Diagnostic accuracy and comparison of two assays for *Borrelia*-specific IgG and IgM antibodies: proposals for statistical evaluation methods, cut-off values and standardization

Ram Benny Dessau

Department of Clinical Microbiology, Slagelse Hospital, Region Sjælland 18 Ingemannsvej, DK4200 Slagelse, Denmark

Two assays (Liaison, Diasorin; IDEIA, Oxoid) for detection of *Borrelia*-specific antibodies were compared. A case–control design using patients with neuroborreliosis (*n* = 48), laboratory defined by a positive *Borrelia*-specific antibody index in the spinal fluid, was available and was intended to represent the serological response of disseminated early Lyme borreliosis in general. Serum samples were obtained from 216 Danish blood donors as controls. By comparing sensitivity and specificity using pre-specified cut-off values, significant differences were found. However, using receiver operating characteristic (ROC) curves to optimize and standardize test interpretation, it was shown that testing with both IDEIA IgG and IgM was comparable to testing with Liaison IgG alone by comparing the area under the curve of the diagnostically relevant 25 % partial ROC curve (*P* = 0.1). When using the Liaison OspC/VlsE IgM assay, the specificity was decreased without a gain in sensitivity. This study proposes standardizing of reporting by using a control population as the reference and choosing decision thresholds guided by the risk of false-positive results at 2 and 8 %. The sensitivities for IDEIA (IgG and IgM combined) were 85 and 95 % and for the Liaison (VlsE IgG) method were 67 and 96 %, respectively. Methods for test evaluation, test interpretation and statistical testing are presented and discussed. In conclusion, Liaison VlsE IgG alone and IDEIA IgG/IgM combined showed a high and comparable discriminatory ability to distinguish serum samples from patients with neuroborreliosis from blood donor controls. However, cut-off values should be adjusted for a proper comparison.

INTRODUCTION

Lyme borreliosis (LB) is caused by infection with the tick-borne bacterium *Borrelia burgdorferi sensu lato*. The most common clinical manifestation is a rash called erythema migrans. The disseminated forms of LB are rarer and include Lyme neuroborreliosis (NB), arthritis, multiple erythema migrans, lymphocytoma and carditis (Stanek et al., 2011, 2012). To support clinical diagnosis of disseminated LB, the standard laboratory tests use detection of the antibody response in blood or spinal fluid. *Borrelia*-specific antibody detection is an excellent tool, as sensitivities are high in disseminated disease. The analytical specificity will be high, as *B. burgdorferi* possesses diagnostically relevant antigens that are distinct from other related bacteria. The clinical specificity will also be high, except in certain subpopulations with continuous exposure to ticks.

As with the use of serology for diagnostic purposes for any infectious disease, interpretation requires knowledge of the risk of background immunity, and the fact that sensitivity may be low in early clinical disease but should be detectable in all patients by 6–8 weeks after onset of symptoms. Laboratory support is not necessary or useful for diagnosis of localized erythema migrans, as some patients may not develop a detectable antibody response at all.

Studies on diagnostic accuracy for *Borrelia*-specific IgG and IgM antibodies are most often presented in the published literature by counting positive and negative results, without analysis of the quantitative results. The cut-off recommendation provided by the developer of the assay has been accepted uncritically in many studies without analysis or discussion (Busson et al., 2012; Cerar et al., 2010; Marangoni et al., 2006, 2008; Petersen et al., 2008; Riesbeck & Hammas, 2007; Tjernberg et al., 2007). It has, however, been generally recommended that an explanation of the definition of and rationale for the units and cut-off values is given when reporting studies of diagnostic accuracy (Bossuyt et al., 2003a, b). The purpose of the
current study was to compare the diagnostic accuracy of two assays that are commonly used in Scandinavian laboratories (Dessau et al., 2011) and to propose a choice of cut-off values. Statistical and graphical methods to report the results are presented and discussed. The aim of the study was to show that analysis of raw quantitative measurements is essential to establish conclusions on assay performance. Additionally, it is proposed that standardization of results of *Borrelia* serology is necessary using the controls as a reference population.

**METHODS**

All test samples were Danish. Serum samples from 48 patients with a pleocytosis (≥3 × 10³ per litre) and a positive antibody index in the cerebrospinal fluid were included. The antibody index was measured using an IDEIA Lyme Neuroborreliosis IgG/IgM assay (Oxoid). This assay was developed in Denmark (Hansen, 1994; Hansen & Lebech, 1991, 1992). The NB samples were collected consecutively from 2006 to 2010. Serum samples from 216 blood donors (BDs) were collected in January 2011 as controls. This is another set of samples from a previous publication (Dessau et al., 2010b). Forty-four consecutive routine samples from September 2011 were used in this study to illustrate the proposed standardized reporting.

The *Borrelia burgdorferi* IDEIA (Oxoid) serum assay is based on purified native flagella antigen from a cultured strain of *Borrelia afzelii* (strain DK1) (Hansen et al., 1988, 1991). This assay was run on a Siemens BEP2000 automated ELISA instrument. The cut-off value for this assay was originally adjusted by choosing 2% seropositivity for IgG or IgM in a population of Danish BDs. The 2% specificity has not changed over time, as shown in a recent evaluation (Dessau et al., 2010b).

The Liaison *Borrelia* IgM Quant and the *Borrelia* IgG assays (Diasorin) are based on recombinant OspC and VlsE antigens in the IgM assay and on VlsE antigens in the IgG assay. These assays run on a dedicated automated instrument (Liaison). Antigens are coated on magnetic beads and chemiluminescence is used for signal detection. The recombinant proteins used as antigens are expressed in *Escherichia coli*. This study compared the IDEIA and the Liaison assays. A second assay, for example an immunoblot, was not performed, but considerations about threshold adjustment for the purpose of two-tier testing are presented in the discussion. It should be noted that many laboratories in Scandinavia do not use a two-tier procedure on most routine samples (Dessau et al., 2011). This study focused on describing, comparing and optimizing the discriminatory power by quantitative analysis.

**Definitions.** The term decision threshold is used to denote a value based on the distribution of the measurement values in the BD controls. This value may be based on a single assay or a combination of assays scored together.

The term cut-off is used to denote a chosen measurement value (e.g. calibrated arbitrary units) based on the quantitative results provided by the laboratory equipment for each assay.

The term specificity in this study is defined as 1 − (fraction of BDs who were test positive at a chosen cut-off value or decision threshold).

The term sensitivity in this study is defined as the fraction of NB patients who were test positive at a chosen cut-off value or decision threshold.

**Statistics.** To assess the statistical significance of the sensitivity and specificity, a regression method for paired samples was used (Pepe, 2003a). This method may be seen as an extension of the McNemar test for paired samples. The regression methodology has the advantage of estimating relative positive fractions with confidence intervals and multiple tests may be compared. Receiver operating characteristic (ROC) curves were used for comparison of quantitative data. Logistic regression was used to combine the results of the four biomarkers. Comparison of the area under the curve (AUC) was performed using a bootstrap method (Pepe et al., 2009; Robin et al., 2011). A concept of cumulative distributions, which is similar to the ROC curve, is proposed to standardize the results for evaluation and reporting (Huang & Pepe, 2009). All statistical analyses were performed using R statistical software (R Development Core Team, 2012). Further details are provided in the supplementary material (available in JMM Online). Confidence intervals were chosen at 95%.

**RESULTS**

**Qualitative analysis**

Traditional qualitative results using the cut-off values recommended by the manufacturer are shown in Table 1. For statistical significance, the relative sensitivity and specificity of each test was compared using the positive fraction of IDEIA IgG or IgM as reference. The most sensitive test was Liaison IgG or IgM at 96%, but the specificity was low at 89%. Liaison IgG alone had a sensitivity of 94% and the specificity was increased to 94%. Thus, it would seem obvious to choose Liaison IgG alone and not use IgM. In contrast, IDEIA IgG and IgM supplemented each other and both had to be used to gain a sensitivity of 81%, whilst the specificity remained high at 98%. By choosing to use Liaison IgG VlsE alone, the sensitivity was higher (borderline significant), but the false-positive rate was three times higher compared with IDEIA. However, is this an adequate description and comparison of assay performance? To answer this question, an analysis of the quantitative data was necessary.

**Quantitative analysis**

Strip plots as shown in Fig. 1 may be useful, as they show the individual raw data values and give a visual impression of the provided cut-off values. In this case, as a compromise, the many quite negative results mainly in the BDs are indicated instead by their number. Using Liaison IgG, all NB samples were reactive, with just two samples in the indeterminate zone; however, among the 14 reactive BDs, nine were more strongly reactive. Using Liaison IgM, there were 13 seropositive BDs. For the IDEIA assay, the cut-off values were chosen at the upper tail of the distribution of BD controls. This was in contrast to the Liaison assays where sensitivity was maximized for IgG. For Liaison IgM, the choice of cut-off value appeared to be a difficult compromise to yield at least some sensitivity.

An ROC curve uses ranking of the samples, and thus tests with different numerical scales can be compared. As shown in Fig. 2, the results of the 216 BD samples were ranked on
The number positive and the relative fraction of seropositivity in 48 patients with NB and 216 BDs for the Liaison and IDEIA assays

A regression model of relative positive fractions was used to assess statistical significance. For the comparison, IDEIA IgG or IgM was chosen as the reference and given a value of 1. 'Indeterminate' results have been counted as negative.

<table>
<thead>
<tr>
<th>Assays and combinations</th>
<th>Patients with NB (n=48)</th>
<th>Danish BDs (n=216)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number positive (%)</td>
<td>Relative positive fraction (95 % CI)</td>
</tr>
<tr>
<td>IDEIA IgG or IgM</td>
<td>39 (81 %)</td>
<td>1</td>
</tr>
<tr>
<td>Liaison IgG or IgM</td>
<td>46 (96 %)</td>
<td>1.20 (1.0–1.4)*</td>
</tr>
<tr>
<td>IDEIA IgG</td>
<td>21 (44 %)</td>
<td>0.54 (0.38–0.76)*</td>
</tr>
<tr>
<td>IDEIA IgM</td>
<td>29 (60 %)</td>
<td>0.74 (0.57–0.97)*</td>
</tr>
<tr>
<td>Liaison IgG</td>
<td>45 (94 %)</td>
<td>1.15 (0.99–1.35)</td>
</tr>
<tr>
<td>Liaison IgM</td>
<td>22 (46 %)</td>
<td>0.56 (0.40–0.79)*</td>
</tr>
</tbody>
</table>

*95% confidence intervals (CI) greater or smaller than 1 are significant.

The x-axis whilst the 48 NB samples were ranked on the y-axis. ROC curves allow assessment of the discriminatory power of the biomarker and the choice of cut-off value. Visual inspection of the ROC curve in Fig. 2(a) showed that the higher sensitivity of the Liaison IgG and better specificity of IDEIA IgG or IgM were due mainly to differences in the cut-off values chosen by the manufacturers. When combining IDEIA IgG and IgM in a regression model, the classifier performance of IDEIA IgG or IgM appeared slightly more specific compared with Liaison IgG (Fig. 2b); however, this difference was not statistically significant (P=0.1) when the 75–100 % specificity partial AUCs were compared.

Decision thresholds to maximize specificity are proposed at 2 % and to maximize sensitivity at 8 % false-positive rate (Fig. 2b, Table 2). The difficulties in setting a decision threshold may be appreciated by looking at Fig. 2(b). Had the threshold been shifted to 3 %, the true positive rate of Liaison and IDEIA would be nearly identical. A small tuning of the threshold may therefore have a large influence on the apparent sensitivity. Comparison of the quantitative data showed that the discriminatory performance of Liaison IgG was very similar to the combination of IDEIA IgG and IgM.

Another characteristic to assess is the positive predictive value (PPV) using the chosen decision thresholds. The pretest probability of disease has a large impact on test performance, and low pre-test probabilities may be problematic (Habbema et al., 2002; Sackett & Haynes, 2002). In Fig. 3(a), the PPV is shown as a function of the pre-test prevalence of disease using the formula shown in the supplementary material. Sensitivity had little impact on PPV compared with specificity when the pre-test prevalence was at the lower range of 10 % or less. At higher pre-test prevalence, the sensitivity became more important for negative predictive value (NPV).

A specificity of 98 % was chosen as an important criterion to set as the cut-off value, as at this level the PPV may reach 50 %, even at a low pre-test prevalence around 2 %, and a level of 80 % PPV may be reached at 10 % pre-test prevalence. The NPV remained above 95 % at the 2 or 10 % pre-test prevalence, regardless of the sensitivity level (Fig. 3b). At a higher pre-test prevalence, the NPV clearly became very low. Choosing decision thresholds at the beginning and end of the shoulder of the ROC curve maximizes the specificity and sensitivity, respectively, with an area of greater diagnostic uncertainty in between.

Agreement

All correlations were positive (Fig. 4) and the patients with NB tended to be positive pairwise for the four biomarkers but with a wide variability in magnitude. The two IgG assays had a relatively high overall correlation of 0.84, whereas Liaison IgM has less correlation with the other three biomarkers. The BDs resided mostly in the negative cloud (Fig. 4, black circles). Smoothed non-linear regression lines were added to the graphs to visualize the mean tendency to positive correlation in each data pair (Fig. 4).

Standardization using controls as the reference population

It has been argued above that the PPV of a test depends mainly on the pre-test probability of disease and the specificity. Thus, it has been proposed to standardize a biomarker by using the controls as a reference distribution (Huang & Pepe, 2009). The result of the measurement in a patient is then expressed as the ‘percentile value’ of the empirical cumulative distribution in the control group. This percentile value is exactly the same as the false-positive rate on the ROC curve. Thus, the methodology is basically the same, but the presentation is different. The important implication is that this framework indicates the possibility of biomarker standardization for reporting of results. Seropositivity is thus expressed relative to the control
distribution. With regard to the present data, the 48 NB patients had a high reactivity compared with the healthy background population (Fig. 5) represented by the BDs.

The Liaison IgG assay contained 181 measurements censored below 5 units. This gave 181 ties in the cumulative distribution and the percentile values could only start at 84 % (Fig. 5). The observations at or below the 75 % percentile value were censored in the plot anyway, as they had little diagnostic value. Thus, the results of the different assays were conveniently standardized and expressed on the same scale. To illustrate standardized reporting (Fig. 5), 44 consecutive routine patients of unknown clinical status were analysed. Ten (22 %) were Liaison IgG positive and four (9 %) were IDEIA IgG and IgM positive above the 92 % percentile (McNemar’s test $P=0.04$). This can be compared with Fig. 1 where the four different plots each had their own scale of measurement and levels of cut-offs. Furthermore, the scale (y-axis on Fig. 5) is clinically informative as the risk of false or natural background seroreactivity is inherent.

**DISCUSSION**

Using the pre-specified cut-off values, the present study falls within the range of sensitivities found in previous studies. In two studies using the Liaison assay with a total of 28 and 34 patients with NB, the reported sensitivities were 7 and 56 % for IgM and 86 and 74 % for IgG, respectively (Cerar et al., 2006, 2010). This is similar to the present study where 46 % of NB patients were found to be...
Liaison IgM positive and 94% IgG positive. In the first study, only two of 28 patients were found to be IgM positive, but this can be explained as the NB patients were not defined and the IgM assay was a previous version of the assay containing only OspC and not VlsE as antigen. In previous publications with NB patients, the sensitivity and specificity of the IDEIA assay ranged from 44 to 84% (median 65%) for IgG and from 41 to 74% (median 57%) for IgM (Bennet et al., 2008; Cerar et al., 2010; Dessau et al., 2010b; Ekerfelt et al., 2004; Hansen & Lebech, 1991, 1992; Lebech et al., 2000). In the present study, the sensitivity of IgG was 44% and of IgM was 60%, which is within the range of published studies. In this study, 81% of NB patients were IgG or IgM positive, which is comparable to the published sensitivities of 76–93% (median 88%), which confirms the utility of combining IDEIA IgG and IgM (Bennet et al., 2008; Dessau et al., 2010b; Hansen & Lebech, 1992; Lebech et al., 2000).

In this group of 48 patients with NB, the VlsE-based IgG antigen alone was sensitive enough to render IgM testing unnecessary. ‘Single-tier’ testing with the C6 fragment of VlsE has been found to be comparable to various other test combinations in a study based on North American patients (Wormser et al., 2013). A multicenter European study evaluating the Siemens Enzygnost assay found that, after addition of recombinant VlsE, an increased IgG sensitivity was observed, mainly in the IgM-positive samples (Hunfeld et al., 2005).

Agreement between two assays using different Borrelia-specific antigens should not be expected, only a tendency to co-reactivity. There is no known biological mechanism to co-regulate development of individual antibodies to different specific antigens. In contrast, variable antibody expression both in time and in magnitude should be expected due to complex regulation of antibody responses with feedback regulation (Hjelm et al., 2006). In this study, some positive correlation was found but certainly no linear relationship. The data corroborate an assumption of variability concerning antibody responses in patients with LB. There is heterogeneity of the humoral response to different target antigens in the same patient and between patients after bacteraemia with Staphylococcus aureus (Verkaik et al., 2010). This natural variation among biomarkers detecting antibodies to the same disease has been found previously and is troublesome for proficiency testing programmes and assay comparison (Ang et al., 2011; Brandenburg et al., 2011; Müller et al., 2012; Robertson et al., 2000). It is possible that the detection system containing somewhat denatured antigens will contribute to the variability in signal strength. This means that a given patient may not necessarily express detectable antibodies to both the flagella and the VlsE antigen at the same time. However, a tendency for a stronger response to most immunodominant antibodies as a function of disease

---

**Table 2.** Sensitivity using decision thresholds set at a false-positive rate of 2 and 8%  

<table>
<thead>
<tr>
<th></th>
<th>IDEIA IgG and IgM</th>
<th>Liaison VlsE IgG</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB (n=48)</td>
<td>85%</td>
<td>67%</td>
</tr>
<tr>
<td>Chosen specificity</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>Cut-off point</td>
<td>0.52</td>
<td>117</td>
</tr>
</tbody>
</table>

**Fig. 2.** Partial ROC curves up to a 25% rate of seropositivity in the BDs. (a) Individual assays with the position of cut-off points as recommended by the manufacturers. (b) Combination of IDEIA IgG and IgM using a logistic regression model compared with Liaison IgG. The partial AUCs in (b) are 23.8 and 22.7% (P=0.10, bootstrap method). Cut-off values are proposed at 2 and 8% false-positive rate (dashed vertical lines).
development (time, dissemination and intensity of the inflammatory response) could be expected. In this study, we did not have clinical data on the patients with NB. It is likely that these patients had variability in clinical presentation concerning the duration of disease and severity. This variability was evident in the data, as shown in the matrix plot (Fig. 4).

ROC curves provide a description of performance and may be used for comparing tests and for guiding the choice of threshold for clinical applications (Pepe, 2003b). In this study, a partial 25 % AUC was chosen as a clinically relevant summary measure. The choice of the restricted region may be arguable, but in this study it was chosen to safely include the maximal obtainable specificity.

ROC analysis has been used occasionally for evaluating *Borrelia* assays (Burbelo et al., 2010; Huppertz et al., 1998; Porwancher et al., 2011; Wagner et al., 2011). The author has not been able to find previously published papers discussing the choice of cut-off for detecting *Borrelia*-specific antibodies for clinical application. Major reviews do not consider this issue and just list the range of sensitivities and specificities (Aguero-Rosenfeld, 2008; Aguero-Rosenfeld et al., 2005). As shown in the present study, this heterogeneity could mainly be explained by the choice of cut-off value. The cut-off point of the IDEIA was originally established as a decision threshold by defining that 98 % of Danish healthy controls should be seronegative (Dessau et al., 2010b). The strategy of cut-off determination for the Liaison assay is not documented, but it appears from the present study that sensitivity was maximized for IgG and a compromise between sensitivity and specificity was chosen for IgM. The statistical comparison of ROC curves gives a gross impression if the AUC, in this case the 25 % partial AUC, is different in the two datasets. As the curves were nearly parallel (e.g. without crossover), this model should be fairly valid for testing of statistical significance.
The concept of standardizing using a disease-negative or normal population as reference and expressing the ‘abnormal’ result relative to this distribution is a basic issue concerning the use of any biomarker for diagnostic purposes (Huang & Pepe, 2009; Sackett & Haynes, 2002). Using the control group to determine the decision threshold is practical, as samples from patients without LB are readily available in large numbers in any diagnostic laboratory. In this study, we used BDs routinely screened as healthy at the time of sampling. Thus, the lowest possible background seroreactivity for both IgG and IgM should be expected especially when sampled during winter, as in this case. More reasons to base the decision threshold mainly on the distribution of a non-LB control group are that samples are quite conveniently available in larger numbers and, if chosen locally, will represent the seroreactivity in the adult population.

The choice of decision threshold is intended to guide the clinician in diagnosing a patient with active Borrelia infection. The author would like to propose two decision thresholds for use in different clinical situations where LB is suspected:

1. Cut-off at 98 % specificity for patients with longer (e.g. >30 days) duration of clinical disease considering the IgG response alone or combined with IgM as appropriate according to recommendations (Stanek et al., 2011), and in any case where the pre-test probability of LB is low. This applies to the ‘rule out’ situation in patients where the clinical suspicion of LB is low.

2. A lower decision threshold at 92 % specificity for patients with shorter duration of clinical disease and where the clinical presentation is typical of LB, consistent with case definitions and when the pre-test probability is higher than 10 %. Both IgG and IgM response should be considered or VlsE IgG alone.

This could make clinical sense as the diagnostic situations differ as a function of disease development. Between these two decision thresholds is the shoulder of the ROC curve where the assays are neither safely negative nor safely positive. Thus, this is the clinically important indeterminate or grey-zone interval. The more specific cut-off could be applied when diagnosing acrodermatitis or arthritis where high levels of antibody reactivity are expected. However, when diagnosing earlier diseases like lymphocytoma or Borrelia carditis, the lower measurement values could be evaluated as positive, but the clinician has to be aware of an 8 % positive rate in the background population. If an ELISA is to be used as a screening assay before a second test, the lower cut-off value should probably always be used. However, it is beyond the scope of this study to optimize a two-tier combination. It is probably more important to optimize the co-interpretation of IgG and IgM as proposed in the present study and in a previous study (Dessau et al., 2010b). However, the principles of data analysis presented in this study could serve as a basis for adding more biomarkers to the test strategy.

It is relevant to consider a combination of test results when choosing decision thresholds, as most laboratories perform both IgG and IgM testing and some perform supplementary tests such as immunoblotting. It is possible that cut-offs may have to be adjusted to a representative local or regional serum panel of controls as background immunity may vary. Another possibility is to avoid using thresholds. The author has previously recommended reporting quantitative measurements with probability thresholds (risk scores) for guidance (Dessau et al., 2010b).
This study had a case–control design. This design was chosen because of the practical availability of samples and a lower risk of misclassification. In principle, this is not ideal, as it is not representative for a consecutive cohort of patients. Cases of NB were included by virtue of being reactive to the flagella antigen in the IDEIA antibody index test (and cerebrospinal pleocytosis). Thus, the sensitivity of IDEIA could be overestimated compared with Liaison. Thus, the overall sensitivity could be overestimated. In the already cited Slovenian study, the inclusion criteria for NB were a current or recent erythema migrans and spinal fluid pleocytosis (Cerar et al., 2010). In this study, the sensitivity was found to be lower in both Liaison (79 %) and IDEIA (59 %) for IgG or IgM. However, the selection of cases is in principle not so important for the choice of cut-off or estimated predictive values. Even if the IDEIA assay is less sensitive than estimated in the present study, the impact on PPV or NPV would be small, as discussed. We can assume that patients without Borrelia-specific antibodies would belong to a group with lower values compared with Borrelia patients who have converted to higher values, and that few patients are caught in the phase of conversion. Thus, the means and variation of the two distributions would not be affected by selection bias in the NB group and the choice of decision threshold should not be changed. Selection bias would affect the number of persons counted above and below the decision threshold.

Other manifestations of LB (e.g. Borrelia carditis, lymphocytoma) are too rare to collect serum panels within a practical time-span. Thus, it is necessary to use patients with NB instead of patients with other disseminated clinical manifestations of LB for test evaluations. Hypothetically, antibody detection should depend more on the disease development in general (duration and extent of inflammation) and less on the anatomical site as such, and the dynamics of the humoral immune response in patients with NB should be similar to patients with early disseminated LB involving other organ systems.

The risk of including non-LB patients should be low. The IDEIA antibody index test is documented to be very specific and sensitive in the Danish context and may be considered a reference standard based on evaluation of a large panel of well-characterized patients with NB (Dessau et al., 2010b; Hansen, 1994; Hansen & Lebech, 1992). In a highly endemic area like Slovenia, this index assay was 90 % specific (Cerar et al., 2010). In this study, the NB samples were collected as a part of the consecutive clinical routine.

Pre-test prevalence is more important for the predictive value compared with sensitivity and specificity, as shown in Fig. 3 (Fletcher et al., 1996). This emphasizes that LB diagnosis should be based on the clinical presentation and an assessment of tick-exposure risk and not on laboratory results alone (Stanek et al., 2011). One reason is that pre-test prevalence, depending on the individual clinical situation, will vary over a wide range from much less than 1 % to near 100 % diagnostic certainty.

Detection of serum antibodies has been recommended for diagnosing disseminated cases of LB except for erythema migrans and NB (Stanek et al., 2011). However, Borrelia serology is widely used in various clinical situations where the clinician may instead want to rule out LB and where the pre-test probability of LB is low (Dessau et al., 2010a). This argues for the importance of selecting decision thresholds based on the distribution of LB-negative controls.

CONCLUSION

The study has shown the importance of assessing quantitative data so that false conclusions concerning performance of the test antigens are not reached. It is important to assess the cut-off strategy independently and not just accept the provided thresholds. A newer development is the availability of tests allowing the assessment of statistical significance when comparing ROC curves (Pepe et al., 2009; Robin et al., 2011). The use of non LB controls (for example BDs) as a reference population to standardize interpretation of results using empirical percentile values instead of measurement units has been discussed.

It was shown that the Liaison IgG assay using recombinant VlsE antigen had an excellent discriminatory capability comparable to the combination of IDEIA IgG and IgM when evaluated in a serum panel from patients with NB compared with Danish BDs. Two clinical decision thresholds have been proposed defined by the specificity: a specificity of 98 % for patients with a longer duration of clinical disease, and an intermediate seroreactivity between the lower 92 % and upper 98 % decision threshold in clinical situations where there is suspicion of early LB and where the pre-test probability of LB is high at around 10 % or more.

ACKNOWLEDGEMENTS

The author is a member of the ESCMID Study Group for Lyme Borreliosis – ESGBOR (https://www.escmid.org/esgbor). Discussions I had with participants at ESGBOR workshops held in Vienna in November 2011 and 2012 were very helpful as background to prepare this manuscript. I thank Oxoid Thermostisher and Diasorin for providing technical support. This study was approved by the ethics committee in Region Sjælland (1-01-83-0002-07, December 2010).

REFERENCES


