Commercial test kits for detection of Lyme borreliosis: a meta-analysis of test accuracy

Michael J Cook1
Basant K Puri2
1Independent researcher, Dorset, UK; 2Department of Medicine, Hammersmith Hospital, Imperial College London, London, UK

Abstract: The clinical diagnosis of Lyme borreliosis can be supported by various test methodologies; test kits are available from many manufacturers. Literature searches were carried out to identify studies that reported characteristics of the test kits. Of 50 searched studies, 18 were included where the tests were commercially available and samples were proven to be positive using serology testing, evidence of an erythema migrans rash, and/or culture. Additional requirements were a test specificity of \( \geq 85\% \) and publication in the last 20 years. The weighted mean sensitivity for all tests and for all samples was 59.5%. Individual study means varied from 30.6% to 86.2%. Sensitivity for each test technology varied from 62.4% for Western blot kits, and 62.3% for enzyme-linked immunosorbent assay tests, to 53.9% for synthetic C6 peptide ELISA tests and 53.7% when the two-tier methodology was used. Test sensitivity increased as dissemination of the pathogen affected different organs; however, the absence of data on the time from infection to serological testing and the lack of standard definitions for “early” and “late” disease prevented analysis of test sensitivity versus time of infection. The lack of standardization of the definitions of disease stage and the possibility of retrospective selection bias prevented clear evaluation of test sensitivity by “stage”. The sensitivity for samples classified as acute disease was 35.4%, with a corresponding sensitivity of 64.5% for samples from patients defined as convalescent. Regression analysis demonstrated an improvement of 4% in test sensitivity over the 20-year study period. The studies did not provide data to indicate the sensitivity of tests used in a clinical setting since the effect of recent use of antibiotics or steroids or other factors affecting antibody response was not factored in. The tests were developed for only specific \textit{Borrelia} species; sensitivities for other species could not be calculated.

Keywords: test sensitivity, 2 tier test, two-tier test, ELISA test, Western Blot, test specificity

The disease
Lyme borreliosis (LB) is an arthropod-borne zoonosis caused by several species of \textit{Borrelia} bacteria. It can present with a wide range of symptoms, which are common to a number of other diseases. Soon after infection, the symptoms can be nonspecific and include fatigue. A characteristic erythema migrans (EM) dermatological lesion presents sometimes soon after infection in some cases of infection. Later symptoms reflect inflammation caused to specific organs and include cognitive dysfunction, arthralgia, myalgia, neuropathy, and ophthalmic and auditory symptoms.

The pathogen
LB is caused by bacteria of the genus \textit{Borrelia}, of the family Spirochaetaceae in the phylum Spirochaetes, and is named after the French physician and biologist Amédée...
Borrel (1867–1936) who worked on the classification of spirochetes and carried out studies on, in particular, *Borrelia gallinarum* (then named *Spirillum gallinarum*).1 At the time of writing, 21 species of *Borrelia* in the *Borrelia burgdorferi sensu lato* complex related to LB have been identified, many of which are known or potential pathogens to humans.2 They are adapted to be able to move between arthropod and mammalian hosts and express different outer surface proteins to avoid the immune systems of each host.

**Disease stages**

LB is a multisystemic infection, and dissemination from the site of a tick bite occurs rapidly by transportation of the spirochetes through the bloodstream. Invasion of specific organs and tissues may occur randomly and with no definite order of progression. Inflammation of specific organs and expression of symptoms can be dependent upon preexisting conditions. Each study assessed for the present meta-analysis classified samples into disease stages, sometimes including acute, chronic, early disseminated, early and late neurological, and early and late arthritis. There was no standard definition for any of these stages. The length of time for early and late stages was not uniformly defined and was frequently not defined at all. A discussion of LB transmission and difficulties in defining time of infection has been discussed in a prior publication by one of the authors (MJC).3

The possibility of bias resulting from the retrospective nature of the studies has been discussed in a recent analysis of both commercial and developmental test kits, including those in this analysis; all case–control studies (59 out of 59) had a high probability of bias in patient sampling and a high risk of bias in the timing and flow of the disease.4 Since there was no standardization for definitions for the disease stages, including timing for early and late disease, in the present study, the definitions provided by each author have been used together with summaries generated by combining symptoms where possible. However, because of these uncertainties, data for overall sensitivity for specific test kits and methodologies will have more validity than the data by disease stages.

**Test methods for detection of Borrelia infections**

**Microscopy**

Microscopy has been the “gold standard” method for detecting and classifying bacterial infections since the father of microbiology, Antonie van Leeuwenhoek described bacteria viewed through his microscope in 1676. The method is still widely used for the diagnosis of disease and is widely used in the field of LB research during preparation of cultures for experimental work and for the production and control of materials for use in the manufacture of antigens, and occasionally in clinical investigations.5 With culture enhancement based on a polymerase chain reaction (PCR) technique, the yield of blood cultures with microscopic detection can be increased to over 70%.6 The method is not generally used for clinical diagnosis as it requires highly trained technicians. A new fluorescent assay has been developed that makes the method more definitive using a molecular beacon-based multiplex real-time quantitative PCR assay to detect specific bacterial species.7

The majority of tests carried out for diagnosis of LB depend upon detecting antibodies created in response to the infection.

**Enzyme-linked immunosorbent assay and immunofluorescence assay**

The enzyme-linked immunosorbent assay (ELISA) is the primary test normally used to evaluate blood samples; it is used sometimes as a standalone test or as the first stage of a two-tier test. There are a number of embodiments of the basic technology used in research and commercial test kits. They are based on using native antigens derived from “whole cell” extracts, recombinant antigens, or synthetic peptides based on a conserved region of the VlsE (the variable surface antigen) of *B. burgdorferi* for the detection of immunoglobulins.8 The tests are quantitative and give an indication of the concentration of antibodies in the sample. The Centers for Disease Control and Prevention (CDC) recommend that an immunofluorescence assay can be used instead of an ELISA as the first test in two-tier testing for Lyme disease. This indirect method has a number of limitations compared with the ELISA and is rarely used in human diagnostic testing for LB, although it is still used in veterinary practice in a number of European countries to detect antibodies against *Borrelia* species.9

**Western blot**

A second test commonly in use is the Western blot (WB), originally developed by Towbin et al.10,11 The majority of commercial *Borrelia* test kits use electrophoresis to separate antigens from known positive samples across a polymer strip. This is then processed with serum samples, and *Borrelia* antibodies that match the antigens on the strip will bind. After incubation using a substrate that will be dyed, the antibody/antigen complex can be visualized.
The dark bands on the strip can be compared with a standard provided by the manufacturer, and the density and scoring bands can be judged as positive if the density is greater than a reference band. There are many antibodies that can be expressed after infection with *Borrelia*. The probability that a given one is present and detectable by the test can be very low. Of more than 12 antibodies that can be present, the probability of detection of specific bands varies from a low of 0% to a high of 68%.12 As a result, positive interpretation of a WB strip requires more than one band to be present. For IgG test interpretation, some manufacturers specify that five or more of 10 critical bands must be positive for the test to be considered positive. Guidelines for IgM tests usually require that two or more bands are positive; however, a modified MiQ 12 2000 standard for Germany required only one IgM band for a positive diagnosis.13 There have been attempts to standardize bands and proposals to increase the sensitivity.14 Another adaptation of the technology uses antigens that are printed in strips on the polymer substrates, which are then reacted with patient serum and are more easily read manually or using a scanning densitometer. The selection of optical density thresholds is empirical and governed by the need to avoid cross-reactivity with other diseases while attempting to maintain a viable sensitivity. Variations of electroblotting include double blotting, pressure double blotting, electro-double blotting, slice blotting, tissue printing, native electrophoresis and Western blot (NEWB), grid-immunoblotting, multiple antigen blot assay, multiple tissue WB, dot-immunobinding assay, the use of polymer immunocomplexes, blotting from PhastGel via ultrasound, and Eastern blotting.15 The WB kits are used as standalone tests or as the second stage of the two-tier protocol.

Selection of *Borrelia* species for extraction of antigens is based on “local” sources and a limited number of species. Test kits used in the USA typically use antigens from the B31 strain of *B. burgdorferi sensu stricto*, whereas those in Europe usually also include antigens derived from *Borrelia afzelii* and *Borrelia garinii*. There are now at least nine species assigned to the LB group that are actual or probable pathogens.5,6,16–20 Sensitivity of WB tests is known to vary depending on the species present in the sample. One study indicates that when an in-house test was prepared from local *Borrelia* species, the test was significantly more sensitive than a commercial test kit in use in that region.21 No published studies have been identified that characterize the sensitivity of commercial tests for species other than those included by the manufacturer.

**Two-tier testing**

At a meeting held by the Association of State and Territorial Public Health Laboratory Directors in 1994, the CDC adopted a two-step methodology for defining LB cases for epidemiological studies.22 This requires a first-stage ELISA or immunofluorescence assay, and positive and equivocal samples are then tested using a WB confirmatory test (IgM and IgG WB if signs or symptoms have been present for up to 30 days [inclusive]; IgG WB if signs or symptoms have been present for more than 30 days). This was designed to standardize the methodology for testing and reporting LB throughout the USA and to minimize false positives. During one of the workshop sessions at this meeting, this methodology was recommended for diagnostic testing in clinical cases22 and has been widely accepted for that use, including by the Infectious Diseases Society of America, British Infection Association, European Union Concerted Action on Lyme Borreliosis, and many others.23–25

**Polymerase chain reaction**

This technique is based on direct detection of DNA by amplification of small quantities of DNA fragments from a sample and comparison with a known DNA sequence.26,27 There are a number of modifications including quantitative PCR and real-time PCR. The test is moving out of the research laboratory into use to support clinical diagnosis. The search profiles did not identify any independent evaluation data. Some are shown in this analysis for comparison, but not included in the calculations of overall test sensitivities.

**New testing technologies**

There are a number of tests currently in use by laboratories for detecting LB infections that do not use antibody detection, including the enzyme-linked immunospot, lymphocyte transformation test, and culture-enhanced microscopy. These are based on detecting immune system activation or direct visualization of the bacteria. Another identifies *Borrelia* DNA fragments using electrospray ionization mass spectrometry, and others are in development.

There is preliminary evidence suggesting that newer technologies might have some advantages. For example, in a study of 54 consecutively presenting patients carried out by one of the authors (BKP), while 45 had negative WB results by the CDC criteria, 19 (42%) of these patients were positive according to the lymphocyte transformation test-memory lymphocyte immunostimulation assay.28 None of these newer tests is included in this analysis owing to a lack of sufficient independent studies of sensitivity and specificity at the time of preparation of this manuscript.
Study selection
PubMed and Google Scholar were used with the search terms “Lyme disease OR borreliosis AND testing” to identify studies. All papers published since 1995 were selected for consideration. Data analysis and charting were carried out using software based on the Microsoft Excel platform.29
The studies were separated into two groups:

Group 1
Studies were included in the analysis where the following criteria were met:

1. Samples were proven to be positive for LB based on one or more of the following: clinical records of an EM rash; positive serology; culture; samples meeting the CDC criteria (generally being an EM rash or being two-tier positive) or CDC-certified panels with samples characterized by them as positive, negative, or equivocal. Full criteria are available for your reference in Centers for Disease Control and Prevention.30
2. The tests were commercially available.
3. The specificity was ≥85%.

These are shown in Table 1 with test method, sample type/disease stage, and performance.31–47

Group 2
Studies were excluded where one or more the inclusion criteria was absent.

These are given in detail in Table 2 which includes performance data and the reasons for exclusion. They included studies that used in-house developed tests or experimental interpretation methodologies, those where samples were from patients with suspected LB but not definitively proven, where there was no specificity data, and one study where interpretation criteria were modified from the manufacturers’ specified method were all excluded.28,46–74

Some studies included a mixture of samples, with some meeting Group 1 inclusion criteria, and other samples selected based on clinical symptoms alone. For example, Porwancher et al41 used acute and convalescent samples that were from culture-confirmed cases, and later-stage samples based solely on clinical symptoms. In such cases, only data meeting Group 1 criteria were included.

Sensitivity and specificity
The most important parameters that define the characteristics of a test are the sensitivity, which is the probability that a positive sample will be defined as positive by the test, and specificity, the probability that a negative sample will be defined as negative, usually defined as 1 – (false positive probability).

In antibody tests, there is a relationship between sensitivity and specificity as described by the receiver operating curve. As a test is made more specific, the sensitivity decreases and vice versa. This is important in the context of the studies selected in this analysis. It is possible to increase the sensitivity of a test by lowering detection thresholds; however, in doing so, the probability of false positives increases. There are many process parameters that affect the sensitivity of ELISA and WB tests including serum dilution, WB incubation times, WB antigen concentration, band selection, etc.

Where the kits provided by the manufacturer had separate IgM and IgG antibody tests, the highest sensitivity reported was used to determine study means. In some studies, the sensitivity based on either IgM or IgG antibody detection was reported, and this value was used.

Specificity data are those reported by the researchers. In some cases, the specificity was determined using blood donor samples from healthy people in LB endemic areas, in some cases from donors in areas where LB was nonendemic (for example, Branda et al),33 whereas in other cases, controls included samples from patients with known diseases that could cross-react with the test kits. Engstrom et al,44 for example, included a small group of samples from patients who had been diagnosed with other diseases including Epstein–Barr viral infection, syphilis, rheumatoid arthritis, multiple sclerosis, leptospirosis, systemic lupus erythematosus, group A streptococcal infection, and relapsing fever. The possibility exists that samples from patients with LB were included in the groups used to determine specificity. For example, many studies indicate that healthy controls from general populations and at-risk groups are seropositive for LB75,76 and that blood donor samples contain significant numbers that are seropositive.77,78 This could result in unnecessary desensitization of tests to meet specificity targets.

Data extraction and analysis
There was no standardized method for carrying out the evaluations or recording data, and so data were extracted manually from the documents and entered into Microsoft Excel worksheets. This allowed computation and preparation of a standard format giving sample size, positive samples, and percentage of positive results. This was used to define the sensitivity for each stage of disease and for each test method. To compute the overall sensitivity for all studies and for all subgroups, the weighted average of sample size and positive samples of subgroups was used.
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Patient enrollment</th>
<th>Disease stage</th>
<th>Test</th>
<th>Mean of subgroup positives</th>
<th>Subgroup sample size</th>
<th>Sens</th>
<th>Spec</th>
<th>Number of test kits</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacon et al&lt;sup&gt;31&lt;/sup&gt;</td>
<td>2003</td>
<td>CDC</td>
<td>All</td>
<td>Two-tier</td>
<td>186.3</td>
<td>280</td>
<td>66%</td>
<td>99%</td>
<td>8</td>
<td>Multiple samples from some subjects. Only IgM result included. IgG sensitivity was 31%.</td>
</tr>
<tr>
<td>Klempner et al&lt;sup&gt;37&lt;/sup&gt;</td>
<td>2001</td>
<td>CDC</td>
<td>Chronic</td>
<td>WB IgG</td>
<td>14.0</td>
<td>21</td>
<td>67%</td>
<td>100%</td>
<td>1</td>
<td>C6 data excluded with specificity of 73%. Test runs with low specificity (60.0%, 68.2%) excluded.</td>
</tr>
<tr>
<td>Mogilyansky et al&lt;sup&gt;33&lt;/sup&gt;</td>
<td>2004</td>
<td>CDC panel</td>
<td>All</td>
<td>WB</td>
<td>14.3</td>
<td>18</td>
<td>79%</td>
<td>89%</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Tilton et al&lt;sup&gt;35&lt;/sup&gt;</td>
<td>1997</td>
<td>CDC</td>
<td>All</td>
<td>WB</td>
<td>30.7</td>
<td>46</td>
<td>67%</td>
<td>Note</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Smit et al&lt;sup&gt;38&lt;/sup&gt;</td>
<td>2015</td>
<td>Seropositive</td>
<td>All</td>
<td>IgM/IgG</td>
<td>8.5</td>
<td>21</td>
<td>40%</td>
<td>99%</td>
<td>4</td>
<td>Only two-tier results included. ELISA test “indeterminate” was included with “positive” results for WB.</td>
</tr>
<tr>
<td>Branda et al&lt;sup&gt;34&lt;/sup&gt;</td>
<td>2010</td>
<td>CDC/EM</td>
<td>All</td>
<td>ELISA/WB/two-tier</td>
<td>28.2</td>
<td>45</td>
<td>69%</td>
<td>98%</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Wormser et al&lt;sup&gt;37&lt;/sup&gt;</td>
<td>2012</td>
<td>EM and/or seropositive</td>
<td>EM</td>
<td>Two-tier</td>
<td>83.0</td>
<td>158</td>
<td>53%</td>
<td>99%</td>
<td>5</td>
<td>Only two-tier results included. ELISA test “indeterminate” was included with “positive” results for WB.</td>
</tr>
<tr>
<td>Dessau&lt;sup&gt;40&lt;/sup&gt;</td>
<td>2013</td>
<td>CSF positive</td>
<td>Neuroborreliosis</td>
<td>ELISA</td>
<td>37.0</td>
<td>48</td>
<td>69%</td>
<td>97%</td>
<td>2</td>
<td>CSF-positive samples. IgG data only used.</td>
</tr>
<tr>
<td>Branda et al&lt;sup&gt;34&lt;/sup&gt;</td>
<td>2013</td>
<td>EM/culture/seropositive</td>
<td>All</td>
<td>ELISA, C6, WB</td>
<td>51.3</td>
<td>64</td>
<td>80%</td>
<td>99%</td>
<td>8</td>
<td>Data for European samples tested with US B31 test kits not included</td>
</tr>
<tr>
<td>Goossens et al&lt;sup&gt;36&lt;/sup&gt;</td>
<td>1999</td>
<td>EM</td>
<td>EM</td>
<td>ELISA, WB</td>
<td>13.9</td>
<td>26</td>
<td>48%</td>
<td>88%</td>
<td>15</td>
<td>Only IgM result included. IgG sensitivity was lower at 16%</td>
</tr>
<tr>
<td>Johnson et al&lt;sup&gt;80&lt;/sup&gt;</td>
<td>1996</td>
<td>EM</td>
<td>All</td>
<td>Two-tier</td>
<td>37.1</td>
<td>58</td>
<td>64%</td>
<td>96%</td>
<td>2</td>
<td>Only two-tier study included. See excluded studies</td>
</tr>
<tr>
<td>Tjernberg et al&lt;sup&gt;34&lt;/sup&gt;</td>
<td>2007</td>
<td>EM</td>
<td>Acute</td>
<td>ELISA, C6</td>
<td>55.8</td>
<td>158</td>
<td>35%</td>
<td>87%</td>
<td>3</td>
<td>Only commercial test included and only samples with clear definition of status of sample. Data are for samples at enrollment. Follow-up data not included.</td>
</tr>
<tr>
<td>Trevejo et al&lt;sup&gt;39&lt;/sup&gt;</td>
<td>1999</td>
<td>EM</td>
<td>All</td>
<td>Two-tier</td>
<td>37.0</td>
<td>121</td>
<td>31%</td>
<td>100%</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Binnicker et al&lt;sup&gt;12&lt;/sup&gt;</td>
<td>2008</td>
<td>EM and/or culture</td>
<td>All</td>
<td>WB IgM</td>
<td>16.3</td>
<td>28</td>
<td>58%</td>
<td>95%</td>
<td>4</td>
<td>Only commercial test included and only samples with clear definition of status of sample. Data are for samples at enrollment. Follow-up data not included.</td>
</tr>
<tr>
<td>Gomes-Solecki et al&lt;sup&gt;13&lt;/sup&gt;</td>
<td>2002</td>
<td>EM and culture positive</td>
<td>EM</td>
<td>ELISA</td>
<td>56.0</td>
<td>65</td>
<td>86%</td>
<td>100%</td>
<td>2</td>
<td>Only commercial test included and only samples with clear definition of status of sample. Data are for samples at enrollment. Follow-up data not included.</td>
</tr>
<tr>
<td>Marangoni et al&lt;sup&gt;38&lt;/sup&gt;</td>
<td>2005</td>
<td>Culture</td>
<td>EM</td>
<td>ELISA, C6</td>
<td>27.7</td>
<td>45</td>
<td>40%</td>
<td>90%</td>
<td>5</td>
<td>Only commercial test included and only samples with clear definition of status of sample. Data are for samples at enrollment. Follow-up data not included.</td>
</tr>
<tr>
<td>Marangoni et al&lt;sup&gt;39&lt;/sup&gt;</td>
<td>2008</td>
<td>Culture</td>
<td>All</td>
<td>ELISA</td>
<td>44.0</td>
<td>66</td>
<td>67%</td>
<td>97%</td>
<td>4</td>
<td>Only commercial test included and only samples with clear definition of status of sample. Data are for samples at enrollment. Follow-up data not included.</td>
</tr>
<tr>
<td>Porwancher et al&lt;sup&gt;11&lt;/sup&gt;</td>
<td>2011</td>
<td>Culture</td>
<td>All</td>
<td>All</td>
<td>45.3</td>
<td>80</td>
<td>57%</td>
<td>96%</td>
<td>4</td>
<td>Only commercial test included and only samples with clear definition of status of sample. Data are for samples at enrollment. Follow-up data not included.</td>
</tr>
</tbody>
</table>

Notes: Enrollment criteria – CDC: samples met CDC criteria; EM rash and/or seropositive: using a two-tier test with approved test kits; EM: samples from patients with documentation of EM rash in clinical records; culture: samples were positive using culture and microscopy; CDC panel: samples obtained from the CDC and certified to meet the criteria defined.

Abbreviations: sens, weighted mean test sensitivity; spec, test specificity; WB, Western blot; EM, erythema migrans; CDC, Centers for Disease Control and Prevention; EIA, immunofluorescence assay; ELISA, enzyme-linked immunosorbent assay; CSF, cerebrospinal fluid; C6, synthetic C6 peptide ELISA.
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Patient enrollment</th>
<th>Disease stage</th>
<th>Test</th>
<th>Sens</th>
<th>Spec</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ang et al⁴⁸</td>
<td>2011</td>
<td>Clinical</td>
<td>All</td>
<td>ELISA, VhE, C6</td>
<td>44%</td>
<td>86%</td>
<td>Samples from patients with suspected Lyme disease. In-house developed test.</td>
</tr>
<tr>
<td>Burbelo et al⁴⁹</td>
<td>2010</td>
<td>Clinical</td>
<td>All</td>
<td>VOOVO, C6, ELISA</td>
<td>98%</td>
<td>99%</td>
<td>Samples from patients with suspected Lyme disease. Duplicate samples from patients. Kits supplied by manufacturers.</td>
</tr>
<tr>
<td>Busson et al⁵²</td>
<td>2012</td>
<td>All</td>
<td>ELISA, WB, two-tier</td>
<td>98%</td>
<td>91%</td>
<td>In-house tests. Work carried out in 1992, tests not commercially available. Interlaboratory study. Cross-reactive samples not used for some specificity characterization.</td>
<td></td>
</tr>
<tr>
<td>Coulter et al⁵¹</td>
<td>2005</td>
<td>Suspected</td>
<td>All</td>
<td>ELISA</td>
<td>45%–72%</td>
<td>ND</td>
<td>No specificity data. Study of culture sensitivity and algorithm using culture plus PC.</td>
</tr>
<tr>
<td>Craven et al⁵³</td>
<td>1996</td>
<td>CDC</td>
<td>All</td>
<td>ELISA/WB</td>
<td>57%–100%</td>
<td>71%–98%</td>
<td>In-house tests. Work carried out in 1992, tests not commercially available. Interlaboratory study. Cross-reactive samples not used for some specificity characterization.</td>
</tr>
<tr>
<td>Dessau et al⁵⁵</td>
<td>2010</td>
<td>Clinical</td>
<td>All</td>
<td>ELISA local antigens</td>
<td>9.2%</td>
<td>ND</td>
<td>Data are best-case IgM tests. IgG was lower sensitivity, specificity was 98%.</td>
</tr>
<tr>
<td>Engstrom et al⁵⁴</td>
<td>1995</td>
<td>EM documented</td>
<td>Early and EM</td>
<td>In house</td>
<td>43%</td>
<td>97%</td>
<td>Test development.</td>
</tr>
<tr>
<td>Eshoo et al⁵⁵</td>
<td>2012</td>
<td>EM/CDC</td>
<td>All</td>
<td>PCR/EIS-MS</td>
<td>62%</td>
<td>100%</td>
<td>PCR ESI-MS.</td>
</tr>
<tr>
<td>Goetnner et al⁵⁶</td>
<td>2005</td>
<td>EM</td>
<td>Western blot</td>
<td>81%</td>
<td>99%</td>
<td></td>
<td>In-house developed test.</td>
</tr>
<tr>
<td>Heikkilä et al⁵⁷</td>
<td>2003</td>
<td>Seropositive</td>
<td>Arthritis</td>
<td>ELISA</td>
<td>37%–96%</td>
<td>ND</td>
<td>Results from experiments with eleven different antigens.</td>
</tr>
<tr>
<td>Johnson et al⁶⁰</td>
<td>1996</td>
<td>Clinical</td>
<td>All</td>
<td>Culture/two-tier</td>
<td>81%</td>
<td>90%</td>
<td>Possible sample bias described by authors.</td>
</tr>
<tr>
<td>Koidl et al⁵⁹</td>
<td>2013</td>
<td>Clinical</td>
<td>All</td>
<td>IgM</td>
<td>32%</td>
<td>ND</td>
<td>Samples from patients with suspected Lyme disease. In-house developed test.</td>
</tr>
<tr>
<td>Lawrenz et al⁵¹</td>
<td>1999</td>
<td>All</td>
<td>EM</td>
<td>ELISA</td>
<td>70%</td>
<td>98%</td>
<td>In-house developed test.</td>
</tr>
<tr>
<td>Liang et al⁵⁸</td>
<td>1999</td>
<td>Seropositive</td>
<td>C6</td>
<td>89%</td>
<td>99%</td>
<td></td>
<td>In-house developed test.</td>
</tr>
<tr>
<td>Liang et al⁵⁹</td>
<td>2000</td>
<td>All</td>
<td>C6</td>
<td>84%</td>
<td>99%</td>
<td></td>
<td>In-house developed test.</td>
</tr>
<tr>
<td>Liu et al⁵³</td>
<td>2013</td>
<td>All</td>
<td>ELISA, WIB</td>
<td>67%</td>
<td>94%</td>
<td></td>
<td>In-house developed test.</td>
</tr>
<tr>
<td>Nordberg et al⁵⁵</td>
<td>2012</td>
<td>Well characterised</td>
<td>Neuroborreliosis</td>
<td>Elispot</td>
<td>36%</td>
<td>82%</td>
<td>Samples from patients with suspected Lyme disease. Culture test not commercially available.</td>
</tr>
<tr>
<td>Nowakowski et al⁵⁶</td>
<td>2001</td>
<td>EM/CDC</td>
<td>EM</td>
<td>Culture</td>
<td>48%</td>
<td>ND</td>
<td>Culture test not commercially available.</td>
</tr>
<tr>
<td>Nowakowski et al⁵⁶</td>
<td>2001</td>
<td>EM/CDC</td>
<td>EM</td>
<td>PCR</td>
<td>72%</td>
<td>ND</td>
<td>PCR tests not generally available.</td>
</tr>
<tr>
<td>Nowakowski et al⁵⁶</td>
<td>2001</td>
<td>EM/CDC</td>
<td>EM</td>
<td>Two-tier</td>
<td>68%</td>
<td>ND</td>
<td>Specificity not determined in the study.</td>
</tr>
<tr>
<td>Puri et al⁵⁹</td>
<td>2014</td>
<td>Clinical</td>
<td>Consecutive presentation</td>
<td>LTT MELISA</td>
<td>ND</td>
<td>ND</td>
<td>Study compared WB and LTT test results. Absolute sensitivity and specificity were not investigated.</td>
</tr>
<tr>
<td>Rebman et al⁵⁵</td>
<td>2015</td>
<td>EM rash &lt;5 cm</td>
<td>EM or Clinical</td>
<td>All</td>
<td>41%</td>
<td>ND</td>
<td>Commercial laboratory. No data for specificity. 2-, 3-, and 4-band interpretation criteria, not 5 band.</td>
</tr>
<tr>
<td>Robertson et al⁵⁶</td>
<td>2000</td>
<td>EM or Clinical</td>
<td>EM</td>
<td>All</td>
<td>19%–100%</td>
<td>44%–100%</td>
<td>Protein-based study for Borrelia species.</td>
</tr>
<tr>
<td>Ryffel et al⁵⁷</td>
<td>1999</td>
<td>Seropositive</td>
<td>All</td>
<td>WB</td>
<td>ND</td>
<td>ND</td>
<td>Test development.</td>
</tr>
<tr>
<td>Schulte-Spetchel et al⁵⁵</td>
<td>2003</td>
<td>Seropositive</td>
<td>Neuro</td>
<td>Whole cell blot</td>
<td>86%</td>
<td>86%</td>
<td>Test development.</td>
</tr>
<tr>
<td>Seriburi et al⁵⁹</td>
<td>2012</td>
<td>Clinical</td>
<td>All</td>
<td>two-tier</td>
<td>ND</td>
<td>ND</td>
<td>False-positive study of clinical cases. No sensitivity data.</td>
</tr>
<tr>
<td>Skarpaas et al⁵⁰</td>
<td>2007</td>
<td>Definite LNB</td>
<td>Neuro</td>
<td>Quick C6 Sera</td>
<td>98%</td>
<td>61%</td>
<td>Excluded for low specificity (61% with serum). Only data for in-house test is available.</td>
</tr>
<tr>
<td>Skogman et al⁵¹</td>
<td>2008</td>
<td>CSF positive</td>
<td>NB</td>
<td>ELISA</td>
<td>82%</td>
<td>100%</td>
<td>Borderline results counted positive. Specificity less than 85%.</td>
</tr>
<tr>
<td>Smimsans et al⁵⁶</td>
<td>2006</td>
<td>Seropositive</td>
<td>All</td>
<td>Study mean</td>
<td>77%</td>
<td>84%</td>
<td>Borderline results counted positive. Specificity less than 85%.</td>
</tr>
<tr>
<td>Steere et al⁵⁷</td>
<td>2008</td>
<td>CDC/EM</td>
<td>All</td>
<td>C6/two-tier</td>
<td>45%</td>
<td>100%</td>
<td>First-stage test was in-house ELISA.</td>
</tr>
<tr>
<td>Vermeesch et al⁵⁸</td>
<td>2009</td>
<td>Seropositive/CSF</td>
<td>Neurological</td>
<td>C6</td>
<td>68%</td>
<td>ND</td>
<td>No specificity data.</td>
</tr>
<tr>
<td>Wormser et al⁵⁹</td>
<td>2008</td>
<td>CDC/EM</td>
<td>All</td>
<td>C6/two-tier</td>
<td>69.5%/38.9%</td>
<td>ND</td>
<td>No specificity data.</td>
</tr>
<tr>
<td>Wormser et al⁵⁹</td>
<td>2012</td>
<td>EM</td>
<td>All</td>
<td>All</td>
<td>63%</td>
<td>ND</td>
<td>No specificity data.</td>
</tr>
<tr>
<td>Study mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>69.3%</td>
<td>91.5%</td>
<td>Specificity mean based on limited reported data.</td>
</tr>
</tbody>
</table>

**Abbreviations:** ESI-MS, electrospray ionization mass spectrometry; sens, weighted mean test sensitivity; spec, test specificity; WB, Western blot; EM, erythema migrans; CDC, Centers for Disease Control and Prevention; ELISA, enzyme-linked immunosorbent assay; CSF, cerebrospinal fluid; PCR, polymerase chain reaction; LTT, lymphocyte transformation test; LNB, Lyme neuroborreliosis; ND, not determined; C6, synthetic C6 peptide ELISA.
Results and discussion

A summary of the tests, including sensitivity and specificity data, is shown in Figures 1–5 and Tables 3 and 4. Included are the sample sources, disease stage and test methods used, and the sensitivity confidence intervals.

The weighted mean sensitivity for all studies and for all disease stages is 59.5%, and the weighted mean specificity is 96.1%. In comparison, the mean sensitivity of the excluded Group 2 studies was 69.3%; however, 13 studies did not report specificity data, and others demonstrated specificities as low as 61.5%. Low specificity is associated with increased test sensitivity.

The mean sensitivity for individual studies varied from a low of 30.6% (Trevejo et al45) to a high of 80.2% (Mogilyansky et al40). The lowest sensitivity for a specific test was 7.4% (Branda et al33). This was a WB test used to identify IgG antibodies in samples defined by the authors as acute stage.

The most sensitive test methodology was WB with a weighted mean of studies of 62.4%, and varied from a low of 53.5% to a high of 76.6%. The weighted mean of all ELISA tests was 62.3%, which was not statistically different from the WB test, with a low of 45.0% and a high of 82.2%. The mean sensitivity of the six studies using synthetic C6 peptide ELISA tests was 53.9%, with a low of 42.1% and a high of 53.9%. Finally, the mean of the two-tier tests was 53.7%, with a low of 38.9% and a high of 67.5%. These data are shown in Table 3 and Figures 3–6.

Studies were excluded where the specificity was <85%. Four studies were excluded based on this. Smismans et al72 was excluded as the study mean specificity was 74% and sensitivity was 77%. High specificity was obtained using a two-tier test with the first-stage test optical density threshold lowered. Nordberg et al63 was excluded as this study had a specificity of 82%; the sensitivity was also low at 36% and would have depressed the study mean sensitivity. Craven et al’s52 study had a low specificity of 71% with a sensitivity of 51%, which would not significantly bias the overall results of this study. Finally, Skarpaas et al’s70 study demonstrated a high sensitivity of 98% but a very low specificity of 61%, which if used would lead to very high levels of false positives.

It is frequently stated that IgM antibody levels will be higher in early-stage disease and IgG antibody levels will be higher in later-stage disease. This was not always supported by the data. Branda et al34 reported a sensitivity of an IgM test as 80% for neuroborreliosis with the corresponding IgG test having a sensitivity of 60%. Some studies did demonstrate

Figure 1 Studies included in the analysis. Weighted sensitivity for each study.
a higher sensitivity for IgM tests with early-stage disease, for example, Goossens et al\textsuperscript{36} showed IgM sensitivity of 59% and IgG sensitivity of 48% using a whole-cell ELISA, and IgM sensitivity of 48% and IgG of 15.5% using a WB test. However, Bacon et al\textsuperscript{31} reported a sensitivity of 43.8% with acute stage disease using an IgG test and 18.8% using an IgM test, and Liu et al\textsuperscript{62} for early-stage EM rash samples showed an IgG test sensitivity of 78.8%, much higher than the IgM test at 57.7%.

Some studies defined a positive result based on either an IgM or an IgG response, whereas others reported sensitivity based on a single test, IgM or IgG. There was a significantly higher test sensitivity when the IgM and IgG positives were combined. Branda et al\textsuperscript{34} found that the sensitivity of a WB test was 53.1% based on IgM response, 67.2% for an IgG response, and 81.3% when either the IgM or IgG response was used.

In the case of early-stage disease where there was a record of an EM rash, the sensitivity based on IgM was 35%, 35% for an IgG response, and 55% for combined IgM or IgG response. Similarly, for later-stage neuroborreliosis, the sensitivity for IgM was 80%, for IgG 60%, and for combined IgM or IgG 87%. Marangoni et al\textsuperscript{38} demonstrated for culture confirmed samples an IgM sensitivity of 67%, for IgG 33%, and for combined IgM or IgG 71%.

These results indicate that laboratories where the selection of an IgM or IgG test is based on clinical records that suggest early- or late-stage disease have a lower sensitivity.
than those running both tests and accepting either test as indicative of disease.

Within the limits already discussed regarding the lack of definitions and standardization of disease stage and the high risk of bias in all studies, the tests were most sensitive at 89.7% when neurological and/or arthritic symptoms were present. The sensitivity for samples from patients at the convalescent stage of the disease was 64.5%, and the sensitivity for samples at the acute stage was 35.4%. It should be noted that these results are based on samples that were defined as positive by the definitive existence of an EM rash, culture of *Borrelia*, or positive with a prior serology test, or combinations of these. These data and confidence intervals are shown in Table 4 and Figures 6–8.

The lack of antibody response in early-stage disease is well recognized by the main guidelines, all of which define...
Figure 6 Sensitivity for acute stage samples.

Figure 7 Sensitivity of convalescent samples.

Figure 8 Sensitivity for neurological/arthritis/carditis samples.
the need to diagnose and treat LB if an EM rash is present, usually giving specific instructions that serology tests are not necessary. Two studies gave details of length of time between EM rash onset and sample extraction/testing. Marangoni et al indicate that patients enrolled in the study had a duration of EM rash ranging from 5 to 106 days with a mean of 16 days. However, only the mean sensitivity of 60.1% (95% confidence interval 38.6%–84.4%) with testing carried out at enrollment was presented. Wormser et al gave details of the “days postonset of symptoms”; however, the sensitivity data presented did not show sensitivity versus postonset time, and only overall averages were shown. The sensitivity achieved with a history of an EM rash using a synthetic C6 peptide ELISA and WB test in a two-tier protocol was 34.5%, for a false-negative rate of 65.5%.

There was no evidence that the commercial test kit sensitivity has improved significantly over time. Linear regression of data from all included studies is shown in Figure 9, which demonstrates an increase in test sensitivity of 4% points over the 20-year study period. These results lend support to the recently published conclusion of Stricker and Johnson to the effect that “FDA-cleared commercial serological testing for Lyme disease is inadequate for the diagnosis of the disease”.

**Conclusion**

All studies included in this analysis used test samples that were predefined as positive for LB infection either by a prior serological test that was positive, or a clinical record showing a history of an EM rash, or culture confirmed, or a combination of these. Based on this, the sensitivity determined for the tests with these samples would be expected to be close to 100%. This is not the case, as has been demonstrated in this analysis.

The sensitivity of an individual test was as low as 7.4%. The mean sensitivity of all test kits with all samples was 59.5%, and ranged from 30.6% to 86.2%.

The technologies with the highest sensitivity were WB, with a mean sensitivity of all test kits of 62.4%, and ELISA at 62.3%. The mean of tests using the synthetic C6 peptide ELISA was 53.9%, followed by the two-tier methodology with a sensitivity of 53.7%.

There was a lack of standardization of disease “stages” and of the definition of early and late disease. Along with the potential for bias in sample categorization, the sensitivity of the tests presented for disease stages should be considered as indicative and not definitive. The studies do demonstrate that sensitivity increases with severity of symptoms and dissemination to joints, the heart, and the central nervous system.

The sensitivities achieved by these studies do not represent test performance in clinical settings. The methods for selection of samples used in these studies eliminated samples from patients demonstrating a weak antibody response. Also, clinical samples will include those taken soon after infection before antibodies have developed; those from patients already treated with antibiotics or steroids, which suppress antibody production and depress test sensitivity; and those from patients with weakened immune systems. Quantification of the impact on test sensitivity was not possible since no data were presented by the individual study authors for the effects of any of these variables. Additional issues include quality control practices of testing facilities used by clinicians and issues such as sample shipment and storage. Currently, there are 21 named species assigned to the LB group, of which nine are proven or suspected pathogens, and other *Borrelia* species that generate Lyme-like symptoms such as *B. miyamotoi*. It was not possible to determine from the studies identified in this analysis, nor from other sources, the sensitivity of commercial test kits when patient samples were infected by species other than *B. burgdorferi sensu stricto*, *B. afzelii*, or *B. garinii*. It is recommended that studies be carried out to determine test sensitivities for other species and improved test methods actively pursued and approved for general use.

An important clinical implication of our conclusion that current Lyme testing lacks sensitivity is that many genuine cases of LB may be underdiagnosed. Based on the findings of our meta-analysis, we would recommend that clinicians do not assume that negative laboratory investigation results exclude a diagnosis of Lyme disease.

**Acknowledgment**

The first author (MJC) was diagnosed with Lyme disease in 2009.
Disclosure
The authors have no financial competing interests and report no conflicts of interest in this work.

References


