SCIENTIFIC ADVICE

A systematic literature review on the diagnostic accuracy of serological tests for Lyme borreliosis

www.ecdc.europa.eu

ECDC SCIENTIFIC ADVICE

A systematic literature review on the diagnostic accuracy of serological tests for Lyme borreliosis

This report is based on work commissioned by the European Centre for Disease Prevention and Control (ECDC), (Contracts ECD.3700 and ECD.4892-RIVM) and was coordinated by Hervé Zeller and Wim Van Bortel. It is based on a report produced by by Mariska M. G. Leeflang and Hein Sprong, with contributions from C. Wim Ang, Hanneke Berkhout, Henk Bijlmer, Afke H. Brandenburg, Nathalie van Burgel, Alje van Dam, Ram Dessau, Volker Fingerle, Joppe W. R. Hovius, Benoit Jaulhac, Mariska M. Leeflang, Bart Meijer, Wilfrid van Pelt, Joop FP Schellekens, René Spijker, Hein Sprong, Foukje F. Stelma, Gerold Stanek, and Frans Verduyn-Lunel. The original report is available upon request from ECDC.

Suggested citation: European Centre for Disease Prevention and Control. A systematic literature review on the diagnostic accuracy of serological tests for Lyme borreliosis. Stockholm: ECDC; 2016.

Stockholm, April 2016

ISBN 978-92-9193-724-0

doi 10.2900/309479

Catalogue number TQ-04-15-826-EN-N

Contents

Figures

Tables

Abbreviations

ACA	Acrodermatitis chronica atrophicans
AI	Antibody cerebrospinal fluid-serum index
CI	Confidence interval
CSF	Cerebrospinal fluid
CXCL-13	Chemokine belonging to the CXC family
DOR	Diagnostic odds ratio
EFNS	European Federation of Neurological Societies
EIA	Enzyme immunoassay
ELISA	Enzyme-linked immunosorbent assay
EM	Erythema migrans
EUCALB	European Concerted Action on Lyme borreliosis
HSROC	Hierarchical summary ROC model
IB	Immunoblot
IFA	Indirect fluorescent antibody
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IgT	Total immunoglobulin (immunoglobulin G and immunoglobulin M)
IQR	Interquartile range
LA	Lyme arthritis
LB	Lyme borreliosis
LTT	Lymphocyte transformation test
LST	Lymphocyte stimulation test
NB	Lyme neuroborreliosis
ROC	Receiver operating characteristics
SROC	Summary ROC model
WB	Western blot

Glossary

Case definition	Definition used to indicate someone with one of the forms of Lyme borreliosis
Cross reactivity controls	Controls with a condition that may cause cross-reactivity of the test
Diagnostic accuracy	Ability of a test to discriminate between persons with the disease or target condition from those without
Healthy controls	Controls without any forms of disease
Index test	Indicates the tests of interest, under evaluation
Reference standard	Formerly known as the 'gold standard'; the best test available to determine the target condition
Sensitivity	Sensitivity is the proportion of truly diseased persons who are correctly identified as diseased by the screening test
Specificity	Specificity is the proportion of truly non-diseased people who are so identified correctly by the screening test
Setting	Healthcare setting where the patients were recruited from
Target condition	Illness which the test aims to diagnose

Executive summary

Background: Any interpretation of laboratory diagnostic assays for Lyme borreliosis requires an understanding of the indications and the limitations of the currently available tests. Since the accuracy of serological tests for Lyme borreliosis varies, a critical appraisal of the current available laboratory tests for Lyme borreliosis in the EU was performed.

Aim: To make inferences about the role that serological tests may play in the diagnosis of Lyme borreliosis based on their diagnostic accuracy. A secondary aim was to investigate sources of heterogeneity in test accuracy.

Methods: The available literature on sensitivity and specificity of serological tests and lymphocyte transformation/stimulation tests used in Europe was systematically reviewed. Inclusion criteria were the evaluation of enzyme immunoassay, immunoblot or lymphocyte transformation/stimulation tests against a reference standard, and the usage of established clinical case definitions. All studies were assessed for quality, using QUADAS-2 (a tool for the systematic review of diagnostic accuracy studies). For meta-analyses, a hierarchical meta-regression method was used that incorporated both sensitivity and specificity, while taking into account the possible correlation between the two. For investigation of sources of heterogeneity, test type (commercial or in-house), immunoglobulin type, antigen type and study quality were added as covariates to the model in order to assess their effect on test accuracy.

Results: Seventy-eight of the 8026 unique titles found in the initial search were included in the study. The summary estimates of sensitivity for any Enzyme immunoassay (EIA) or Immunoblot (IB) in case-control studies were as follows: erythema migrans 0.50 (95% CI 0.40–0.61); neuroborreliosis 0.77 (95% CI 0.67–0.85); Lyme arthritis 0.96 (95% CI 0.89–0.98); acrodermatitis chronica atrophicans 0.97 (95% CI 0.94–0.99) and Lyme borreliosis-unspecified 0.73 (95% CI 0.53–0.87). The estimates for specificity were around 95%. A large heterogeneity was found in sensitivity and specificity. The heterogeneity could only be partially explained by the covariates. In the cross-sectional studies, sensitivity was similar compared to the case-control studies, whereas specificity was remarkably lower, at around 80% for both neuroborreliosis and Lyme borreliosis-unspecified. None of the other tests – two-tiered algorithms, specific antibody index, LTT or LST – outperformed either EIA or IB.

Conclusions: This review provides a systematic overview of test accuracy of serological tests used for Lyme borreliosis. The overall estimates of sensitivity and specificity should be interpreted with caution, as the results showed much variation and the included studies were at high risk of bias. The data in this review do not provide sufficient evidence to make inferences about the value of the tests for clinical practice. More information is needed on prevalence of Lyme borreliosis among those tested and the clinical consequences of a negative or positive test result. The sensitivity and specificity estimates from this review might be used to provide a first idea of the possible ranges in predictive values when the test is being used in different patient groups. Interpretation of serological tests for the diagnosis of Lyme borreliosis needs to be done with caution and is only supportive of the diagnosis in combination with a clinical presentation compatible with the established case definitions. Future research should primarily focus on more targeted clinical validation of these tests and research into appropriate use of these tests.

Background

Lyme borreliosis

Lyme borreliosis (LB), one of the most prevalent vector-borne diseases in Europe, is caused by spirochetes of the *Borrelia burgdorferi* sensu lato (s.l.) species complex, which are transmitted by several species of Ixodid ticks [1]. In Europe at least five genospecies of the *Borrelia burgdorferi* s.l. complex can cause disease, leading to a wide variety of clinical manifestations.

The most common clinical manifestation of Lyme borreliosis is erythema migrans (EM). Other symptoms in the early stage of disease may be malaise and flu-like symptoms. If the infection remains unnoticed and untreated in this early localised stage, *Borrelia burgdorferi* s.l. can spread to other tissues and organs. The second, so-called early disseminated stage of the disease causes more severe manifestations that can involve the skin, nervous system, joints, or heart. This stage is mainly characterised by neurological signs (neuroborreliosis (NB)) or joint aches (Lyme arthritis (LA)). Neuroborreliosis presents as meningitis, facial paralysis and/or severe pain in limbs or body. Lyme arthritis causes a swelling of the joints. The third stadium is late disseminated Lyme borreliosis. Typical presentations are acrodermatitis chronica atrophicans (ACA), arthritis, or more severe stages of neuroborreliosis.

Each of these clinical presentations can be seen as a distinct target condition, i.e. the disorder that a test aims to determine, as they affect different body parts and organ systems. Further, the patients suffering from these conditions may enter and move through the healthcare system following different clinical pathways. For example, patients with neuroborreliosis might be diagnosed by neurologists, and patients with acrodermatitis chronica atrophicans by dermatologists.

Available tests

Serological tests: Serology is the test of first choice, both in primary care and in more specialised settings. It measures specific antibody response which may take some weeks to develop after infection. It does not directly detect the presence of the bacteria. Further, antibodies may be present when the bacteria are not (or no longer) there. After a positive serology result, patients will generally be treated with antibiotics, while after negative serology patients will not be treated for Lyme borreliosis, followed up or referred for further diagnosis.

Assays available for serology are ELISAs (or EIAs) and indirect fluorescent antibody (IFA) assay. ELISAs are available as first, second or third generation tests. Third generation ELISAs use recombinant or synthetic antigens like C6, OspC, p100, p18, p41 and VlsE. Other assays for serology are immunoblots (IB), or western blots. These are not routine diagnostic tests as they require a specialised laboratory setting. IBs are mainly used as confirmatory tests and in case the ELISA was positive.

Apart from testing for antibody response in serum, antibody response may also be measured in cerebrospinal fluid (CSF). This can be done by standard ELISAs, but this is not a routine diagnostic procedure because invasive techniques are needed to collect CSF.

Immunological tests for cellular response: These are tests that measure cellular immunological response. Examples are T-cell activity and decrease or increase of certain lymphocyte subpopulations. These tests are labour intensive, require laboratory expertise and are still under development.

Culture: Culturing of bacteria is only being done in specialised laboratory settings. Sensitivity of culture varies considerably because the bacteria only grow under specific circumstances and are not always present in tissue or liquids. Its specificity is close to 100% (if the bacteria are isolated, it is unlikely that there will be false-positive results). Culture may be used as reference standard in accuracy studies.

PCR: PCR aims to detect the DNA of the bacteria. Positive results do not distinguish between viable or dead bacteria. PCR is used in more specialised laboratory settings. PCR can be done on skin tissue, cerebrospinal fluid, blood and serum, synovial fluid (from joints), heart tissue and urine.

Purpose of the review

European Concerted Action on Lyme borreliosis[1] (EUCALB) [2] and European Federation of Neurological Societies (EFNS guidelines) [3] have reviewed clinical presentations and laboratory diagnostic support. The EUCALB case definitions and EFNS guidelines for Lyme neuroborreliosis recommend that laboratory support should be sought for

[1] The European Concerted Action on Lyme borreliosis was succeeded by the ESCMID Study Group for Lyme Borreliosis (ESGBOR). ESGBOR provides a pan-European information resource for Lyme borreliosis based on the network of physicians and scientists that was established during EUCALB. See also: http://www.escmid.org/research_projects/study_groups/esgbor/

the clinical diagnosis of all manifestations of Lyme borreliosis other than erythema migrans, as clinical characteristics of later stage presentations are not unique to *Borrelia burgdorferi* infection (Table 1). Patients who present in primary care settings with clear EM can be treated directly with antibiotics. The diagnosis is more complicated in patients with stage two or three, and laboratory confirmation can be of added value. In all cases the clinical presentation and tick exposure risk should be carefully evaluated. Tests should only be performed on patients in whom there is a significant likelihood of Lyme borreliosis, i.e. the pre-test likelihood of infection.

Interpretation of laboratory diagnostic assays in Lyme borreliosis requires an understanding of the use and the limitations of the currently available tests. Since the reliability of serological laboratory tests for Lyme borreliosis is not always adequately evaluated, a critical appraisal of the current available serological test for Lyme borreliosis in the EU is needed as a first step to improve the diagnosis of this disease.

Table 1. Manifestations and recommended approach for the diagnosis of Lyme borreliosis in routine practice

Clinical manifestations	Primary diagnostic testing	Supporting testing and findings	Differential diagnosis
Erythema migrans			
Expanding red or bluish-red patch (≥5 cm in diameter). Advancing edge is typically distinct, often intensely coloured, and not noticeably raised	Testing is conducted on the basis of history and visual inspection of the skin lesion. If lesion is atypical, acute-phase and convalescent-phase serological testing is recommended.	Culture or PCR is not needed for routine clinical practice.	Tick-/insect-bite hypersensitive reaction, bacterial cellulitis, erysipelas, erythema multiforme, tinea, nummular eczema, granuloma annulare, contact dermatitis, urticaria, fixed drug eruption, pityriasis rosea, or parvovirus B19 infection in children
Lyme neuroborreliosis			
Mainly meningo-radiculitis, meningitis and peripheral facial palsy; rarely encephalitis, myelitis; very rarely cerebral vasculitis. In children, mainly meningitis and peripheral facial palsy.	Pleocytosis and demonstration of synthesis of intrathecal antibodies to Lyme borrelia. Serological testing usually positive at time of presentation; if negative, convalescent phase sera should be tested.	Detection of *Borrelia burgdorferi* s.l. by culture or PCR in CSF Intrathecal synthesis of total immunoglobulin Recent or concomitant erythema migrans	Other causes of facial palsy, viral meningitis, mechanical radiculopathy, first episode of relapsing-remitting multiple sclerosis, or primary progressive multiple sclerosis
Lyme arthritis			
Recurrent attacks or persisting objective joint swelling in one or more large joints. Alternative explanations should be excluded.	Serological testing. As a rule, high concentrations of specific serum IgG antibodies are present.	Detection of *Borrelia burgdorferi* s.l. by culture or PCR in synovial fluid Previous well-defined Lyme borreliosis manifestations	(Pseudo-)gout, septic arthritis, viral arthritis, psoriatic arthritis, HLA B27-positive juvenile oligoarthritis, reactive arthritis in adults, sarcoid arthritis, early rheumatoid arthritis, or seronegative spondyloarthropathies
Acrodermatitis chronica atrophicans			
Long-standing red or bluish-red lesions, usually on the extensor surfaces of extremities; initial doughy swelling; lesions eventually become atrophic; possible skin induration and fibroid nodules over bony prominences.	Serological testing. As a rule, high concentrations of specific serum IgG antibodies are present.	Histology, culture or PCR are not needed for routine clinical practice. Previous well-defined Lyme borreliosis manifestations	Consequence of old age, chilblains, (chronic) venous insufficiency, superficial thrombophlebitis, hypostatic eczema, arterial obliterative disease, acrocyanosis, livedo reticularis, lymphoedema, erythromelalgia, scleroderma lesions, rheumatoid nodules, gout (tophi), or erythema nodosum

Source: Adapted from [1]

Review methods

Review questions

The primary research questions of this systematic review were:

- What is the sensitivity and specificity of serology tests for the target conditions EM, NB, LA, ACA and LB-unspecified (i.e. Lyme borreliosis not differentiated as one of the target conditions)?
- What is the diagnostic accuracy of immunoblot (IB) tests or tests done on CSF for these target conditions?

The Secondary review question was:

- What are the sources of heterogeneity?

Search strategy

EMBASE and Medline databases were searched for eligible studies (see Annex 1 for full electronic search strategy). Grey literature and possibly missed titles were retrieved through experts.

Eligibility criteria and selection process

Target condition, reference standard and case definitions

In this study EM, NB, LA, and ACA were included. Each of these presentations may be seen as a distinct target condition and was included and analysed separately. If a study included multiple target conditions, the data were separated in the analyses. If a study did not distinguish between the different target conditions, the data of the study were included in an analysis for the target condition 'Lyme borreliosis-unspecified' (LB-unspecified).

Studies focusing on specific risk groups (e.g. forest workers as cases, compared with non-endemic controls) were excluded. Studies focusing on specific symptoms or syndromes as target condition, without any reference to how these symptoms may relate to Lyme borreliosis (e.g. uveitis or meningitis patients compared with a group of healthy controls) were excluded as well. Studies focusing on Bannwarth's syndrome were included, as this is a manifestation of NB.

The reference standard is the test or testing algorithm used to define whether someone has the target condition or not. As there is no gold standard for diagnosing Lyme borreliosis, most diagnoses are made based on clinical criteria. Studies were included regardless of reference standard. The use of case definitions as a reference standard – for example the definitions advocated by WHO, EUCALB, EFNS, or as described by Stanek et al. [1] (see also Table 1) – received a positive score in the quality assessment.

Design of the eligible studies

Cross-sectional studies would be the ideal study type to answer the review questions [4,5]. Such studies would provide valid estimates of sensitivity and specificity and would also directly provide estimates of prevalence and predictive values. Any of these studies would be included.

It was anticipated that most of the studies in the search would be case-control studies [6]. These studies estimate the sensitivity of a test in a group of cases, i.e. patients with a high likeliness of having Lyme borreliosis. The specificity is estimated in a group of controls, i.e. patients with a low likeliness of having Lyme borreliosis. The control group can be healthy volunteers (healthy controls) or patients with other diseases than Lyme (cross-reacting controls). The prevalence of Lyme borreliosis cannot be estimated based on case-control studies, and the estimates of sensitivity and specificity may not be representative for the sensitivity and specificity of the test when used in practice. As cross-sectional studies were anticipated to be very sparse, case-control studies were included despite their shortcomings.

Survey studies investigating seroprevalence were excluded, as well as studies based on samples used by the laboratories, e.g. for technical validation.

Index tests

The following types of index tests, i.e. tests under evaluation, were included:

- Enzyme-linked immunosorbent assay (ELISA, subsequently referred to as EIA)
- Immunoblot (IB)
- Two-tiered tests (algorithm, usually consisting of an ELISA and an IB)

- Specific antibody index measurement (usually done by EIA)
- Lymphocyte transformation test (LTT) or lymphocyte stimulation test (LST).

Tests performed on serum or CSF were included as were those done on the lymphocyte fraction of the blood for the LTT tests. The specific antibody index measurements are done by either EIA or IB and compare the antibody titres in serum and CSF to calculate the antibody cerebrospinal fluid-serum index (AI). AI is not always calculated the same way in different studies, but this was considered a form of threshold effect.

Both commercial and in-house based tests were included. If a study reported a number of results from a single test for several different antigens or proteins, then the study was considered to be a technical evaluation of the test and excluded.

Setting and patient population

Preferably only studies that recruited both cases and controls from the same healthcare setting should be included in the review. For example, for Lyme neuroborreliosis, the ideal study would recruit patients from a neurological department. Studies on ACA would recruit from a dermatological department. Because the studies were suspected to recruit from a variety of healthcare settings, it was decided to include studies from all healthcare settings and patient populations.

This study focusses only on European variants of Lyme borreliosis, so all studies recruiting from non-European populations were excluded.

Other considerations and exclusion criteria

Studies from which a 2x2 table containing true positives, false positives, false negatives and true negatives could not be drawn, were excluded. Further, studies that were too unspecific in their reporting to ensure that they fulfilled the above criteria, were excluded.

Assessment of quality and risk of bias

The quality of all included studies was assessed using QUADAS-2, a tool recommended by the Cochrane Collaboration to assess the quality of diagnostic test accuracy studies [7,8]. QUADAS-2 consists of four domains:

- patient selection
- index test
- reference standard
- flow and timing.

Each of these domains has a subdomain for 'risk of bias' and 'applicability', except for the last one. A number of signalling questions were used to guide the evaluation of bias [7]. See Annex 2 on how the QUADAS-2 items were applied. QUADAS-2 was scored by two assessors per study, independently from each other.

Data extraction

The data were extracted independently by two assessors. While extracting data, assessors also had to decide whether a study was a case-control or a cross-sectional study.

Data synthesis and analysis

Diagnostic accuracy

Diagnostic accuracy can be defined as the ability of a test to discriminate between persons with the disease or target condition from those without. Quantitative indicators for accuracy are sensitivity, specificity, predictive values, likelihood ratios, and diagnostic odds ratio.

In this study sensitivity, specificity and diagnostic odds ratio were used. Sensitivity is defined as the proportion of positive test results among the diseased. Specificity is the proportion of negative test results among the non-diseased. The diagnostic odds ratio summarises the diagnostic accuracy of the index test as a single indicator and describes how many times higher the odds are of obtaining a test positive result in a diseased rather than in a non-diseased person. The diagnostic odds ratio ranges from 0 to infinity, with higher values indicating better discriminatory test performance. A value of one means that the test does not discriminate between diseased and non-diseased persons.

A ROC curve of a test represents the change of sensitivity and specificity by varying positivity thresholds (cut-off values). The graph plots sensitivity (true positive rate) against 1-specificity (false-positive rate). The position of the

ROC-curve depends on the degree of overlap of the distributions of the diseased and non-diseased and helps to estimate the level of discriminatory power of a test. The closer the ROC curve is to the upper-left corner of the graph, the better the tests discriminates between diseased and non-diseased.

It is likely that a systematic review includes test results that will be at a mixture of different positivity thresholds. Therefore it is likely to assume that there is an underlying summary ROC curve to the study results. Summary ROC plots display the results of individual studies in ROC space.

- The ROC scatter plot displays the results of individual tests/studies in ROC space: each included study is plotted as a single sensitivity–specificity point. The plot depicts the scatter of the study results.
- The fitted summary ROC curve is obtained by meta-analysis methods and displays an estimated ROC curve based on the included studies, thus providing summary information on the discriminatory power of the included tests.

Meta-analysis

The hierarchical summary ROC (HSROC) model was used for the meta-analyses [9]. The HSROC model is a hierarchical meta-regression method that incorporates both sensitivity and specificity while taking into account the possible correlation between the two. The HSROC model assumes that there is an underlying summary ROC curve to the study results. This curve is defined by:

- the accuracy of a test defined in terms of the diagnostic odds ratio (DOR);
- the threshold at which the test operates; and
- the shape of the curve which provides information about how the DOR varies when the threshold varies.

From these estimates, it is possible to derive an average sensitivity and specificity which will be presented for the ease of interpretation.

The following approach was taken:

- Studies that included a 'suspected' or 'possible' category were initially included in the 'diseased' group. Excluding these groups would overestimate sensitivity and/or specificity and is therefore not recommended. However, this approach may lead to an underestimation of sensitivity. Therefore, a sensitivity analysis was conducted by comparing the results of both approaches to determine whether this decision made a difference.
- If a study reported both a 'suspected' and a 'possible' group, both were considered 'possible'. A separate analysis was done for all target conditions.
- A separate analysis was conducted for healthy controls and for cross-reacting controls, i.e. controls with a condition that may cause cross-reactivity in the test. The sensitivity was predicted to be the same for both groups, but specificity was predicted to be higher in the healthy control groups. This was checked by comparing the confidence intervals of the two results.
- One data row was entered per test. If multiple Ig types were tested, IgT was included where possible, otherwise IgM was included because IgM is considered to be more sensitive than IgG in early disease. The analyses accounted for the fact that multiple tests could have been evaluated in the same study.
- Blaauw et al. (1999) reported 'previous Lyme'. This was considered to be a cross-reactivity control group.
- One study classified NB cases by 'days after onset of neurological symptoms': <20 days; 21 to 40 days; 41 to 160 days (Hansen et al. 1988). The entire NB group from this study was included in the analysis, regardless of the number of days after onset.

Investigation of heterogeneity

Heterogeneity was investigated by adding covariates to the HSROC model. Covariates added to the HSROC model may explain the variation in the following parameters: the actual accuracy (balance between sensitivity and specificity), the threshold at which the tests operates or the shape of the curve. It does not directly explain the effect of the covariates on the diseased (and thus on sensitivity) and on the non-diseased group (and thus on specificity).

The following approach was taken:

- First, the variation in accuracy between tests detecting IgM, IgG and tests detecting both IgM and IgG (Ig total, IgT) was analysed. It was assumed that Ig type could have influenced all three parameters.
- Then, the effect of the test type (EIA or IB) and test origin (commercial or in-house) was investigated. It was assumed that differences in test type could have caused variation in all three parameters (accuracy, threshold and shape), but that the test origin (commercial or in-house) could have affected accuracy or threshold, but not shape because both the commercial and the in-house tests are based on the same test principles and may operate at different thresholds – or be more sensitive in detecting antibody –, but this will not have an effect on how test accuracy changes with varying thresholds.

- Other investigated covariates were the effect of antigens, the effect of publication year (and whether there was a relation between antigen type and publication year), and if possible, the effect of late versus early disease. Where possible, a subgroup analysis for individual tests was performed.
- The effect of the quality of the studies on the accuracy estimates was verified.

Sensitivity analyses were performed to evaluate the robustness of the results and the effect of the choices we made. Sensitivity analyses were carried out with regard to borderline results and possible Lyme cases.

Review results

Summary of the results of the search and selection

The results of the search and selection process are presented in Figure 1. The full text of 486 titles was evaluated (of an initial total of 8 026 unique titles), 118 of which were eligible for data extraction. Eventually, 76 studies were included in the first round. The search was updated on February 2014, which revealed 418 new studies. Of these, two were included in the review. In total, this review describes the results of 78 studies.

Of the 78 studies included in the review (see Annex 3 for detailed references), 60 had a case-control design which compared a group of well-defined cases with a group of healthy controls or a group of cross-reacting controls. Sixteen had a cross-sectional design. Two studies could be classified either as a case-control or cross-sectional design (Ruzic-Sabljic et al. 2002; Skogman et al. 2008): one study was included as cross-sectional design; the second study was included as a case-control study because the selection of patients differed significantly from the other cross-sectional designs.

Figure 1. Search and selection process

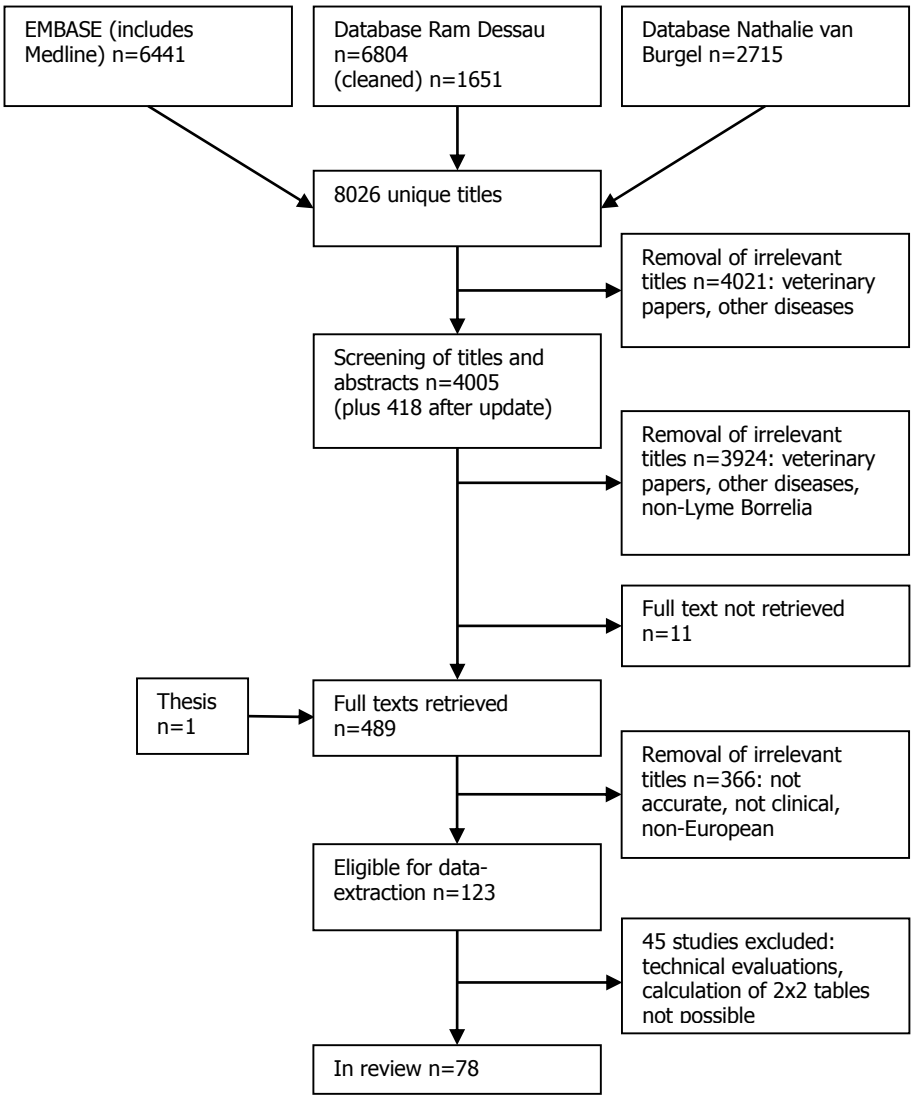


Figure 2 and Figure 3 show the general characteristics of the included studies. The majority of the studies were done in Germany (n=19), Sweden (n=12) and the Netherlands (n=10). The studies from these three countries amount to 53% of all included studies. Most studies evaluated diagnostic tests for NB patients. Erythema migrans patients formed the second largest group of patients.

Figure 2. Countries in which the studies were conducted

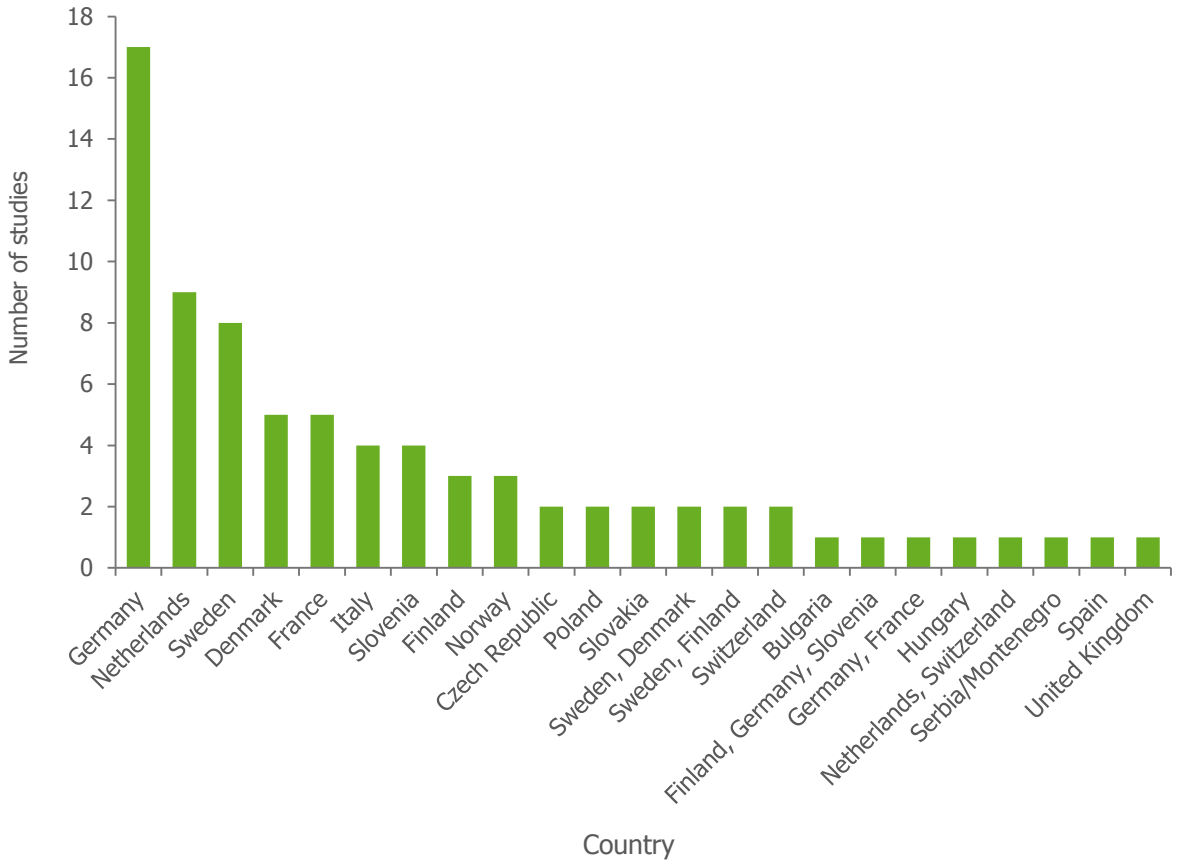
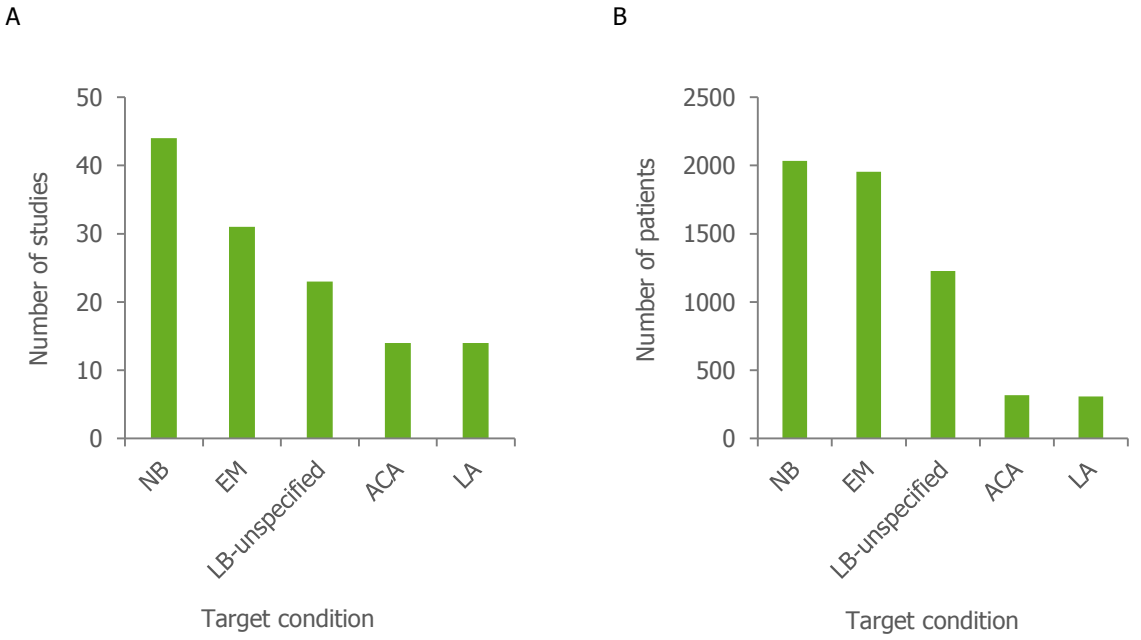


Figure 3. Number of studies (A) and number of included patients (B), by target condition



Results of the assessment of quality and risk of bias

Reliability of data extraction

On average, 59% of the items (QUADAS-2 signalling questions and study characteristics) per study received an identical score from both assessors (minimum 29%, maximum 86% per study) for cross-sectional design studies. The three items that were rated least consistently (21% agreement) were:

- 'Were the persons applying the reference standard blinded to all other test results?'
- 'Were all eligible patients enrolled consecutively or in a randomised way?'
- 'Was there an appropriate interval between the index test(s) and reference standard?'

The two items that were rated most consistently were:

- 'Do you have any concerns that these authors have a conflict of interest?' (95% agreement)
- 'Does the study involve only paediatric patients?' (100% agreement)

In three cross-sectional studies, both assessors evaluated the study as if it was a case-control study.

The two assessors disagreed about the content of 12 of the 17 2x2 tables of the cross-sectional studies. These disagreements ranged from a simple typographical error to the extraction of tables for different tests (e.g. extracting data for all separate antigens reported).

With regard to case-control studies, an average of 68% of all items per study received an identical score from both assessors (minimum 25%, maximum 94% per study). The two items that were rated least consistently were:

- 'Where was the study done?' (38% agreement)
- 'What was the stage of the disease?' (43% agreement)

The two items which were rated most consistently were:

- 'Is this a diagnostic case-control study' (100% agreement)
- 'Does the study involve only paediatric patients?' (100% agreement)

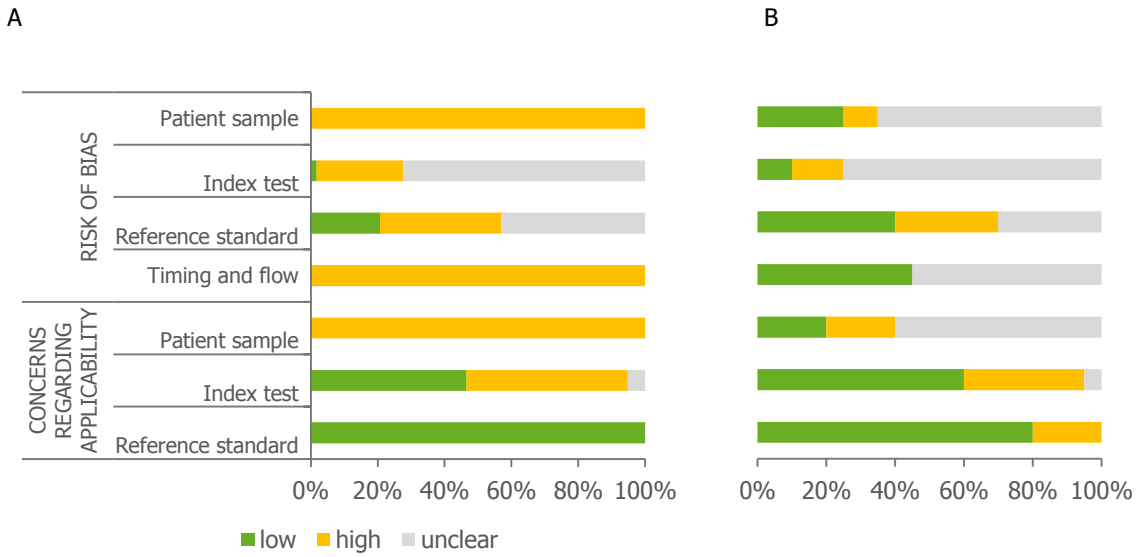
In the 27 of the 61 case-control studies there was no agreement about the 2x2 tables. Disagreements varied from a simple typo to the extraction of tables for different tests (e.g. when extracting data for all reported antigens) or different patient populations (for example by lumping all case groups together instead of analysing them separately).

Quality assessment

All case-control studies had a high risk of bias in their 'patient sampling' and there was high concern regarding 'applicability of their sample' (Figure 4, Annex 4). Although this is inherent to the way the criteria were defined, it does reflect the disadvantages of case-control studies. None of the case-control studies were rated to have a low risk of bias in all other domains. Also, none of the cross-sectional studies were rated to have a low risk of bias in all domains.

The largest problems were the selection of the patients, the reporting of study characteristics (how the study was conducted, where patients had been recruited, and whether the assessment of tests was blinded), and the flow of the patients.

Figure 4. Overview of the assessors' judgements about each methodological quality item, presented as percentages across all included studies



A: Case-control studies
B: Cross-sectional studies
Low: low risk of bias; high: high risk of bias; unclear: bias is unclear

Erythema migrans

Summary of the study characteristics

Thirty-one studies evaluated erythema migrans (EM). Thirty studies had a case-control design of which 12 included both a healthy control group (usually blood donors) and a cross-reacting control group. The results for both groups were reported separately. Seven studies only included a healthy control group and 11 included a cross-reacting control group.

An overview of the study characteristics is provided in Table 2:

- Twenty-four case control studies included EIA tests, one study included 10 different EIAs. Seven of these studies also included an IB, including two that used these tests in a two-tiered algorithm.
- None of the EM studies included a 'possible EM' category; nine included borderline test results.

The quality assessments of the studies are presented in Annex 4.

Table 2. Characteristics of the EM studies

										Number			
Study (see Annex 3 for references)	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Two-tiered test	Acceptable case definition	Serology in case definition	Setting	EM	DDxTot	HC_Tot	CC_Tot
Ang 2012	Netherlands	CC	No	10	0	No	Unclear	Unclear	Laboratory	105	0	228	212
Bergstrom 1991	Sweden	CC	Possibly	1	0	No	Unclear	Yes	Departments of microbiology and bacteriology	30	0	64	161
Branda 2013	Slovenia	CC	Possibly	3	2	Yes	Yes	Yes	University hospital	20	0	100	0
Cerar 2006	Slovenia	CC	No	1	0	No	Yes	No	Department of infectious diseases; laboratory	76	0	49	0
Christova 2003	Bulgaria	CC	No	1	1	Yes	No	No	Not reported	105	0	90	0
Flisiak 1996	Poland	CC	No	3	0	No	Yes	No	Laboratory	18	0	0	69
Goettner 2005	Germany	CC	No	0	3	No	No	Yes	Laboratory	15	0	0	110
Hansen 1989	Sweden/Denmark	CC	No	2	0	No	Yes	No	Hospital	107	0	200	98
Hernandez 2003	Spain	CC	No	0	1*	No	Yes	Unclear	Hospital laboratory	24	0	0	129
Hofmann 1990	Germany	CC	No	2	0	No	No	Unclear	Laboratory; department of dermatology	112	0	0	205
Hofmann 1996	Germany	CC	No	2	0	No	No	No	Laboratory; department of dermatology	52	0	106	50

										Number			
Study (see Annex 3 for references)	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Two-tiered test	Acceptable case definition	Serology in case definition	Setting	EM	DDxTot	HC_Tot	CC_Tot
Hunfeld 2002	Germany/France	CC	Yes	1	0	No	No	Unclear	NR	148	0	1107	275
Jovivic 2003	Serbia/Montenegro	CC	No	1	1	No	Yes	Unclear	Department of microbiology	40	0	0	120
Karlsson 1989siid	Sweden	CC	No	2	0	No	Yes	No	Laboratory	30	0	0	73
Lahdenne 2003	Finland, Germany, Slovenia	CC	No	3	0	No	No	No	Laboratory	65	0	40	0
Lencakova 2008	Slovakia	CC	No	1	1	No	Yes	Unclear	Laboratory	54	0	0	60
Marangoni 2005jmm	Italy	CC	No	3	0	No	Yes	No	Not specified	45	0	234	40
Marangoni 2005new	Italy	CC	No	2	3	No	Yes	No	Not specified	30	0	0	65
Marangoni 2008	Italy	CC	no	2	0	No	Unclear	No	Laboratory	66	0	300	100
Mathiesen 1996	Denmark	CC	No	1	1	No	Yes	Yes	Laboratory	47	0	100	29
Mathiesen 1998	Sweden/Denmark	CC	No	2	0	No	Yes	No	Laboratory	80	0	0	138
Olsson 1991	Sweden	CC	No	2	1	No	Unclear	No	Department of dermatology	100	0	100	0
Putzker 1995	Germany	CC	No	4	2*	No	Unclear	Yes	Laboratory	24	0	93	0
Rauer 1995	Germany	CC	No	1	0	No	Unclear	Yes	Laboratory	118	0	154	136
Ruzic 2002	Slovenia	CC	No	0	1	No	Unclear	No	Hospital	117	0	96	0
Ryffel 1998	Switzerland	CC	No	0	1*	No	Unclear	Yes	Laboratory	35	0	180	50
Smismans 2006	Netherlands	CC	No	3	0	No	Unclear	Yes	Laboratory	23	0	0	40
Tjernberg 2007	Sweden	CC	No	3	0	No	Yes	Yes	Hospital and laboratory	158	0	55	200
Wilske 1993	Germany	CC	Yes	2	0	No	Yes	Yes	Laboratory	31	0	100	42
Wilske 1999	Germany	CC	Possibly	0	1	No	Yes	Yes	Laboratory	66	0	0	139
Barrial 2011	France	CS	No	0	4*	No	Yes	Unclear	Laboratory	12	33		

*CC=case control design; CS=cross-sectional design; DDxTot=total number of controls in a cross-sectional design; HC_Tot=total number of healthy controls; CC_Tot=total number of cross-reacting controls; acceptable case definition=an acceptable case definition was used in accordance with international standards; serology in case definition= serology included as part of the reference standard; *=IB was done on a pre-selected sample of patients, in most cases on samples with a positive EIA.*

Erythema migrans – case-control studies with healthy controls

Overall results and methodological quality of the studies

The analyses were based on 19 studies with a total of 1 449 persons with EM and 3 396 healthy controls. The median number of cases per study was 66 (range 20 to 158), and the median number of controls was 100 (range 40 to 1 107).

The following methodological quality issues were observed:

- For most studies it was unclear whether the reference standard posed a high or low risk of bias, but this is probably less relevant for EM as it is considered to be the clearest target condition in Lyme borreliosis;
- Flow and timing of all studies posed a high risk of bias, due to the fact that case-control design tends to exclude all 'difficult-to-diagnose' patients;
- It was impossible to assess whether the execution of the index test may lead to bias in all studies except the one by Mathiesen et al. (1996) which reported that the assessment of the (commercial) index test was blinded to the disease status of the participants.
- Another problem in the execution of the index tests was that, especially for the in-house tests, the cut-off value for positive or negative test results was decided after the study was completed.

The studies by Hunfeld et al. (2002) and Wilske et al. (1993) reported potential conflict of interest because the authors worked for the company that produced the tests, i.e. Biotest AG, Behringwerke AG and Mikrogen GmbH. The studies were included in the analyses.

Results specific to Ig type

In most studies, IgM tests had a higher sensitivity compared to IgG tests for detecting EM. This was expected because EM is typically seen early in the disease (Annex 5). IgT tests had the highest sensitivity in all studies, but this was combined with the lowest specificity in most studies.

Overall, when meta-analysing the diagnostic accuracy of the studies that evaluated tests for more than one antibody type, the IgT tests are significantly more accurate than the IgM or IgG tests (P-value=0.0029 for accuracy). This is mainly due to a generally higher sensitivity (Table 3).

Table 3. Summary estimates of test accuracy by antibody type for EM case-control studies with healthy controls

Antibody type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	15.1 (9.3–24.4)	0.43 (0.36–0.49)	0.95 (0.92–0.97)
IgG	13.8 (8.8–21.6)	0.36 (0.29–0.43)	0.96 (0.94–0.98)
IgT	17.6 (11.0–28.1)	0.61 (0.50–0.70)	0.92 (0.89–0.94)

Results specific to test type, commercial versus in-house

Figure 5A represents the ROC scatter plot which showed much heterogeneity but the data points seem to follow a curve-like shape.

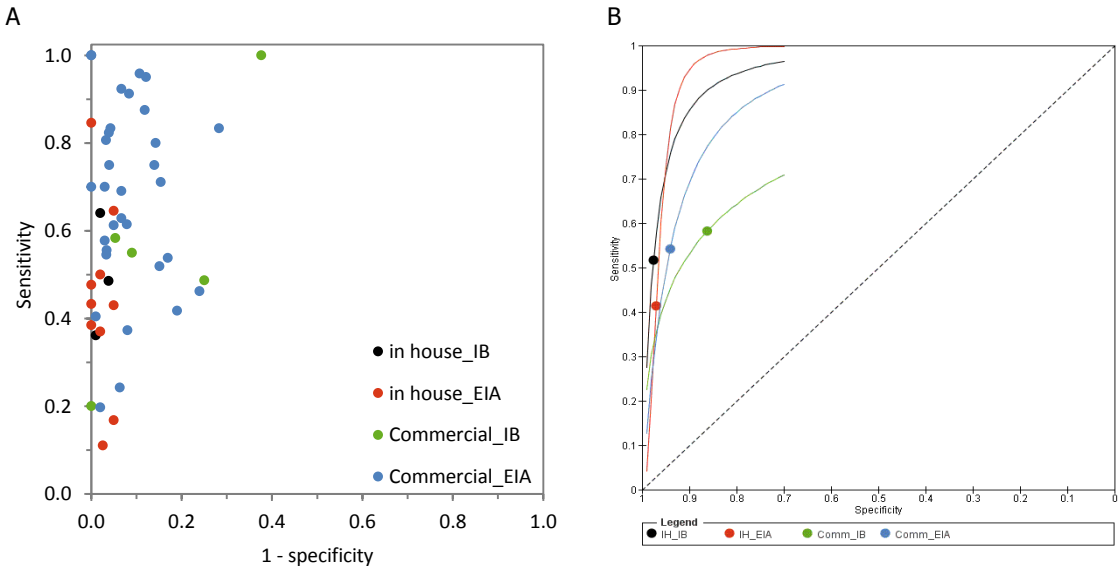
On average, for any EIA or IB test detecting EM patients, the diagnostic odds ratio (DOR) was 19.2 (95% CI 11.2 to 32.8), which coincides with a sensitivity of 0.50 (95% CI 0.40 to 0.61) and a specificity of 0.95 (95% CI 0.90 to 0.97).

When adding test type (EIA or IB) to the analyses, the model fit improved, and test type turned out to have a significant effect on accuracy (P-value=0.008) and threshold (P-value=0.03), which means that EIA may operate at a different threshold than IB tests. Addition of a covariate to the model which accounted for commercial or in-house test further improved the model fit, but did not seem to have a significant effect on either model parameter (P-values above 0.05) (Table 4 and Figure 5B).

According to the ROC scatter plot (Figure 5A), there seems to be less heterogeneity in the in-house tests than the commercial tests, even though the in-house tests were expected to be more variable. One reason for this may be that in-house tests are optimally tailored to the laboratory and setting in which they are used. Another explanation may be publication bias: results of in-house tests with a poor specificity are not published. Instead, tests are continued until a higher specificity is found, and only then the results are published.

In the analyses mentioned above, specificity between the EIAs and the IB tests was significantly different. However, these comparisons were based on both comparative studies (including IB and EIA) and non-comparative studies (containing only IB or EIA; the majority). This impacts the interpretation, as the comparison may be confounded by other factors. Perhaps the IBs were analysed in slightly different patients than the EIAs. More valid comparisons could be made if only comparative studies had been included. However, only three studies included both EIA and IB, which is not enough to compare the two test types.

Figure 5. ROC scatter plot (A) and fitted summary ROC curves (B) for EM case-control studies with healthy controls



Graph A: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Table 4. Summary estimates of test accuracy for commercial and in-house IB and EIAs for EM case-control studies with healthy controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	22.9 (9.4–55.6)	0.41 (0.25–0.60)	0.97 (0.95–0.98)
In-house IB	42.7 (13.9–131.0)	0.52 (0.38–0.66)	0.98 (0.94–0.99)
Commercial EIA	16.7 (9.6–29.0)	0.54 (0.44–0.65)	0.93 (0.90–0.96)
Commercial IB	8.7 (4.0–19.3)	0.58 (0.49–0.67)	0.86 (0.75–0.93)

Sources of heterogeneity

- **Generation of antigens.** There were three categories of antigen-generations: (1) whole-cell lysate or sonicate; (2) purified antigens; and (3) recombinant or synthetic antigens. The antigen type of one test was not known. No differences in test accuracy were found between the three antigen types.
- **Year of publication.** Studies were published between 1989 and 2013. There was no correlation between year of publication and antigen type used (P-value=0.76). Including the year of publication as a continuous covariate in the analysis had no effect on any of the model parameters (P-values above 0.25).
- **Specific tests.** The only individual test that could be evaluated specifically in the meta-analysis was Enzygnost, with six data rows. Its diagnostic odds ratio was 38.1 (95% CI 9.62–151), which was significantly higher than that of all other tests (P-value=0.04). Its sensitivity was 0.79 (95% CI 0.48–0.94), its specificity 0.91 (95% CI 0.83–0.96). The results of other tests were not subjected to meta-analysis, but are presented in Annex 5.
- **Effect of methodological quality.** None of the studies were of high quality – they all posed a risk of bias in patient sampling and flow-and-timing bias; studies were at best unclear for risk of bias in reference standard and index test. The effect of certain specific quality items on the sensitivity and specificity of the tests were investigated. The items concerning case definition (i.e. acceptable case definition and serology in case definition) show some variation over the different studies (Table 2). Therefore, case definition items were included as dichotomous variables in the models (acceptable case definition versus no or unclear case definition; and serology in the case definition versus no or unclear case definition). None had an effect on accuracy (P-values above 0.29).

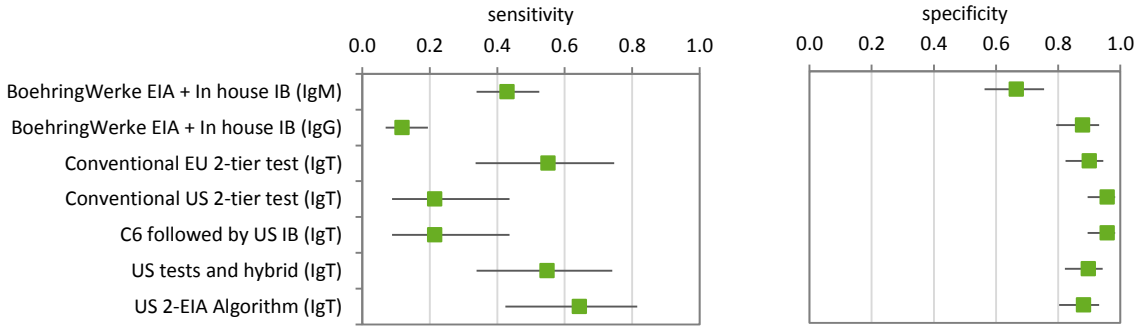
Sensitivity analyses

In the above analyses, it was assumed that borderline test results were considered positive test results. When borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Other tests: two-tiered tests

Other tests in this set of studies with healthy controls were two-tiered tests. Only two studies evaluated two-tiered tests. Their sensitivity did not exceed 65%, and their specificity was between 85% and 95%, with one outlier (specificity 67%) (Figure 6).

Figure 6. Evaluation of two-tiered tests on EM patients and healthy control



Erythema migrans – case-control studies with cross-reacting controls

Overall results and methodological quality of the studies

The analyses were based on 23 studies with a total of 1 434 persons with EM and 2 541 cross-reacting controls (Table 2). The latter were usually patients with syphilis, other infectious diseases or auto-immune disease. The median number of cases per study was 47 (range 15 to 158), whereas the median number of controls was 100 (range 29 to 275).

The following methodological quality issues were observed:

- There was a high risk of selection bias as enrolment did not occur randomly or consecutively. Cross-reacting controls were selected because of their potential for false-positive results; they were not representative of patients suspected of Lyme borreliosis.
- For most studies it was unclear whether the reference standard posed a high or low risk of bias. Even though the controls may be cross-reacting, other 'difficult-to-diagnose' patients were not selected, which made it easier to distinguish between cases and controls regardless of test method.
- Whether the execution of the index test may lead to bias was impossible to assess for most studies, although this led to high risk of bias in six studies due to the post-hoc selection of the cut-off value. Lencakova et al. (2008) and Mathiesen et al. (1996) reported that the assessment of the index test was blinded to the disease status of the participants, but only for one of the two tests they evaluated.
- The two studies reporting potential conflict of interest were also included in this analysis (Hunfeld et al. 2002, and Wilske et al. 1993).

Results for specific Ig type

Table 5 provides the summary estimates of the DOR, sensitivity and specificity for studies evaluating tests for more than one antibody type (for details see Annex 6). The meta-analysis showed that all three Ig types had a different accuracy (P-value=0.01 for IgG and 0.03 for IgT; IgM was the reference category) and operated at different thresholds (P-value<0.001 for both IgG and IgT; IgM was reference category).

Table 5. Summary estimates of test accuracy by antibody type for EM for case-control studies with cross-reacting controls

Antibody type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	6.7 (3.0–14.9)	0.42 (0.31–0.54)	0.90 (0.82–0.95)
IgG	10.1 (4.6–22.2)	0.38 (0.26–0.52)	0.94 (0.90–0.97)
IgT	9.5 (4.0–22.3)	0.67 (0.53–0.78)	0.82 (0.70–0.90)

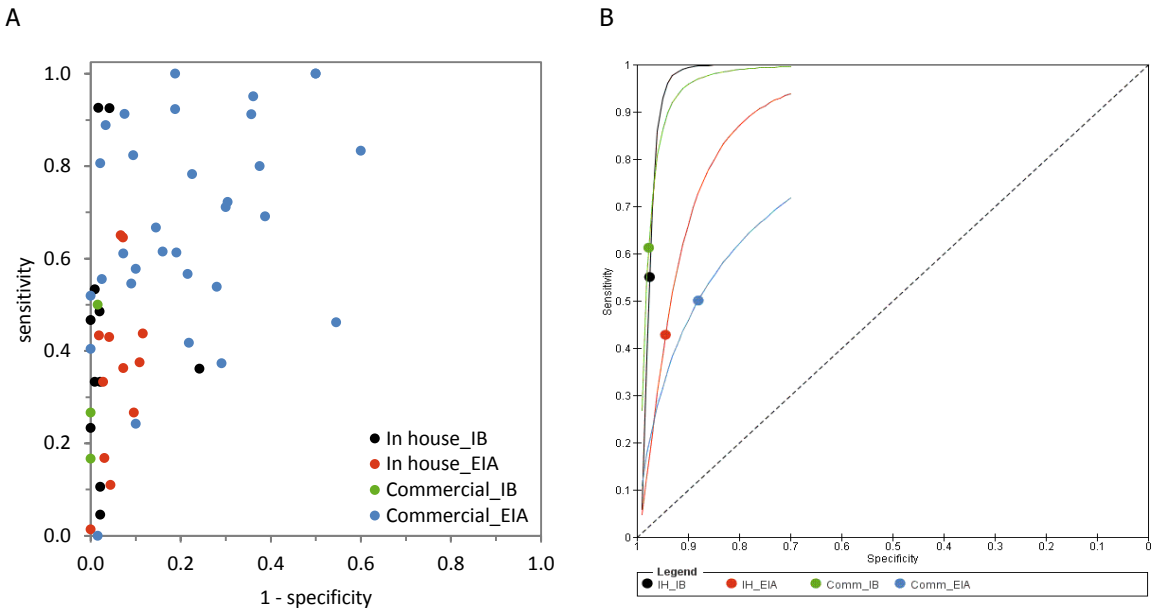
Results for specific test type, commercial versus in-house tests

The ROC scatter plot shows much heterogeneity, especially for the commercial EIA tests (Figure 7A).

The overall diagnostic odds ratio for any EIA or IB test detecting EM patients was 11.8 (95% CI 4.81 to 18.8), with a sensitivity of 0.47 (95% CI 0.33 to 0.61) and a specificity of 0.93 (95% CI 0.88 to 0.98). This was in line with the overall results from the studies with healthy controls.

When adding test type (EIA or IB) to the analyses, the model fit improved and IB tests turned out to have a significantly higher accuracy than EIA tests (P-value<0.01), operated at a lower threshold (P-value=0.04). In addition, the shape of the curve was different (P-value=0.04). Addition of a covariate to the model which accounted for commercial or in-house test further improved the model fit. The analyses demonstrated that IB and EIA tests have different accuracy levels, but that the in-house and commercial tests merely operate at different thresholds, which may point to ad hoc selection of optimal cut-offs in the in-house tests (Figure 7B and Table 6). These comparisons were based on both comparative studies (including IB and EIA) and non-comparative studies (containing only IB or EIA; the majority). Only four studies included both EIA and IB, which is not sufficient to compare the two test types directly.

Figure 7. ROC scatter plot (A) and fitted summary ROC curves (B) for EM case-control studies with cross-reacting controls



Note: Graph A, every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Table 6. Summary estimates of test accuracy for commercial and in-house IB and EIAs for EM case-control studies with cross-reacting controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	12.6 (6.1–26.0)	0.43 (0.27–0.61)	0.94 (0.90–0.97)
In-house IB	48.1 (17.9–130)	0.55 (0.31–0.77)	0.98 (0.96–0.99)
Commercial EIA	7.34 (3.9–13.8)	0.50 (0.38–0.63)	0.88 (0.78–0.94)
Commercial IB	62.1 (18.9–204.0)	0.61 (0.39–0.80)	0.98 (0.94–0.99)

Sources of heterogeneity

- **Generation of antigens.** The following categories were tested: whole-cell lysate or sonicate, purified antigens, and recombinant or synthetic antigens. Purified antigens tended to have a higher sensitivity than the two other types, although the confidence intervals overlap for all three (Table 7). There were not enough data to estimate the accuracy of combinations of antigens.
- **Year of publication.** Studies were published between 1989 and 2013. There was no relation between year of publication and antigen type used (P-value=0.70). Including year of publication as a continuous covariate in the analysis had an impact on the threshold at which these tests operated (P-value-0.02), which coincided with an increasing sensitivity over the years. This did not have a significant effect on the overall accuracy.
- **Specific tests.** The Enzygnost and C6 ELISA tests could be evaluated in a meta-analysis, both with five data rows.
  - The DOR of the Enzygnost was 13.2 (95% CI 1.10–60), with a sensitivity of 0.77 (95% CI 0.41–0.94) and a specificity of 0.78 (95% CI 0.41–0.96).
  - The DOR of the C6 ELISA was 8.80 (95% CI 0.66–118), with a sensitivity of 0.51 (95% CI 0.03–0.97) and a specificity 0.89 (95% CI 0.17–0.99).
  - The results of other tests were not meta-analysed, but are presented in Annex 6.
- **Effect of methodological quality.** As none of the studies was of high quality, only the effect of the quality items for case definition was evaluated (Table 2). Case definition items were included as dichotomous variables in the models (acceptable case definition versus no or unclear case definition; serology in the case definition versus no or unclear). None had an effect on any parameter of the models (P-values above 0.60).

Table 7. Summary estimates of sensitivity and specificity for generation of antigen

Antigen generation	Sensitivity (95% CI)	Specificity (95% CI)
Whole cell	0.36 (0.21–0.55)	0.96 (0.88–0.99)
Purified	0.62 (0.49–0.73)	0.87 (0.70–0.95)
Recombinant	0.48 (0.27–0.69)	0.93 (0.83–0.97)

Sensitivity analyses

In the above analyses, it was assumed that borderline test results were considered positive test results. When these borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Other tests

None of the other tests were evaluated in cross-reacting controls.

Erythema migrans – cross-sectional studies

Overall results and methodological quality of the studies

There was one cross-sectional study with patients suspected to have EM (Barrial et al. 2011, Table 2). In this study, the researchers included 75 patients over a period of eight months who had a positive or unclear ELISA result and tried to confirm these patients status with an IB. The final diagnosis of these patients was made on the basis of the EUCALB criteria: 33 patients were regarded as not having Lyme borreliosis. The prevalence of EM in this study was 56%. Because the selection process and the characteristics of the included study was poorly described, the study was rated 'unclear' both on 'risk of bias' and 'concerns regarding applicability for patient sampling'. It was also rated to have a 'low risk of bias' for the index test, reference standard, and flow and timing. There were also low concerns regarding the applicability of the reference standard and index test.

A meta-analysis was not possible. Sensitivity varied between 33% and 92%, while specificity varied between 27% and 70% when borderline results were considered positive. These results were in line with the case-control studies.

Neuroborreliosis

Summary of the study characteristics

Forty-four studies evaluated neuroborreliosis (NB): 34 case-control studies and ten cross-sectional studies. Cerar et al. (2006) investigated NB as part of a larger study with early disseminated Lyme and analysed the NB groups separately. For this review, only the NB-only groups from Cerar et al. were included. Hansen et al. (1988) differentiated between early, intermediate and late NB. These categories were combined into one group of patients suspected of NB.

The study characteristics of the case-control studies are presented in Table 8:

- Eight studies included only healthy controls, 15 included only cross-reacting controls and 11 studies included both healthy controls and cross-reacting controls. The latter reported the results for both groups separately.
- The case-control studies evaluated EIA tests (n=24), IB (n=12), a two-tiered test (n=1) and the specific antibody index (n=8). The EIA tests and IB were evaluated in serum and one study evaluated EIA in CSF.
- Cerar et al. (2006) included tick-borne encephalitis patients as controls and included an intermediate group of patients suspected of (non-confirmed) NB. For the main analyses, this group of suspected patients was considered a control group rather than a group of cases.

The study characteristics of the cross-sectional studies are presented in Table 9:

- The cross-sectional studies evaluated EIA tests (n=6), IB (n=3), two-tiered tests (n=2), the specific antibody index (n=4) and a lymphocyte stimulation test (LST, n=1).
- Different sample types were used for the EIA and IB tests, the specific antibody index is by definition in CSF and serum, the LST is by definition done in lymphocytes.

The quality assessment results are presented in Annex 4.

Table 8. Characteristics of NB case-control studies

									Number		
Study (see Annex 3Error! eference source not found. for references)	Country	Conflict of interest	Sample type	Number of EIAs tested	Number of IBs tested	Other tests	Acceptable case definition	Serology in case definition	NB	HC_Tot	CC_Tot
Ang 2012	Netherlands	No	Serum	10	0	0	Unclear	Unclear	102	228	212
Branda 2013	Slovenia	Possibly	Serum	3	2	5 two-tiered	Yes	Yes	15	100	0
Cerar 2006	Slovenia	No	Serum	1	0	0	Yes	No	28	49	0
Cerar 2010	Slovenia	No	CSF+serum	2	0	2 AI	Yes	Unclear	61	0	32
Cinco 2006	Italy	No	Serum	1	0	0	No	No	6	57	0
Dessau 2010	Denmark	No	Serum	1	0	0	Unclear	Yes	117	815	0
Dessau 2013	Denmark	No	Serum	2	0	0	Unclear	Yes	48	216	
Flisiak 1996	Poland	No	Serum	3	0	0	Yes	No	17	0	69
Goettner 2005	Germany	No	Serum	0	3*	0	No	Yes	50	0	110
Hansen 1988	Denmark	No	CSF+serum	2	0	0	Unclear	No	24	200	92
Hansen 1991	Denmark	Yes	CSF+serum	1	0	1 AI	Yes	No	100	0	35
Hofstad 1987	Norway	No	CSF+serum	1	0	1 AI	Yes	No	10	0	36
Hunfeld 2002	Germany/France	Yes	Serum	1	0	0	No	Unclear	35	1107	275
Kaiser 1998	Germany	No	CSF+serum	0	0	1 AI	Yes	Yes	67	0	24
Kaiser 1999inf	Germany	No	Serum	3	2*	0	Yes	Yes	96	80	40
Karlsson 1989eur	Sweden	No	Serum	1	1	0	Yes	No	68	0	44
Karlsson 1989siid	Sweden	No	Serum	2	0	0	Yes	No	47	0	73
Lakos 2005	Hungary	No	CSF+serum	0	0	2 AI	Yes	Yes	69	0	85
Lencakova 2008	Slovakia	No	Serum	1	1	0	Yes	Unclear	7	0	60
Mathiesen 1996	Denmark	No	Serum	1	1*	0	Yes	Yes	50	100	29
Mathiesen 1998	Sweden/Denmark	No	Serum	3	0	0	Yes	No	100	0	138
Nicolini 1992	France	No	Serum	0	1	0	Unclear	No	10	18	0
Panelius 2001	Finland	No	Serum	1	0	0	Yes	Yes	14	13	10
Putzker 1995	Germany	No	Serum	4	2*	0	Unclear	Yes	9	93	0
Rauer 1995	Germany	No	Serum	1	0	0	Unclear	Yes	33	154	136
Reiber 2013	Germany	No	CSF+serum	0	0	1 AI	Yes	No	29	16	45
Ryffel 1998	Switzerland	No	Serum	0	1*	0	Unclear	Yes	61	180	50
Schulte 2004	Germany	No	Serum	0	3	0	Yes	Yes	36	67	0
Tjernberg 2007	Sweden	No	Serum	3	0	0	Yes	Yes	26	55	200
Tjernberg 2011	Sweden	No	CSF+serum	0	0	1 AI	Unclear	Yes	124	0	92
VanBurgel 2011	Netherlands	No	CSF+serum	1	0	2 AI	Yes	Yes	118	0	143
Wilske 1993	Germany	Yes	Serum	2	0	0	Yes	Yes	60	100	42
Wilske 1999	Germany	Possibly	Serum	0	3*	0	Yes	Yes	42	0	139
Zoller 1990	Germany	No	Serum	1	0	0	No	Unclear	18	102	37

*Note: DDxTot=total number of controls in a cross-sectional design; HC_Tot=total number of healthy controls; CC_Tot=total number of cross-reacting controls; acceptable case definition=an acceptable case definition was used in accordance with international standards; serology in case definition= serology included as part of the reference standard; *=patients were pre-selected or had intrathecal antibodies.*

Table 9. Characteristics of NB cross-sectional studies

											Number	
Study (see Annex 3 for references)	Country	Conflict of interest	Sample type	Number of EIAs tested	Number of IBs tested	Two-tiered	Reference standard	Acceptable reference standard	Blinding	Children	NB	DDxTot
Albisetti 1997	Switzerland	No	CSF+serum	2	1	1 AI	Serology and CSF findings	Yes	Unclear	Yes	15	8
Barrial 2011	France	No	Serum	0	4*	0	EUCALB criteria	Yes	Unclear	No	19	33
Bednarova 2006	Czech Republic	No	CSF+serum	0	0	1 AI	Clinical criteria, serology, pleocytosis	Unclear	Unclear	No	38	20
Bennet 2008	Sweden	No	Serum	1	0	0	Clinical criteria plus pleocytosis	Unclear	Unclear	Yes	70	197
Blanc 2007	France	No	CSF+serum	0	0	1 AI	Clinical criteria and serology	No	Unclear	No	49	73
Ekerfelt 2004	Sweden	No	CSF+serum	0	0	1 AI	Clinical criteria and lab results	Unclear	Yes	No	59	58
Ljostad 2005	Norway	No	CSF+serum	1	0	0	Two or more of: (1) recent EM; (2) CSF cell count; (3) antibodies in serum or CSF; (4) intrathecal antibody production	Yes	Unclear	No	10	59
Nordberg 2012	Sweden/Finland	No	Blood	0	0	1 LST	Clinical criteria, 'lymphocytic meningitis', or intrathecal antibodies, or IgM in serum	Yes	Unclear	No	14	103
Roux 2007	France	No	CSF+serum	2	1	1 AI; 1 two-tiered	Clinical criteria and serology	Yes	Unclear	No	11	16
Skarpaas 2007	Norway	No	CSF+serum	1	0	1 two-tiered	Clinical criteria, serology, pleocytosis	Unclear	Unclear	No	60	18
Skogman 2008	Sweden/Finland	No	CSF	5	0	0	Clinical criteria, pleocytosis, intrathecal antibodies	Yes	No	Yes	40	36

*Note: DDxTot=total number of controls in a cross-sectional design; *=IB was performed on samples with doubtful or positive EIA.*

Neuroborreliosis – case-control studies with healthy controls

Methodological quality of the studies

The analyses are based on 20 studies with a total of 817 persons with NB and 3 750 healthy controls (Table 8). The median number of cases per study was 31 (range 6 to 117), and the median number of controls was 100 (range 13 to 1107).

The following methodological quality issues were observed:

- All studies had a high risk of bias in the patient selection domain because of their healthy control design.
- Most studies were rated 'unclear' on risk of bias due to the execution of the index test and due to not reporting whether the assessment of the (commercial) index test was blinded to the disease status of the participants.
- Four studies selected the cut-off value with the highest sensitivity and/or specificity based on their results, which poses a high risk of bias.
- There were 'high' concerns about the applicability of all in-house tests as these tests may be different in each laboratory, regardless of what is reported about the execution. Two studies reported potential conflict of interest (Hunfeld et al. 2002, Wilske et al. 1993).

Results specific to Ig type

Annex 7 provides an overview of the sensitivity and specificity by Ig type for the different studies. Overall, IgT tests had the highest sensitivity, but this was combined with the lowest specificity in most studies.

Overall, when meta-analysing the sensitivity and specificity of the studies that evaluated tests for more than one antibody type, the underlying summary ROC curve was significantly different for IgG than for IgM and IgT (P-values for accuracy, threshold and shape of the curve all <0.01) (Figure 8). However, because of the average operating threshold at which IgM and IgG were used, their performance seemed not significantly different from each other (Table 9). Overall, the IgT tests were more sensitive than the IgM and IgG tests.

Figure 8. Fitted summary ROC curves for IgG, IgM and IgG test types for NB case control studies with healthy controls

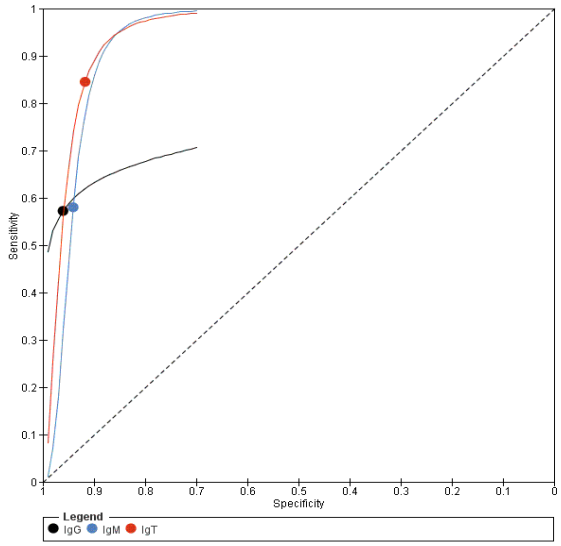


Table 9. Summary estimates of test accuracy by antibody type for NB case-control studies with healthy controls

Antibody type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	21.9 (12.9–37.1)	0.58 (0.46–0.70)	0.94 (0.93–0.95)
IgG	32.8 (17.3–61.9)	0.57 (0.53–0.62)	0.96 (0.93–0.98)
IgT	60.7 (34.3–107.3)	0.85 (0.77–0.90)	0.92 (0.89–0.94)

Results specific to test type and commercial versus in-house

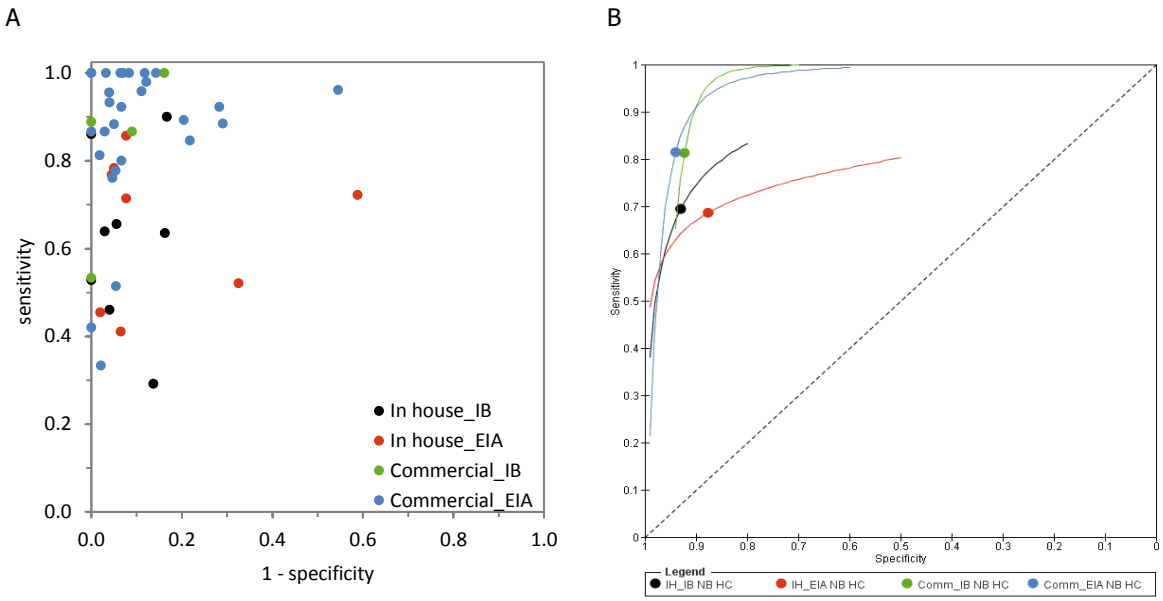
The ROC scatter plot shows much heterogeneity in the different test results (Figure 9A).

The overall diagnostic odds ratio for any EIA or IB test for detecting NB patients was 44.0 (95% CI 23.5–82.6), with a sensitivity of 0.77 (95% CI0.68–0.85) and a specificity of 0.93 (95% CI 0.88–0.96).

When adding test type (EIA or IB) to the analyses, the model fit slightly improved. Test type turned out to have a significant effect on the threshold at which the tests operate (P=0.042). Addition of a covariate to the model which accounted for commercial or in-house test further improved the model fit and revealed that the in-house EIAs tests had a lower accuracy (P=0.0016), operate at different threshold (P=0.0077) and had a different shape of the curve (P=0.0012) than the commercial test (Figure 9B, Table 10).

The above comparisons were based on comparative studies (including both IB and EIA) and non-comparative studies (containing IB or EIA only; the majority). More valid comparisons could have been made if only comparative studies had been included. However, only four studies included both EIA and IB, which was not enough to directly compare the two test types.

Figure 9. ROC scatter plot (A) and fitted summary ROC curves (B) for NB case-control studies with healthy controls



Note: Graph A: every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Table 10. Summary estimates of test accuracy for commercial and in-house IBs and EIAs for NB case-control studies with healthy controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	15.5 (5.6–42.7)	0.69 (0.60–0.76)	0.88 (0.72–0.95)
In-house IB	30.2 (13.3–68.2)	0.69 (0.57–0.80)	0.93 (0.86–0.97)
Commercial EIA	68.5 (36.5–129.0)	0.81 (0.70–0.89)	0.94 (0.91–0.96)
Commercial IB	51.8 (15.4–174.0)	0.81 (0.57–0.96)	0.92 (0.88–0.95)

Sources of heterogeneity

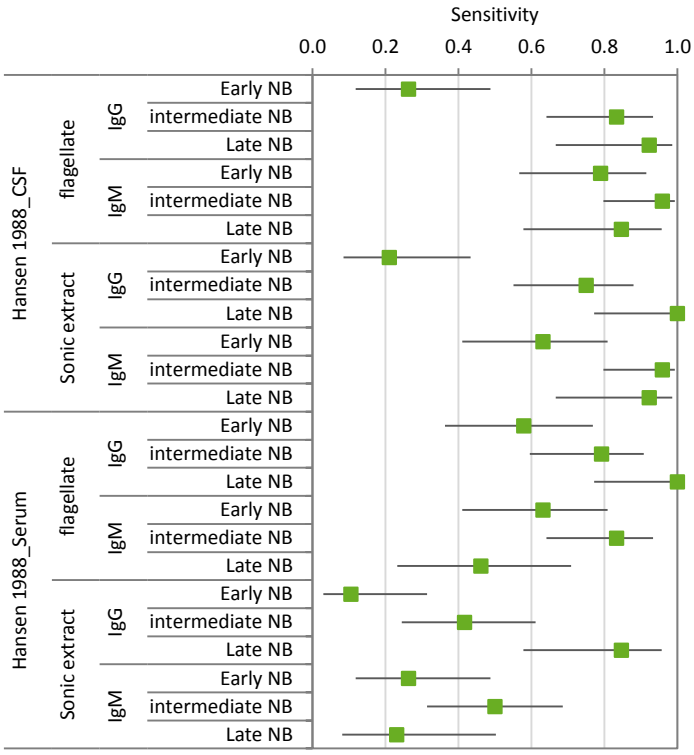
- **Generation of antigens.** Three categories of antigen generations were compared. The DOR of purified and recombinant antigens tended to be higher than the DOR for whole-cell tests, but the confidence intervals overlapped (**Table** 11).
- **Year of publication.** There was no relation between year of publication and antigen type (P-value=0.98). Including year of publication as a continuous covariate in the analyses slightly improved the fit of the model. Studies published after 2000 had a higher accuracy than those published before 2000 (P-value=0.01) (Table 11). The year 2000 was an arbitrary cut-off.

Table 11. Summary estimates of test accuracy taking into account sources of heterogeneity

Sources of heterogeneity	Test categories	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Antigen generation	Whole cell	25.3 (9.2–69.8)	0.72 (0.54–0.85)	0.91 (0.80–0.96)
	Purified antigen	82.3 (49.3–137.0)	0.77 (0.62–0.87)	0.96 (0.93–0.98)
	Recombinant	85.0 (64.8–270.0)	0.84 (0.65–0.94)	0.94 (0.88–0.97)
Year of publication	<2000	14.9 (7.4–29.9)	0.63 (0.52–0.73)	0.90 (0.82–0.94)
	2000 and later	76.2 (25.0–232.2)	0.85 (0.72–0.93)	0.93 (0.83–0.97)

- **Early versus late NB**. The timing of the NB diagnosis (early versus late) was poorly reported and not well-defined.
  - Hansen et al. (1988) included healthy controls and evaluated EIA in serum and CSF in early NB (<20 days post onset of symptoms), intermediate NB (21 to 40 days post onset of symptoms) and late NB (41 to 160 days post onset of symptoms). As there was no difference in specificity (one healthy control group), only sensitivity is reported in Figure 10.
  - Kaiser et al. (1998) discriminated between acute (≤ 6 months' duration of symptoms) and chronic NB (six months' duration of symptoms). Sensitivity of an EIA was 54% (28/52) for IgM in acute NB and 6% (1/15) in chronic NB; sensitivity of IgG was 90% (47/52) in acute NB and 100% (15/15) in chronic NB.

Figure 10. Sensitivity for early, intermediate and late NB



Note: Early NB (<20 days post onset of symptoms), intermediate NB (21 to 40 days post onset of symptoms) and late NB (41 to 160 days post onset of symptoms)

- **Specific tests.** None of the individual tests was evaluated in more than four studies, therefore a meta-analysis was not done. Each test and study are presented individually in Annex 7.
- **Effect of methodological quality.** All studies had high risk of bias in patient sampling, flow and timing. Three separate analyses were done based on:
  - studies with unclear or low risk of bias in index test execution;
  - studies that did not include serology in the case definition;
  - studies that had a case definition in accordance with published guidelines. The separate analyses show only small non-significant differences compared with the overall analysis.

Sensitivity analyses

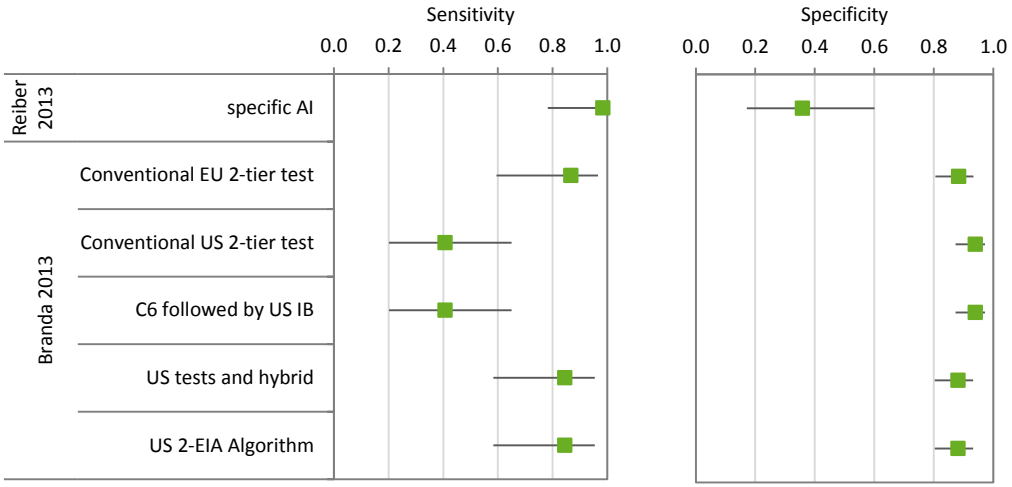
In the above analyses, it was assumed that borderline test results were considered positive test results. When borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Other tests: two-tiered tests and antibody index test

Other tests in this set of studies with healthy controls were two-tiered tests and one specific antibody index test (Figure 11):

- Branda et al. (2013) evaluated two-tiered tests. The sensitivity varied between 41% and 87%, and the specificity ranged between 88% and 94%.
- Reiber et al. (2013) evaluated a specific antibody test and reported a sensitivity of 98% and a specificity of 36%.

Figure 11. Sensitivity and specificity of two-tiered tests and AI algorithm for NB case-control studies with healthy controls



Neuroborreliosis – case-control studies with cross-reacting controls

Overall results and methodological quality of the studies

The analyses were based on 26 studies with a total of 1 428 persons with NB and 2 248 cross-reacting controls (Table 8). The median number of cases per study was 50 (range 7 to 124), and the median number of controls was 65 (range 10 to 275). The controls were usually patients with syphilis, other infectious diseases, auto-immune diseases or neurological conditions.

Cerar et al. (2010) included 'suspected' patients in their case definition and also included a group of tick-borne encephalitis patients as controls. It was not possible to evaluate the 'suspected' patients as a separate control group, they were therefore included in the case group, and the tick-borne encephalitis patients became cross-reacting controls.

Eighteen studies evaluated EIAs in serum, and five studies also evaluated these EIAs in CSF. Eight studies evaluated between IBs (none of the IBs was commercially available). None of the IBs were evaluated in CSF. Eight studies evaluated a specific antibody index, which is based on the antibody titre in serum and the titre in CSF.

The following methodological quality issues were observed:

- There was a high risk of selection bias as enrolment did not occur randomly and consecutively. Cross-reacting controls were selected because of their potential for false-positive results and were not representative of patients suspected of Lyme borreliosis.
- For most studies it was unclear whether the reference standard posed a high or low risk of bias. Even though the controls may be cross-reacting, other 'difficult-to-diagnose' patients were not selected, which made it easier to distinguish between cases and controls regardless of test method.
- Whether the execution of the index test may lead to bias was either highly probable or impossible to assess, except for the studies by Lencakova et al. (2008) and Mathiesen et al. (1996). Especially for the in-house tests, the cut-off value for positive or negative test results was often decided after the study was completed and thus may have led to an overestimation of sensitivity and specificity.
- The three studies reported potential conflicts of interest but are also included in this analysis (Hansen et al. 1991, Hunfeld et al. 2002, and Wilske et al. 1993).

Results specific to Ig type

Annex 8 provides an overview of the sensitivity and specificity by Ig type for the different studies. Overall, IgT tests had the highest sensitivity, but the lowest specificity.

The accuracy of the three Ig types differed significantly from each other, as shown by the meta-analysis of the studies that evaluated tests in serum for more than one antibody type (Table 12, Figure 12). This may be because the shape of the curves (especially IgG) is different, as was the case for the studies with healthy controls.

Figure 12. Fitted summary ROC curves for NB case-control studies with cross-reacting controls

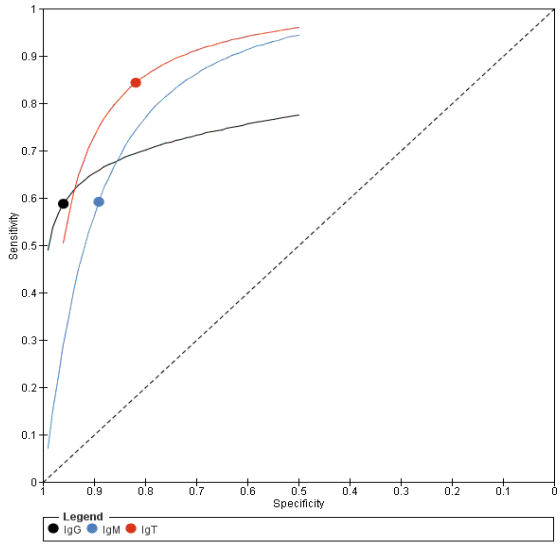


Table 12. Summary estimates of test accuracy by antibody type for NB case-control studies with healthy controls

Antibody type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	11.8 (6.10–22.8)	0.59 (0.47–0.71)	0.89 (0.84–0.93)
IgG	34.2 (13.6–86.1)	0.59 (0.52–0.65)	0.96 (0.91–0.98)
IgT	24.6 (11.6–50.4)	0.84 (0.76–0.90)	0.82 (0.73–0.88)

Results specific to test type and commercial versus in-house

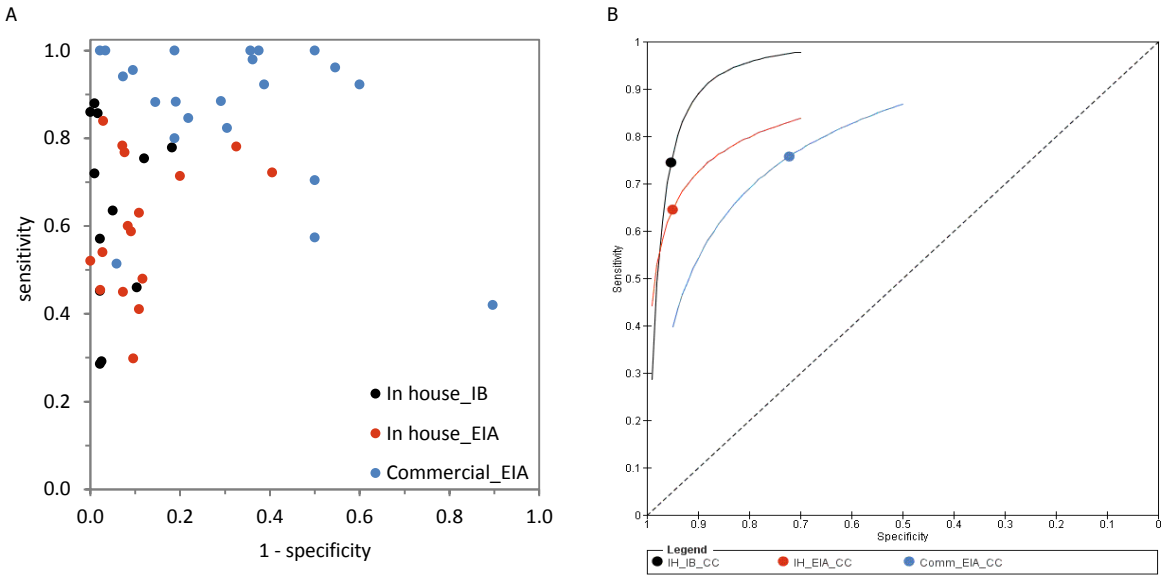
The ROC scatter plot showed much heterogeneity, but the patterns seem to be different for the different tests: in-house tests show more variation in sensitivity while the commercial tests seemed to have more variation in specificity (Figure 13A).

The overall accuracy for any EIA or IB test in serum for detecting NB patients was: DOR 21.7 (95% CI 10.8–43.9), with a sensitivity of 0.71 (95% CI 0.60–0.79) and a specificity of 0.90 (95% CI 0.83–0.94).

When adding test type (EIA or IB) to the analyses, the model fit improved and test type turned out to have a significant effect on accuracy (P-value=0.0061), threshold (P-value=0.007) and shape of the summary ROC curve (P-value=0.0043). Addition of a covariate to the model that accounted for commercial or in-house test further improved the model fit but only had a significant effect on the threshold. The IBs had a significantly higher accuracy than EIAs. The commercial EIAs had a significantly lower accuracy than the in-house tests (Table 13, Figure 13B).

These comparisons were based on comparative studies (including both IB and EIA) and non-comparative studies (only IB, or EIA; the majority). Only four studies included both EIA and IB, which was not enough to compare the two test types directly.

Figure 13. ROC scatter plot (A) and fitted summary ROC curves (B) for NB case-control studies with cross-reacting controls



Note: Graph A: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table. One dot has an extremely low specificity combined with a low sensitivity (Mathiesen et al. 1996). This might have been caused by the fact that the group of cross-reactive controls in this study only consisted of patients with syphilis, which are known to cross-react with the flagellin antigen.

Table 13. Summary estimates of test accuracy for in-house and commercial EIAs and for in-house IB tests of NB case-control studies with cross-reacting controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	34.3 (12.1–96.8)	0.65 (0.54–0.74)	0.95 (0.88–0.98)
In-house IB	58.9 (21.2–163.7)	0.74 (0.58–0.86)	0.95 (0.91–0.98)
Commercial EIA	8.1 (2.9–22.1)	0.76 (0.63–0.85)	0.72 (0.54–0.85)

Sources of heterogeneity

- **Generation of antigens.** Three categories of antigen-generations were tested: whole-cell lysate or sonicate, purified antigens, and recombinant or synthetic antigens. No differences were observed among the three categories, though the recombinant antigens tended to have a higher accuracy than the other two types.
- **Year of publication.** Studies were published between 1989 and 2013. There was no relation between year of publication and antigen type (P-value=0.39). Year of publication was included as a binary variable (before 2000, after 2000). Year had no effect on any of the parameters (P-values all above 0.1).
- **Specific tests.** None of the individual tests was evaluated in more than three studies, hence no meta-analysis was done. Sensitivity and specificity of the different tests are presented in Annex 8.
- **Effect of methodological quality.** Three separate analyses were done based on:
  - studies with unclear or low risk of bias in index test execution;
  - studies that did not include serology in the case definition;
  - studies that had a case definition in accordance with published guidelines. The accuracy estimates were similar for all categories and comparable with the overall results.

Sensitivity analysis

In the above analyses it was assumed that borderline test results were considered positive test results. When these borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

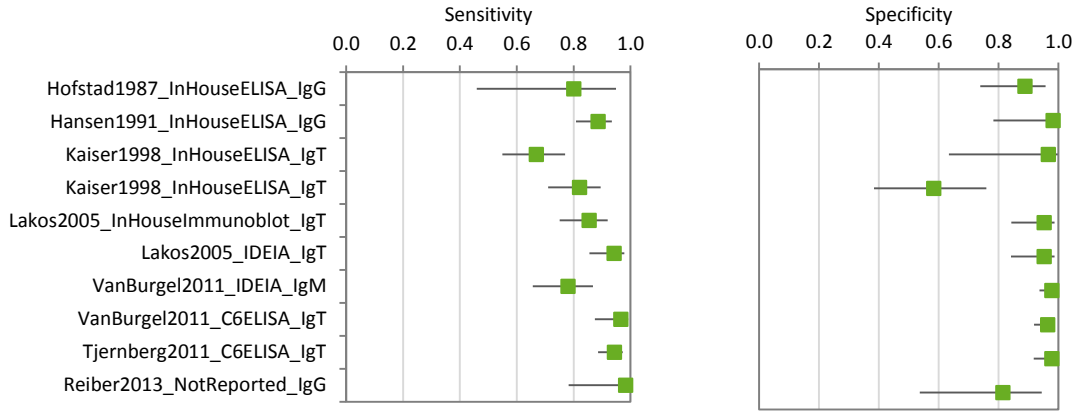
Other tests: EIA in CSF and antibody index tests

Six studies evaluated an EIA in CSF, two of which were commercial tests; the remaining four were in-house tests. The overall accuracy in CSF was DOR 68.8 (95% CI 8.6–551.0), which is considerably higher than the DOR observed for serum (but also less precise). The summary sensitivity was 0.74 (95% CI 0.38–0.93) and the specificity was 0.96 (95% CI 0.85–0.99).

Seven studies evaluated a specific antibody index (Figure 14): one of these studies used IB and EIA; one study did not report which test was used to measure antibody titres; the other studies evaluated AI using EIAs. The summary DOR was 87.3 (95% CI 17.0–447.4), with a sensitivity of 0.86 (95% CI 0.63–0.95) and a specificity of 0.94 (95% CI 0.85–0.97).

Kaiser et al. (1998) also evaluated the effect of early versus late NB. Sensitivity of the EIA was 54% (28/52) for IgM in early NB and 6% (1/15) in late NB; sensitivity of IgG was 60% (47/52) in early NB and 100% (15/15) in late NB.

Figure 14. Sensitivity and specificity of specific antibody index tests in NB case-control studies with cross-reacting controls



Neuroborreliosis – cross-sectional studies

Overall results and methodological quality of the studies

Ten studies used a cross-sectional design to evaluate one or more tests for the diagnosis of NB. Not all studies were clear in the description of their patient sample, and not all studies were clearly cross-sectional designs (Table 14). In total, the studies contained 336 patients with NB with a median of 28 patients per study (range 10–70). The studies contained a total of 515 persons without NB, with a median of 29 per study (range 8 to 197).

Six studies evaluated EIAs and three studies evaluated IBs. Five studies evaluated specific antigen index tests, two a two-tiered test and one study evaluated an LST. Only one IB was evaluated in CSF, the EIAs were evaluated as often in CSF as in serum. The results were analysed separately for serum and CSF.

The quality of the studies varied (Annex 4). Although cross-sectional designs are generally considered to be of higher quality for estimating the clinical sensitivity and specificity, the included studies suffer from some of the following shortcomings:

- Only two studies (Albisetti et al. 1997, Skarpaas et al. 2007) gave a clear description of the inclusion criteria, the included patients, and had a low risk of bias in patient sampling. But both studies failed to clearly describe the reporting of flow and timing which made it difficult to determine the risk of bias in this domain.
- Only two studies (Barrial et al. 2011, Ekerfelt et al. 2004) explicitly stated that the assessment of the index test was blinded to the disease status of the participants; this was only done for one of the two tests they analysed.
- Six out of eleven studies had a high risk of bias in at least one of the quality domains.
- None of the studies reported potential conflicting interests.

Table 14. Overview of cross-sectional studies for NB

Study	Patients' characteristics	Prevalence
Albisetti 1997	Study done in a children's hospital; 21 consecutive children with facial palsy; serology was used in the diagnosis.	71%, 62%*
Barrial 2011	Study conducted in bacteriological laboratory; 52 patients with a positive or unclear ELISA result; diagnosis based on EUCALB.	37%
Bednarova 2006	Study conducted in department of clinical microbiology; 58 patients, 38 of which had NB and 20 had other neurological diseases. It was not clear how the patients were selected or why they were admitted to the hospital. Only evaluated specific antigen index.	66%

Study	Patients' characteristics	Prevalence
Bennet 2008	Study conducted in a children's hospital; 267 children who were tested in CSF and serum; charts were retrospectively examined and diagnosed based on several items (also serology); six patients with EM (but no neuroborreliosis) were included in the control group.	26%
Blanc 2007	Study conducted in bacteriological laboratory; 122 patients who all had antibodies in their CSF; divided into definite, possible and no borreliosis cases. Only evaluated specific antigen index.	33%, 41%*
Ekerfelt 2004	Origin of suspected samples not clearly explained; 117 suspected samples were ranked based on their likelihood to have NB. The patients without Lyme had facial palsy or another explanation for their symptoms.	50%
Ljostad 2005	Authors from department of neurology and ear, nose and throat; 79 adults with acute peripheral facial palsy (10 were excluded), 10 of which had NB (the remainder had Bell's palsy or viral palsy).	14%
Nordberg 2012	This study analysed the ELISPOT, an LST. It is not clear why the 117 (out of an initial 310) patients in the study underwent the ELISPOT. All patients had symptoms compatible with NB.	12%
Roux 2007	Not certain that this is a cross-sectional design; the controls were sampled from the same place and were initially suspected of Lyme meningoradiculitis. Authors from departments of rheumatology and microbiology.	41%
Skarpaas 2007	Authors from department of neurology and microbiology. This study included consecutive patients with clinical signs, symptoms of NB and pleocytosis; group was divided into probable, possible and no Lyme cases.	77%
Skogman 2008	Study conducted in a children's hospital. This study included consecutive children with clinical signs, symptoms of NB and pleocytosis; group was divided into probable, possible and no Lyme cases.	70%, 42%*

** First percentage: prevalence when considering possible cases as cases, second reported percentage: prevalence when considering possible cases as controls*

Results specific to Ig type

Annex 9 provides an overview of the sensitivity and specificity by Ig type for the different studies.

Due to the limited number of studies, the meta-analysis could only estimate the differences between IgG and IgM (and not IgT). The shape of the summary ROC curve was significantly different between the IgG and IgM tests, with a lower accuracy and lower specificity for the IgM tests on average. The sensitivity of both antibody types was similar (Table 15).

Table 15. Summary estimates of test accuracy by antibody type for NB cross-sectional control studies

Antibody type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	4.1 (1.7–10.0)	0.66 (0.50–0.78)	0.68 (0.58–0.77)
IgG	11.4 (0.7–181.4)	0.67 (0.60–0.74)	0.85 (0.27–0.99)

Results specific to test type, commercial versus in-house

The ROC scatter plot showed much heterogeneity (Figure 15), with some of the commercial EIAs having relatively low sensitivity combined with a relatively low specificity. Four of these were from the same study by Ekerfelt et al. (2004), which included all types of Lyme borreliosis, but for this analysis only NB was considered. This implied that the group in which the disease was regarded as absent still contained patients with Lyme borreliosis. This inevitably led to a lower specificity.

The overall accuracy (DOR) for any EIA or IB test in serum or CSF for detecting NB patients was 17.0 (95% CI 5.1–56.8), with a sensitivity of 0.78 (95% CI 0.59–0.92) and a specificity of 0.81 (95% CI 0.54–0.94).

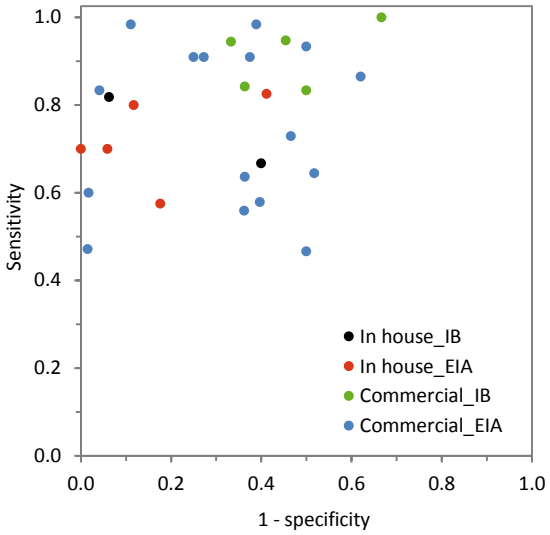
When analysing tests evaluated in serum only, the summary DOR was slightly lower: 13.1 (95% CI 3.2–53.7), with a sensitivity of 0.78 (95% CI 0.53– 0.92) and a specificity of 0.78 (95% CI 0.40–0.95). Addition of test type or the covariable for commercial versus in-house did not improve the model fit and had no significant effect on model parameters.

Repeating the analyses without the results by Ekerfelt et al. (2004) resulted in following accuracy estimates:

- Tests conducted in either serum or CSF: DOR 23.3 (95% CI 7.3–74.2), sensitivity 0.82 (95% CI 0.578–0.935), specificity 0.84 (95% CI 0.56–0.96)
- Tests conducted in serum only: DOR 18.6 (95% CI 4.19–82.7), sensitivity 0.80 (95% CI 0.49–0.95), specificity 0.82 (95% CI 0.37–0.97).

Figure 15. Raw ROC scatter plot (A) and fitted summary ROC curves (B) for NB cross-sectional studies

A



Note: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Sources of heterogeneity

- **Generation of antigens.** Data were insufficient to conduct a subgroup analysis for antigen type.
- **Year of publication.** Studies were published in 1997, 2004, 2005, 2007, 2008 and 2011. The irregular steps between the dates made it impossible to perform a meta-analysis to evaluate the effect of year.
- **Prevalence or pre-test probability.** A potential source of variation is the spectrum of patients included in the studies. In a cross-sectional study, this may be reflected by the prevalence (or pre-test probability) of disease. This prevalence is not the same as the prevalence measured in the general population, it is the proportion of people with the target condition in the study sample. The median prevalence when 'possibles' were considered as cases was 50.2% (interquartile range (IQR) 37.0%–70.2%). Adding prevalence as a continuous covariate to the model (only allowing accuracy to change) showed no effect on accuracy. Removing the study reported by Ekerfelt et al. (2004) from these analyses did not change the results.
- **Specific tests.** None of the individual tests was evaluated in more than three studies; instead, the results are presented individually for each test and study in Annex 9.
- **Study quality.** The studies were rated to have a low or unclear risk of bias in most quality domains. Two studies, however, were judged 'high risk of bias' for the patient domain, due to non-consecutive enrolment. One other study was considered to have high risk of bias in the index test domain due to ad hoc cut-off value selection. Studies with low or unclear risk of bias in all QUADAS-domains were analysed separately. Results were not different from the overall analysis.

Sensitivity analysis

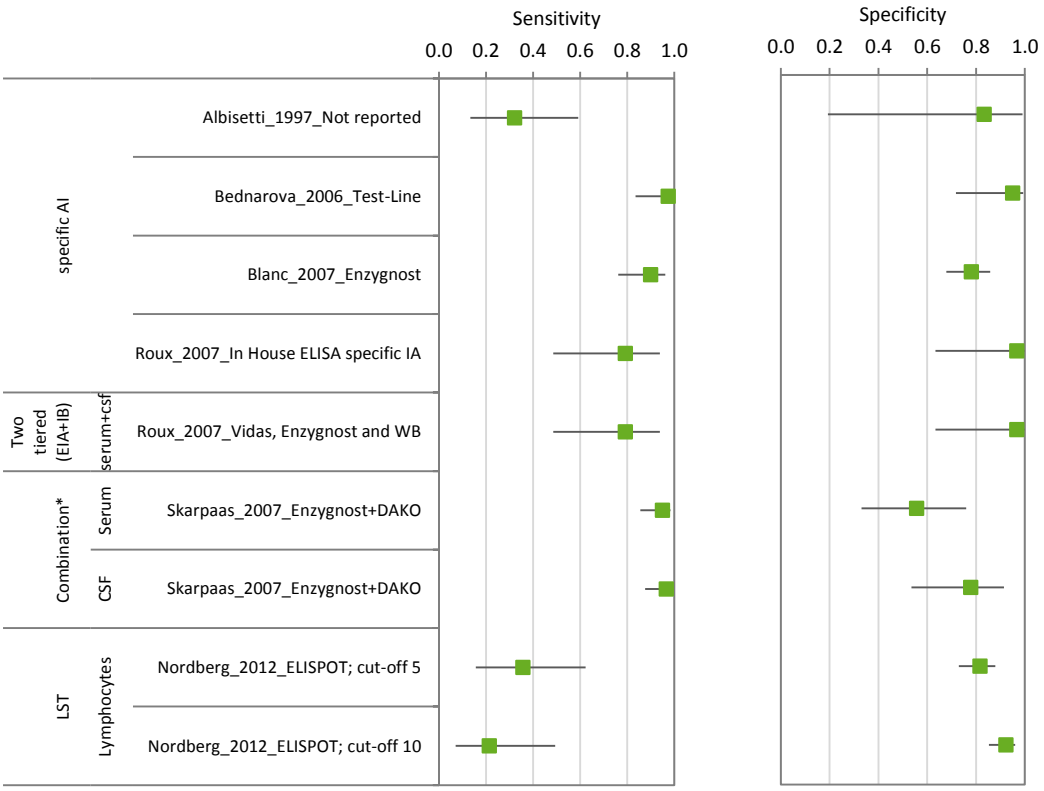
In the above analyses, borderline test results were considered negative and patients who had possibly NB were considered as cases. Following two sensitivity analysis was conducted: patients with possible NB were treated as controls; and borderline test results were considered positive results. The accuracy estimates of the two analyses were similar and comparable with the overall results.

Other tests: antibody index test, two-tiered test and LTT

Figure 16 provides an overview of the sensitivity and specificity of the antibody index, two-tiered and LTT tests.

Four studies evaluated a specific antibody index test. The summary DOR was 97.4 (95% CI 11.9–796.4), with a sensitivity of 0.79 (95% CI 0.34–0.97) and a specificity of 0.96 (95% CI 0.64–0.99).

Figure 16. Sensitivity and specificity of other tests on NB using a cross-sectional design



Lyme arthritis

Summary of the study characteristics

Fourteen studies evaluated Lyme arthritis (LA) (Table 16). Eleven studies had a case-control design and three a cross-sectional design. Five of the 11 case-control studies included both a healthy control group and a cross-reacting control group and reported the results for both groups separately; two case-control studies included a healthy control group only; and three included only cross-reacting controls; one study included children only. Hunfeld et al. (2002) combined patients with ACA and LA in one group and could therefore not be included in the analysis.

The study characteristic of the case-control studies are presented in:

- Eight studies included between one to four EIA tests, one study included eight different EIAs. Two of these studies also included an IB, one used these tests in a two-tiered algorithm. Three studies evaluated only IBs.
- One of the cross-sectional studies included a 'possible' LA category.
- Five studies included borderline test results.

The quality assessment results are presented in Annex 4.

Table 16. Characteristics of LA studies

											Number			
Study	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Other tests	Acceptable case definition	Serology in case definition	Setting	Children	LA	DDxTot	HC_Tot	CC_Tot
Ang 2012	Netherlands	CC	No	8	0	0	Unclear	Unclear	Laboratory	No	25		228	212
Branda 2013	Slovenia	CC	Possibly	3	2	5	Yes	Yes	Hospital	No	15		100	0
Cinco 2006	Italy	CC	No	1	0	0	No	No	laboratory	No	16		57	0
Flisiak 1996	Poland	CC	No	3	0	0	Yes	No	Laboratory	No	7		0	69
Goettner 2005	Germany	CC	No	0	3	0	No	Yes	Laboratory	No	10		0	110
Lencakova 2008	Slovakia	CC	No	1	1	0	Yes	Unclear	Laboratory	No	13		0	60
Putzker 1995	Germany	CC	No	4	2*	0	Unclear	Yes	Laboratory	No	28		93	0
Rauer 1995	Germany	CC	No	1	0	0	Unclear	Yes	Laboratory	No	17		154	136

											Number		
Study	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Other tests	Acceptable case definition	Serology in case definition	Setting	Children	LA	DDxTot	HC_Tot	CC_Tot
Ryffel 1998	Switzerland	CC	No	0	1*	0	Unclear	Yes	Laboratory	No	41		180	50
Tjernberg 2007	Sweden	CC	No	3	0	0	Yes	Yes	Hospital & laboratory	No	3		55	200
Wilske 1993	Germany	CC	Yes	2	0	0	Yes	Yes	Laboratory	No	24		100	42
Barrial 2011	France	CS	No	0	4*	0	Yes	Unclear	Laboratory	No	8	33		
Blaauw 1993	Netherlands	CS	No	1	0	0	Yes	Unclear	Hospital	No	45	57		
Huppertz 1996	Germany	CS	No	0	0	LTT	Unclear	Unclear	Hospital and laboratory	Yes	55	48		

*Note: CC=case control design; CS=cross-sectional design; DDxTot=total number of controls in a cross-sectional design; HC_Tot=total number of healthy controls; CC_Tot=total number of cross-reacting controls; LA=total number of LA cases included; acceptable case definition=an acceptable case definition was used in accordance with international standards; serology; serology in case definition=serology included as part of the reference standard; COI=conflict of interest; *=IB was done on pre-selected samples or samples with positive screening test.*

Lyme arthritis – case-control studies with healthy control

Overall results and methodological quality of the studies

The analyses were based on eight case-control studies with a total of 169 persons with LA and 967 healthy controls. The median number of cases per study was 21 (range 3–41), while the median number of controls was 100 (range 55–228).

The following methodological quality issues were observed:

- Because all studies included healthy controls, they are expected to overestimate clinical specificity and thus have a high risk of bias in the patient selection domain of QUADAS-2.
- Five out of seven studies based their case definition on serology results, leading to high risk of bias in the reference standard domain.
- Flow and timing of all studies did pose a high risk of bias in all case-control studies with healthy controls.
- The execution of the index test led to high or unclear risk of bias.
- Wilske et al. (1993) reported a potential conflict of interest because one author worked for Behringwerke AG and another for Mikrogen Gmbh, the manufacturers of two of the evaluated tests.

Results specific to Ig type

Annex 10 provides an overview of the sensitivity and specificity by Ig type for the different studies. In most cases, the sensitivity of the IgT was higher than the sensitivity of the other Ig types.

Overall, a meta-analysis of the diagnostic accuracy of the studies that evaluated tests for more than one antibody type showed that the IgG and IgT tests were more accurate than the IgM tests, mainly due to a much higher sensitivity (Table 17).

Table 17. Summary estimates of test accuracy by antibody type for LA case-control studies with healthy controls

Antibody tested	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	12.4 (4.96–31.0)	0.39 (0.28–0.52)	0.95 (0.88–0.98)
IgG	488.0 (175.0–1360.0)	0.94 (0.86–0.98)	0.97 (0.94–0.98)
IgT	202.0 (55.2–738.0)	0.95 (0.84–0.98)	0.92 (0.84–0.96)

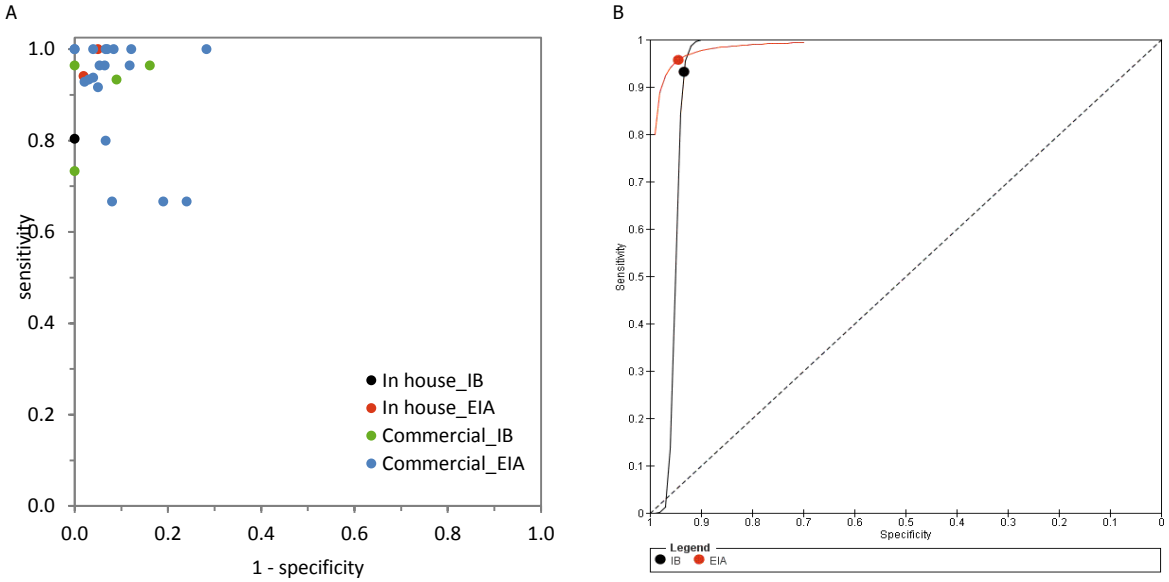
Results specific to test type

The ROC scatter plot shows much heterogeneity in the different test results (Figure 17A).

On average, for any EIA or IB test detecting LA patients, the diagnostic odds ratio was 86.3 (95% CI 45.5–163.0), with a sensitivity of 0.88 (95% CI 0.83–0.92) and a specificity of 0.92 (95% CI 0.88–0.95).

The effect of test type (IB or EIA) was assessed, but not the effect of commercial versus in-house test, as only three data rows were available for in-house tests. The DOR is higher for EIA than for IB (Table 18, Figure 17B). This difference is not significant, but the model fitted much better after adding the covariate. Both test types had a similar average estimated sensitivity and specificity, which is higher than that of the overall estimate mentioned above. This may be unexpected, as one may expect that the overall estimate should be an average of the estimates from both tests, but this is a result of the way the analyses were executed. In both analyses, the best fitting curve was fitted and from this curve the mean estimates were derived.

Figure 17. ROC scatter plot for LA case-control studies with healthy controls



Note: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Table 18. Summary estimates of test accuracy for EIAs and IB tests for LA case-control studies with healthy controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
EIA	377.7 (93.7–1522.0)	0.96 (0.89–0.98)	0.94 (0.89–0.97)
IB	191.6 (11.0–3335.9)	0.93 (0.48–0.99)	0.93 (0.89–0.96)

Sources of heterogeneity

- **Generation of antigens**. The accuracy between recombinant and non-recombinant antigens was evaluated. Addition of recombinant tests had a significant effect on the shape parameter, i.e. accuracy for the recombinant tests changed in a different way depending on the threshold compared to the non-recombinant tests. However, no difference in average accuracy, sensitivity and specificity of the two test types were not present.
- **Year of publication.** Studies were published between 1993 and 2013. Between 1998 and 2006, no studies were published. There was no relation between year of publication and antigen type (P-value=0.99). Including year of publication as a dichotomous covariate (published before/after 2000) in the analyses did improve the model, but had no effect on any of the model parameters (P-values > 0.20).
- **Specific tests.** Only the C6 ELISA test was evaluated (four studies), all other tests were evaluated in three or fewer studies. The model did not converge for the C6 ELISA test and none of the other individual tests were analysed due to low numbers. The raw data are provided in Annex 10.
- **Effect of methodological quality.** All tests posed a risk of bias in patient sampling and flow and timing and they were unclear regarding the risk of bias in reference standard and index tests. It was investigated whether certain specific quality items had an effect on the sensitivity and specificity of the tests. Case definition items were included as dichotomous variables in the models ('acceptable case definition' versus 'no' or 'unclear' and 'serology in the case definition' versus 'no' or 'unclear'). None had an effect on accuracy (P-values above 0.17), although the fit of the model improved.

Sensitivity analyses

In the above analyses, the borderline test results were considered positive test results. The models did not converge for the data when borderline results were considered negative.

Other tests: two-tiered tests

Branda et al. (2013) evaluated two-tiered tests. All but one had a specificity of 100% (one of 99%) and the sensitivity varied between 60% and 93%. This study included only 15 patients with LA. It also included assays developed for use in the USA.

Lyme arthritis – case-control studies with cross-reacting controls

Overall results and methodological quality of the studies

The data from eight studies were included in these analyses, with a total of 140 persons with LA and 879 cross-reacting controls. These were usually patients with syphilis, other infectious diseases or auto-immune disease, but other causes for arthritis were also included. The median number of cases per study was 15 (range 3– 41), while the median number of controls was 90 (range 42–212).

The following methodological quality issues were observed:

- There was a high risk of selection bias as enrolment did not occur randomly or consecutively. Cross-reacting controls were selected because of their potential for false-positive results and were not representative of patients suspected of Lyme borreliosis
- Most studies were at high risk of bias due to serology being part of the case definition.
- Two studies did not specify whether they incorporated serology in the case definition.
- Lencakova et al. (2008) reported that the assessment of the commercial index test was blinded (but not for the in-house test).
- One study reported potential conflict of interest (Wilske et al. 1993).

Results specific to Ig type

Annex 11 provides an overview of the sensitivity and specificity by Ig type for the different studies. IgM had lower sensitivity than IgG and IgT. Meta-analysing the accuracy resulted in instable models, therefore summary estimates are not provided. IgT tests had generally the highest sensitivity and the lowest specificity, while the IgG tests have the highest specificity.

Results specific to test type and commercial versus in-house

The ROC scatter plot showed high heterogeneity (Figure 18). This was mainly due to outlying data points in the study by Tjernberg et al. (2007). The results of that study were based on only three LA patients and 55 cross-reacting controls. However, to retain comparability of the results, the same model will be used here.

The overall diagnostic odds ratio for any EIA or IB test detecting LA patients was 216 (95% CI 36.0–1304), with a sensitivity of 0.95 (95% CI 0.89–0.98) and a specificity of 0.92 (95% CI 0.75–0.98). Adding test type as a covariate did not improve the model fit, nor did it have a significant effect on any of the model parameters. The effect of in-house versus commercial test could not be investigated (Table 19, Figure 18).

Figure 18. Raw scatter ROC plot for LA case-control studies with cross-reacting controls

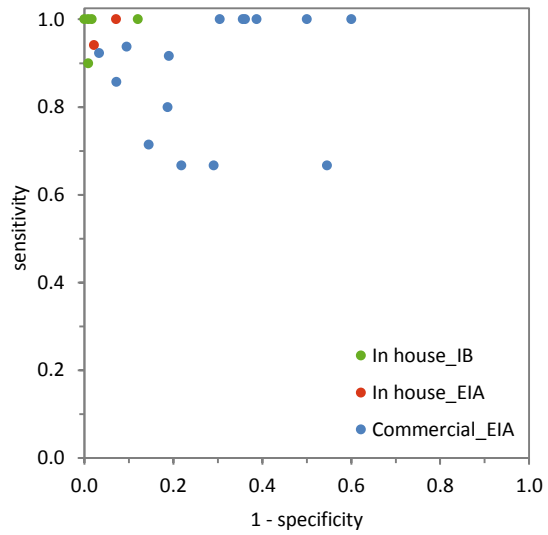


Table 19. Summary estimates of test accuracy for EIA and IB for LA case-control studies including cross-reacting controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
EIA	88.6 (17.7–444.0)	0.92 (0.82–0.97)	0.88 (0.69–0.96)
IB	2408.0 (73.9–78495.0)	0.99 (0.87–0.99)	0.97 (0.76–0.99)

Sources of heterogeneity

- **Generation of antigens.** The accuracy between recombinant and non-recombinant antigens was evaluated. The model fit improved, but addition of recombinant tests had no significant effect on any of the model parameters.
- **Year of publication.** Studies were published between 1993 and 2012, with no studies published between 1998 and 2005. There was no relation between year of publication and antigen type (P-value=0.99). Year of publication did not have a significant effect on any of the parameters when included as a binary covariate (before or after 2000).
- **Specific tests.** There were no individual tests that could be evaluated specifically because the largest group contained only two data points. The crude results per test and per study are presented in Annex 11.
- **Effect of methodological quality.** The effect of quality items could not be assessed: limiting the analyses to high-quality studies and including quality items as covariates in the model both resulted in unstable models.

Sensitivity analyses

In the above analyses, the borderline test results were considered positive test results. When these borderline test results were considered negative no difference in DOR, sensitivity and specificity was observed.

Other tests

None of the other tests were evaluated in cross-reacting controls.

Lyme arthritis – cross-sectional studies

Overall results and methodological quality of the studies

Three cross-sectional studies were included for LA (Table 16):

- Barrial et al. 2011 included 75 patients with a positive or unclear ELISA result and aimed to confirm the patients' status with an IB. The final diagnosis of these patients was made on basis of the EUCALB criteria and 33 patients were regarded not to have Lyme borreliosis. Of the 42 patients with Lyme borreliosis, eight were considered to have LA. This resulted in a prevalence of Lyme (overall, not limited to arthritis) of 56% (95% CI 45%–67%);
- Blaauw et al. 1993 evaluated an in-house EIA and included 102 arthritis patients, 15 of whom turned out to have LA (either very likely or likely) and 30 patients were classified as 'maybe Lyme arthritis'. When these patients were included as LA cases, the prevalence was 44% (95% CI 35%–54%). When they were not considered LA cases, the prevalence was 15% (95% CI 9%–23%).
- The third cross-sectional study evaluated the LTT test (Huppertz et al. 1996). This study included 103 children and adolescents with arthritis, 55 of whom turned out to have Lyme arthritis. This is a prevalence of 53% (95% CI 44%–63%).

No meta-analysis was conducted. The observed sensitivity and specificity are described in Annex 12.

Acrodermatitis chronica atrophicans

Summary of the study characteristics

Fourteen studies evaluated acrodermatitis chronica atrophicans (ACA) (Table 20). Only one had a cross-sectional design. Eight case-control studies included both a healthy control group and a cross-reacting control group and reported the results for both groups separately; one only study included a healthy control group; and four included only cross-reacting controls. Hunfeld et al. (2002) combined patients with ACA and LA in one group and could therefore not be included in the analysis.

The study characteristic of the case-control studies are presented in Table 20:

- Ten studies included between one and three EIA tests, one study included eight different EIAs. Two of these studies also included an IB, but only one used these tests in a two-tiered algorithm. Three studies evaluated between one and four IBs.
- None of the studies included a 'possible' ACA category; two studies included borderline test results.
- None of the studies included children or analysed the results for children separately from adults.

The quality assessment results are presented in Annex 4.

Table 20. Characteristics of ACA studies

										Number			
Study	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Two-tiered test	Acceptable case definition	Serology in case definition	Setting	ACA	DDxTot	HC_Tot	CC_Tot
Ang 2012	Netherlands	CC	No	8	0	0	Unclear	Unclear	Laboratory	28		228	212
Branda 2013	Slovenia	CC	Possibly	3	2	5	Yes	Yes	University hospital	14		100	0
Goettner 2005	Germany	CC	No	0	3	0	No	Yes	Laboratory	10		0	110
Hansen 1989	Sweden/Denmark	CC	No	2	0	0	Yes	No	Hospital	50		200	98
Hofmann 1990	Germany	CC	No	2	0	0	No	Unclear	Laboratory, department of dermatology	25		0	205
Hofmann 1996	Germany	CC	No	2	0	0	No	No	Laboratory, department of dermatology	31		106	50
Karlsson 1989siid	Sweden	CC	No	2	0	0	Yes	No	Laboratory	10		0	73
Mathiesen 1996	Denmark	CC	No	3	0	0	Yes	Yes	Laboratory	20		100	29
Mathiesen 1998	Sweden/Denmark	CC	No	1	1	0	Yes	No	Laboratory	30		0	138
Rauer 1995	Germany	CC	No	1	0	0	Unclear	Yes	Laboratory	42		154	136
Ryffel 1998	Switzerland	CC	No	0	1*	0	Unclear	Yes	Laboratory	27		180	50
Tjernberg 2007	Sweden	CC	No	3	0	0	Yes	Yes	Hospital and laboratory	9		55	200
Wilske 1993	Germany	CC	Yes	2	0	0	Yes	Yes	Laboratory	19		100	42
Barrial 2011	France	CS	No	0	4*	0	Yes	EUCALB criteria	Laboratory	3	33		

*Note: CC=case control design; CS=cross-sectional design; DDxTot=total number of controls in a cross-sectional design; HC_Tot=total number of healthy controls; CC_Tot=total number of cross-reacting controls; ACA=total number of ACA cases included; acceptable case definition=an acceptable case definition was used in accordance with international standards; serology in case definition=serology included as part of the reference standard; *=IB was done on samples with a positive screening test.*

Acrodermatitis chronica atrophicans – case-control studies with healthy controls

Overall results and methodological quality of the studies

The analyses are based on nine case-control studies with a total of 240 persons with ACA and 1 223 healthy controls. The median number of cases per study was 27 (range 9–50), and the median number of controls was 106 (range 55–228).

The following methodological quality issues were observed:

- Because all the studies included healthy controls, they are expected to overestimate clinical specificity and thus were rated as having a high risk of bias in the patient selection domain of QUADAS-2.
- Six studies based their case definition on serology results, thus leading to high risk of bias in the reference standard domain.
- Two had unclear risk of bias and one had a low risk of bias in the reference standard domain.
- Flow and timing of all studies did pose a high risk of bias in all case control studies with healthy controls;
- The execution of the index test led to high or unclear risk of bias, except for one study with low risk of bias in one of the index tests (Mathiesen et al. 1996);
- One study reported potential conflict of interest as one author worked for Behringwerke AG and another for Mikrogen Gmbh, the manufacturers of two of the evaluated tests (Wilske et al. 1993).

Results specific to Ig type

Annex 13 provides an overview of the sensitivity and specificity by Ig type for the different studies. In most cases, the sensitivity of IgT was higher than the sensitivity of the others Ig types, while the specificity was lower. The low sensitivity of the study performed by Mathiesen et al. (1996) is due to the focus on OspC antibodies presence. Since OspC is mainly recognised in the immune response of persons with early Lyme – in late stage disease the antigen is not present on the surface of the spirochete – low sensitivity is to be expected.

The meta-analysis showed that the IgG and IgT tests were more accurate than the IgM tests, mainly due to a much higher sensitivity (Table 21).

Table 21. Summary estimates of test accuracy by antibody type for ACA case-control studies with healthy controls

Antibody tested	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	6.2 (3.0–13.1)	0.18 (0.09–0.34)	0.97 (0.93–0.98)
IgG	2088.0 (149.0–29295.0)	0.99 (0.82–0.99)	0.97 (0.95–0.98)
IgT	609.0 (107.0–3475.0)	0.98 (0.87–0.99)	0.93 (0.88–0.96)

Results specific to test type, commercial versus in-house

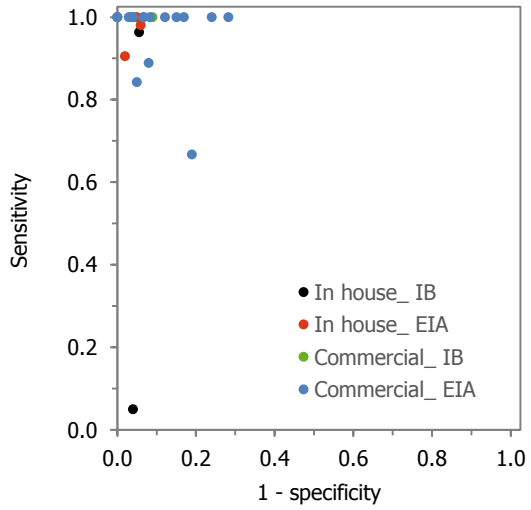
The commercial assays had a relatively high sensitivity, at the cost of a variable specificity. In contrast, the in-house assays have a relatively high specificity with variable sensitivity (Figure 19). One in house assay had an extremely low sensitivity (Mathiesen et al. 1996) because of the antigen (OspC for IgG) used. Another possible outlier (sensitivity 67%) is the Liaison test evaluated in Tjernberg et al. (2007). However, the authors did not provide an explanation for this finding.

On average, for any EIA or IB test detecting ACA patients, the diagnostic odds ratio was high: 632 (95% CI 94.7–4223.0), with a sensitivity of 0.98 (95% CI 0.84–1.00) and a specificity of 0.94 (95% CI 0.90–0.97).

As there were only four data points for the accuracy of IB tests, varying in sensitivity from 5% to 100% and varying in specificity from 91% to 100%, test type (EIA or IB) was not added as a covariate to the analyses. The effect of commercial versus in-house tests could not be assessed because only three data points from in-house EIAs were available.

Data from the EIA tests were analysed separately. For the model to converge, it was assumed that there were no variations in threshold between the studies. The diagnostic odds ratio was 625 (95% CI 189–2062), with a sensitivity of 0.97 (95% CI 0.94–0.99) and a specificity of 0.95 (95% CI 0.88–0.98).

Figure 19. ROC scatter plot for ACA case-control studies with healthy controls



Note: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Sources of heterogeneity

The sources of heterogeneity were investigated when possible, but only for the EIA tests.

- **Generation of antigens**. It was not possible to investigate the effect of recombinant antigens on the results as the model became unstable after adding covariates; the subgroups were also too small to be analysed.
- **Year of publication**. Studies were published between 1989 and 2013. There was no relation between year of publication and antigen type (P-value=0.76). When we included year of publication as a dichotomous covariate (published before or after 2000) in the analyses, the models became unstable.
- **Specific tests**. None of the individual tests was evaluated in more than three studies, hence it was not possible to model summary estimates for any of them. The raw data are provided in Annex 13.
- **Effect of methodological quality**. It was not possible to investigate the effect of quality on the results, as the model became unstable after adding covariates; subgroups were too small to be analysed.

Sensitivity analyses

In the above analyses it was assumed that borderline test results were considered positive test results. When these borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Other tests: two-tiered test

One study evaluated two-tiered tests. All had a sensitivity of 100%. Specificity was 100%, except for one study (99%) (Branda et al. 2013).

Acrodermatitis chronica atrophicans – case-control studies with cross-reacting controls

Overall results and methodological quality of the studies

The data from 12 studies were included in these analyses, with a total of 301 persons with ACA and 1 343 cross-reacting controls. These were usually patients with syphilis, other infectious diseases or auto-immune disease. The median number of cases per study was 26 (range 9 to 50), while the median number of controls was 104 (range 29 to 212).

The following methodological quality issues were observed:

- There was a high risk of selection bias as enrolment did not occur randomly or consecutively. Cross-reacting controls were selected because of their potential for false-positive results and were not representative of patients suspected of Lyme borreliosis.
- Most studies were at high risk of bias because serology was part of the case definition; three studies did not incorporate serology in the case definition.
- Only one study reported that the assessment of the (commercial) index test was blinded to the disease status of the participants.
- One study reported potential conflict of interest.

Results specific to Ig type

Annex 14 shows the sensitivity and specificity for IgG, IgM and IgT tests in the studies that looked at multiple Ig types. Overall, IgM had a lower sensitivity than the other two Ig types.

The Ig types had a significantly different accuracy (P-value=0.001 for IgG and 0.01 for IgT; IgM was reference category). This is mainly due to a lower overall sensitivity for IgM, which can be explained by the fact that ACA is generally seen in a later phase of the disease when the IgM antibodies are diminished (Table 22).

Table 22. Summary estimates of test accuracy by antibody type for ACA case-control studies with cross-reacting controls

Antibody tested	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	3.5 (1.1–11.4)	0.17 (0.09–0.28)	0.95 (0.85–0.98)
IgG	306.0 (37.1–2524.0)	0.97 (0.76–0.99)	0.92 (0.88–0.94)
IgT	178.0 (25.3–1249.0)	0.97 (0.83–0.99)	0.86 (0.69–0.95)

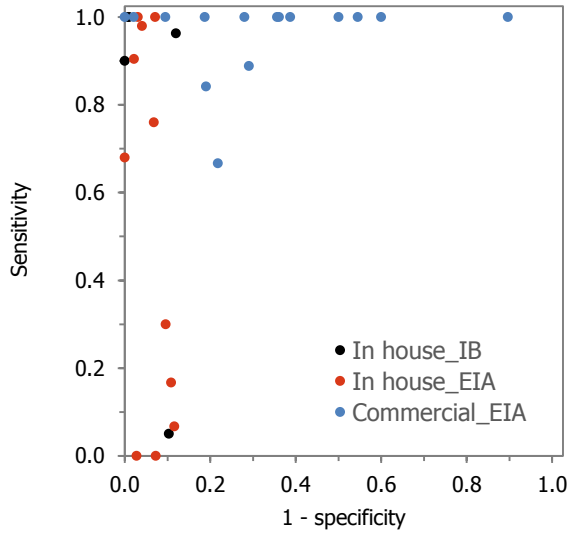
Results specific to test type, commercial versus in-house

The tests included in this analysis are more variable than the once included in the healthy control analyses, both in terms of sensitivity and specificity (Figure 20). One commercial EIA had a considerably lower specificity than any other test (Mathiesen et al. 1996). The low specificity is explained by the fact that syphilis patients were pre-selected based on a high antibody titre for a potentially cross-reacting antigen. The in-house EIAs with relatively low sensitivity were all IgM-assays. One IgG-assay with a low sensitivity used the OspC-antigen already discussed for the healthy control designs.

The overall diagnostic odds ratio for any EIA or IB test detecting ACA patients was 94.9 (95% CI 12.1–743.0), with a sensitivity of 0.91 (95% CI 0.61–0.98) and a specificity of 0.91 (95% CI 0.80–0.96).

With only five data points available for IB tests (varying in sensitivity between 5% and 100% and between 88% and 100% in specificity) test type was not added to the model as a covariate. The model fit for the EIA tests improved when adding the effect of commercial versus in-house test into the model. The summary diagnostic odds ratio is much higher for the in-house tests (including IgM tests) compared to commercial EIAs (Table 23). This is mainly driven by the threshold (P-value for threshold: 0.007): both test-variants seem to follow the same curve, but the in-house tests seem to operate at a higher threshold than the commercial tests. Thus, the researchers investigating the in-house tests seem to focus on a lower proportion of false positives, while the commercial tests seem to aim at a proportion of false negatives as low as possible. Removing the IgM tests caused the models to become unstable.

Figure 20. ROC scatter plot for ACA case-control studies with cross-reacting controls



Note: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Table 23. Summary estimates of test accuracy for commercial and in-house EIAs for ACA and cross-reacting controls

Test type	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	161.0 (2.7–9563.0)	0.88 (0.13–0.99)	0.96 (0.92–0.97)
Commercial EIA	28.6 (2.7 –303.0)	0.95 (0.82–0.99)	0.59 (0.25–0.86)

Sources of heterogeneity

- **Generation of antigens.** The accuracy between recombinant and non-recombinant antigens was evaluated. The recombinant antigens showed a higher DOR. This effect seemed to be mainly driven by the fact that the tests using recombinant antigens were operating at a different threshold (P-value for threshold parameter=0.007).
- **Year of publication.** Studies were published between 1989 and 2012. There was no relation between year of publication and antigen type (P-value=0.26). Year of publication included as a continuous covariate in the analyses did not have a significant effect on any of the parameters.
- **Specific tests.** There were no individual tests that could have been evaluated specifically, as the largest group contained only three data points. The crude results per test and per study are presented in Annex 14.
- **Effect of methodological quality.** The effect of the quality items for 'case definition' was investigated. Case definition items were included as variables in the models ('acceptable case definition' versus 'no' or 'unclear'; and 'serology in the case definition' versus 'no' or 'unclear'). None had an effect on any parameter of the models (P-values above 0.08).

Sensitivity analyses

In the above analyses it was assumed that borderline test results were considered positive test results. When these borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Acrodermatitis chronica atrophicans – cross-sectional studies

Only one cross-sectional study in patients suspected of Lyme borreliosis was found (Table 20). In this study, the researchers included, over a period of eight months, 75 patients with a positive or unclear ELISA result and aimed to confirm the patients' status with an IB. The final diagnosis of these patients was made on the basis of the EUCALB criteria and 33 patients were regarded not to have Lyme borreliosis as they had other diagnoses. The study only contained three patients with ACA, and estimating sensitivity and specificity for ACA for this study would have been pointless.

Lyme borreliosis-unspecified

Summary of the study characteristics

Lyme borreliosis (LB) may present in different ways. If a study did not distinguish between the different target conditions (EM, NB, LA or ACA), the data of this study were included in a separate analysis under the target condition 'Lyme borreliosis-unspecified' (LB-unspecified).

Twenty three studies evaluated LB-unspecified: 18 case-control studies and five cross-sectional studies. Four studies made the distinction between early Lyme borreliosis and late Lyme borreliosis. One of these four studies distinguished between 'early disseminated Lyme' and 'chronic Lyme'. The characteristics of the studies are presented in Table 24:

- Nohlmans et al. (1994) included early Lyme, late Lyme and asymptomatic Lyme. The asymptomatic patients were cross-country runners with high IgG titres. This group was excluded because their case definition was entirely based on these high titres and because they were rather atypical patients.
- Tjenberg et al. (2007) investigated four different case definitions separately and conducted an overall analysis, adding another group of 'possible Lyme patients' to the case group.
- Van Burgel et al. (2011) was excluded from the meta-analysis because all the analyses were done in CSF.
- The case-control studies evaluated EIA tests and IBs; one study evaluated two-tiered tests; one study evaluated a specific antibody index test, and one evaluated the LTT test.
- The cross-sectional studies evaluated EIA tests, IBs, two-tiered tests.
- None of the studies contained separate data on children.

The quality assessment results are presented in Annex 4.

Table 24. Characteristics of LB-unspecified studies

											Number			
Study	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Other tests	Case definition used	Acceptable case definition	Serology in case definition	Setting	LB	DDxTot	HC_Tot	CC_Tot
Ang 2011	Netherlands	CC	No	8	0	0	EM, NB, LA, ACA, LC	Unclear	No	Laboratory	59		14	16
Bergstrom 1991	Sweden	CC	Possibly	1	0	0	EM, ACA	Unclear	Yes	Microbiology Department	22		64	161
Cerar 2006	Slovenia	CC	No	1	0	0	NB, Lyme borreliosis	Yes	No	Infectious disease department/ laboratory	81		49	0
Cinco 2006	Italy	CC	No	1	0	0	EM, NB, LA	No	No	Laboratory	54		57	0
Flisiak 1998	Poland	CC	No	2	3	0	EM, NB, LA, ACA	Yes	No	Infectious disease department	48		26	0
Goossens 2000	Netherlands	CC	No	5	2	8 2-tiered	Early and late Lyme	No	Unclear	Not stated	39		62	128
Goossens 2001	Netherlands	CC	No	2	0	0	EM, NB, ACA	Unclear	Unclear	Laboratory	67		62	140
Gueglio 1996	France	CC	No	1	0	0	NR	No	Unclear	Laboratory	31		30	24
Hernandez 2003	Spain	CC	No	0	1*	0	EM, NB	Yes	Unclear	Hospital laboratory	18		0	129
Jovivic 2003	Serbia/Monte negro	CC	No	1	1	0	EM, NB, LA, LC	Yes	Unclear	Department microbiology	46		0	120
Lange 1991	Germany	CC	Yes	1	1	0	Not specified	No	Unclear	NR	50		100	89
Nohlmans 1994	Netherlands/ Switzerland	CC	No	5	0	0	EM, ACA, LA	Unclear	Unclear	Hospital laboratory	34		84	46
Oksi 1995	Finland	CC	No	3	0	0	Lyme borreliosis	Unclear	No	Hospital and laboratory	41		37	0
Rijpkema 1994	Netherlands	CC	No	2	0	0	EM, atypical symptoms + tick bite/ atypical symptoms	No	No	Public Health laboratory	61		0	41
Smismans 2006	Netherlands	CC	No	3	0	0	EM, NB, LA	Unclear	Yes	Laboratory	22		0	40
Tjernberg 2007	Sweden	CC	No	3	0	0	EM, NB, LA, ACA	Yes	Yes	Hospital and laboratory	227		55	200
VanBurgel 2011	Netherlands	CC	No	1	0	2 specific AI	NB	Yes	Yes	Laboratory	72	74	0	69
VonBaehr 2007	Germany	CC	No	0	0	1 LTT	EM, NB, LA	Unclear	Yes	Laboratory	44	0	136	0

											Number			
Study	Country	Design	Conflict of interest	Number ELISA tested	Number IBs tested	Other tests	Case definition used	Acceptable case definition	Serology in case definition	Setting	LB	DDxTot	HC_Tot	CC_Tot
Bazovska 2001	Slovakia	CS	No	2	2*	2 2-tiered	NB	Unclear	No	Not reported	54	25	0	0
Blaauw 1999	Netherlands	CS	No	1	1*	1 2-tiered	Lyme borreliosis	Yes	No	Hospital, rheumatology department	10	93	0	0
Cermakova 2005	Czech Republic	CS	No	5	0	0	Borreliosis (no further specification)	No	Yes	Routine laboratory	54	36	0	0
Ekerfelt 2004	Sweden	CS	No	5	0	0	EM, NB, ACA	Unclear	Unclear	Laboratory	92	25		0
Kolmel 1992	Germany	CS	No	1	1	0	Bannwarth's syndrome and other borrelioses	No	Unclear	Hospital, neurology department	16	784	0	0

*CC=case-control; CS=cross-sectional; NR=not reported; DDxTot=total number of controls in a cross-sectional design; HC_Tot=total number of healthy controls; CC_Tot=total number of cross-reacting controls; case definition used=case definitions included in the study; acceptable case definition=an acceptable case definition was used in accordance with international standards; serology in case definition=serology included as part of the reference standard; COI=conflict of interest; *=IB was done on samples with a positive screening test.*

Lyme borreliosis-unspecified – case-control studies with healthy controls

Overall results and methodological quality of the studies

The analyses are based on 13 studies with a total of 797 persons with LB-unspecified and 776 healthy controls. The median number of cases per study was 48 (range 22 to 227), and the median number of healthy controls was 57 (range 14 to 136).

The following methodological quality issues were observed:

- All studies were rated as having a 'high risk of bias' for the patient selection domain, as healthy controls are assumed to overestimate clinical specificity;
- All studies were also rated as causing 'high concerns regarding applicability' for the patient selection domain, as healthy controls are not representative of the patients tested in practice;
- Most studies were rated 'unclear' on risk of bias due to the execution of the index test, due to not reporting whether the assessment of the index test was blinded to the disease status of the participants;
- None of the studies posed a 'low risk of bias'; only two studies were rated as having a 'low risk of bias' for the reference standard domain (Cerar et al. 2006, Flisiak et al. 1998); the other studies either included serology in their case definition or did not further elaborate on this topic.
- One study reported potential conflict of interest (Lange et al. 1991). The acknowledgements mention that the study was 'scientifically supported' by the company that provided the tests.

Results specific to Ig type

Annex 15 provides an overview of the sensitivity and specificity by Ig type and study. One study evaluated several tests, both for early and late Lyme borreliosis, and showed that for most tests the IgM tests were more sensitive in the early phase. In the later phases, IgG tests were generally more sensitive (Goossens et al. 2000).

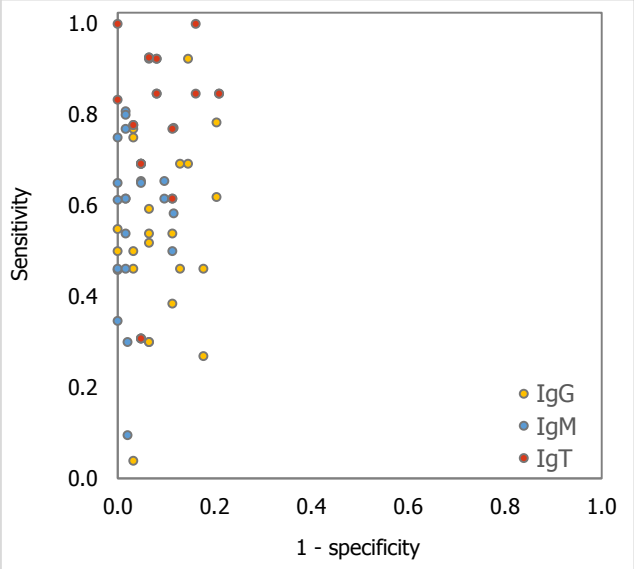
A meta-analysis was not possible, probably due to the variation in the data points. Assuming an underlying summary ROC curve, one would expect that the highest sensitivity has the lowest specificity and vice versa – or that at least the largest variations in specificity can be found among the higher estimates of sensitivity. This is not the case, however, especially not for IgM. Therefore, only the median values for sensitivity and specificity for IgM, IgG and IgT are presented (Table 25). Although different Ig types have a sensitivity of up to 100%, three quarters of all observations for IgM and IgG have a sensitivity below 70%. For IgT, three quarters of the observations are above 75% sensitivity (Figure 21).

Table 25. Median sensitivity and specificity for IgM, IgG and IgT in LB-unspecified case-control studies with healthy controls

Antibody tested	Sensitivity median (P25-P75)	Specificity P25-median-P75
IgM	0.62 (0.49–0.66)	0.98 (0.95–0.98)
IgG	0.54 (0.46–0.69)	0.94 (0.87–0.97)
IgT	0.85 (0.77–0.92)	0.92 (0.89–0.95)

Note: P25: 25th percentile; P75: 75th percentile

Figure 21. ROC scatter plot of IgM, IgG and IgT tests of LB-unspecified case-control studies with healthy controls



Results specific to test type and commercial versus in-house

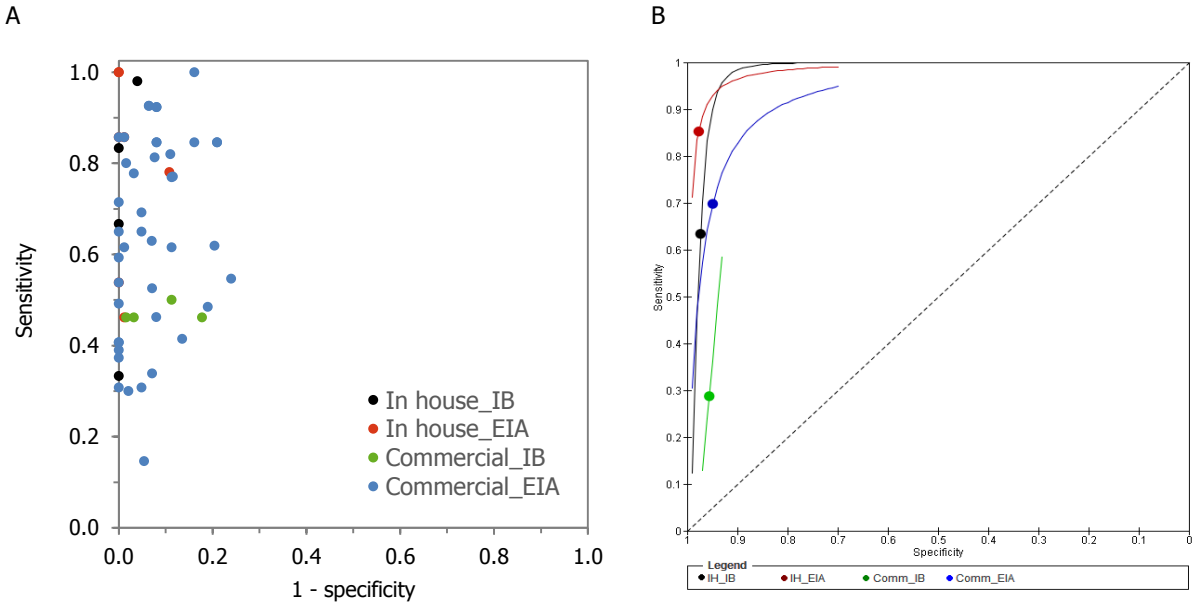
The ROC scatter plot shows the sensitivity and specificity estimates for studies done in serum (Figure 22A).

The overall diagnostic odds ratio for any EIA or IB test for detecting patients with LB-unspecified was 71.3 (95% CI 13.8–369), with a sensitivity of 0.73 (95% CI 0.53–0.87) and a specificity of 0.96 (95% CI 0.91–0.99).

When adding test type (EIA or IB) to the analyses, the model fit improved and had a significant effect on the shape of the summary ROC curve (P=0.017), but not on the other parameters. Addition of a covariate to account for commercial or in-house tests further improved the model fit and showed a significant difference in accuracy between the commercial and in-house tests (P-value=0.008) (Table 26, Figure 22).

The estimates for the commercial IBs are based on four data rows retrieved from only one study and on four data rows from two studies for the in-house IBs. The estimates for the commercial EIA are based on 43 data rows from 10 studies and therefore more precise. Direct comparisons between EIA and IB were not made because only three comparative studies were included.

Figure 22. ROC scatter plot (A) and fitted ROC curves (B) for LB-unspecified case-control studies with healthy controls



Note: Graph A: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

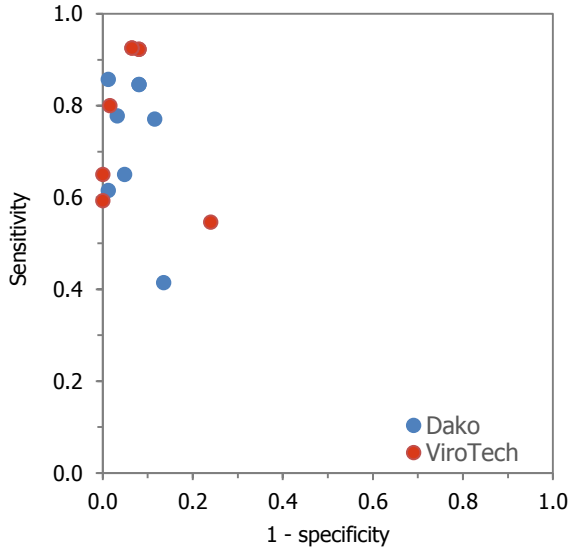
Table 26. Summary estimates of test accuracy for commercial and in-house IB and EIAs for LB-unspecified case-control studies with healthy controls

Label	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	245.0 (53.2–1128.0)	0.85 (0.71–0.93)	0.98 (0.93–0.99)
In-house IB	63.1 (11.3–353.0)	0.63 (0.33–0.86)	0.97 (0.93–0.99)
Commercial EIA	43.1 (12.5–149.0)	0.70 (0.52–0.83)	0.95 (0.89–0.98)
Commercial IB	9.0 (1.0–77.4)	0.29 (0.07–0.68)	0.96 (0.90–0.98)

Sources of heterogeneity

- **Generation of antigens.** Three categories of antigen-generations were compared. The category recombinant or a combination including recombinant antigen had the lowest accuracy mainly due to the low sensitivity, thought the differences were not statistically significant (**Table** 27).
- **Time effect (year of publication).** There was no statistically significant relation between year of publication and antigen type (P-value=0.082). Including year of publication as a binary covariate in the analyses did not improve the fit of the model, but showed a borderline significant effect on accuracy (P-value=0.049) (Table 27).
- **Specific tests.** ViroTech was evaluated in four studies (10 data rows) and the Dako Flagellum in five studies (8 data rows). ViroTech test had a DOR of 17.3 (95% CI1.2–260.7), with a sensitivity of 0.68 (95% CI 0.42–0.86) and a specificity of 0.89 (95% CI 0.59–0.98), while the Dako test had a DOR of 41.5 (95% CI 6.6–258.9), with a sensitivity of 0.71 (95% CI 0.48–0.87) and a specificity of 0.94 (95% CI 0.83–0.98). The two tests showed clear variations in sensitivity and specificity depending on the study (Figure 23). The results of the other tests are presented in Annex 15.

Figure 23. ROC scatter plot for two commercial tests



- **Methodological quality.** All studies had high risk of bias in patient sampling and flow and timing. Fourteen studies had a high risk of bias in the index test execution for at least one included index test. The studies that did not include serology in the case definition were analysed separately. They tended to have lower DOR, but not significantly different from the overall estimates (**Table** 27).
- **Early LB-unspecified versus late LB-unspecified.** Four studies reported results separately for early LB and five studies reported results separately for late LB. Adding late versus early LB to the models did improve the model fit and showed that the tests had, on average, a lower sensitivity in early LB. However, this was not a significant effect (Table 27).

Table 27. Summary estimates of test accuracy taking into account sources of heterogeneity for LB-unspecified case-control studies with healthy controls

Sources of heterogeneity	Test categories	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Antigen generation	Whole cell	45.0 (12.2–165.0)	0.70 (0.56–0.81)	0.95 (0.86–0.98)
	Purified antigen	140.0 (7.13–2751.0)	0.84 (0.46–0.97)	0.97 (0.86–0.99)
	Recombinant or a combination including recombinant antigen	9.7 (3.45–27.2)	0.46 (0.25–0.69)	0.92 (0.81–0.97)
Year of publication	<2000	263.0 (37.2–1863.0)	0.85 (0.63–0.95)	0.98 (0.93–0.99)
	2000 or later	17.8 (3.3–96.6)	0.58 (0.37–0.76)	0.93 (0.81–0.98)
Methodological quality	Overall	71.3 (13.8–369.0)	0.73 (0.53–0.87)	0.96 (0.91–0.99)
	Serology not in case definition	48.2 (13.6–172.0)	0.70 (0.53–0.83)	0.95 (0.90–0.98)
Early versus late LB	Overall studies	82.3 (7.3–926.0)	0.77 (0.47–0.93)	0.96 (0.85–0.99)
	Early LB	45.9 (6.1–345.0)	0.60 (0.32–0.83)	0.97 (0.90–0.99)
	Late LB	88.7 (11.8–664.0)	0.78 (0.55–0.93)	0.96 (0.88–0.99)

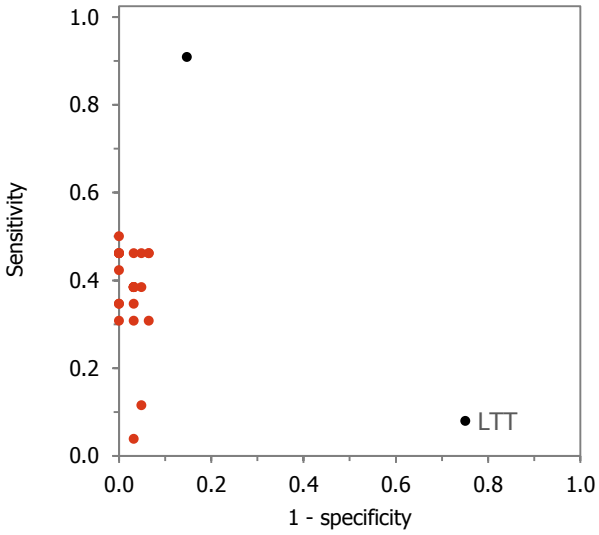
Sensitivity analyses

In the above analyses it was assumed that borderline test results were considered positive test results. When these borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Other tests: two-tiered test and LTT

One study evaluated two-tiered tests, both in early LB and in late LB-unspecified. None of the evaluations showed sensitivity above 50% (one was 50%) (Goossens et al. 2000). Von Baehr et al. (2007) evaluated the LTT with 44 cases and 136 controls and reported a sensitivity of 91% and a specificity of 85% (Figure 24).

Figure 24. ROC scatter plot for LTT and two-tiered test for LB-unspecified



Lyme borreliosis-unspecified – case-control studies with cross-reacting controls

Overall results and methodological quality of the studies

The analyses were based on twelve studies with a total of 676 persons with LB-unspecified and 1 134 cross-reacting controls (Table 24). The median number of cases per study was 43 (range 18 to 227) and the median number of controls was 105 (range 16 to 200). The controls were usually patients with syphilis, other infectious diseases, auto-immune diseases or neurological conditions. Eleven studies evaluated between one and five (one study evaluated eight) EIAs in serum and four studies evaluated one or two IBs. One study evaluated eight different two-tiered tests.

The following methodological quality issues were observed:

- There was a high risk of selection bias as enrolment did not occur randomly consecutively. Cross-reacting controls are selected because of their potential for false-positive results and were not representative of patients suspected of Lyme borreliosis;
- For most studies (n=9) it was unclear whether the used reference standard posed a high or low risk of bias; for the remaining two it was clear that the case definitions included serology and thus that there was a high risk of bias;
- Whether the execution of the index test may lead to bias was either high or impossible to assess, except for three studies with a high risk of bias in the execution of at least one index test;
- One study reported potential conflict of interest (Lange et al. 1991).

Results specific to Ig type

IgM was often more sensitive than IgG in the early phases of disease than in the later phases. IgT had the highest sensitivity in most of the cases, but also the lowest specificity (Annex 16).

The accuracy of the three Ig types differed significantly from each other (Table 28). This is due to the fact that the shape of the curves, and thus the way sensitivity and specificity change when the threshold changes, is different (P-value=0.003 for IgG and P-value=0.0066 for IgT), and because the average threshold at which the tests operate are different (P-value=0.045 for IgG and 0.002 for IgT) (Figure 25).

Figure 25. Fitted summary ROC curves for IgG, IgM and IgG test types

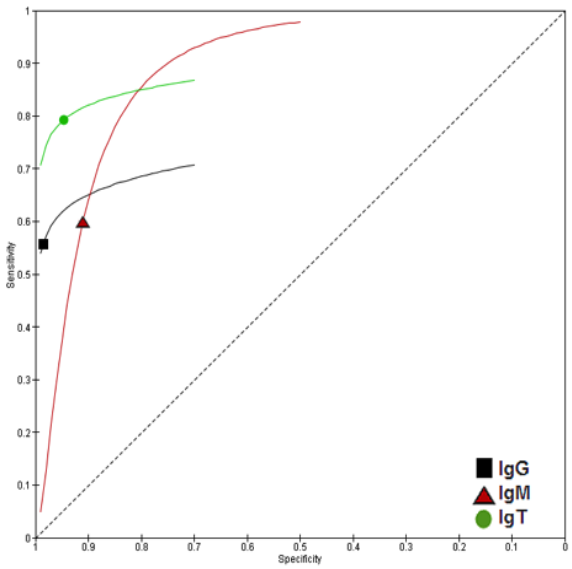


Table 28. Summary estimates of test accuracy by antibody type for LB-unspecified case-control studies with cross-reacting controls

Antibody tested	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
IgM	15.1 (6.5–35.2)	0.60 (0.32–0.82)	0.91 (0.82–0.96)
IgG	85.6 (11.5–636.0)	0.56 (0.45–0.66)	0.99 (0.88–0.99)
IgT	67.5 (13.4–342.0)	0.79 (0.69–0.87)	0.95 (0.73–0.99)

Results specific to test type, commercial versus in-house and Ig type

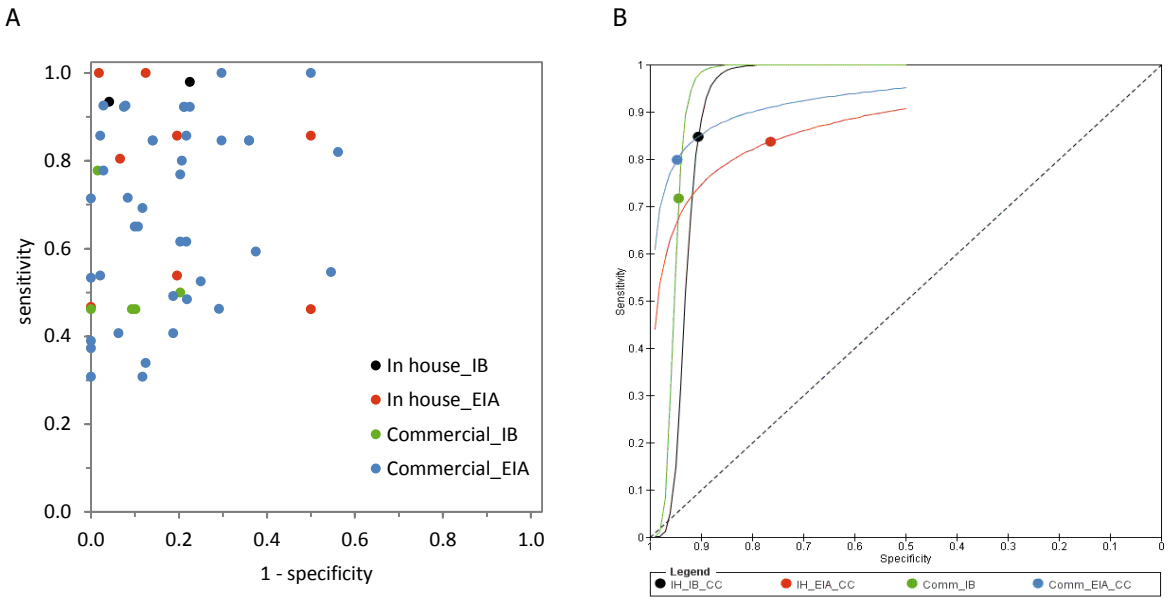
The ROC scatter plot showed a high level of heterogeneity for sensitivity and specificity (Figure 26A).

The overall accuracy for any EIA or IB test in serum for detecting patients with LB-unspecified was DOR 38.3 (95% CI 10.6 to 138), with an average sensitivity of 0.81 (95% CI 0.64–0.91) and a specificity of 0.90 (95% CI 0.79–0.96).

When adding test type (EIA or IB) to the analyses, the model fit improved, and test type had a significant effect on threshold (P-value=0.0044) and shape of the summary ROC curve (P-value=0.0062). Addition of a covariate that accounted for commercial or in-house test further improved the model fit and had a significant effect on the threshold parameter (P-value=0.0013). The IBs had a significantly higher accuracy than EIAs, and the commercial EIAs had a significantly higher accuracy than the in-house tests (Figure 26, Table 29).

These comparisons were based on comparative studies (including both IB and EIA) and non-comparative studies (containing only IB or EIA; the majority). Only two studies included both EIA and IB, which was not enough to compare the two test types directly.

Figure 26. ROC scatter plot (A) and fitted summary ROC curves (B) for LB-unspecified case-control studies with cross-reacting controls



Note: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

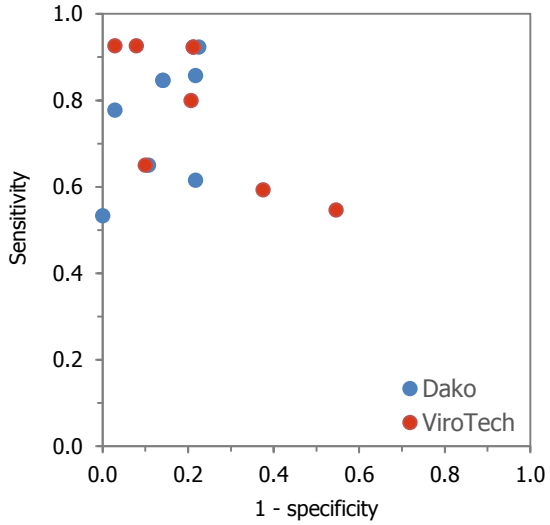
Table 29. Summary estimates of test accuracy for commercial and in-house IB and EIAs for LB-unspecified case-control studies with cross-reacting controls

Label	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
In-house EIA	16.6 (2.8–99.6)	0.84 (0.68–0.93)	0.76 (0.42–0.93)
In-house IB	53.1 (3.23–873.0)	0.85 (0.27–0.99)	0.91 (0.85–0.94)
Commercial EIA	70.5 (12.0-413)	0.80 (0.63–0.90)	0.95 (0.80–0.99)
Commercial IB	42.0 (2.5–698.0)	0.72 (0.15–0.98)	0.94 (0.90–0.97)

Sources of heterogeneity

- **Generation of antigens.** The following categories were analysed separately: whole-cell lysate or sonicate, purified antigens, recombinant or synthetic antigens, and combined categories of antigen generations from four studies. The antigen types were similar but the purified antigens tended to have a higher accuracy than the other types (**Table** 30).
- **Year of publication.** Studies were published between 1991 and 2011. There was no relation between year of publication and antigen type (P-value=0.12). Year of publication was included as a binary variable, before/after 2000. Year had no effect on either of the parameters (P-values all above 0.1) (Table 30).
- **Specific tests.** The ViroTech test was evaluated in three studies (10 data rows) and the Dako Flagellum test in four studies (8 data rows). The two tests showed clear variations in sensitivity and specificity depending on the study (Figure 27). The results of all tests are presented in Annex 16.

Figure 27. ROC scatter plot for two commercial tests



- **Quality of the studies.** There were not enough high-quality studies for the analyses. The only subgroup that could be analysed separately were studies that did not include serology in their case definition or that did not report if serology was part of the case definition. This had no impact on the accuracy.
- **Early LB-unspecified versus late LB-unspecified.** Four studies reported results separately for early LB and five studies reported results separately for late LB. Adding late versus early LB to the models did improve the model fit and showed that the tests had on average a lower sensitivity in early LB than in late LB. However, this was not a significant effect (Table 30).

Table 30. Summary estimates of test accuracy for LB-unspecified case-control studies with cross-reacting controls taking into account sources of heterogeneity

Sources of heterogeneity	Test categories	Diagnostic odds ratio (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Antigen generation	Whole cell	11.9 (3.4–42.1)	0.70 (0.53–0.83)	0.84 (0.63–0.94)
	Purified antigen	56.1 (8.3–380.0)	0.85 (0.52–0.97)	0.91 (0.78–0.96)
	Recombinant antigen	18.2 (1.4–236.0)	0.70 (0.32–0.92)	0.89 (0.55–0.98)
	Any combination	14.4 (1.0–199.0)	0.77 (0.13–0.99)	0.81 (0.35–0.97)
Early versus late Lyme	Timing not reported	16.7 (2.8–98.2)	0.73 (0.44–0.91)	0.86 (0.62–0.96)
	Early Lyme	35.9 (8.0–162.0)	0.79 (0.56–0.92)	0.91 (0.77–0.97)
	Late Lyme	123.0 (27.1–560.0)	0.91 (0.76–0.97)	0.93 (0.82–0.97)

Sensitivity analysis

In the above analyses it was assumed that borderline test results were considered positive test results. When these borderline test results were considered negative, no significant differences in DOR, sensitivity and specificity were observed.

Other tests: two-tiered test

Goossens et al. (2000) evaluated two-tiered tests in early and in late LB. The results were comparable to those for the study with healthy controls, except that there was more variation in specificity. None of the evaluations revealed sensitivity above 50% (one was 50%), and specificity varied between 88% and 100%.

Lyme borreliosis-unspecified – cross-sectional studies

Overall results and methodological quality of the studies

Four studies used a cross-sectional design to evaluate LB-unspecified. In total, the studies covered 226 patients with Lyme borreliosis (median 54, range 10–92) and a total of 963 persons without Lyme borreliosis (median of 35, range 25–784) (Table 31).

Two studies focused primarily on NB (Bazovska et al. 2001, Kolmel et al. 1992). Because they also included other forms of Lyme borreliosis, it was decided to analyse them under LB-unspecified. Ekerfelt et al. (2004) was included in the previous analyses on NB, as this study reported two sufficiently large groups of patients: one group of NB patients and one group of other borreliosis patients.

The following methodological quality issues were observed:

- Bazovska et al. (2001) did not report sufficient information to rate the risk of bias in most quality domains; only flow and timing were rated as having a low risk of bias; there were low concerns regarding applicability.
- Blaauw et al. (1999) were rated as having a low risk of bias in patient selection and reference standard, but raised high concerns regarding applicability because the study mainly included arthritis patients.
- Cermakova et al. (2005) were rated as having a low risk of bias in patients and flow and timing, but included serology as part of the reference standard.
- Kölmel et al. (1992) included serology in the reference standard and was rated 'unclear' for the other sources of bias.
- None of the studies reported potential conflicts of interest.

Table 31. Overview of cross-sectional studies for Lyme borreliosis-unspecified

Study	Patients' characteristics	Prevalence
Bazovska 2001	Study conducted in neurology department; most included patients had neurological symptoms. Seventy-nine patients included; classified into five categories: Lyme borreliosis, suspected Lyme borreliosis or unknown inflammatory disease, potentially different disease, improbable diagnosis, or other confirmed diagnosis. One EIA test, one IB test and a two-tiered combination of the two were analysed.	Definitive LB: 0.24 (19/79) Definite and suspected: 0.51 (40/79)
Blaauw 1999	Study conducted in departments of rheumatology and immunology; 103 patients with persisting musculoskeletal complaints or who believed they had chronic Lyme borreliosis; 49 were classified as previous Lyme borreliosis, 10 were classified as active Lyme borreliosis and 44 patients were classified as no Lyme borreliosis. One EIA test, one IB test and a two-tiered combination of the two were analysed.	0.097 (10/103)
Cermakova 2005	ELISA tests conducted on samples from 90 patients from a teaching hospital or general practice. Five different EIA tests were analysed.	0.60 (54/90)
Ekerfelt 2004	Origin of suspected samples not clearly explained; 117 suspected samples were ranked based on their likelihood to have NB or another form of borreliosis. The patients without Lyme borreliosis had facial palsy or other symptoms. Five different EIA tests were analysed for IgG and for IgM.	0.79 (92/117)
Kolmel 1992	All samples (n=800) from neurology department were tested. One EIA test and one IB test were analysed. No further information about symptomology, but study included a wide range of final diagnoses, not all neurological.	0.02 (16/800)

Results specific to Ig type

Only one study evaluated more than one Ig type (Ekerfelt et al. 2004). For all tests, IgM had a lower sensitivity than IgG, but a higher or equal specificity. Meta-analyses were not possible due to limited data.

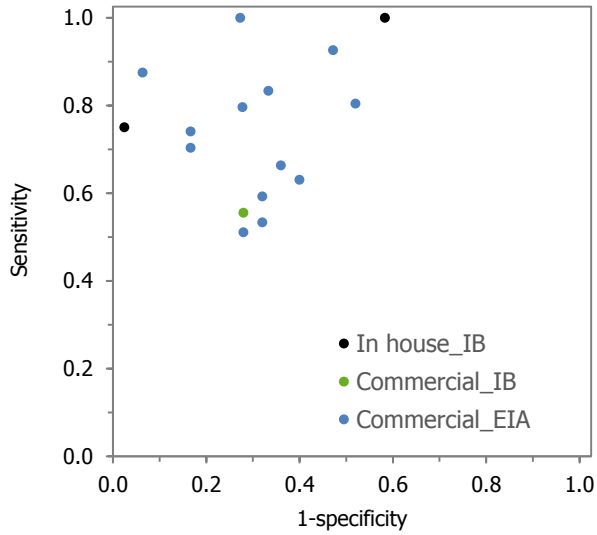
Results specific to test type, commercial versus in-house

The ROC scatter plot showed much heterogeneity (Figure 28).

The overall accuracy for any EIA or IB test in serum for detecting LB was DOR 11.6 (95% CI 1.47–91.2), with a sensitivity of 0.77 (95% CI 0.48–0.93) and a specificity of 0.77 (95% CI 0.46–0.93).

No covariable test type was added as there were only three scattered data points for IB tests (Figure 28). Subgroup analyses of the EIA tests returned an overall accuracy of DOR 12.6 (95% CI 0.77 to 208), with a sensitivity of 0.80 (95% CI 0.49–0.92) and a specificity of 0.76 (95% CI 0.49–0.92).

Figure 28. ROC scatter plot for LB-unspecified cross-sectional studies



Note: Every dot represents a 2x2 table, one for each test. One study may have contributed more than one 2x2 table.

Sources of heterogeneity

There were not enough data to produce valid models for any of the potential sources of heterogeneity. Studies were published in 1992, 1999, 2001, 2004 and 2005. There was no significant relationship between year of publication and the antigen type (P-value=0.24).

Results for specific tests and situations can be found in Annex 17.

Other tests: two-tiered test

Two studies evaluated a two-tiered approach (Bazovska et al. 2001, Blaauw et al. 1999). In one study, the sensitivity was 0.46 and the specificity 0.84, while in the other study the sensitivity was 1.00 and the specificity 0.84.

Discussion

Main findings

This study aims to assess the sensitivity and specificity of serology tests for the different target conditions of Lyme borreliosis and the diagnostic accuracy of tests like IBs or those done on CSF. Sources of heterogeneity were evaluated for each target condition.

A total of 78 of the 8026 unique studies found in the initial search were included in this study. Three types of study design were considered eligible for inclusion: case-control studies with healthy controls, case-control studies with cross-reacting controls, and cross-sectional surveys. Case-control studies, especially those using healthy controls, can be seen as an evaluation of the maximum capacity of a diagnostic test's performance. It can be assumed that the risk of misclassification in case-control studies is very low, i.e. cases have a high probability of having Lyme borreliosis and healthy controls have a very low probability of having the disease. The case-control studies with cross-reacting controls provide more information about situations where the test's ability to discriminate between Lyme borreliosis and other diseases that may mimic Lyme borreliosis is difficult. Cross-sectional studies are ideal to answer the review questions [4,5]. If well designed, these studies provide valid estimates of sensitivity and specificity and can also directly provide estimates of prevalence and predictive values.

Lyme borreliosis encompasses several clinical syndromes. Each of these presentations may be seen as a separate target condition for laboratory testing because they affect different body parts and different organ systems and because patients suffering from these conditions may enter and move through the healthcare system differently. Hence, the target conditions included in the study (EM, NB, LA, and ACA) were analysed separately. If a study did not distinguish between the different target conditions, its data were included in a separate analysis of 'LB-unspecified'.

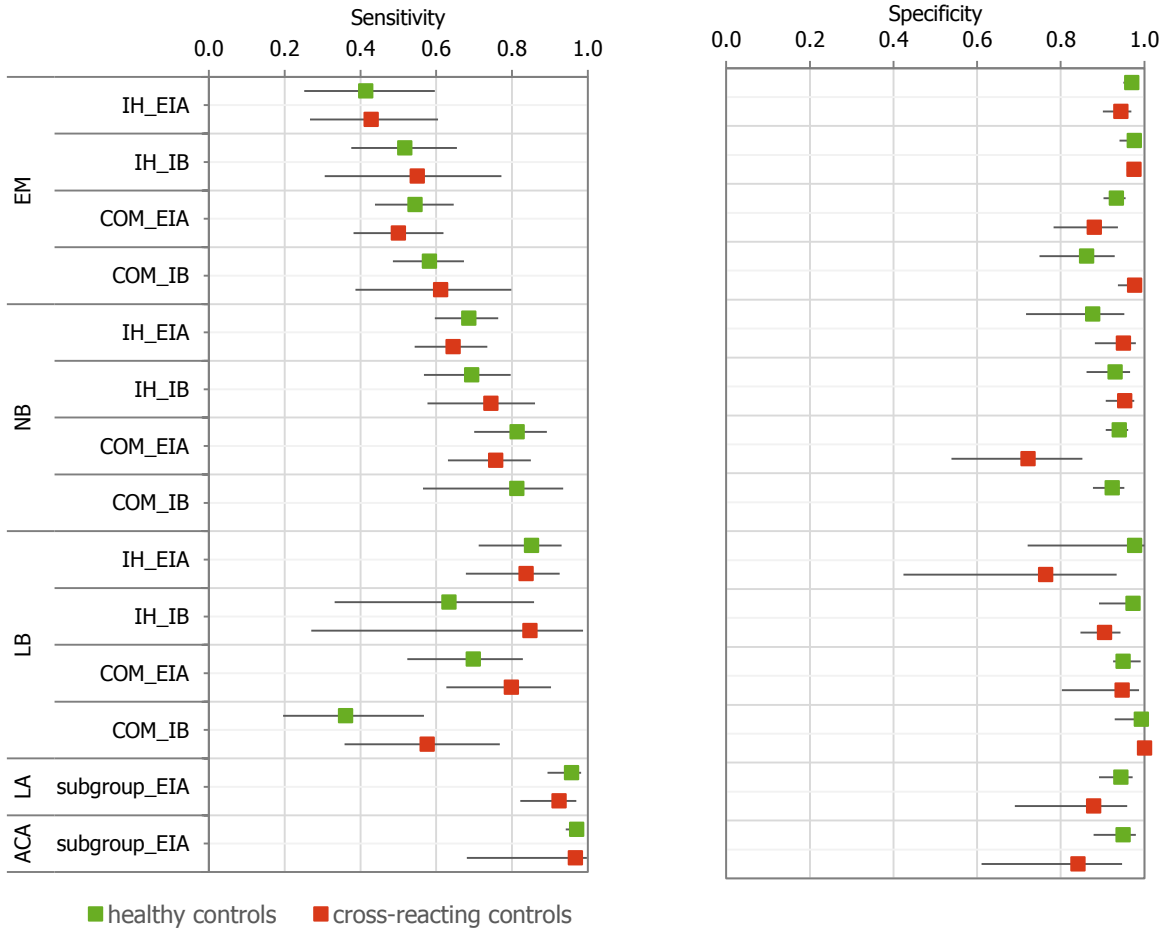
The main finding of this study can be summarised as follows (Figure 29):

- The overall accuracy in case-control studies with healthy controls was the lowest for EM, with an overall sensitivity of 0.50 (95% CI 0.40–0.61) and an overall specificity of 0.95 (95% CI 0.92–0.97). The EIA tests performed less well than the IB tests, and commercial tests did not perform significantly different from in-house tests. The 23 case-control studies on EM with cross-reacting controls had similar results.
- For NB, the overall sensitivity was 0.77 (95% CI 0.67–0.85), with a specificity above 0.90, except for the in-house EIAs (specificity 0.88, 95% CI 0.72–0.95) and for studies in which serology did not form part of the reference standard (specificity 0.87, 95% CI 0.74–0.94). Twenty-six case-control studies with cross-reacting controls showed similar results.
- For LB-unspecified the overall sensitivity was 0.73 (95% CI 0.53–0.87), with summary estimates for sensitivity between 0.25 and 0.90. This may be explained by the variable nature of the included conditions, i.e. the study did not distinguish between the different target conditions and could have been a combination of the other conditions or no specifications whatsoever.
- The highest overall estimates for sensitivity were observed for LA and ACA. For these conditions, only reliable estimates for EIA tests (commercial and in-house combined) were available in this review. The overall sensitivity for LA was 0.96 (95% CI 0.89–0.98) and for ACA 0.97 (95% CI 0.94– 0.99). The specificity was above 0.90 for most tests and for both the healthy and cross-reacting controls tests. However, there were some outliers, and there was more variation in the cross-reacting controls than in the healthy controls.

The above-mentioned results were obtained when considering the borderline results as positive. When borderline results were considered negative, overall sensitivity decreased and specificity increased, but not significantly.

In general, the IgG tests had a comparable sensitivity to the IgM tests, except for EM, with IgM showing a slightly higher sensitivity, which was still below 0.50. For LA and ACA, IgM showed much lower sensitivity (below 0.40). IgT had the highest sensitivity and the lowest specificity, which can be explained by the fact that the IgT test is positive when either IgM or IgG is positive. The specificity of the IgT tests was still above 0.80 in most cases.

Figure 29. Summary estimates of sensitivity and specificity by target conditions and test type for case-control studies with healthy and cross-reacting controls



Regarding cross-sectional studies, a meta-analysis could only be conducted for NB and LB-unspecified because only for these target conditions a sufficient number of cross-sectional studies was available. As expected, the specificity in these studies was considerably lower than in the case-control designs. For NB, the overall specificity in case-control studies with cross-reacting controls was 0.90 (95% CI 0.83–0.94), whereas in cross-sectional studies it was 0.81 (95% CI 0.54–0.94). For LB-unspecified, the specificity in the cross-sectional studies was 0.77 (95% CI 0.46–0.93), whereas it was 0.90 (95% CI 0.78–0.96) in the studies with cross-reacting controls. The sensitivity was not very different: slightly higher in the case-control studies for NB and slightly lower for LB-unspecified. The main challenge for the cross-sectional studies was the definition of the target condition and the availability of a reference standard.

One study evaluated two-tiered testing in several target conditions. The sensitivity of the two-tiered tests varied, but was below 60% for EM and LB-unspecified and above 60% for LA and ACA. In NB, two-tiered tests had a sensitivity of around 40% and three others of around 85%. The specificity for all two-tiered tests was above 90%.

Some of the studies that evaluated tests for NB – mainly studies with cross-reacting controls – also evaluated specific antibody index tests. The summary estimate for sensitivity was 0.87 (95% CI 0.72–0.95) and 0.94 (95% CI 0.83–0.98) for specificity. In NB studies, CSF was sometimes tested instead of serum. The sensitivity was lower than for serum, but the specificity was higher.

Sources of heterogeneity

For the evaluation of antigen types as source of heterogeneity, tests were categorised in three groups – whole-cell lysate, purified antigens, recombinant or synthetic antigens – which is a simplification of the wide variety of antigens, Borrelia strains, test set-up, and manufacturer. Only in NB, recombinant antigens had a higher accuracy, with both a higher sensitivity and a higher specificity in serum. This group was the only group where year of publication had a significant impact on the results, with the newer studies showing higher sensitivity and specificity. For LB-unspecified, the recombinant antigen had a lower sensitivity (0.46) and specificity (0.92) than the whole cell

and purified antigen assays. Here, the newer studies had a lower accuracy, with a lower sensitivity and specificity and with broad confidence intervals.

Only for the LB-unspecified group it was possible to directly compare the early stages of disease with the later stages. On average, the serological assays showed a lower accuracy in the early stages of the disease, combined with lower sensitivity and slightly higher specificity. The difference in sensitivity was at least 10 percentage points.

Summary estimates for specific tests were difficult to obtain. The results per manufacturer varied considerably. For example, the Enzygnost test was evaluated in six studies with healthy controls including EM cases, and its sensitivity varied between 50% and 95%, whereas the specificity varied between 80% and 100%. The ViroTech test was evaluated in NB studies with healthy controls. Its sensitivity varied between 40% and 90%, its specificity was between 80% and 100%.

Another issue potentially influencing the results might be the fact that the study by Ang et al. (2012) contained at least eight different 2x2 tables for each target condition and therefore potentially influenced the results significantly. A sensitivity analysis from which a 2x2 table from that study was removed showed that its effect was only marginal (Annex 18). In fact, the largest difference in sensitivity or specificity was 4 percentage points. For ACA the difference observed in the DOR was larger but coincided with a difference in sensitivity of 1 percentage point and 0 percentage points in specificity.

The quality of the studies varied as well. The main problems were absence of blinding of reference standard and test assessments and ad-hoc choice of the cut-off value. On average, one would expect the studies of higher quality to show results with less overestimation. For EM, NB, LA and ACA, studies with higher quality or lower risk of bias had a similar accuracy to the lower quality studies or the overall estimates. The analyses for LB-unspecified showed an effect, especially when including serology in the case definitions; if serology was not included in the case definitions, than the accuracy was lower.

Limitations of the evidence

According to the GRADE approach [10,11], five issues need to be considered when evaluating the overall body of evidence, namely limitations (risk of bias), indirectness, imprecision, inconsistency and publication bias.

- Limitations: Of the 58 case-control studies, 46 could not be rated 'low risk of bias' and exceeded this criterion in all four domains. Seven of the case-control studies were rated 'high risk of bias' in all four domains. Seven of the cross-sectional studies exceeded 'low risk of bias' in all of the domains, and none was rated 'low risk of bias' in all domains. The largest problems were encountered in the selection and flow of patients.
- Indirectness: Indirectness in accuracy reviews plays a role at two levels: 1) the indirectness of the study estimates of sensitivity and specificity is related to concerns regarding applicability; 2) the translation from accuracy estimates to important outcomes for patients is complex, as sensitivity and specificity give no direct answer to the question if the patient actually benefits from testing. In order to make inferences about this question (and thus about important outcomes for patients), one has to make several assumptions, which lower the confidence in the final result.
- Imprecision: Imprecision depends on the number of studies included, but also on the number of diseased and non-diseased people. In the forest plots this is reflected as the confidence interval around the estimates. In general, precision for the specificity in the case-controls studies is high, but for the cross-sectional designs and for the sensitivity it varies.
- Inconsistency: This review shows inconsistency on multiple levels. First, there is much variation between the results. Although this variation can be partly explained by the use of different Ig types, some variation and thus inconsistency in the results remains, even within the same Ig type. Second, there is inconsistency in the comparison between EIA and IB. In general, IB has a lower sensitivity than EIA, but this is often not significant and in some cases it is reversed. In general, IB has a higher specificity than EIA, but not in the cross-sectional studies for NB. This inconsistency lowers the credibility of the results.
- Publication bias: No direct evidence of publication bias was found other than that the in-house tests show less variation than the commercial tests. Testing for publication bias is not recommended for accuracy reviews [12]. Considering the wide network of experts in the field involved in this review, it is not expected that published work has been missed.

In conclusion, the overall estimates of accuracy are not very reliable as their quality may be downgraded based on the above-mentioned considerations and should therefore be used with caution and seen only as indicative.

Implications for clinical practice and surveillance

The implications for clinical practice depend on the role that the tests fulfil in practice and the setting in which they will be used. The average sensitivity of commercial EIA for EM was 0.54 (95% CI 0.44 –0.65). This means that of all tested people who have EM, an average of 46% of cases will not be detected. The corresponding specificity of 0.93 (95% CI 0.90–0.95) means that of the people without EM, 7% will test positive and be regarded as having EM. In clinical practice, this test does not provide an added value, as the predictive value of the test is too low to be useful. These results are in line with the current guidelines which do not recommend testing and suggest that patients with EM are treated immediately [1].

When testing for NB, the question is what manifestation of Lyme borreliosis is anticipated: do clinicians test to specifically rule in or rule out NB or do they test for or against Lyme borreliosis in general? The cross-sectional studies included in this review mostly included patients with neurological symptoms, for example facial palsy, which makes NB the target condition. This implies that at a sensitivity of 0.80 (95% CI 0.59–0.92), as found for all EIAs or IBs, 20% of the neurological patients with NB will test negative. At a corresponding specificity of 0.81 (95% CI 0.54–0.94), 19% of the neurological patients without NB will test positive. The practical implications of these numbers depend on the prevalence or pre-test probability in the tested group. In the cross-sectional studies included in this review, the prevalence (or pre-test probability) varied between 14% and 77%. The median prevalence was 50% when 'possible' Lyme cases were considered actual cases, and 41% when they were considered controls. This is a high prevalence, but the group of patients tested is a highly specific group because they presented with neurological symptoms, and it was expected that prevalence would be higher in this group than in a more heterogeneous group of patients with different types of symptoms suggestive of Lyme borreliosis. At a prevalence of 41% and a sensitivity of 0.80, the post-test probability of having Lyme after a positive test result, or positive predictive value, will be 75% (95% CI 70%–78%) and the post-test probability of not having NB after a negative test result (the negative predictive value) will be 85% (95% CI 82%–88%) at a specificity of 0.81. That means that of all positive EIA or IB tests, an average of 75% will indeed have NB, and that of all negative test results, 15% will actually have NB. The implications of these percentages depend on the clinical consequences of testing.

In patients suspected of NB, other test algorithms and samples are also possible. When testing in CSF, the specificity may be somewhat higher. When using the specific antibody index, the specificity may be higher while the sensitivity is not necessarily lower than in tests done in serum alone. This may lead to a decrease in false positives without a decrease in false negatives.

Whether two-tiered testing should be recommended depends on the algorithm used and the correlation between the test results. We have not found evidence that the accuracy of two-tiered approaches is higher than that of single tests.

When these tests are used to detect Lyme borreliosis in general, the summary estimates for LB-unspecified may be more relevant. Patients with the target condition are patients with any form of Lyme borreliosis, while the controls are patients without any form of the disease. The sensitivity of any EIA or IB in the cross-sectional studies for this target condition was 0.77 (95% 0.48–0.93). This means that of all patients with EM, NB, ACA, LA or any other form of Lyme borreliosis, 23% will test negative. The specificity was also 0.77 (95% CI 0.46–0.93), meaning that of all patients without Lyme borreliosis, 23% will test positive. Here as well, the practical implications depend on the pre-test probability. In practice, this pre-test probability may be far more diverse as indicated by the prevalence of any form of Lyme borreliosis in the cross-sectional studies which varied from 2% to 79%. At a prevalence of 2%, the positive predictive value will be 6%, meaning that of all positive tests only 6% will actually have Lyme borreliosis and, the negative predictive value will be 99%, meaning that of all negative tests only 1% will actually have Lyme borreliosis. When the prevalence is 24%, the positive predictive value is 51% (51% of all positive tested patients will have LB) and the negative predictive value is 91% (9% of all negatively tested patients will actually have LB). These predictive values were calculated with the same sensitivity and specificity in both situations, but the accuracy of a test used in a 2% prevalence situation may be different from that of a test used in a 24% prevalence situation. Furthermore, the included studies were very variable and not always representative of routine clinical practice. Therefore, these numbers should be interpreted with caution. They show that for a better understanding of the practical usefulness of these tests, more information is needed on the population that will be tested, the target condition that will be tested for, and the pre-test probability of Lyme borreliosis.

The sensitivity and specificity for LA and ACA were very high, and clinicians testing for these target conditions would probably do so in a relatively specific group of patients in which the prevalence might also be higher than in patients with less specific symptoms. However, these sensitivity and specificity estimates come from case-control studies that used healthy controls. Cross-sectional studies showed more variation. Here as well, more information is needed on the population that will be tested, the target condition that will be tested for, and the pre-test probability of ACA and LA, in order to be able to assess the practical value of these tests.

For the surveillance of Lyme disease clear case definitions are essential. The availability of several laboratory tests and the uncertainty over the overall estimates of accuracy, points to the need for a better evaluation,

standardisation, and understanding of the usability and quality of the different tests in Europe. The situation is further complicated by different patient populations in different regions (e.g. the presence of other diseases resembling Lyme borreliosis) and differences in the healthcare systems.

Implications for further research

New studies should focus on the place of the test in the clinical pathway and include those patients that are tested in practice. Prospective studies with a consecutive sample of patients are preferable and should be conducted in the same location where patients are tested (either a neurology department, a general practice, or another clinical department). The patient follow-up should run long enough to obtain a sufficient degree of certainty about their actual condition and the final diagnosis.

One of the hurdles for setting up these studies is the lack of a gold standard. The gold standard, or reference standard, is essential in test accuracy research. Sensitivity and specificity are defined in terms of the probability that a test provides the correct results, i.e. the reference standard. However, the need for an absolute gold standard may be debateable [13] and alternatives might be explored, for example a clinically relevant definition of disease; the use of an ordinal reference standard with multiple categories of certainty such as no LB, probably no LB, possibly LB, certainly LB; statistical approaches such a latent class analyses; or expert opinion or composite reference standards consisting of multiple tests and measurements.

Finally, the link between test accuracy and outcomes in patient management has to be reviewed (for example if non-treatment results in disease progression). A perfectly accurate test may still be of little use if it does not change clinical management decisions, e.g. if the clinical picture is sufficiently discriminating between people who would and who would not benefit from treatment. On the other hand, a moderately accurate test may be very useful in a situation where clinical picture alone is not enough to guide therapeutic decisions.

Conclusions

This systematic review provides an overview of diagnostic accuracy of serological tests for different target conditions of Lyme borreliosis. Overall, the diagnostic accuracy of EIA and IBs for Lyme borreliosis in Europe varies widely.

The translation from accuracy estimates to important outcomes for patients is complex, as sensitivity and specificity do not cover all aspects of this task. The performance of diagnostic tests depends on the population in which the test will be used because, for example, the prevalence of the condition in question, which is needed to estimate predictive values, differs. Tests should therefore be evaluated in the population where they are used. The performance of a test and the implications of testing may be different in different healthcare settings, e.g. in a general practice or a tertiary clinic. The same applies to the differences between hospital departments, for example a dermatology department (for ACA) and a neurological department (for NB). Performance, implications and prevalence may differ substantially in these settings.

Key messages

- Based on the available literature, the serological tests for the different clinical target conditions of Lyme borreliosis had a sensitivity of around 80% and a specificity of around 95%, which in some cases could reach 100%. For LA and ACA, the sensitivity was around 95%. For EM, the sensitivity was around 50%.
- The sensitivity and specificity estimates from this review may be used to provide a first idea of the possible ranges in predictive values when a test is used in different patient groups. Readers should interpret these predictive values with caution because the results showed a high level of variation and the included studies were at high risk of bias.
- The data in this review do not provide sufficient evidence to make inferences about the value of the tests for clinical practice. More information is needed, including prevalence of Lyme borreliosis among those tested and the clinical consequences of a negative or positive test result. The latter depend on the place of the test in the clinical pathway and the clinical decisions that are driven by the test results. The performance of diagnostic tests depends strongly on the population in which the test is being used.
- Serological test results for the diagnosis of Lyme borreliosis needs to be interpreted with caution and are only supportive of the diagnosis in combination with a clinical presentation compatible with the established case definitions.
- Better designed diagnostic accuracy studies will provide more valid estimates of the tests' accuracy (e.g. by supplying predictive values), but the actual added value of testing for Lyme disease requires information about subsequent actions and consequences of testing.
- Future research should primarily focus on more targeted clinical validation of these tests and research into appropriate usage. The lack of a gold standard for most manifestations of LB may be solved in future studies by using a reference standard with multiple levels of certainty or by statistical approaches like latent class analysis, use of expert opinion, and/or response to treatment.

References

1. Stanek G, Wormser GP, Gray J, Strle F. Lyme borreliosis. Lancet. 2012 Feb 4;379(9814):461-73.
2. Stanek G, Fingerle V, Hunfeld KP, Jaulhac B, Kaiser R, Krause A, et al. Lyme borreliosis: Clinical case definitions for diagnosis and management in Europe. Clinical Microbiology and Infection. 2011 Feb 2;17:69-79.
3. Mygland A, Ljostad U, Fingerle V, Rupprecht T, Schmutzhard E, Steiner I. EFNS guidelines on the diagnosis and management of European Lyme neuroborreliosis. Eur J Neurol. 2010 Jan;17(1):8-16.
4. Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis - Evaluation of diagnostic procedures. Brit Med J. 2002 Feb 23;324(7335):477-80.
5. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. Journal of clinical epidemiology. 2003 Nov;56(11):1118-28.
6. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005 Aug;51(8):1335-41.
7. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine. 2011 Oct 18;155(8):529-36.
8. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 100: The Cochrane Collaboration; 2009.
9. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001 Oct 15;20(19):2865-84.
10. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. Evidence Based Medicine. 2008 Dec;13(6):162-3.
11. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. BMJ. 2008 May 17;336(7653):1106-10.
12. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. Journal of clinical epidemiology. 2005 Sep;58(9):882-93.
13. Lord SJ, Staub LP, Bossuyt PM, Irwig LM. Target practice: choosing target conditions for test accuracy studies that are relevant to clinical practice. Brit Med J. 2011;343:d4684.

Annexes

Annex 1. Search strategy

Search strategy in OvidSP: database Embase, 1980 to present, date search 10 Jan 2013		
Line#	**Term**	**Results**
1	exp serology/	171111
2	serolog*.ti,ab,ot.	99141
3	antibod*.ti,ab,ot.	739175
4	exp antibody/	745964
5	immunoglobin*.ti,ab,ot.	1034
6	IgG.ti,ab,ot.	119346
7	IgM.ti,ab,ot.	58826
8	exp enzyme linked immunosorbent assay/	185612
9	ELISA.ti,ab,ot.	136569
10	exp immunoassay/	343973
11	EIA.ti,ab,ot.	10028
12	immunosorbent.ti,ab,ot.	64523
13	immunofluorescent.ti,ab,ot.	16475
14	immunofluorescence.ti,ab,ot.	87513
15	immunoblot*.ti,ab,ot.	67674
16	'western blot'.ti,ab,ot.	98174
17	immunoassay.ti,ab,ot.	45568
18	exp lymphocyte transformation test/	1200
19	'lymphocyte transformation test'.ti,ab,ot.	826
20	LTT.ti,ab,ot.	574
21	((t-cell* or lymphocyte) adj15 (diagnostic or diagnosis or diagnosing or screen* or test*)).mp.	41706
22	VIDAS.ti,ab,ot.	709
23	liason.ti,ab,ot.	104
24	Enzygnost.ti,ab,ot.	233
25	Serion.ti,ab,ot.	58
26	recomline.ti,ab,ot.	21
27	(virotech adj5 (europline or 'line blot')).ti,ab,ot.	1
28	euroimmunoblot.ti,ab,ot.	0
29	diacheck.ti,ab,ot.	2
30	euroimmun.ti,ab,ot.	191
31	Medac.ti,ab,ot.	136
32	mikrogen$.ti,ab,ot.	78
33	virotech.ti,ab,ot.	53
34	ELISPOT.ti,ab,ot.	5408
35	exp enzyme linked immunospot assay/	4884
36	(c6 adj3 immunetics).ti,ab,ot.	5
37	or/1-36	1629954
38	lyme.ti,ab,ot.	10067
39	Lyme borreliosis/	11571
40	tick borne disease/	1803
41	exp tick bite/	1956
42	(tick adj2 bite).ti,ab,ot.	1400
43	Neuroberreliosis.ti,ab,ot.	0
44	exp erythema chronicum migrans/	1641
45	erythema migrans.ti,ab,ot.	1187
46	'erythema chronicum migrans'.ti,ab,ot.	433
47	exp Acrodermatitis chronica atrophicans/	155
48	'Acrodermatitis chronica atrophicans'.ti,ab,ot.	426
49	meningoradiculitis.ti,ab,ot.	269
50	lyme.ti,ab,ot.	10067
51	Lyme borreliosis/	11571
52	tick borne disease/	1803
53	exp tick bite/	1956

Search strategy in OvidSP: database Embase, 1980 to present, date search 10 Jan 2013		
54	(tick adj2 bite).ti,ab,ot.	1400
55	Neuroborreliosis.ti,ab,ot.	1091
56	exp erythema chronicum migrans/	1641
57	erythema migrans.ti,ab,ot.	1187
58	'erythema chronicum migrans'.ti,ab,ot.	433
59	exp Acrodermatitis chronica atrophicans/	155
60	'Acrodermatitis chronica atrophicans'.ti,ab,ot.	426
61	meningoradiculitis.ti,ab,ot.	269
62	lyme.ti,ab,ot.	10067
63	Lyme borreliosis/	11571
64	tick borne disease/	1803
65	exp tick bite/	1956
66	(tick adj2 bite).ti,ab,ot.	1400
67	exp erythema chronicum migrans/	1641
68	erythema migrans.ti,ab,ot.	1187
69	'erythema chronicum migrans'.ti,ab,ot.	433
70	exp Acrodermatitis chronica atrophicans/	155
71	'Acrodermatitis chronica atrophicans'.ti,ab,ot.	426
72	meningoradiculitis.ti,ab,ot.	269
73	Neuroborreliosis.ti,ab,ot.	1091
74	or/38-73	16588
75	exp Borrelia/	10262
76	borrelia.ti,ab,ot.	8824
77	burgdorferi.ti,ab,ot.	7378
78	Borrelia infection/	2668
79	or/75-78	12748
80	VLsE.ti,ab,ot.	142
81	OspC.ti,ab,ot.	446
82	or/80-81	551
83	74 or 79	20529
84	82 or 83	20551
85	37 and 84	7578
86	37 or 82	1630187
87	83 and 86	7789
88	85 or 87	7790
89	animal/ not human/	1348171
90	88 not 89	7369
91	review.pt.	1907269
92	90 not 91	6510

Annex 2. Rating of QUADAS-2 items

1. Patient selection

1a. Risk of bias, signalling questions

- Was a case-control design avoided?
  - Case-control designs, especially if they include healthy controls, carry a high risk of bias. Therefore, all case-control studies are automatically rated to be of high risk of bias in the overall judgement.
- Was a consecutive or random sample of patients enrolled?
- Did the study avoid inappropriate exclusions?
- Overall judgement:
  - Case-control studies were always rated as having a high risk of bias.
  - Cross-sectional studies: only low risk of bias if the other two signalling questions are answered with 'yes'. If one of the questions was answered 'no', then high risk of bias. Otherwise 'unclear'.

1b. Concerns regarding applicability: this concerns the extent to which the patients (both cases and controls) that were included in a study are representative for the patients which will receive these serology tests.

- Is there concern that the included patients do not match the review question?
  - All case-control studies are by default rated 'high concern'. All cross-sectional studies are by default 'low concern', except when the used case definition was not very clear.
  - One study only included facial palsy patients → high concern: applicable, but not a representative group
  - One study only included arthritic patients → high concern: applicable, but not a representative group

2. Index test

2a. Risk of bias, signalling questions

- Were the index test results interpreted without knowledge of the results of the reference standard?
  - This was very poorly reported in the studies, hence almost all cases were rated 'unclear'.
- If a threshold was used, was it pre-specified? By selecting the cut-off value with the highest sensitivity and/or specificity, researchers artificially optimise the accuracy of their tests, which may lead to an overestimation of sensitivity and specificity.
  - The responses to this question vary. Sometimes the authors state that the 95% value of the controls is used as threshold, or that the mean of the controls plus 2 or 3 SD is used as threshold. Both variations were rated as post-hoc.
- Overall judgement:
  - If the second question is answered with 'yes', overall judgement is automatically 'yes'. This is because the first question is usually not reported or answered with 'yes'.
  - If the latter is rated as 'unclear', then overall is also 'unclear'; if the latter receives a 'no', the overall risk is considered 'high'.

2b. Concerns regarding applicability: this concerns the extent to which the index test evaluated is representative for the tests that will be used in practice.

- Are there concerns that the index test, its application or interpretation deviate from the review question?
  - All in-house tests are automatically rated as 'high concern'.

Risk of bias and concerns regarding applicability should be assessed for each test separately.

3. Reference standard

3a. Risk of bias, signalling questions

- Is the reference standard likely to correctly classify the target condition?
  - Assumption: This will likely be the case for case-control studies that use the 'correct' case definitions (e.g. Stanek [1], WHO)
  - This is also likely for cross-sectional studies which use the 'correct' case definitions.
  - Studies using Western blot as reference standard will receive a 'no' in response to this question.
- Were the reference standard results interpreted without knowledge of the results of the index test?
  - Assumption: This will likely be the case for most case-control studies, but only if serology was not part of the case definition.
  - For cross-sectional studies, this should be explicitly stated.
- Overall judgement risk of bias:
  - Case-control studies with clear case definitions were rated as having a 'low' risk of bias.
  - case-control studies with unclear/wrong case definitions rated as 'unclear' Or 'high risk' of bias respectively
  - Cross-sectional studies with a clear case definition and the second question answered with 'yes': low risk of bias.
  - Otherwise 'unclear',

3b. Concerns regarding applicability: Are there concerns that the target condition as defined by the reference standard does not match the review question?

- Western blots measure antibody response, while this review focuses on Lyme borreliosis, regardless of antibody status. Consequently, Western blots are considered to have high concerns regarding applicability.
- If serology is included in the case definition, there is an incorporation bias and thus a high risk of bias.
- If a case-control study uses clear criteria and does not include serology in these criteria: 'low' concern of bias.

4. Risk of bias regarding flow and timing, signalling questions

- Was there an appropriate interval between index test(s) and reference standard?
  - We expected that studies with a cross-sectional design conducted most tests on a date sufficiently close to the final diagnosis. If we had reason to suspect that the patient status changed between the time of testing and the time of diagnosis, we rated this as 'no'.
  - For case-control studies, this was always rated as 'no', because serology was always determined after the case definitions were defined, sometimes with a long delay.
- Did patients receive the same reference standard?
  - We rated this as 'no' for all case-control studies, as the controls were often from different settings, different departments and had to meet different criteria.
- Were all patients included in the analysis?
  - This was rated 'no' for all case-control studies.
- Overall judgement:
  - Case-control studies were always rated as having a high risk of bias.
  - For cross-sectional studies, we perceived a low risk of bias if all three questions were answered with 'yes'; a high risk of bias was perceived if at least one of them was answered 'no'. All other possibilities were rated as 'unclear'.

Annex 3. Studies included in the analysis

Albisetti M, Schaer G, Good M, Boltshauser E, Nadal D. Diagnostic value of cerebrospinal fluid examination in children with peripheral facial palsy and suspected Lyme borreliosis. Neurology. 1997;49:817-24.

Ang CW, Brandenburg AH, van Burgel ND, Bijlmer HA, Herremans T, Stelma FF, et al. Nationale vergelijking van serologische assays voor het aantonen van Borrelia-antistoffen. Nederlands tijdschrift voor Medische Microbiologie. 2012;20:111-9.

Ang CW, Notermans DW, Hommes M, Simoons-Smit AM, Herremans T. Large differences between test strategies for the detection of anti-Borrelia antibodies are revealed by comparing eight ELISAs and five immunoblots. European Journal of Clinical Microbiology and Infectious Diseases. 2011;30:1027-32.

Barrial K, Roure-Sobas C, Carricajo A, Boibieux A, Tigaud S. Performance of four commercial immunoblots for serological confirmation of Lyme borreliosis in Rhone-Alpes, France. [French] Evaluation des performances de quatre immunoblots commercialises pour la confirmation de la borreliose de Lyme dans une population de patients suivis en Rhone-Alpes. Immuno-Analyse et Biologie Specialisee. 2011;26:194-200.

Bazovska S, Kondas M, Simkovicova M, Kmety E, Traubner P. Significance of specific antibody determination in Lyme borreliosis diagnosis. Bratislavske lekarske listy. 2001;102:454-7.

Bednarova J. Cerebrospinal-fluid profile in neuroborreliosis and its diagnostic significance. Folia Microbiologica. 2006;51:599-603.

Bennet R, Lindgren V, Zweygberg Wirgart B. Borrelia antibodies in children evaluated for Lyme neuroborreliosis. Infection. 2008;36:463-6.

Bergstrom S, Sjostedt A, Dotevall L, Kaijser B, Ekstrand-Hammarstrom B, Wallberg C, et al. Diagnosis of Lyme borreliosis by an enzyme immunoassay detecting immunoglobulin G reactive to purified Borrelia burgdorferi cell components. European Journal of Clinical Microbiology and Infectious Diseases. 1991;10:422-7.

Blaauw AAM, Van Loon AM, Schellekens JFP, Bijlsma JWJ. Clinical evaluation of guidelines and two-test approach for Lyme disease. Rheumatology. 1999;38:1121-6.

Blaauw I, Dijkmans B, Bouma P, Van Der Linden S. Rational diagnosis and treatment in unclassified arthritis: How clinical data may guide requests for Lyme serology and antibiotic treatment. Annals of the Rheumatic Diseases. 1993;52:206-10.

Blanc F, Jaulhac B, Fleury M, De Seze J, De Martino SJ, Remy V, et al. Relevance of the antibody index to diagnose Lyme neuroborreliosis among seropositive patients. Neurology. 2007;69:953-8.

Branda JA, Strle F, Strle K, Sikand N, Ferraro MJ, Steere AC. Performance of United States serologic assays in the diagnosis of Lyme borreliosis acquired in Europe. Clin Infect Dis. 2013 Aug;57(3):333-40.

Cerar T, Ogrinc K, Strle F, Ruzic-Sabljic E. Humoral immune responses in patients with lyme neuroborreliosis. Clinical and Vaccine Immunology. 2010;17:645-50.

Cerar T, Ruzic-Sabljic E, Cimperman J, Strle F. Comparison of immunofluorescence assay (IFA) and liaison in patients with different clinical manifestations of Lyme borreliosis. Wiener Klinische Wochenschrift. 2006;118:686-90.

Cermakova Z, Ryskova O, Honegr K, Cermakova E, Hanovcova I. Diagnosis of Lyme borreliosis using enzyme immunoanalysis. Medical Science Monitor. 2005;11:BR121-BR5.

Christova I. Enzyme-linked immunosorbent assay, immunofluorescent assay, and recombinant immunoblotting in the serodiagnosis of early Lyme borreliosis. Int J Immunopathol Pharmacol. 2003 Sep-Dec;16(3):261-8.

Cinco M, Murgia R. Evaluation of the C6 enzyme-linked immunoadsorbent assay for the serodiagnosis of Lyme borreliosis in north-eastern Italy. New Microbiol. 2006 Apr;29(2):139-41.

Davidson MM, Ling CL, Chisholm SM, Wiseman AD, Joss AW, Ho-Yen DO. Evidence-based diagnosis of Lyme disease. European Journal of Clinical Microbiology and Infectious Diseases. 1999 Jul;18(7):484-9.

Dessau RB. Diagnostic accuracy and comparison of two assays for Borrelia-specific IgG and IgM antibodies: proposals for statistical evaluation methods, cut-off values and standardization. Journal of Clinical Microbiology. 2013 Dec;62(Pt 12):1835-44.

Dessau RB, Ejlertsen T, Hilden J. Simultaneous use of serum IgG and IgM for risk scoring of suspected early Lyme borreliosis: graphical and bivariate analyses. APMIS. 2010 Apr;118(4):313-23.

Ekerfelt C, Ernerudh J, Forsberg P, Jonsson AL, Vrethem M, Arlehag L, et al. Lyme borreliosis in Sweden - Diagnostic performance of five commercial Borrelia serology kits using sera from well-defined patient groups. APMIS. 2004;112:74-8.

Flisiak R, Kalinowska A, Bobrowska E, Prokopowicz D. Enzyme immunoassay in the diagnosis of Lyme borreliosis. Rocz Akad Med Bialymst. 1996;41(1):83-9.

Flisiak R, Wierzbicka I, Prokopowicz D. Western blot banding pattern in early Lyme borreliosis among patients from an endemic region of north-eastern Poland. Rocz Akad Med Bialymst. 1998;43:210-20.

Goettner G, Schulte-Spechtel U, Hillermann R, Liegl G, Wilske B, Fingerle V. Improvement of lyme borreliosis serodiagnosis by a newly developed recombinant immunoglobulin G (IgG) and IgM line immunoblot assay and addition of VlsE and DbpA homologues. Journal of Clinical Microbiology. 2005;43:3602-9.

Goossens HA, van den Bogaard AE, Nohlmans MK. Serodiagnosis of Lyme borreliosis using detection of different immunoglobulin (sub)classes by enzyme-linked immunosorbent assay and Western blotting. Clin Lab. 2001;47(1-2):41-9.

Goossens HAT, Van den Bogaard AE, Nohlmans MKE. Reduced specificity of combined IgM and IgG enzyme immunoassay testing for Lyme borreliosis. European Journal of Clinical Microbiology and Infectious Diseases. 2000;19:400-2.

Gueglio B, Poinsignon Y, Berthelot JM, Marjolet M. Western blot specificity in the diagnosis of Lyme Borreliosis. Med Maladies Infect. 1996 Mar;26(3):332-7.

Hansen K, Asbrink E. Serodiagnosis of erythema migrans and acrodermatitis chronica atrophicans by the Borrelia burgdorferi flagellum enzyme-linked immunosorbent assay. Journal of Clinical Microbiology. 1989 Mar;27(3):545-51.

Hansen K, Hindersson P, Pedersen NS. Measurement of antibodies to the Borrelia burgdorferi flagellum improves serodiagnosis in Lyme disease. Journal of Clinical Microbiology. 1988 Feb;26(2):338-46.

Hansen K, Lebech AM. Lyme neuroborreliosis: a new sensitive diagnostic assay for intrathecal synthesis of Borrelia burgdorferi--specific immunoglobulin G, A, and M. annals of Neurology. 1991 Aug;30(2):197-205.

Hernandez-Novoa B, Orduna A, Bratos MA, Eiros JM, Fernandez JM, Gutierrez MP, et al. Utility of a commercial immunoblot kit (BAG-Borrelia blot) in the diagnosis of the preliminary stages of Lyme disease. Diagnostic Microbiology and Infectious Disease. 2003 Sep;47(1):321-9.

Hofmann H. Lyme borreliosis--problems of serological diagnosis. Infection. 1996 Nov-Dec;24(6):470-2.

Hofmann H, Meyerkonig U. Serodiagnosis of dermatological manifestations of Borrelia burgdorferi infection. Hautarzt. 1990 Aug;41(8):424-31.

Hofstad H, Matre R, Nyland H, Ulvestad E. Bannwarth's syndrome: serum and CSF IgG antibodies against Borrelia burgdorferi examined by ELISA. Acta Neurologica Scandinavica. 1987 Jan;75(1):37-45.

Hunfeld KP, Ernst M, Zachary P, Jaulhac B, Sonneborn HH, Brade V. Development and laboratory evaluation of a new recombinant ELISA for the serodiagnosis of Lyme disease. Wiener Klinische Wochenschrift. 2002 Jul 31;114(13-14):580-5.

Huppertz HI, Mosbauer S, Busch DH, Karch H. Lymphoproliferative responses to Borrelia burgdorferi in the diagnosis of Lyme arthritis in children and adolescents. European Journal of Pediatrics. 1996 Apr;155(4):297-302.

Jansson C, Carlsson SA, Granlund H, Wahlberg P, Nyman D. Analysis of Borrelia burgdorferi IgG antibodies with a combination of IgG ELISA and VlsE C6 peptide ELISA. Clinical Microbiology and Infection. 2005;11:147-50.

Jovicic VL, Grego EM, Lako BL, Ristovic BM, Lepsanovic ZA, Stajkovic NT. Improved serodiagnosis of early Lyme borreliosis: Immunoblot with local Borrelia afzelli strain. APMIS. 2003;111:1053-9.

Kaiser R, Rauer S. Analysis of the intrathecal immune response in neuroborreliosis to a sonicate antigen and three recombinant antigens of Borrelia burgdorferi sensu stricto. European Journal of Clinical Microbiology & Infectious Diseases. 1998 Mar;17(3):159-66.

Kaiser R, Rauer S. Serodiagnosis of neuroborreliosis: comparison of reliability of three confirmatory assays. Infection. 1999 May-Jun;27(3):177-82.

Karlsson M, Granstrom M. An IgM-antibody capture enzyme immunoassay for serodiagnosis of Lyme borreliosis. Serodiag Immunother Infect Dis. 1989;3:413-21.

Karlsson M, Mollegard I, Stiernstedt G, Wretlind B. Comparison of Western blot and enzyme-linked immunosorbent assay for diagnosis of Lyme borreliosis. European Journal of Clinical Microbiology & Infectious Diseases. 1989 Oct;8(10):871-7.

Kolmel HW, Neumann P, Lange R. Prevalence of Borrelia-burgdorferi antibodies in 800 neurological patients. [German] Korrelation zwischen neurologischer erkrankung und Borrelia-burgdorferi-antikorpern in 800 serum/liquorpaaren. Nervenarzt. 1992;63:619-24.

Lahdenne P, Panelius J, Saxen H, Heikkila T, Sillanpaa H, Peltomaa M, et al. Improved serodiagnosis of erythema migrans using novel recombinant borrelial BBK32 antigens. Journal of Medical Microbiology. 2003;52:563-7.

Lakos A, Ferenczi E, Komoly S, Granstrom M. Different B-cell populations are responsible for the peripheral and intrathecal antibody production in neuroborreliosis. International Immunology. 2005 Dec;17(12):1631-7.

Lange R, Schneider T, Topel U, Mater-Bohm H, Beck A, Kolmel HW. Borrelia burgdorferi antibody-tests: Comparison of IFT, ELISA and Western-blot. [German]. Laboratoriums Medizin. 1991;15:128-9.

Lencakova D, Fingerle V, Stefancikova A, Schulte-Spechtel U, Petko B, Schreter I, et al. Evaluation of recombinant line immunoblot for detection of Lyme disease in Slovakia: Comparison with two other immunoassays. Vector-Borne and Zoonotic Diseases. 2008;8:381-90.

Ljostad U, Okstad S, Topstad T, Mygland A, Monstad P. Acute peripheral facial palsy in adults. Journal of Neurology. 2005;252:672-6.

Marangoni A, Moroni A, Accardo S, Cevenini R. Borrelia burgdorferi VlsE antigen for the serological diagnosis of Lyme borreliosis. European Journal of Clinical Microbiology and Infectious Diseases. 2008;27:349-54.

Marangoni A, Sparacino M, Cavrini F, Storni E, Mondardini V, Sambri V, et al. Comparative evaluation of three different ELISA methods for the diagnosis of early culture-confirmed Lyme disease in Italy. Journal of Medical Microbiology. 2005;54:361-7.

Marangoni A, Sparacino M, Mondardini V, Cavrini F, Storni E, Donati M, et al. Comparative evaluation of two enzyme linked immunosorbent assay methods and three Western Blot methods for the diagnosis of culture-confirmed early Lyme Borreliosis in Italy. New Microbiol. 2005;28:37-43.

Mathiesen MJ, Christiansen M, Hansen K, Holm A, Asbrink E, Theisen M. Peptide-based OspC enzyme-linked immunosorbent assay for serodiagnosis of Lyme borreliosis. Journal of Clinical Microbiology. 1998;36:3474-9.

Mathiesen MJ, Hansen K, Axelsen N, Halkier-Sorensen L, Theisen M. Analysis of the human antibody response to outer surface protein C (OspC) of Borrelia burgdorferi sensu stricto, B. garinii, and B. afzelii. Medical Microbiology and Immunology. 1996;185:121-9.

Nicolini P, Imbs P, Jaulhac B, Piemont Y, Warter JM, Monteil H. Comparison of indirect immunofluorescence assay with immunoblot for serodiagnosis of neuroborreliosis. [French] Comparaison de l'immunofluorescence indirecte et du Western Blot pour le serodiagnostic des infections neurologiques à Borrelia burgdorferi. Med Maladies Infect. 1992;22:722-6.

Nohlmans MKE, Blaauw AAM, Van den Bogaard AEJ, Van Boven CPA. Evaluation of nine serological tests for diagnosis of Lyme borreliosis. European Journal of Clinical Microbiology and Infectious Diseases. 1994;13:394-400.

Nordberg M, Forsberg P, Nyman D, Skogman BH, Nyberg C, Ernerudh J, et al. Can ELISPOT Be Applied to A Clinical Setting as A Diagnostic Utility for Neuroborreliosis? Cells. 2012;1(2):153-67.

Oksi J, Uksila J, Marjamaki M, Nikoskelainen J, Viljanen MK. Antibodies against whole sonicated Borrelia burgdorferi spirochetes, 41- kilodalton flagellin, and P39 protein in patients with PCR- or culture- proven late Lyme borreliosis. Journal of Clinical Microbiology. 1995;33:2260-4.

Olsson I, Von Stedingk LV, Hanson HS, Von Stedingk M, Asbrink E, Hovmark A. Comparison of four different serological methods for detection of antibodies to Borrelia burgdorferi in erythema migrans. Acta Dermato-Venereologica. 1991;71:127-33.

Panelius J, Lahdenne P, Saxen H, Heikkila T, Seppala I. Recombinant flagellin A proteins from *Borrelia burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii* in serodiagnosis of Lyme borreliosis. Journal of Clinical Microbiology. 2001;39:4013-9.

Popperl AF, Noll F. Diagnostic value of serological test-systems used in diagnosis of infection with Borrelia burgdorferi. [German] Diagnostische Wertigkeit serologischer Verfahren bei der Diagnostik einer Infektion mit Borrelia burgdorferi. LaboratoriumsMedizin. 2000;24:190-7.

Putzker M, Zoller L. Comparative evaluation of commercially available hemagglutination, enzymeimmuno and immunoblot assays in the serodiagnosis of Lyme borreliosis. [German] Vergleichende bewertung kommerzieller hamagglutinations-, enzymimmun- und immunblottests in der serodiagnostik der Lyane-Borreliose. Klinisches Labor. 1995;41:655-66.

Rauer S, Kayser M, Neubert U, Rasiah C, Vogt A. Establishment of enzyme-linked immunosorbent assay using purified recombinant 83-kilodalton antigen of Borrelia burgdorferi sensu stricto and Borrelia afzelii for serodiagnosis of lyme disease. Journal of Clinical Microbiology. 1995;33:2596-600.

Reiber H, Ressel CB, Spreer A. Diagnosis of neuroborreliosis - Improved knowledge base for qualified antibody analysis and cerebrospinal fluid data pattern related interpretations. . Neurol Psych Brain Res. 2013;19:159-69.

Rijpkema S, Groen J, Molkenboer M, Herbrink P, Osterhaus A, Schellekens J. Serum antibodies to the flagellum of Borrelia burgdorferi measured with an inhibition enzyme-linked immunosorbent assay are diagnostic for Lyme borreliosis. Serodiag Immunother Infect Dis. 1994;6:61-7.

Roux F, Boyer E, Jaulhac B, Dernis E, Closs-Prophette F, Puechal X. Lyme meningoradiculitis: Prospective evaluation of biological diagnosis methods. European Journal of Clinical Microbiology and Infectious Diseases. 2007;26:685-93.

Ruzic-Sabljic E, Maraspin V, Cimperman J, Lotric-Furian S, Strle F. Evaluation of immunofluorescence test (IFT) and immuno (western) blot (WB) test in patients with erythema migrans. Wiener Klinische Wochenschrift. 2002;114:586-90.

Ryffel K, Peter O, Binet L, Dayer E. Interpretation of immunoblots for Lyme borreliosis using a semiquantitative approach. Clinical Microbiology and Infection. 1998;4:205-12.

Schulte-Spechtel U, Lehnert G, Liegl G, Fingerle V, Heimerl C, Johnson B, et al. Significant improvement of the recombinant Borrelia IgG immunoblot for serodiagnosis of early neuroborreliosis. International Journal of Medical Microbiology, Supplement. 2004;293:158-60.

Skarpaas T, Ljostad U, Sobye M, Mygland A. Sensitivity and specificity of a commercial C6 peptide enzyme immuno assay in diagnosis of acute Lyme neuroborreliosis. European Journal of Clinical Microbiology and Infectious Diseases. 2007;26:675-7.

Skogman BH, Croner S, Forsberg P, Ernerudh J, Lahdenne P, Sillanpaa H, et al. Improved laboratory diagnostics of lyme neuroborreliosis in children by detection of antibodies to new antigens in cerebrospinal fluid. Pediatric Infectious Disease Journal. 2008;27:605-12.

Smismans A, Goossens VJ, Nulens E, Bruggeman CA. Comparison of five different immunoassays for the detection of Borrelia burgdorferi Igm and IgG antibodies. Clinical Microbiology and Infection. 2006;12:648-55.

Tjernberg I, Henningsson AJ, Eliasson I, Forsberg P, Ernerudh J. Diagnostic performance of cerebrospinal fluid chemokine CXCL13 and antibodies to the C6-peptide in Lyme neuroborreliosis. Journal of Infection. 2011;62:149-58.

Tjernberg I, Kruger G, Eliasson I. C6 peptide ELISA test in the serodiagnosis of Lyme borreliosis in Sweden. European Journal of Clinical Microbiology and Infectious Diseases. 2007;26:37-42.

van Burgel ND, Brandenburg A, Gerritsen HJ, Kroes ACM, van Dam AP. High sensitivity and specificity of the C6-peptide ELISA on cerebrospinal fluid in Lyme neuroborreliosis patients. Clinical Microbiology and Infection. 2011;17:1495-500.

Von Baehr V, Liebenthal C, Gaida B, Schmidt FP, Von Baehr R, Volk HD. Evaluation of the diagnostic significance of the lymphocyte proliferation test in patients with Lyme borreliosis. [German] Untersuchungen zur diagnostischen wertigkeit des lymphozytentransformationstestes bei patienten mit borreliose. LaboratoriumsMedizin. 2007;31:149-58.

Wilske B, Fingerle V, Herzer P, Hofmann A, Lehnert G, Peters H, et al. Recombinant immunoblot in the serodiagnosis of Lyme borreliosis. Comparison with indirect immunofluorescence and enzyme-linked immunosorbent assay. Medical Microbiology and Immunology. 1993;182:255-70.

Wilske B, Habermann C, Fingerle V, Hillenbrand B, Jauris-Heipke S, Lehnert G, et al. An improved recombinant IgG immunoblot for serodiagnosis of Lyme borreliosis. Medical Microbiology and Immunology. 1999;188:139-44.

Zoller L, Haude M, Sonntag HG. Validity of the standard assays in the serodiagnosis of Lyme borreliosis and effects of methodical modifications. [German] Validitat der standardverfahren in der serodiagnostik der Lyme-Borreliose und einfluss methodischer varianten. LaboratoriumsMedizin. 1990;14:404-11.

Annex 4. Overview of the quality assessment of the studies included in the analysis

		Risk of bias				Concerns regarding applicability		
Author year	Design	Patients	Index test	Reference standard	Timing and flow	Patients	Index test	Reference standard
Ang 2011	Case control	High	Unclear	Unclear	High	High	Low	Low
Ang 2012	Case control	High	Unclear	Unclear	High	High	Low	Low
Bergstrom 1991	Case control	High	High	High	High	High	High	Low
Branda 2013	Case control	High	Unclear	High	High	High	Low	Low
Cerar 2006	Case control	High	Unclear	Low	High	High	Low	Low
Cerar 2010	Case control	High	Unclear	Unclear	High	High	Low	Low
Christova 2003	Case control	High	Unclear	Unclear	High	High	Low	Low
Cinco 2006	Case control	High	Unclear	Unclear	High	High	Low	Low
Dessau 2010	Case control	High	Unclear	High	High	High	Low	Low
Dessau 2013	Case control	High	Unclear	High	High	High	Low	Low
Flisiak 1996	Case control	High	Unclear	Low	High	High	Low	Low
Flisiak 1998	Case control	High	Unclear	Low	High	High	Low	Low
Goettner 2005	Case control	High	Unclear	High	High	High	High	Low
Goossens 2000	Case control	High	Unclear	Unclear	High	High	Low	Low
Goossens 2001	Case control	High	Unclear	Unclear	High	High	Low	Low
Gueglio 1996	Case control	High	Unclear	Unclear	High	High	Low	Low
Hansen 1988	Case control	High	High	Unclear	High	High	High	Low
Hansen 1989	Case control	High	High	Low	High	High	High	Low
Hansen 1991	Case control	High	High	Low	High	High	Low	Low
Hernandez 2003	Case control	High	Unclear	Unclear	High	High	Low	Low
Hofmann 1990	Case control	High	Unclear	Unclear	High	High	Low	Low
Hofmann 1996	Case control	High	Unclear	Unclear	High	High	Low	Low
Hofstad 1987	Case control	High	High	Low	High	High	High	Low
Hunfeld 2002	Case control	High	Unclear	Unclear	High	High	Low	Low
Jovivic 2003	Case control	High	High/unclear*	Unclear	High	High	High	Low
Kaiser 1998	Case control	High	High	High	High	High	High	Low
Kaiser 1999inf	Case control	High	High	High	High	High	High	Low
Karlsson 1989eur	Case control	High	High/unclear*	Low	High	High	High	Low
Karlsson 1989siid	Case control	High	High	Low	High	High	High	Low
Lahdenne 2003	Case control	High	High	Unclear	High	High	Low	Low
Lakos 2005	Case control	High	Low	High	High	High	High/low*	Low
Lange 1991	Case control	High	Unclear	Unclear	High	High	High/low*	Low
Lencakova 2008	Case control	High	Low/unclear*	Unclear	High	High	High/low*	Low
Marangoni 2005jmm	Case control	High	Unclear	Low	High	High	Low	Low
Marangoni 2005new	Case control	High	Unclear	Low	High	High	Low*	Low
Marangoni 2008	Case control	High	Unclear	Unclear	High	High	Unclear**	Low
Mathiesen 1996	Case control	High	High/low*	High	High	High	High/low*	Low
Mathiesen 1998	Case control	High	High/unclear*	Low	High	High	High/low*	Low
Nicolini 1992	Case control	High	High	Unclear	High	High	High	Low
Nohlmans 1994	Case control	High	High/unclear*	Unclear	High	High	High/low*	Low
Oksi 1995	Case control	High	High/unclear*	Unclear	High	High	High/low*	Low
Olsson 1991	Case control	High	High	Unclear	High	High	High	Low
Panelius 2001	Case control	High	High	High	High	High	High	Low
Putzker 1995	Case control	High	Unclear	High	High	High	Low	Low

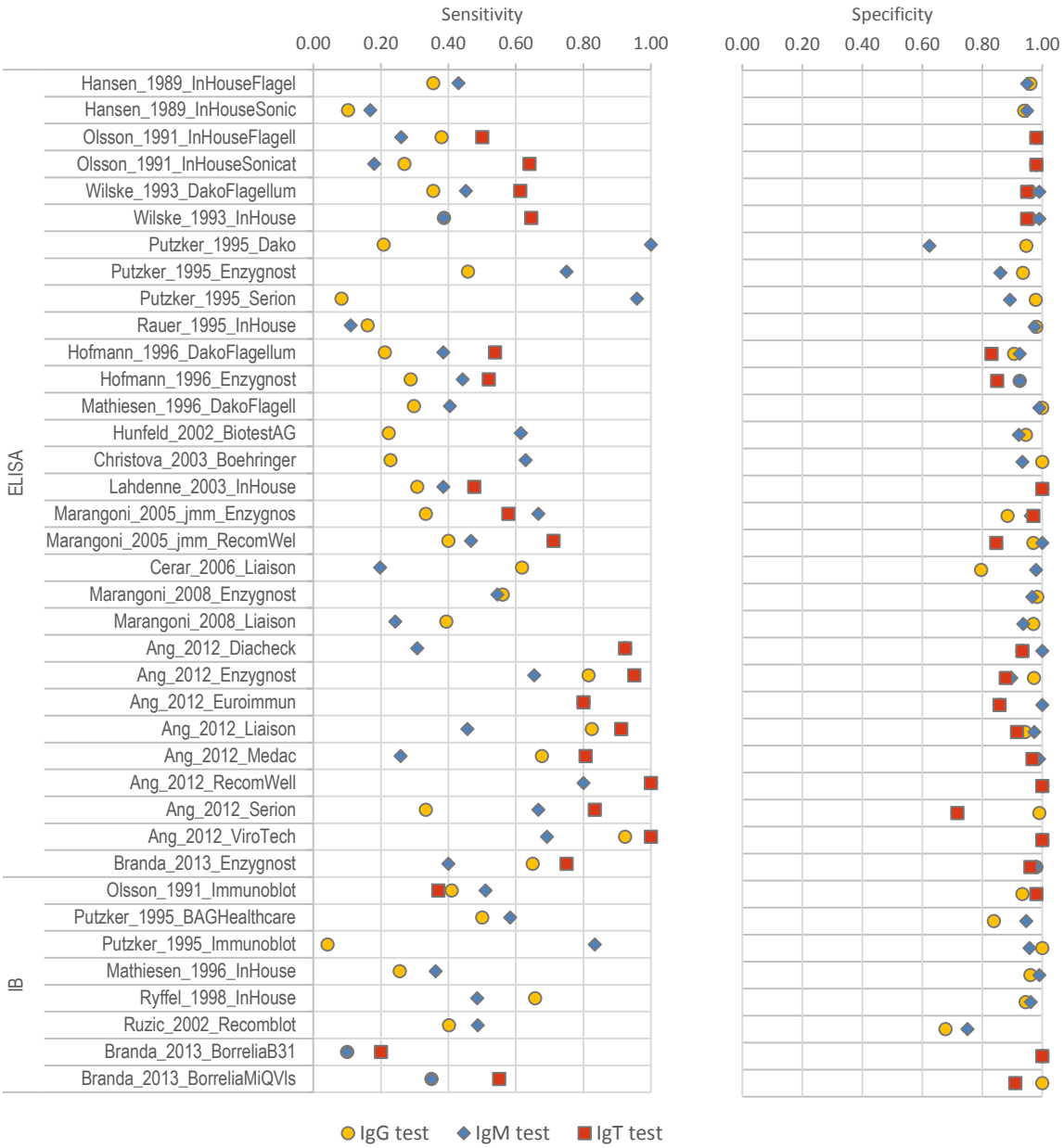
		Risk of bias				Concerns regarding applicability		
Author year	Design	Patients	Index test	Reference standard	Timing and flow	Patients	Index test	Reference standard
Rauer 1995	Case control	High	High	High	High	High	High	Low
Reiber 2013	Case control	High	High	Low	High	High	Unclear	Low
Rijpkema 1994	Case control	High	Unclear	Unclear	High	High	High/low*	Low
Ruzic 2002	Case control	High	Unclear	Unclear	High	High	Low	Low
Ryffel 1998	Case control	High	Unclear	High	High	High	High	Low
Schulte 2004	Case control	High	Unclear	High	High	High	High	Low
Smismans 2006	Case control	High	Unclear	High	High	High	Low	Low
Tjernberg 2007	Case control	High	Unclear	High	High	High	Low/unclear*	Low
Tjernberg 2011	Case control	High	Unclear	High	High	High	Low	Low
VanBurgel 2011	Case control	High	Unclear	High	High	High	Low	Low
VonBaehr 2007	Case control	High	Unclear	High	High	High	High	Low
Wilske 1993	Case control	High	High/unclear*	High	High	High	High/low*	Low
Wilske 1999	Case control	High	Unclear	High	High	High	High	Low
Zoller 1990	Case control	High	Unclear	Unclear	High	High	High	Low
Albisetti 1997	Cross sectional	Low	Unclear	Low	Unclear	Low	Low	Low
Barrial 2011	Cross sectional	Unclear	Low	Low	Low	Unclear	Low	Low
Bazovska 2001	Cross sectional	Unclear	Unclear	Unclear	Low	Unclear	Low	Low
Bednarova 2006	Cross sectional	Unclear	Unclear	Unclear	Unclear	Unclear	Low	Low
Bennet 2008	Cross sectional	Unclear	Unclear	Unclear	Unclear	High	Low	Low
Blaauw 1993	Cross sectional	Low	High	Low	Low	Low	High	Low
Blaauw 1999	Cross sectional	Low	Unclear	Low	Unclear	High	High/low*	Low
Blanc 2007	Cross sectional	Unclear	Unclear	High	Unclear	Unclear	Low	Low
Cermakova 2005	Cross sectional	Low	Unclear	High	Low	Low	Low	Low
Davidson 1999	Cross sectional	Unclear	Unclear	High	Low	Unclear	High/low*	High
Ekerfelt 2004	Cross sectional	Unclear	Low	Unclear	Unclear	Unclear	Low	Low
Huppertz 1996	Cross sectional	Unclear	Unclear	Unclear	Unclear	High	High	Low
Jansson 2005	Cross sectional	Unclear	Unclear	High	Low	Unclear	Low	High
Kolmel 1992	Cross sectional	Unclear	Unclear	High	Unclear	Unclear	Low	Low
Ljostad 2005	Cross sectional	High	Unclear	Low	Low	High	Unclear	Low
Nordberg 2012	Cross sectional	Unclear	High	Low	Unclear	Unclear	High	Low
Popperl 2000	Cross sectional	Unclear	Unclear	High	Low	Unclear	Low	High
Roux 2007	Cross sectional	High	Unclear	Low	Unclear	Unclear	High/low*	High
Skarpaas 2007	Cross sectional	Low	High/unclear*	Unclear	Unclear	Low	Low	Low
Skogman 2008	Cross sectional	Unclear	High	Low	Low	Unclear	High	Low

** Study evaluated more than one test in different ways resulting in different scores, e.g. cut-off values were sometimes pre-specified, sometimes based on results.*

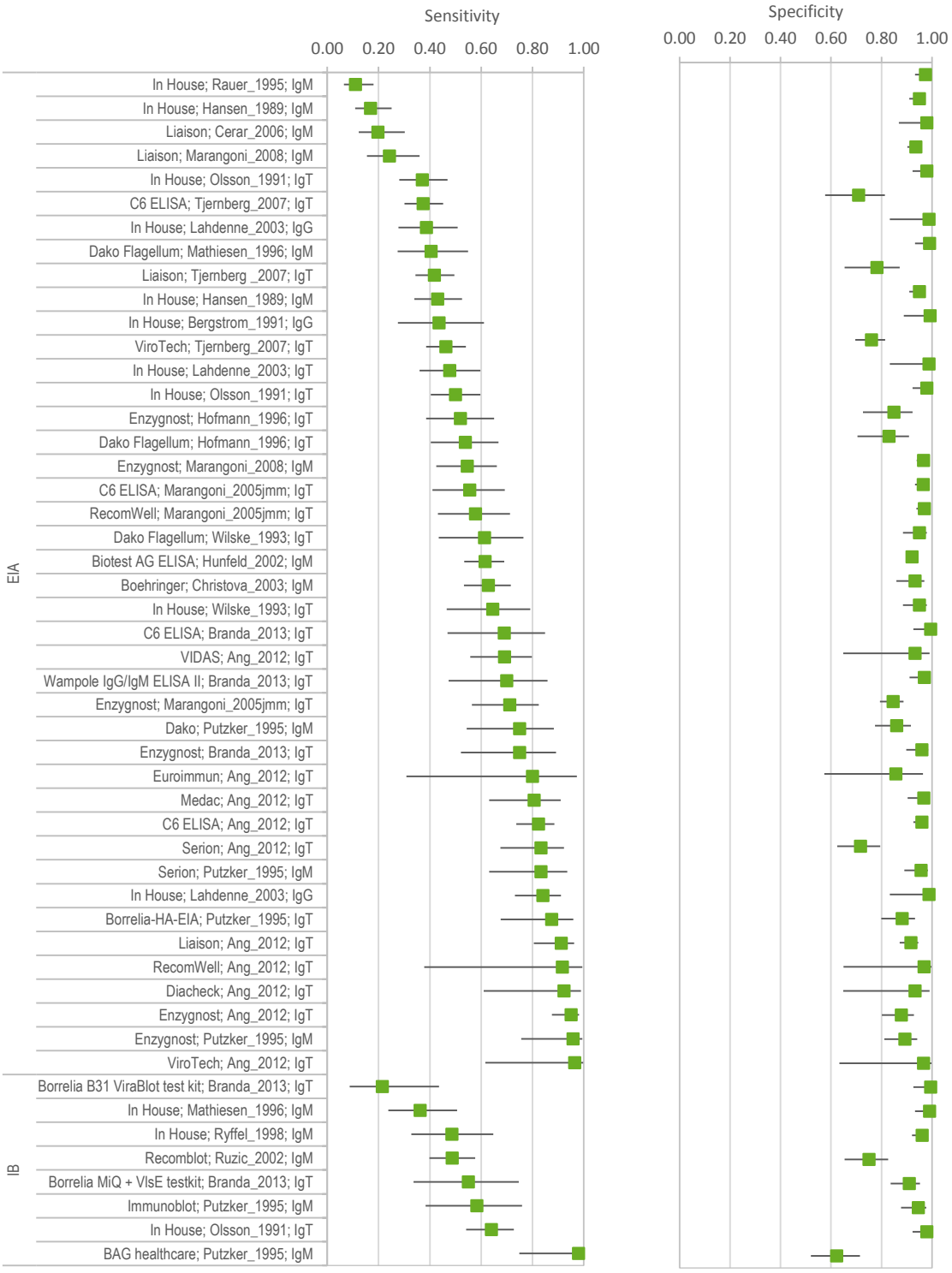
*** All tests were rated 'unclear' for concerns regarding the applicability of the index test, except for one ('high concern').*

Annex 5. Erythema migrans: case-control studies with healthy controls

Sensitivity and specificity of IgM, IgG and IgT tests for EM case-control studies with healthy controls. Of the 38 tests, 30 are ELISAs and 8 are IB. Studies are sorted according to year of publication.

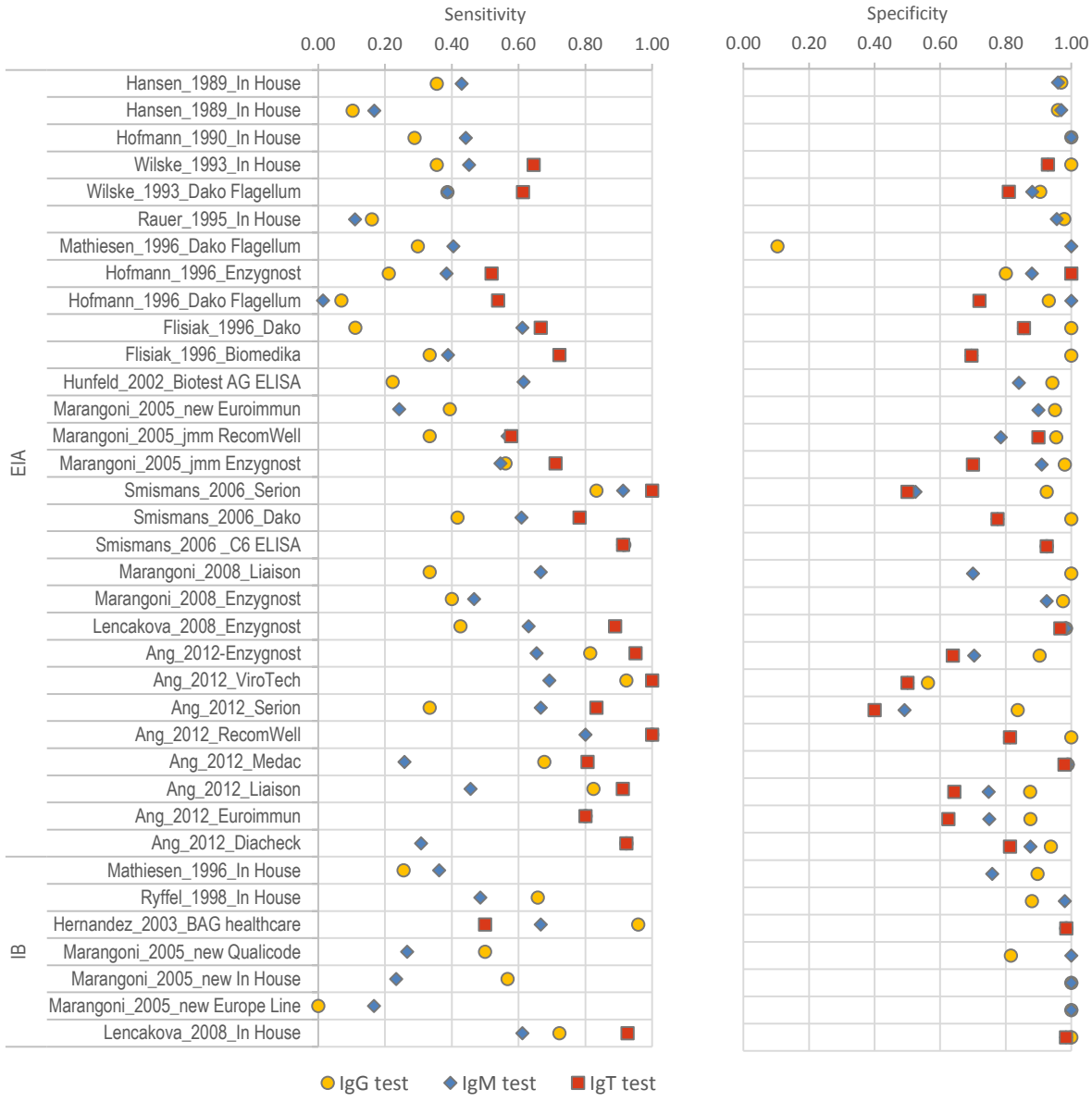


Sensitivity and specificity of EM case-control studies with healthy controls

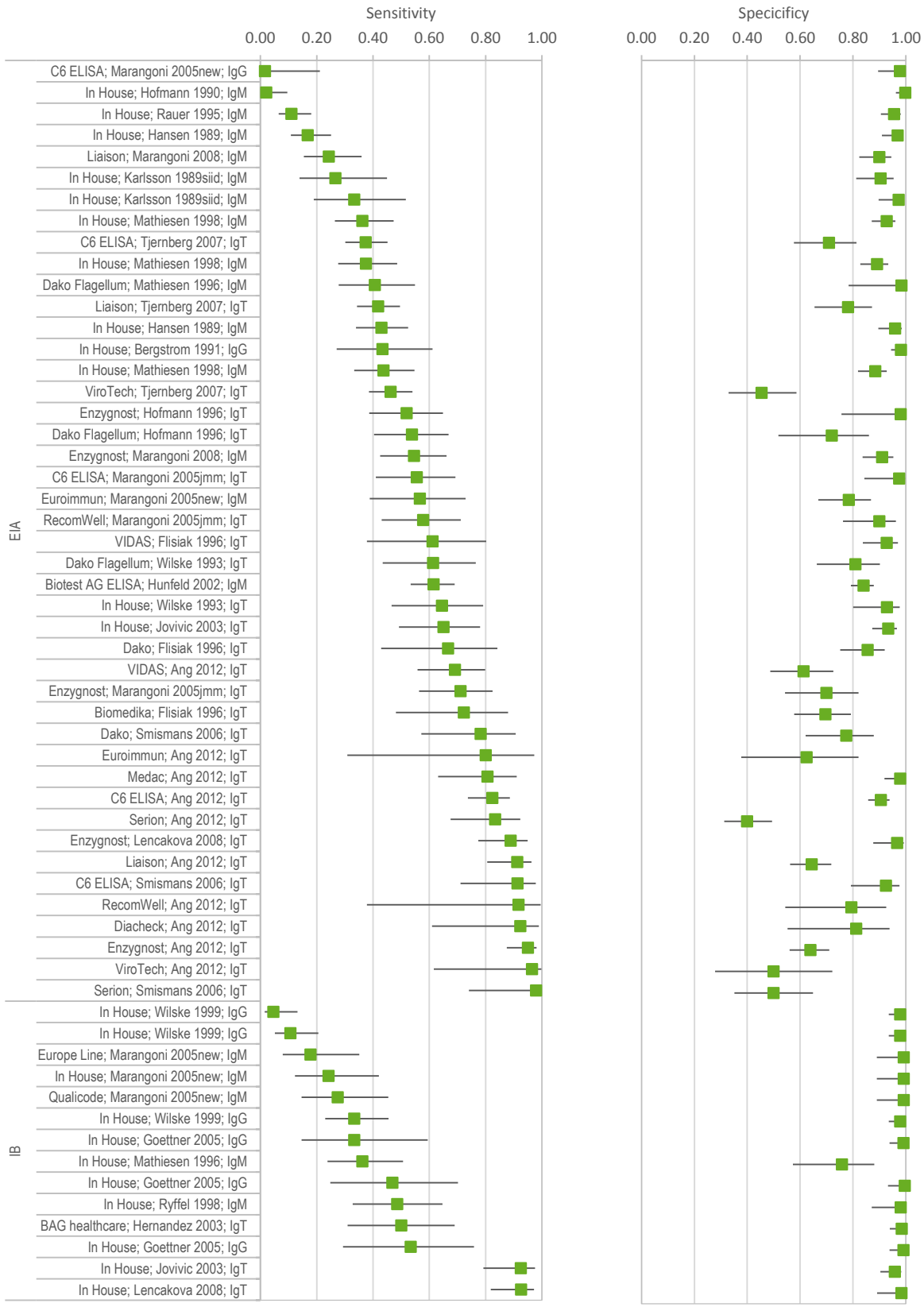


Annex 6. Erythema migrans: case-control studies with cross-reacting controls

Sensitivity and specificity of IgM, IgG and IgT tests for EM case-control studies with cross-reacting controls. Of the 36 tests, 29 are EIA and 7 are IB. Studies are sorted according to year of publication.

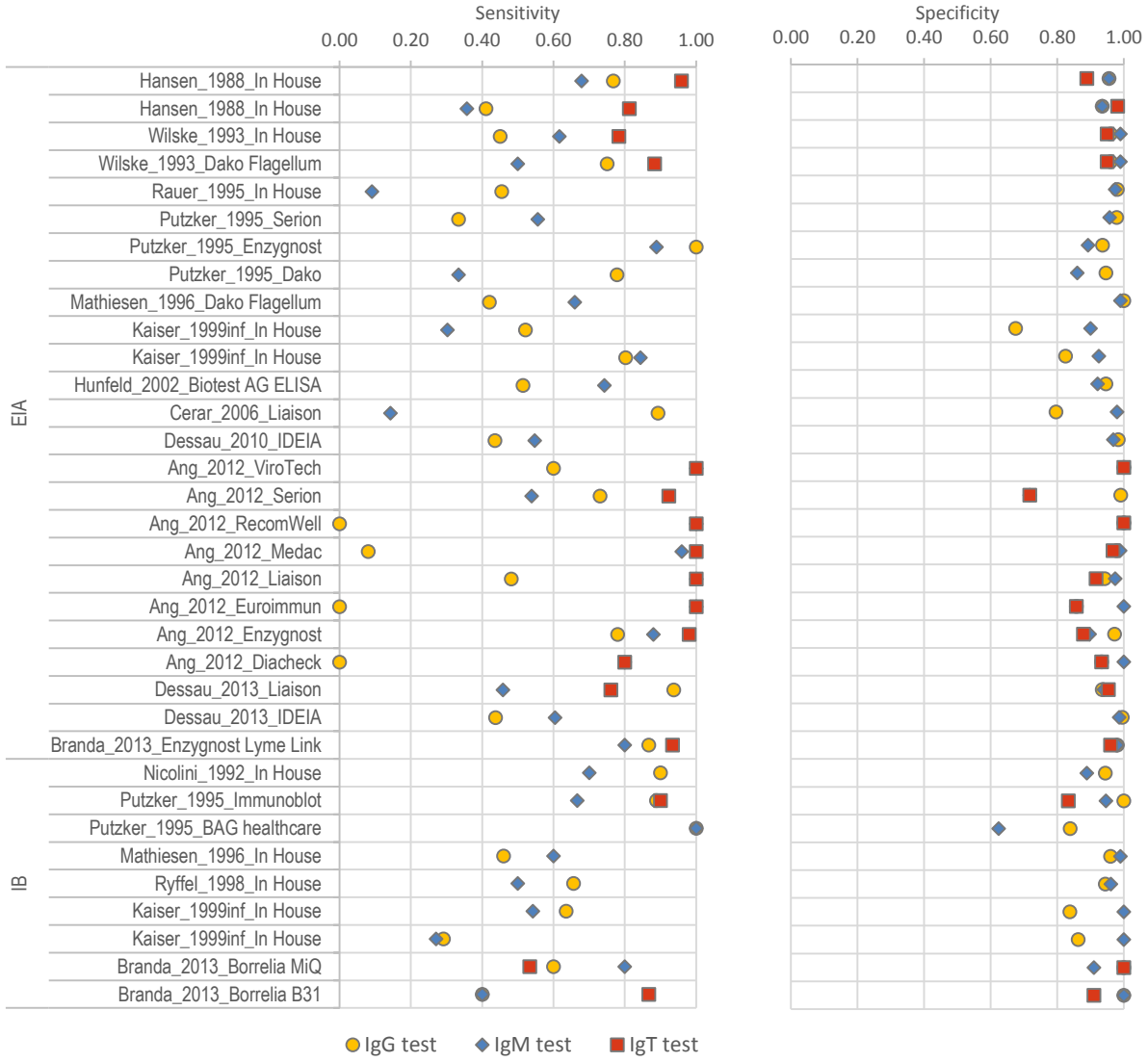


Sensitivity and specificity of EM case-control studies with cross-reacting controls

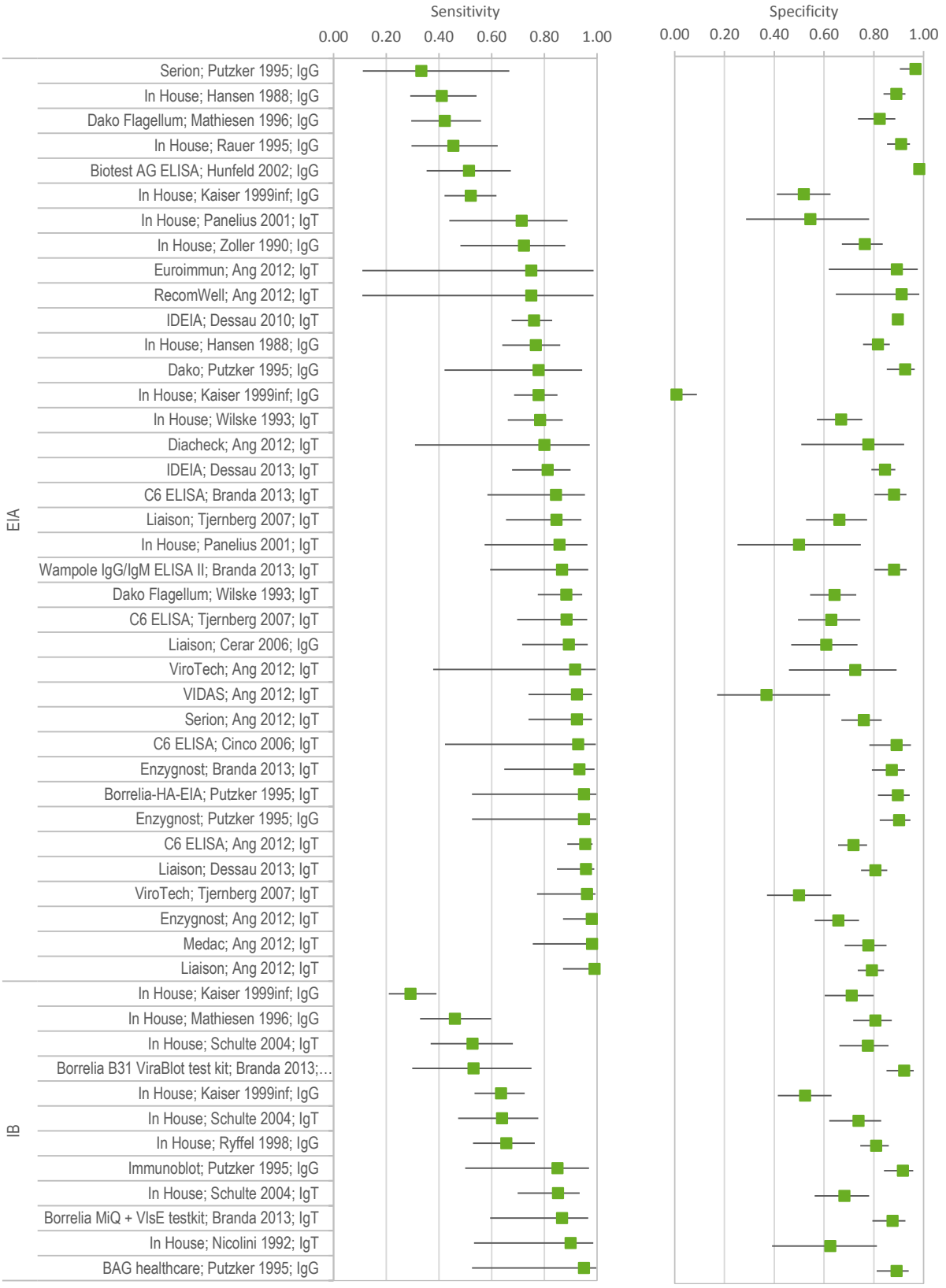


Annex 7. Neuroborreliosis: case-control studies with healthy controls

Sensitivity and specificity of IgM, IgG and IgT tests for NB case-control studies with healthy controls. Of the 34 tests, 25 are EIA and 9 are IB. Studies are sorted according to year of publication.

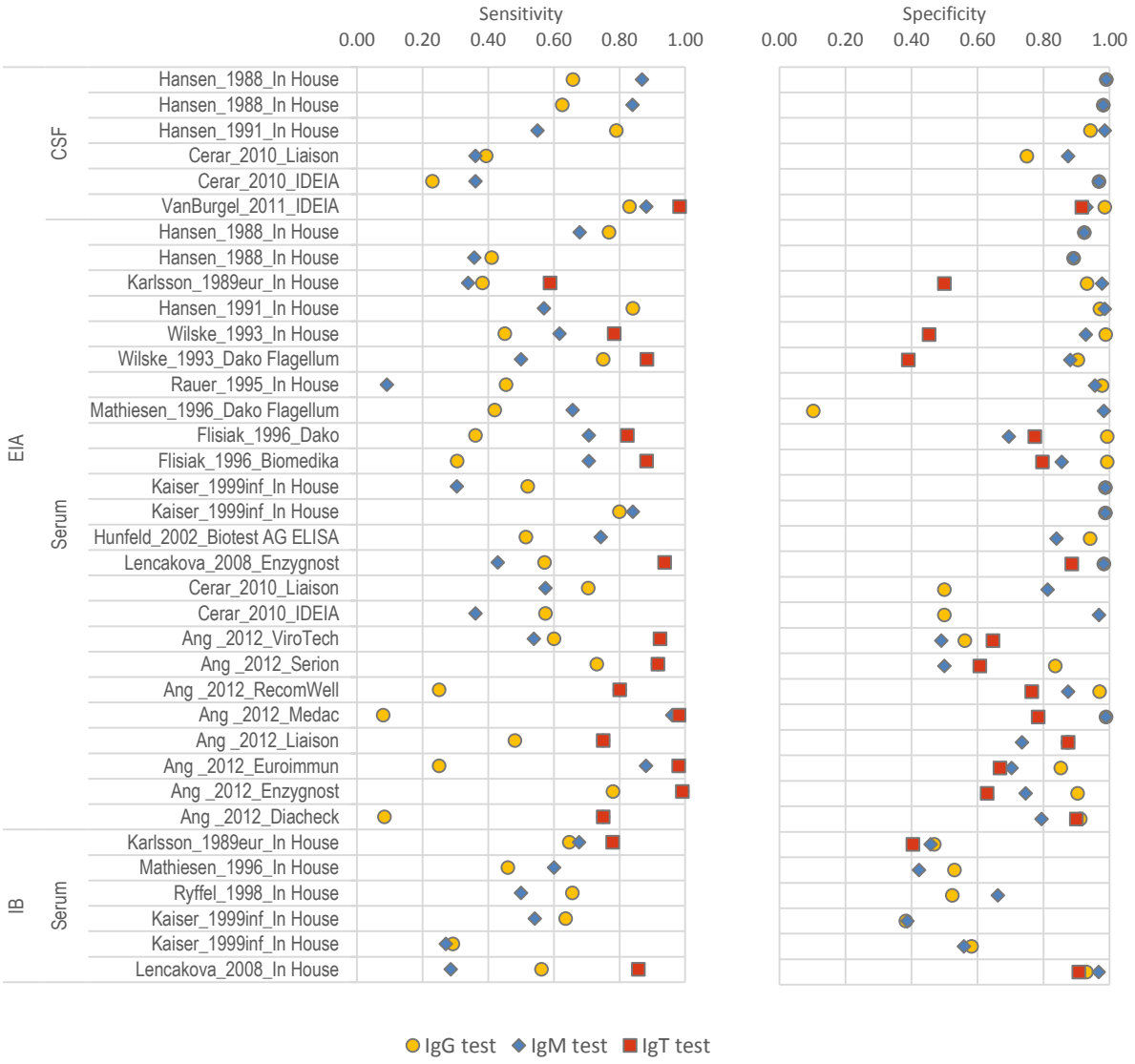


Sensitivity and specificity for NB case control studies with healthy controls

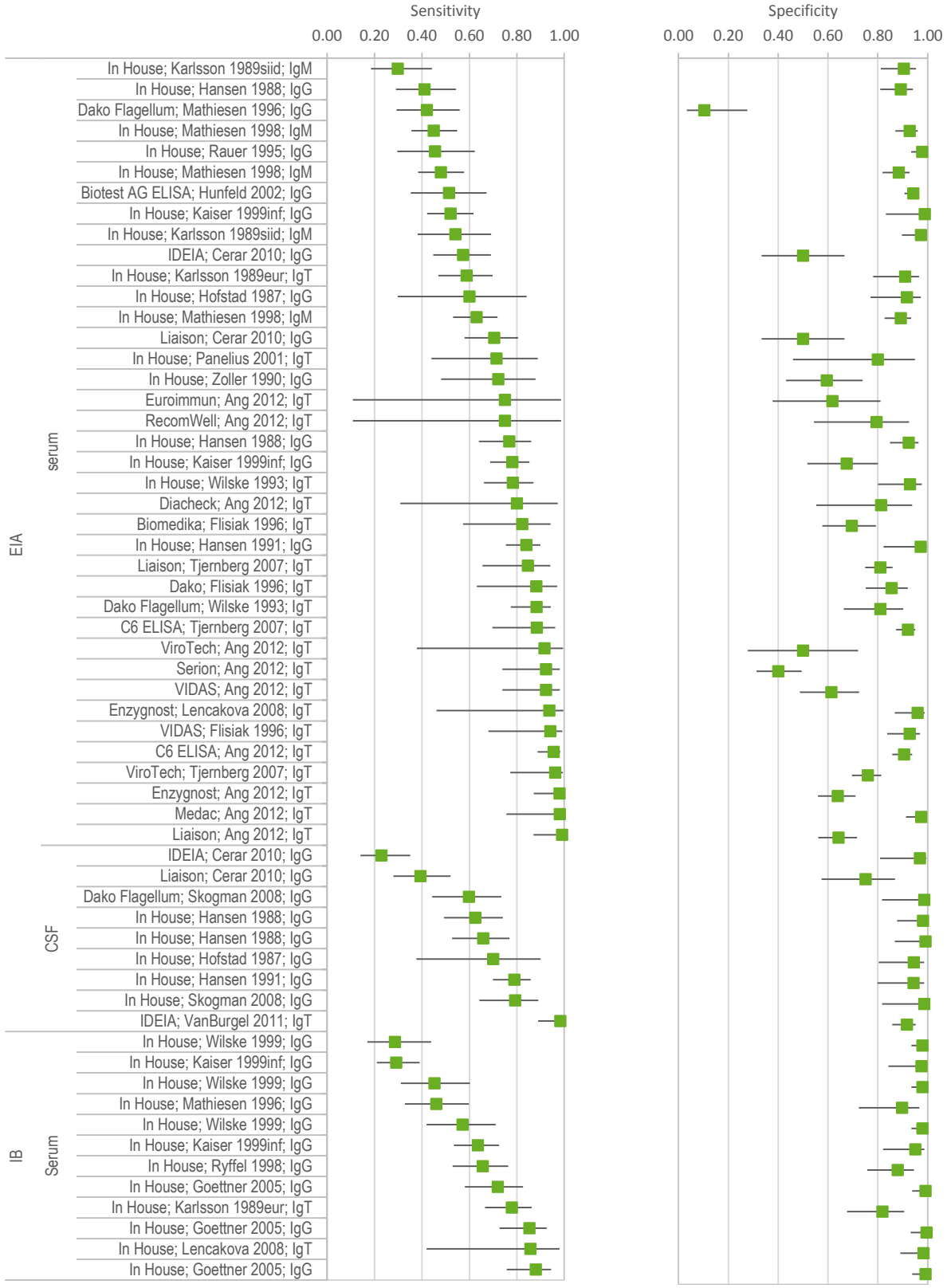


Annex 8. Neuroborreliosis: case-control studies with cross-reacting controls

Sensitivity and specificity of IgM, IgG and IgT tests for NB case-control studies with cross-reacting controls. Of the 36 tests, 30 are EIA and 6 are IB. Studies are sorted according to year of publication.

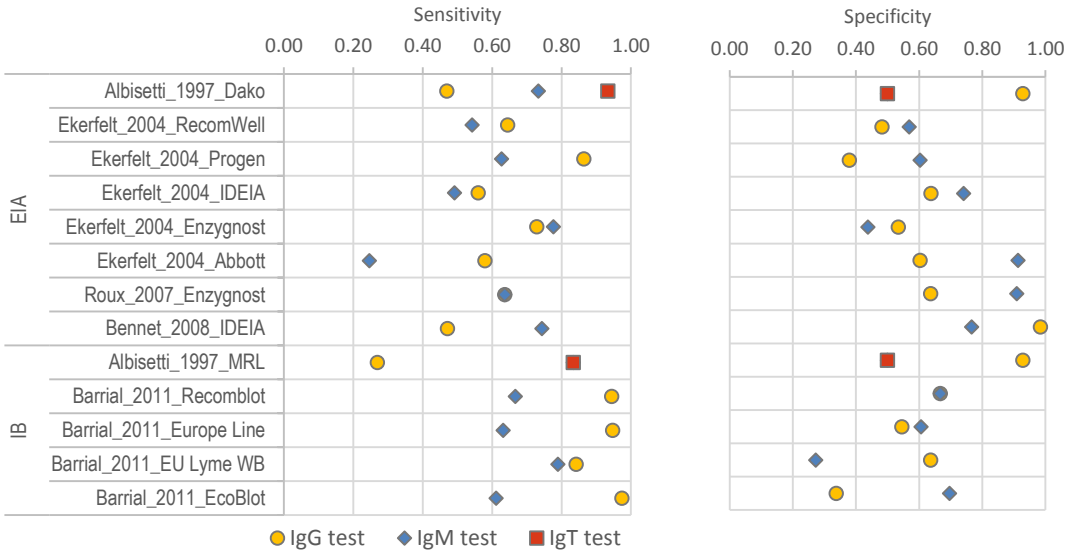


Sensitivity and specificity for NB case-control studies with cross-reacting controls

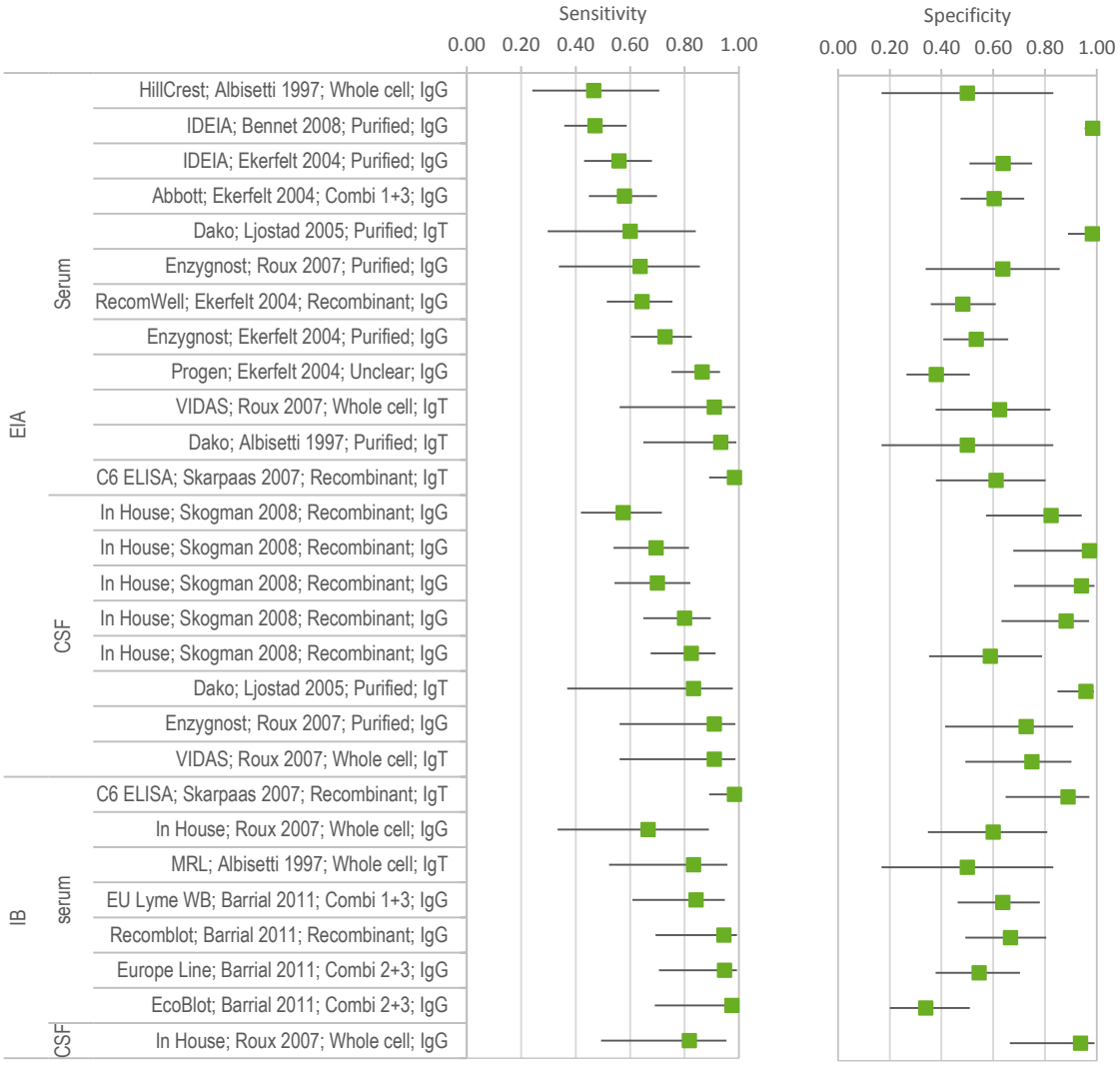


Annex 9. Neuroborreliosis: cross-sectional studies

Sensitivity and specificity of IgM, IgG and IgT tests for NB cross-sectional studies. Of the 13 tests, 8 are EIA and 5 are IB. Studies are sorted according to year of publication.

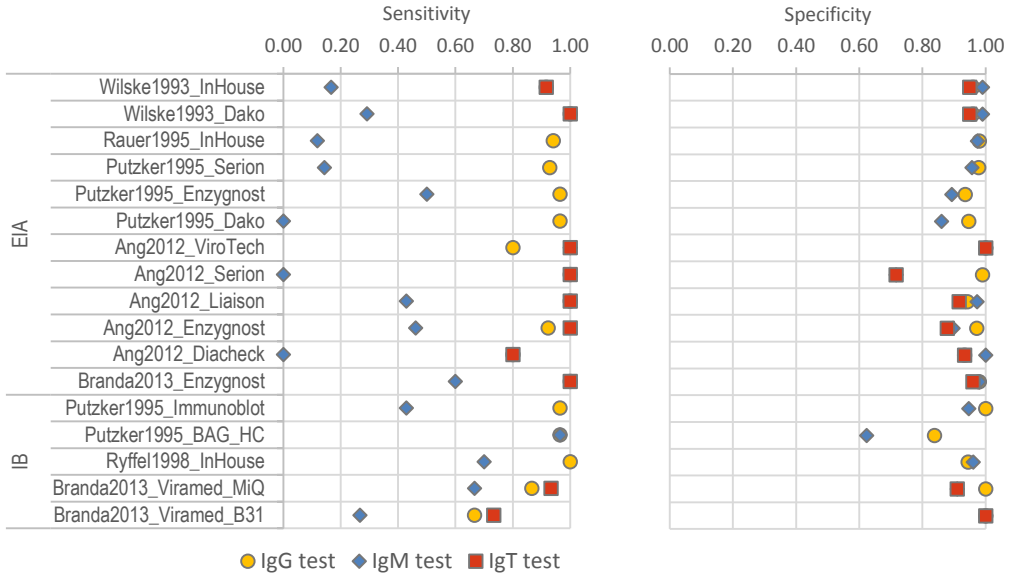


Sensitivity and specificity of NB cross-sectional studies



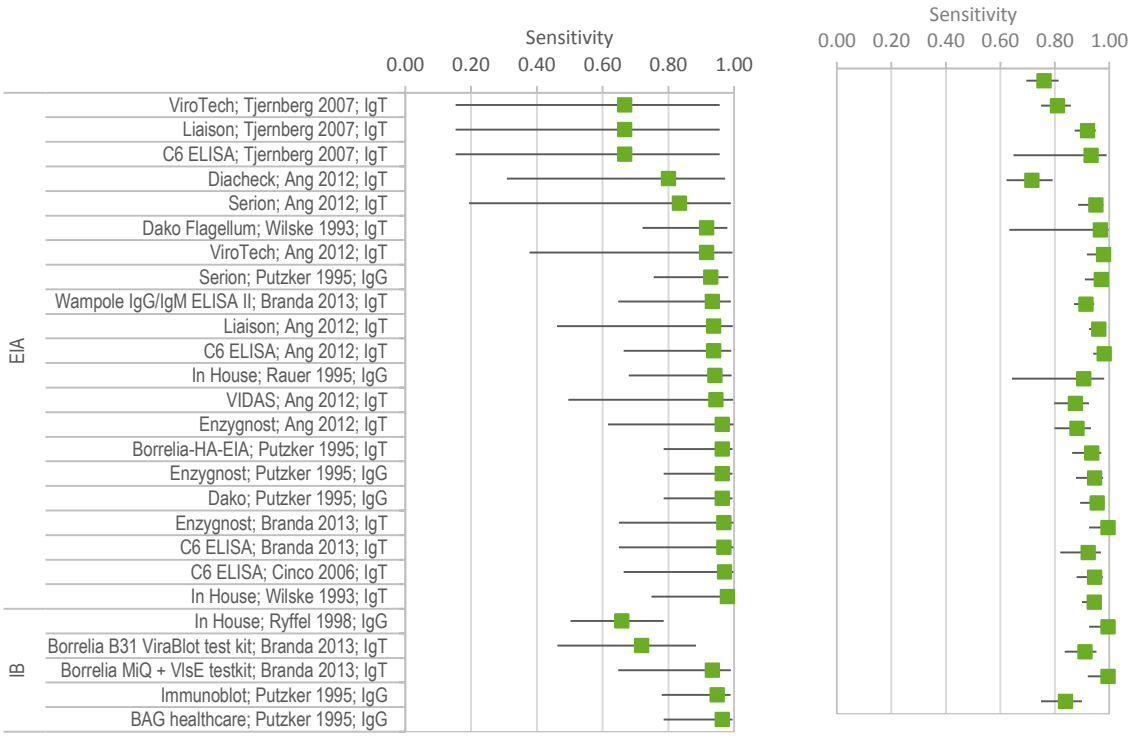
Annex 10. Lyme arthritis: case-control studies with healthy controls

Sensitivity and specificity of IgM, IgG and IgT tests for LA case-control studies with healthy controls. Of the 17 tests, 12 are EIA and 5 are IB. Studies are sorted according to year of publication.



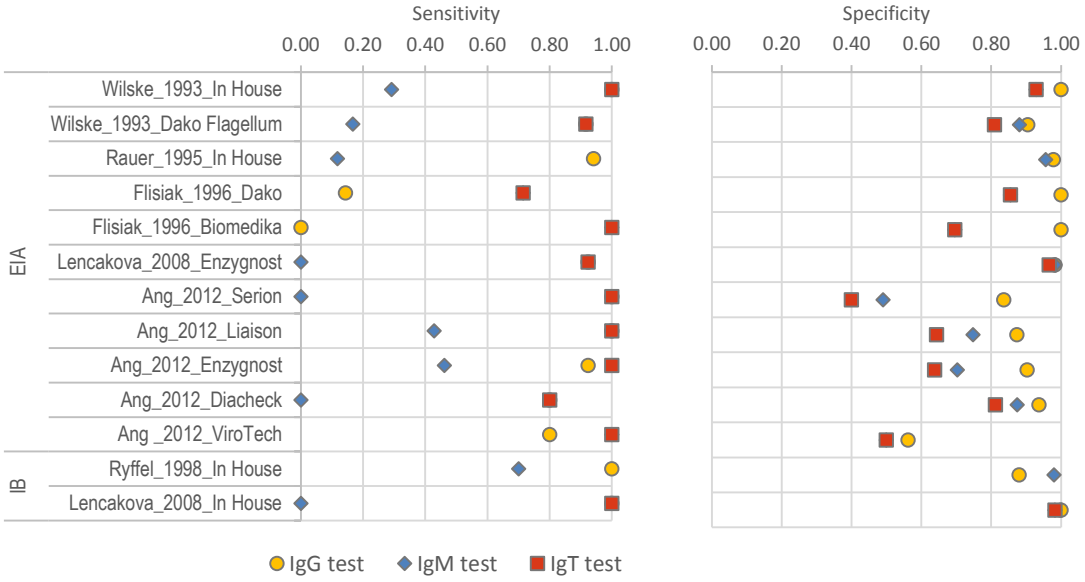
Note: It should be noted that the Viramed B31 assay in Branda et al. (2013) is an assay used in the USA and designed to detect infections in the USA. The aim of their study was to assess whether such assays were useful to detect patients who had acquired the infection in Europe.

Sensitivity and specificity in LA case-control studies with healthy controls

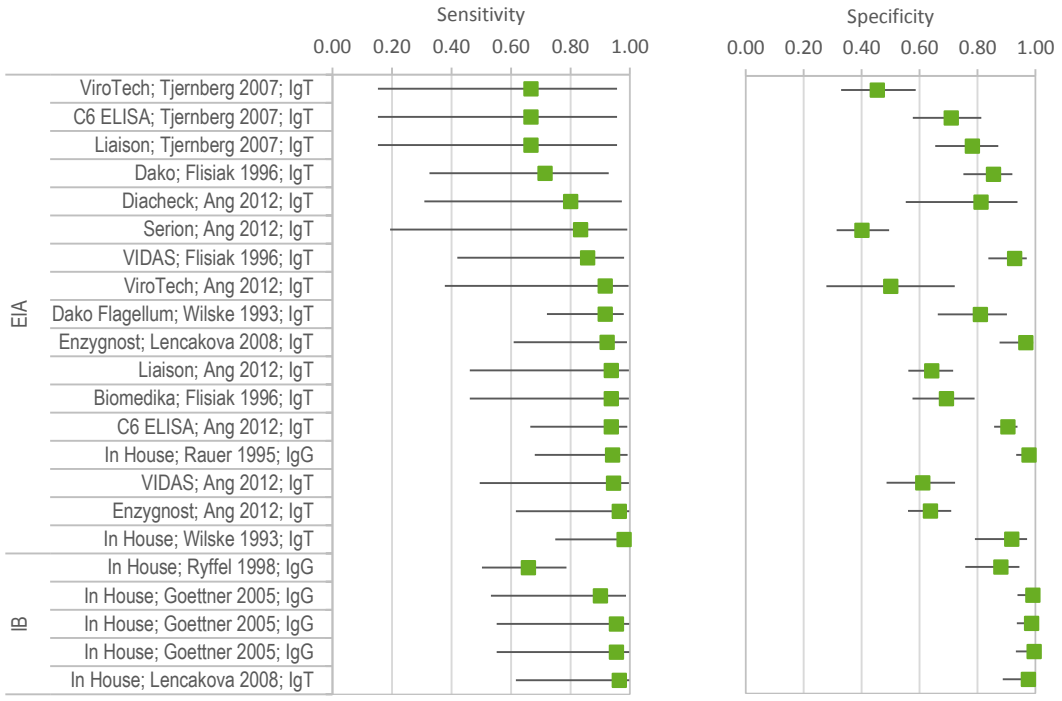


Annex 11. Lyme arthritis: case-control with cross-reacting controls

Sensitivity and specificity of IgM, IgG and IgT tests for LA case-control studies with cross-reacting. Of the 13 tests, 11 are EIA and 2 are IB. Studies are sorted according to year of publication.

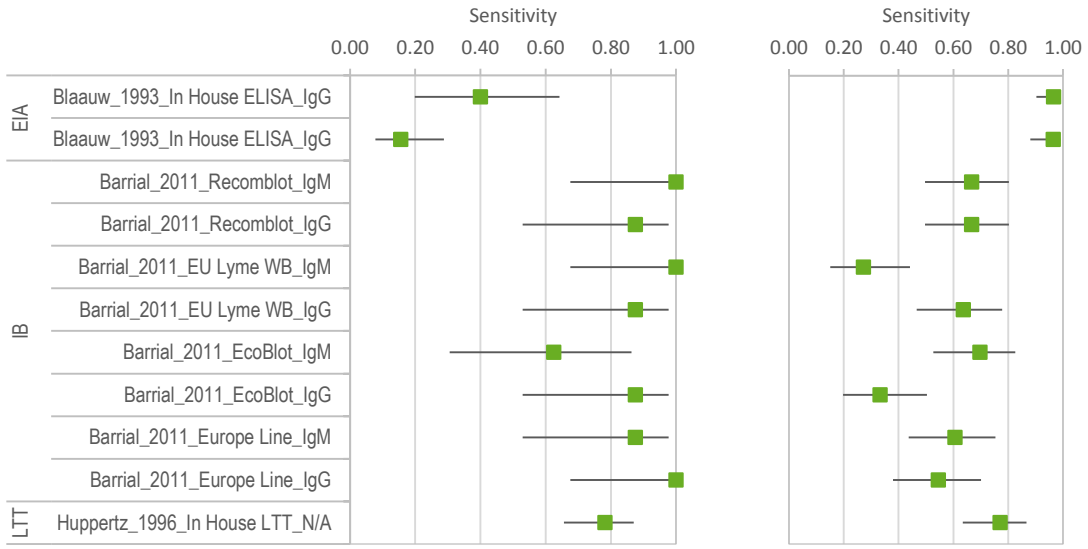


Sensitivity and specificity of LA case-control studies with cross-reacting controls



Annex 12. Lyme arthritis: cross-sectional design studies

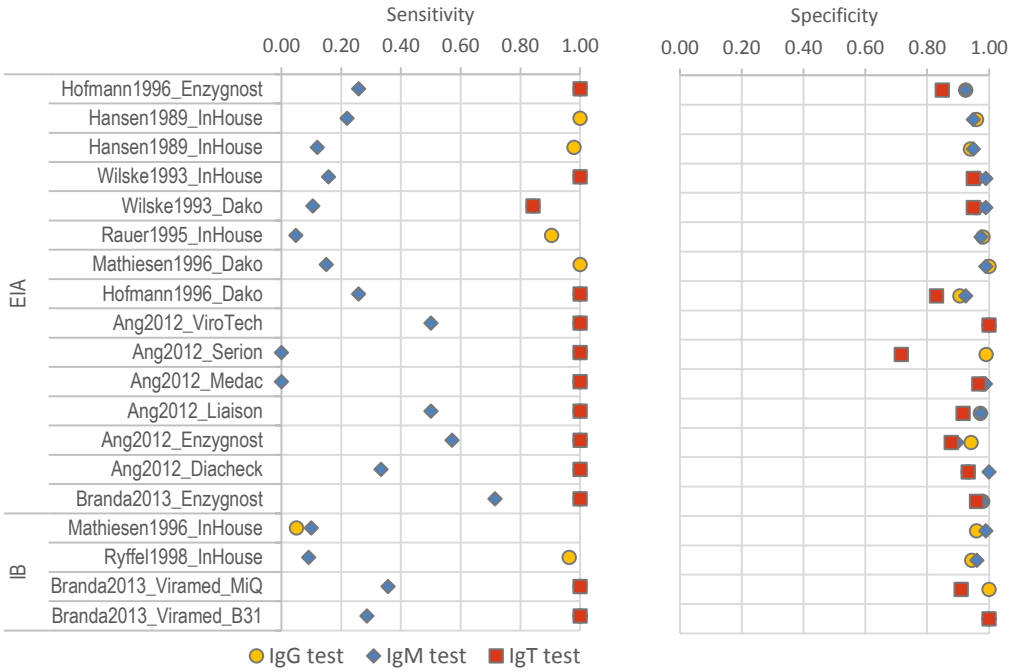
Sensitivity and specificity of LA cross-sectional studies



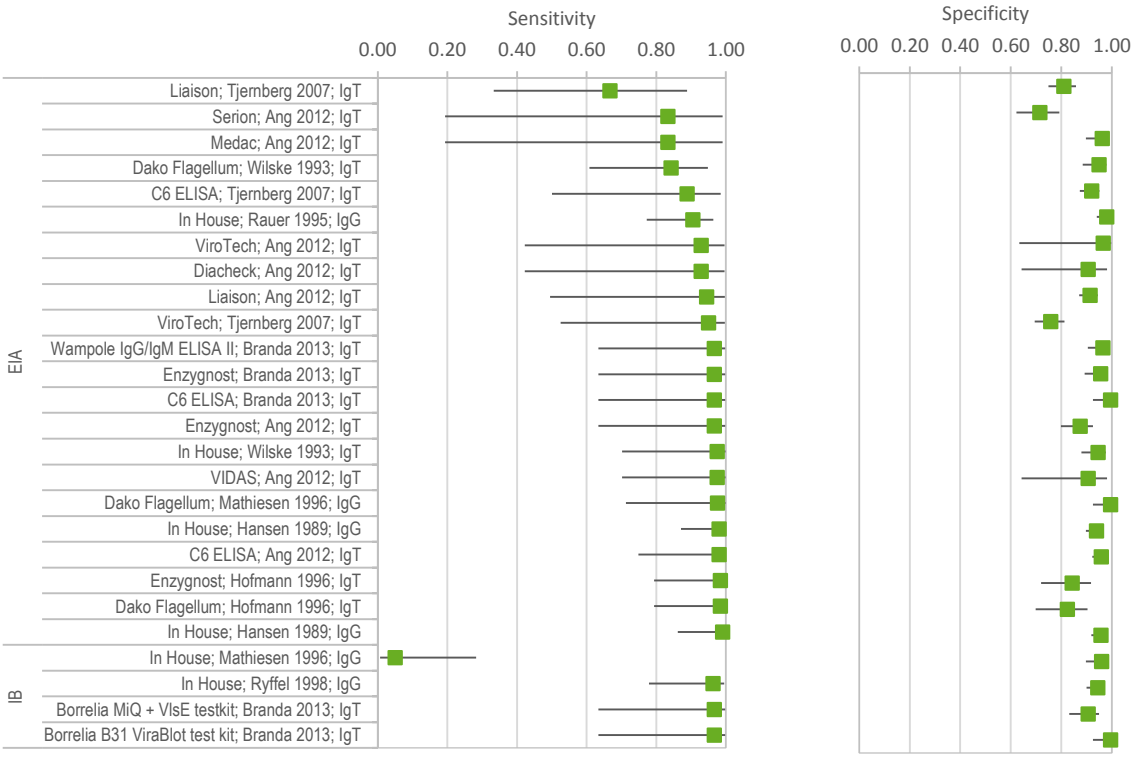
Note: For the tests evaluated by Barrial et al. (2011) , the results are presented with borderline results considered positive test results. When the borderline test results were considered negative, sensitivity was lower (63%–100%). The two results for Blaauw et al. (1993) are for possible Lyme patients considered as cases (prevalence 40%, sensitivity 16%). When 'possibles' are considered to be controls, results are: prevalence 15%, sensitivity 40%.

Annex 13. Acrodermatitis chronica atrophicans: case-control studies with healthy controls

Sensitivity and specificity of IgM, IgG and IgT tests for ACA case-control studies with healthy controls. Of the 19 tests, 15 are EIA and 4 are IB. Studies are sorted according to year of publication.

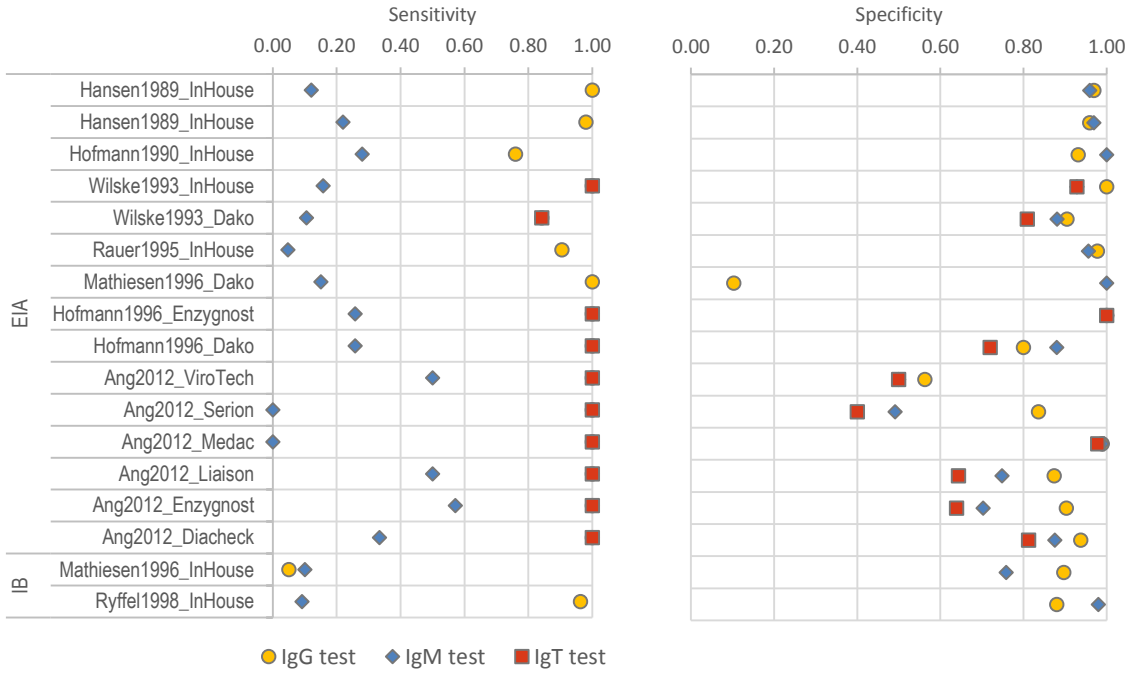


Sensitivity and specificity of ACA case-control studies with healthy controls

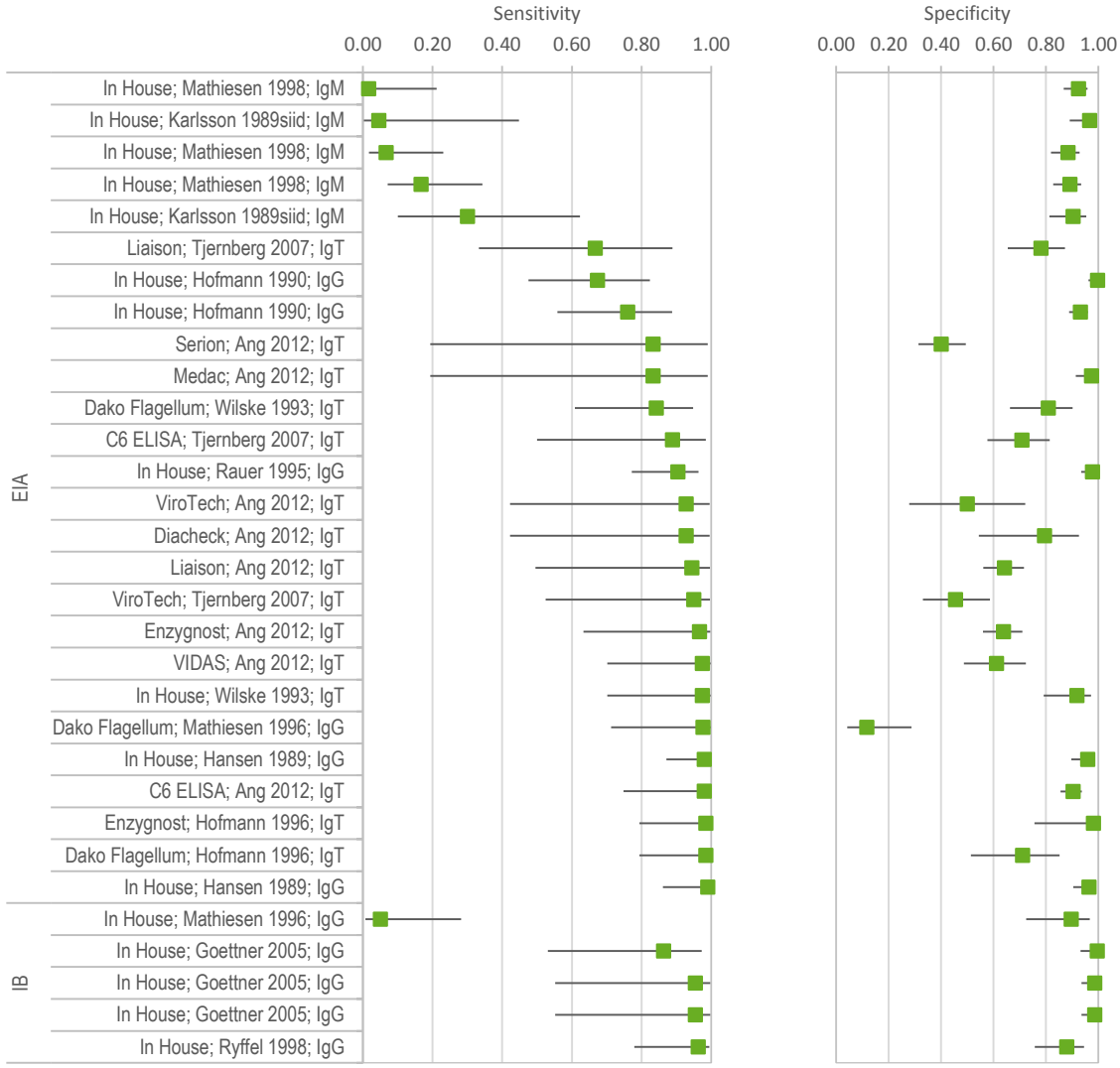


Annex 14. Acrodermatitis chronica atrophicans: case-control studies with cross-reacting controls

Sensitivity and specificity of IgM, IgG and IgT tests for ACA case-control studies with cross-reacting controls. Of the 17 tests, 15 are EIA and 2 are IB. Studies are sorted according to year of publication.

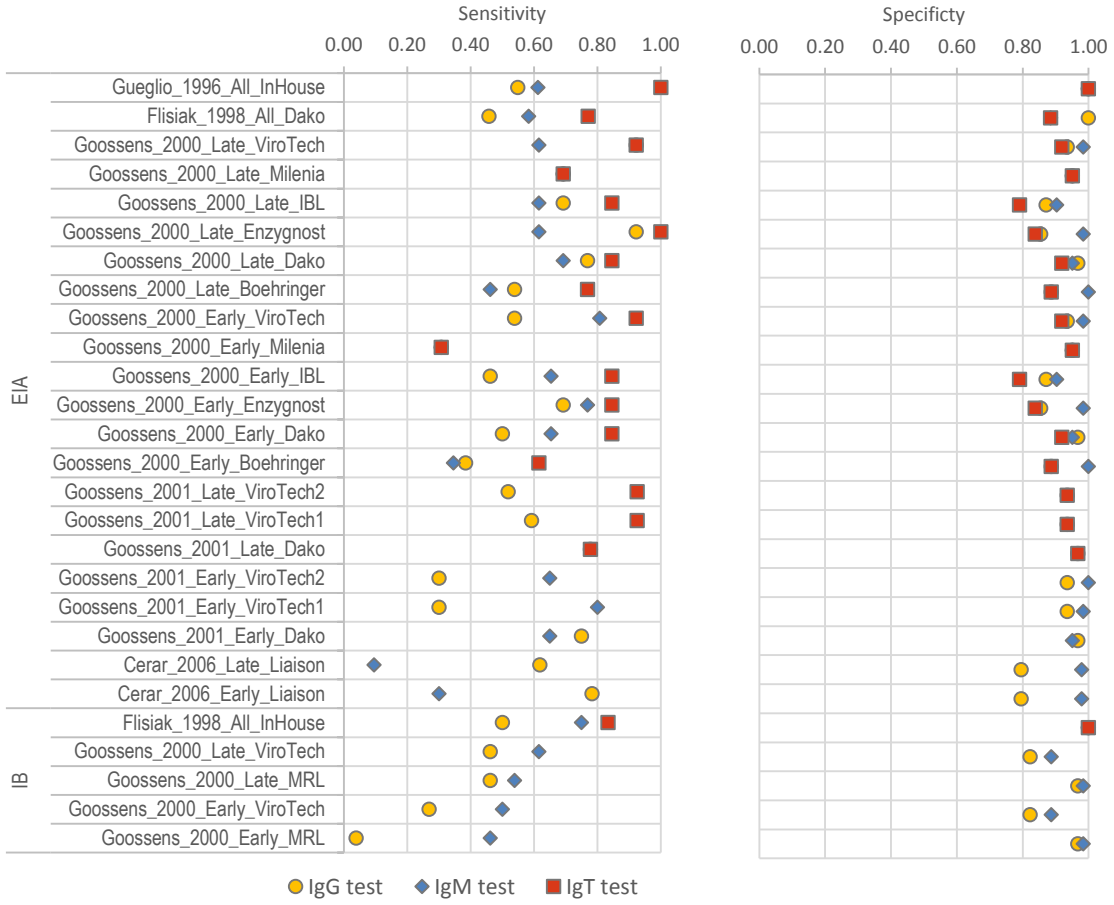


Sensitivity and specificity of ACA case-control studies with cross-reacting controls

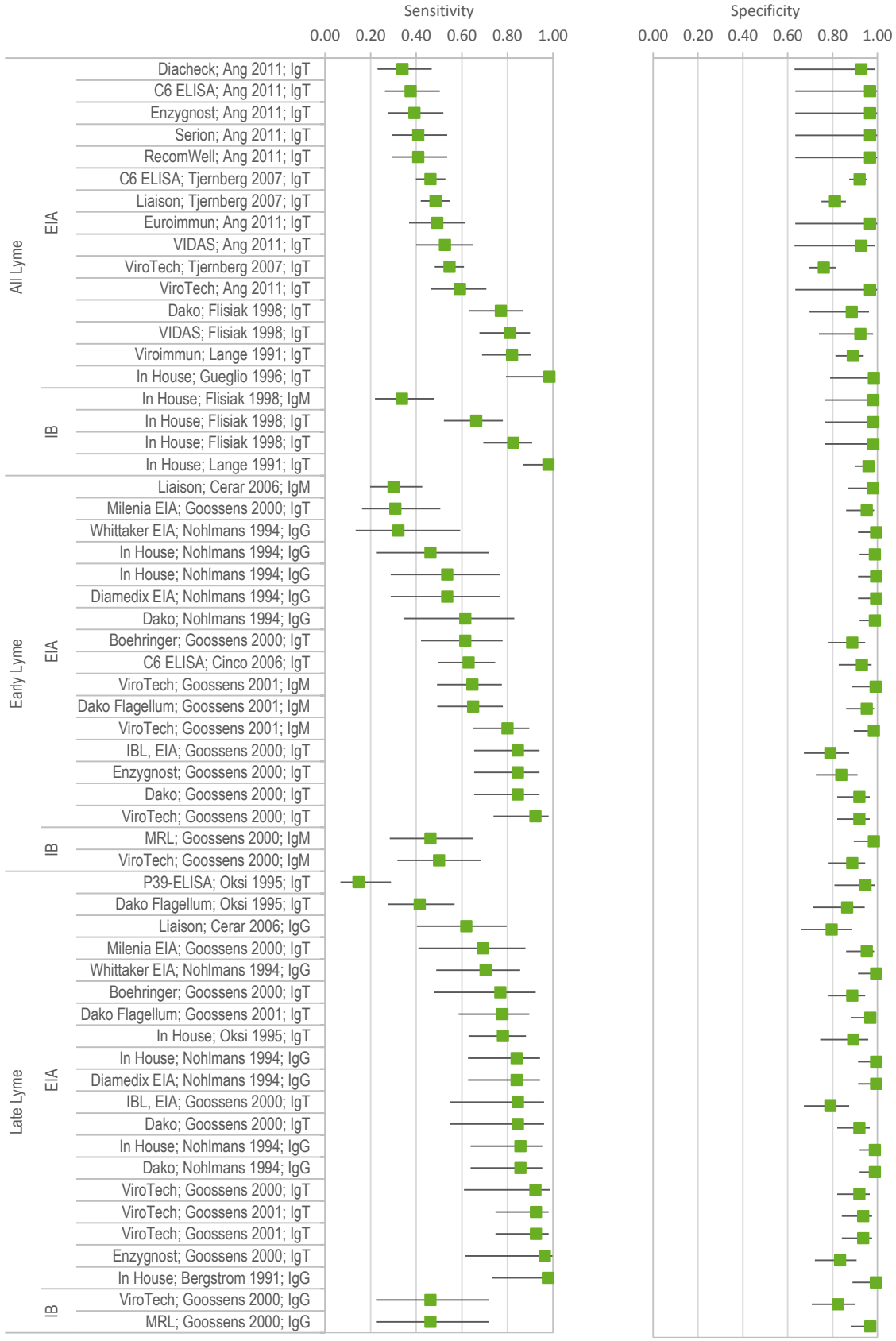


Annex 15. Lyme borreliosis-unspecified: case-control studies with healthy controls

LB-unspecified: sensitivity and specificity of IgM, IgG and IgT tests for case-control studies with healthy controls

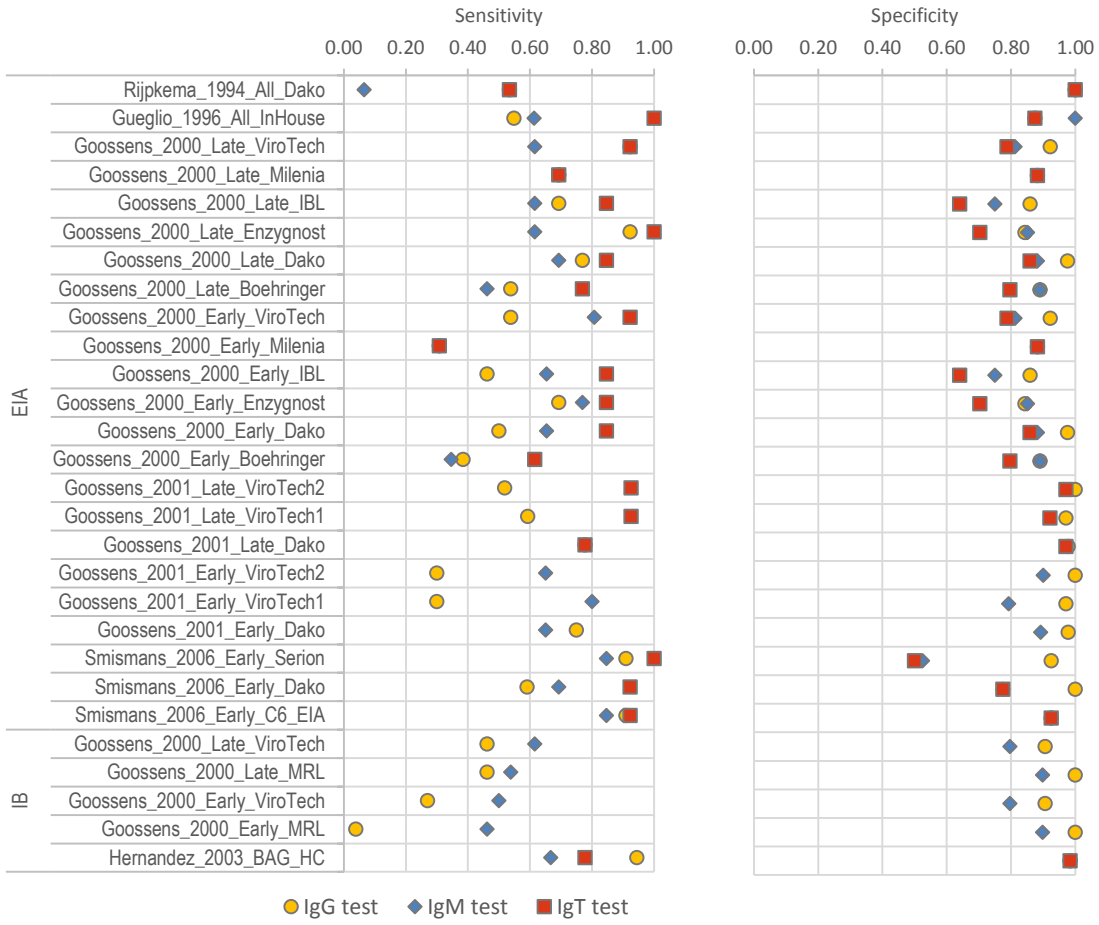


Sensitivity and specificity for LB-unspecified case-control studies with healthy controls.

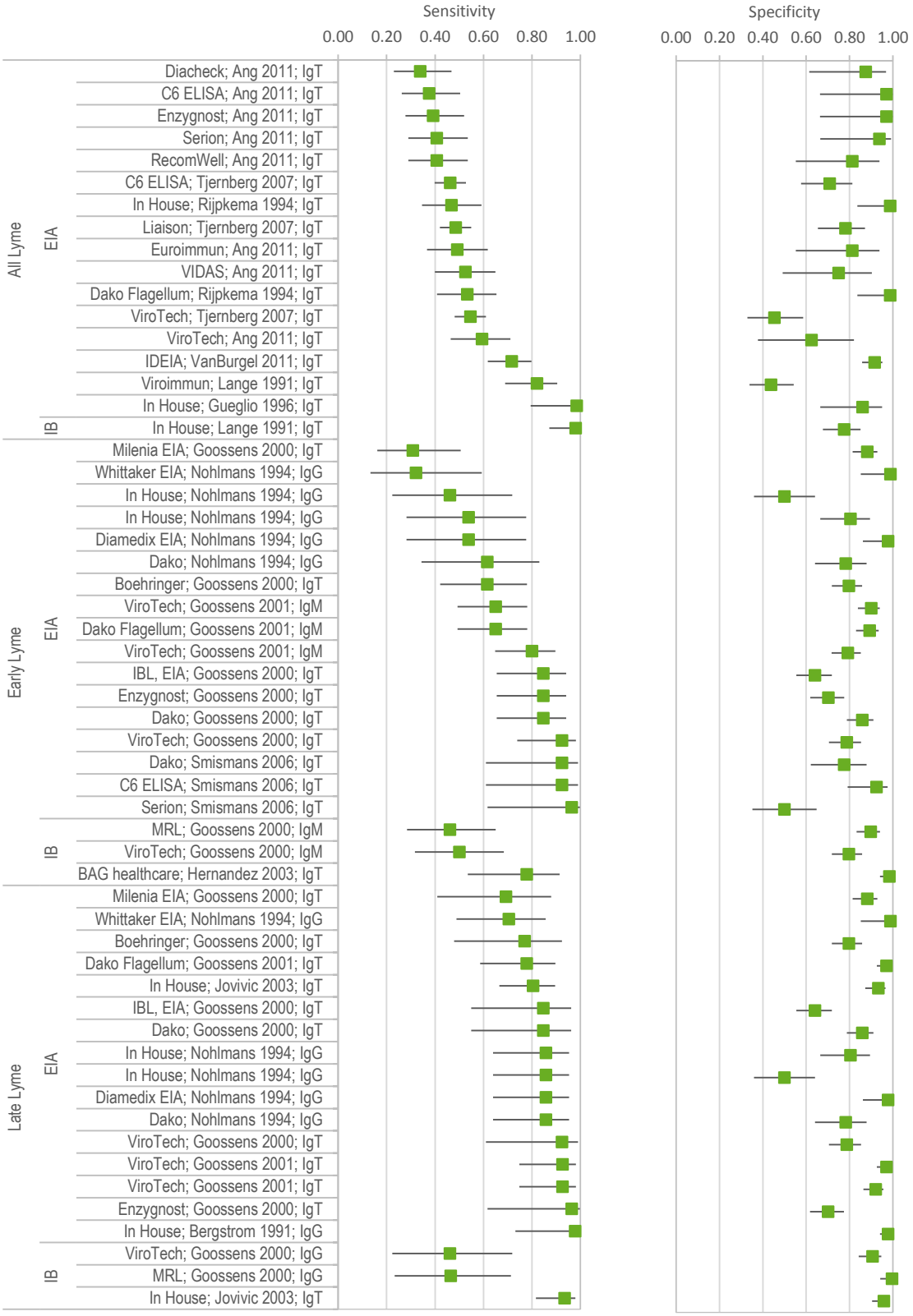


Annex 16. Lyme borreliosis-unspecified: case-control studies with cross-reacting controls

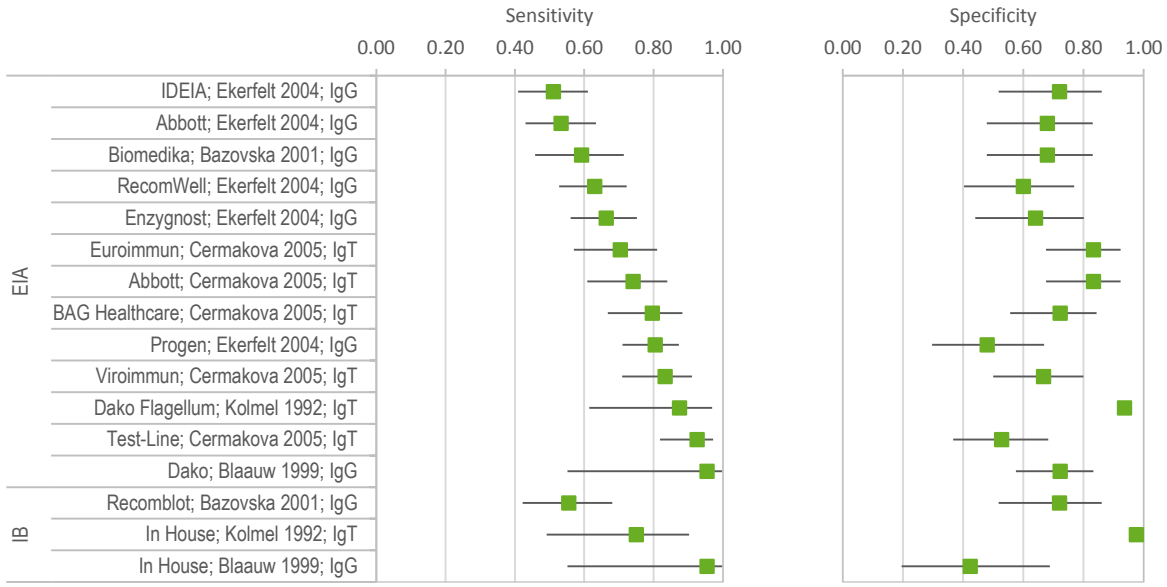
Sensitivity and specificity of IgM, IgG and IgT for LB-unspecified case-control studies with cross-reacting controls.



Sensitivity and specificity of LB-unspecified case-control studies with cross-reacting controls.



Annex 17. Lyme borreliosis-unspecified: cross-sectional studies



Annex 18. Sensitivity analyses with and without the study by Ang et al. 2012

Target condition, study type	Parameter	Estimate (95% CI) including Ang 2012	Estimate (95% CI) excluding Ang 2012	Delta
EM, healthy controls	Diagnostic odds ratio	19.2 (11.2–32.8)	18.2 (10.4–31.9)	0.97
	sensitivity	0.50 (0.40–0.61)	0.48 (0.38–0.58)	0.02
	specificity	0.95 (0.92–0.97)	0.95 (0.92–0.97)	0.00
EM, cross-reacting controls	Diagnostic odds ratio	13.0 (7.6–22.4)	13.0 (7.3–23.1)	0.02
	sensitivity	0.47 (0.33–0.61)	0.45 (0.31–0.59)	0.02
	specificity	0.94 (0.89–0.96)	0.94 (0.90–0.97)	0.00
NB, healthy controls	Diagnostic odds ratio	39.2 (19.8–77.3)	36.8 (18.8–72.2)	2.33
	sensitivity	0.77 (0.67–0.85)	0.75 (0.66–0.82)	0.02
	specificity	0.92 (0.86–0.96)	0.93 (0.87–0.96)	0.00
NB, cross-reacting controls	Diagnostic odds ratio	21.0 (10.4–42.3)	19.1 (9.2–39.5)	1.90
	sensitivity	0.70 (0.60–0.79)	0.68 (0.59–0.76)	0.02
	specificity	0.90 (0.83–0.94)	0.90 (0.83–0.94)	0.00
ACA, healthy controls	Diagnostic odds ratio	632.2 (94.7–4222.6)	448.6 (73.4–2743.3)	183.57
	sensitivity	0.97 (0.84–1.00)	0.96 (0.80–0.99)	0.01
	specificity	0.95 (0.89–0.97)	0.95 (0.89–0.97)	0.00
ACA, cross-reacting controls	Diagnostic odds ratio	94.9 (12.1–743.0)	73.2 (8.7–614.6)	21.64
	sensitivity	0.91 (0.61–0.98)	0.87 (0.53–0.97)	0.04
	specificity	0.91 (0.80–0.96)	0.92 (0.82–0.97)	–0.01
LA, healthy controls	Diagnostic odds ratio	86.3 (45.5–163.9)	85.5 (41.1–178.1)	0.84
	sensitivity	0.88 (0.83–0.92)	0.89 (0.83–0.93)	–0.01
	specificity	0.92 (0.88–0.95)	0.91 (0.87–0.94)	0.01
LA, cross-reacting controls	Diagnostic odds ratio	216.7 (36.0–1304.3)	273.8 (24.7–3038.1)	–57.13
	sensitivity	0.95 (0.90–0.98)	0.95 (0.84–0.99)	0.00
	specificity	0.92 (0.75–0.98)	0.93 (0.78–0.98)	–0.02
LB-unspecified, healthy controls	Diagnostic odds ratio	71.3 (13.8–368.5)	71.3 (13.8–368.5)	0.00
	sensitivity	0.73 (0.53–0.87)	0.73 (0.53–0.87)	0.00
	specificity	0.96 (0.91–0.99)	0.96 (0.91–0.99)	0.00
LB-unspecified, cross-reacting controls	Diagnostic odds ratio	38.3 (10.6–138.1)	36.1 (11.5–113.0)	2.21
	sensitivity	0.81 (0.64–0.91)	0.80 (0.65–0.90)	0.01
	specificity	0.90 (0.79–0.96)	0.90 (0.80–0.95)	0.00

European Centre for Disease Prevention and Control (ECDC)

Postal address:
Granits väg 8, SE-171 65 Solna, Sweden

Visiting address:
Tomtebodavägen 11A, SE-171 65 Solna, Sweden

Tel. +46 858601000
Fax +46 858601001
www.ecdc.europa.eu

An agency of the European Union
www.europa.eu

Subscribe to our monthly email
www.ecdc.europa.eu/en/publications

Contact us
publications@ecdc.europa.eu

Follow us on Twitter
@ECDC_EU

Like our Facebook page
www.facebook.com/ECDC.EU

ECDC is committed to ensuring the transparency and independence of its work

In accordance with the Staff Regulations for Officials and Conditions of Employment of Other Servants of the European Union and the ECDC Independence Policy, ECDC staff members shall not, in the performance of their duties, deal with a matter in which, directly or indirectly, they have any personal interest such as to impair their independence. Declarations of interest must be received from any prospective contractor(s) before any contract can be awarded.
www.ecdc.europa.eu/en/aboutus/transparency

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

Publications Office