with the Boolean operators AND and OR would not lead to any meaningful increase in sensitivity without an unacceptable loss of specificity. Our ROC analyses did not produce any usable optimisation strategies either. It must be noted that the manufacturers' interpretation algorithms for the iSpot and LTT complicate ROC analyses, as we describe in the appendix (pp 17–19).

Another hypothesis with respect to cellular tests for Lyme borreliosis was their potential use as a test of cure.[13,19,23] However, the poor diagnostic parameters found in our study do not allow for such analyses. The low sensitivities and specificities of cellular tests imply a substantial chance that a given test result at follow-up is either a false-positive or false-negative, thus precluding any reliable conclusions about persistence or clearance of the infection. The chance of a false-positive result is a particularly important factor in weighing the value of a diagnostic test for Lyme borreliosis in general because cellular tests are frequently requested for patients with a low chance of having Lyme borreliosis.[24] Thus, the specificity weighs more heavily in the positive predictive value than the sensitivity.

Although we do not want to downplay the importance of having a sufficiently sensitive test, these considerations underscore the need for a test with adequate specificity in particular. Therefore, we conclude that there is no place in the diagnostic repertoire for the three cellular tests in their current form. This conclusion especially relates to LTT-MELISA and iSpot tests given that they have been provided to patients for years despite having unacceptably low specificities. At some commercial laboratories, patients might request these tests without any consultation with a physician and be given an expectation of health or illness on the basis of their results. Depending on the precise interpretation, the chance that a patient without Lyme borreliosis will get a false-positive result might be worse than 50%. This false-positive rate has serious consequences for patients (eg, unnecessary treatment, delayed diagnosis of other conditions, and false expectation of disease or prognosis) and the health-care system (eg, health-care seeking behaviour for a condition the patient does not have).

The observed diagnostic parameters are directly dependent on the true disease status of the participants of this study. A limitation of this study is the absence of a universally accepted reference standard for Lyme borreliosis. Even though a positive culture is definitive proof of *Borrelia* infection, culture can have low sensitivity or can be impractical to obtain, depending on the patient material used to culture[25] and the disease duration at the time the sample is obtained. The same applies to *Borrelia* PCR.[26] For these reasons, we used strict case definitions based on a combination of physician-confirmed clinical criteria, serological tests, or direct detection with culture or PCR.[19] This approach could introduce some circular reasoning into the analyses, because detection of antibodies was both an inclusion criterion and an outcome with respect to patients with disseminated Lyme borreliosis. However, seropositivity was not an inclusion criterion for patients with erythema migrans, who were all diagnosed on the basis of clinical characteristics only (sometimes supported by positive PCR or culture). Similarly, the inclusion of patients with disseminated Lyme borreliosis was based more on criteria than just serology (ie, an extensive clinical assessment, Lyme borreliosis-specific laboratory tests, and exclusion of other potential causes).

It could be argued that the observed serological reactivity of about 50% for the C6-ELISA in patients with erythema migrans was quite low compared with previous studies.[3,27–29] However, the median duration of symptoms at the first phlebotomy was 10 days for the cohort of patients with erythema migrans, whose diagnoses were all confirmed by a physician. In this cohort, the vast majority of sera (75%) were obtained within 3 weeks after onset of symptoms, when serology has been known to yield a low sensitivity.[4] Finally, it could be argued that the setting in which we studied these tests does not match their current clinical use, because patients requesting commercially available cellular tests often report long-lasting symptoms attributed to Lyme borreliosis. This argument holds some sway as far as sensitivity is concerned, as our results were based primarily on patients with early localised Lyme borreliosis. However, the substantial false-positive rate in healthy controls does not bode well for the tests' performance in patients with chronic symptoms attributed to Lyme borreliosis, as our two-gate case-control design using healthy controls has more likely overestimated than underestimated these tests' specificity.[30]

Additional limitations concern the different recall periods for tick bites between patients with Lyme borreliosis and healthy controls, and the relatively low

power of our study due to missing data. With respect to the missing data, cellular assays are generally more difficult to do than serological assays, making them more susceptible to errors. Cellular tests depend on functional immune cells, which are vulnerable to excessive movement or shaking, temperature fluctuations, and delayed processing, making them prone to non-valid results. In particular, the iSpot test had a large number of non-valid results, but our analyses showed that this was not solely related to shipping delays because non-valid results also occurred in samples that were processed on time. Although these arguments related to practicality would be unconvincing if the diagnostic parameters of these tests were superior to other test modalities, under the current circumstances, these arguments form additional reasons not to use cellular tests for Lyme borreliosis in the diagnostic repertoire.

In conclusion, the Spirofind, LTT-MELISA, and iSpot cellular tests are unsuitable for use in clinical practice in their current forms, primarily because of a low specificity. Low specificity leads to an unacceptable risk of false-positive results, particularly with respect to the LTT-MELISA and iSpot assays that are already commercially available to patients. Unfortunately, our findings suggest that patients with suspected Lyme borreliosis are no closer than before this study to an assay that can reliably diagnose Lyme borreliosis across all disease stages and can differentiate between an active and cured infection. Although we realise that patients are suffering and that our study does not provide them with a solution, our findings underscore the need to improve Lyme borreliosis diagnostics and might serve as a stepping-stone to research to that end.

Contributors

MEB and FRvdS were primarily responsible for data collection, data analyses, and writing of the manuscript. SAG, HDV, JU, NG, CDP, and HJMtH assisted in data collection. MMGL served as the lead statistician and consulted on statistical analyses. KK did various laboratory measurements. CCvdW, B-JK, LABJ, and JWH supervised the study project. All authors have read, revised, and approved the final draft of the manuscript. All authors had access to all of the data and can take responsibility for the integrity of the data, its analysis, and the Article.

Declaration of interests

The assays under study were supplied by AID/GenID (Strassberg, Germany), InVitaLab (Neuss, Germany), QIAGEN (Germantown, MD, USA), and Oxford Immunotec (Oxford, UK), either free of charge or at a reduced price; none of the authors have received any direct financial compensation from any of these companies for this project or any other project. MEB and JWH collaborate with Bio-Rad Laboratories, ZEUS Scientific, and Pfizer on unrelated projects on Lyme borreliosis. JWH collaborates with Antigen Discovery on unrelated projects on Lyme borreliosis; JWH has an application for a provisional patent related to *Borrelia* antigens pending. FRvdS and LABJ collaborate with Hycult Biotech on developing novel diagnostic tests for Lyme borreliosis. B-JK and LABJ are coinventors of the Spirofind, an experimental in-house assay for Lyme borreliosis, which is owned by Radboudumc and was licensed for development to Boulder Diagnostics (Boulder, Colorado, USA) and subsequently Oxford Immunotec (Oxford, UK) until 2018. The other authors report no competing interests.

Data sharing

Individual participant data that underlie the results can be obtained from the corresponding author on reasonable request for a period of 36 months after publication of this Article. Data will only be shared after anonymisation, or in deidentified form (subject to relevant institutional regulations and applicable law). Supporting information, such as the study protocol, statistical analysis plan, data dictionary, and informed consent forms, can be shared as well (subject to relevant institutional regulations). Metadata and contact information can be found at https://doi.org/10·21942/uva.17113355.v1/.

Acknowledgments

We thank the patient representatives associated with this project: Els Duba, Kees Niks, Gert van Dijk, and Koen van Kempen for their efforts in setting up this study (eg, design and choice of the tests to be studied) and their perspectives as patients during the study and analysis, which were greatly appreciated. We also thank the following colleagues for their valuable assistance in setting up or doing the study: Anna Tulen, Ingrid Friesema, Margriet Harms, Carla Nijhuis, Mark Jonker (National Institute for Public Health and the Environment, Bilthoven, Netherlands); Jasmin Ersöz, Dieuwertje Hoornstra, Marga Goris, Bregtje Lemkes, and Carolien Duetz (Amsterdam UMC, Amsterdam, Netherlands); Michelle Brouwer and Fidel Vos (Radboudumc, Nijmegen, Netherlands); Tizza Zomer, Yolande Vermeeren, Barend van Kooten, and Renske Wieberdink (Gelre Ziekenhuizen, Apeldoorn, Netherlands); and Henry de Vries, Jacqueline Woutersen, and Titia Heijman (GGD Amsterdam, Amsterdam, Netherlands). This study was funded by the Netherlands Organization for Health Research and Development (522050001), and cofunded by the Ministry of Health, Welfare and Sports of the Netherlands and by the charitable contributions raised by Rood voor Altijd and Minke Verstrepen, donated through the AMC Foundation (Amsterdam UMC). Neither the funding organisations or the participating commercial partners had any role in the design of the study, or the analysis and interpretation of data. JWH's work on an unrelated project was supported by a grant from the EU through the European Regional Development Fund and the Interreg North Sea Region Programme 2014–2020 as part of the NorthTick project (38-2-7-19).

References

1 Stanek G, Fingerle V, Hunfeld KP, et al. Lyme borreliosis: clinical case definitions for diagnosis and management in Europe. *Clin Microbiol Infect* 2011; **17:** 69–79.

2 Lantos PM, Rumbaugh J, Bockenstedt LK, et al. Clinical practice guidelines by the Infectious Diseases Society of America, American Academy of Neurology, and American College of Rheumatology: 2020 guidelines for the prevention, diagnosis, and treatment of Lyme disease. *Neurology* 2021; **96:** 262–73.

3 Baarsma ME, Schellekens J, Meijer BC, et al. Diagnostic parameters of modified two-tier testing in European patients with early Lyme disease. *Eur J Clin Microbiol Infect Dis* 2020; **39:** 2143–52.

4 Leeflang MM, Ang CW, Berkhout J, et al. The diagnostic accuracy of serological tests for Lyme borreliosis in Europe: a systematic review and meta-analysis. *BMC Infect Dis* 2016; **16:** 140.

5 Jaulhac B, Saunier A, Caumes E, et al. Lyme borreliosis and other tick-borne diseases. Guidelines from the French scientific societies (II). Biological diagnosis, treatment, persistent symptoms after documented or suspected Lyme borreliosis. *Med Mal Infect* 2019; **49:** 335–46.

6 Hammers-Berggren S, Lebech AM, Karlsson M, Svenungsson B, Hansen K, Stiernstedt G. Serological follow-up after treatment of patients with erythema migrans and neuroborreliosis. *J Clin Microbiol* 1994; **32:** 1519–25.

7 Kalish RA, McHugh G, Granquist J, Shea B, Ruthazer R, Steere AC. Persistence of immunoglobulin M or immunoglobulin G antibody responses to Borrelia burgdorferi 10–20 years after active Lyme disease. *Clin Infect Dis* 2001; **33:** 780–85.

8 Strle K, Drouin EE, Shen S, et al. Borrelia burgdorferi stimulates macrophages to secrete higher levels of cytokines and chemokines than *Borrelia afzelii* or *Borrelia garinii*. *J Infect Dis* 2009; **200:** 1936–43.

9 Salazar JC, Duhnam-Ems S, La Vake C, et al. Activation of human monocytes by live *Borrelia burgdorferi* generates TLR2-dependent and -independent responses which include induction of IFN-beta. *PLoS Pathog* 2009; **5:** e1000444.

10 Bachmann M, Horn K, Rudloff I, et al. Early production of IL-22 but not IL-17 by peripheral blood mononuclear cells exposed to live *Borrelia burgdorferi*: the role of monocytes and interleukin-1. *PLoS Pathog* 2010; **6:** e1001144.

11 Baarsma ME, van de Schoor FR, van den Wijngaard CC, Joosten LAB, Kullberg BJ, Hovius JW. The initial QuantiFERON-Lyme prototype is unsuitable for European patients. *Clin Infect Dis* 2021; **73:** 1125–26.

12 Raffetin A, Saunier A, Bouiller K, et al. Unconventional diagnostic tests for Lyme borreliosis: a systematic review. *Clin Microbiol Infect* 2020; **26:** 51–59.

13 Arnaboldi PM, D'Arco C, Hefter Y, et al. Detection of IFN-γ secretion in blood samples collected before and after treatment of varying stages of Lyme disease. *Clin Infect Dis* 2021; **73:** 1484–91.

14 Dessau RB, Fingerle V, Gray J, et al. The lymphocyte transformation test for the diagnosis of Lyme borreliosis has currently not been shown to be clinically useful. *Clin Microbiol Infect* 2014; **20:** 786–87.

15 Coumou J, Herkes EA, Brouwer MC, et al. Ticking the right boxes: classification of patients suspected of Lyme borreliosis at an academic referral center in the Netherlands. *Clin Microbiol Infect* 2015; **21:** 368.

16 Gezondheidsraad. Lyme onder de loep. June 26, 2013. www.gezondheidsraad.nl/documenten/adviezen/2013/06/26/lyme-onder-de-loep (accessed Jan 4, 2021).

17 Lymedisease. Number one research priority for Lyme disease? Better testing. Dec 12, 2017. https://www.lymedisease.org/top-ten-lyme-priorities/ (accessed on Dec 2, 2021)

18 Cruickshank M, O'Flynn N, Faust SN. Lyme disease: summary of NICE guidance. *BMJ* 2018; **361:** k1261.

19 van de Schoor FR, Baarsma ME, Gauw SA, et al. Validation of cellular tests for Lyme borreliosis (VICTORY) study. *BMC Infect Dis* 2019; **19:** 732.

20 Bouwmans C, De Jong K, Timman R, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). *BMC Health Serv Res* 2013; **13:** 217.

21 Waddell LA, Greig J, Mascarenhas M, Harding S, Lindsay R, Ogden N. The accuracy of diagnostic tests for Lyme disease in humans, a systematic review and meta-analysis of North American research. *PLoS One* 2016; **11:** e0168613.

22 van Gorkom T, Voet W, Sankatsing SUC, et al. Prospective comparison of two enzyme-linked immunosorbent spot assays for the diagnosis of Lyme neuroborreliosis. *Clin Exp Immunol* 2019; **199:** 337–56.

23 Callister SM, Jobe DA, Stuparic-Stancic A, et al. Detection of IFN-γ secretion by T cells collected before and after successful treatment of early Lyme disease. *Clin Infect Dis* 2016; **62:** 1235–41.

24 Coumou J, Hovius JW, van Dam AP. *Borrelia burgdorferi* sensu lato serology in the Netherlands: guidelines versus daily practice. *Eur J Clin Microbiol Infect Dis* 2014; **33:** 1803–08.

25 Lohr B, Fingerle V, Norris DE, Hunfeld KP. Laboratory diagnosis of Lyme borreliosis: Current state of the art and future perspectives. *Crit Rev Clin Lab Sci* 2018; **55:** 219–45.

26 van Dam AP. Molecular diagnosis of Borrelia bacteria for the diagnosis of Lyme disease. *Expert Opin Med Diagn* 2011; **5:** 135–49.

27 Ang CW, Brandenburg AH, van Burgel ND, et al. A Dutch nationwide evaluation of serological assays for detection of Borrelia antibodies in clinically well-defined patients. *Diagn Microbiol Infect Dis* 2015; **83:** 222–28.

28 Smismans A, Goossens VJ, Nulens E, Bruggeman CA. Comparison of five different immunoassays for the detection of *Borrelia burgdorferi* IgM and IgG antibodies. *Clin Microbiol Infect* 2006; **12:** 648–55.

29 Branda JA, Strle K, Nigrovic LE, et al. Evaluation of modified 2-tiered serodiagnostic testing algorithms for early Lyme disease. *Clin Infect Dis* 2017; **64:** 1074–80.

30 Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem* 2005; **51:** 1335–41.