Michelle Elaine Orme*, Carmen Andalucia, Sigrid Sjölander and Xavier Bossuyt

A hierarchical bivariate meta-analysis of diagnostic test accuracy to provide direct comparisons of immunoassays vs. indirect immunofluorescence for initial screening of connective tissue diseases

<https://doi.org/10.1515/cclm-2020-0094>

Received November 1, 2019; accepted April 4, 2020

Abstract

Objectives: To compare indirect immunofluorescence (IIF) for antinuclear antibodies (ANA) against immunoassays (IAs) as an initial screening test for connective tissue diseases (CTDs).

Methods: A systematic literature review identified cross-sectional or case-control studies reporting test accuracy data for IIF and enzyme-linked immunosorbent assays (ELISA), fluorescence enzyme immunoassay (FEIA), chemiluminescent immunoassay (CLIA) or multiplex immunoassay (MIA). The meta-analysis used hierarchical, bivariate, mixed-effect models with random-effects by test.

Results: Direct comparisons of IIF with ELISA showed that both tests had good sensitivity (five studies, 2321 patients: ELISA: 90.3% [95% confidence interval (CI): 80.5%, 95.5%] vs. IIF at a cut-off of 1:80: 86.8% [95% CI: 81.8%, 90.6%]; $p=0.4$) but low specificity, with considerable variance across assays (ELISA: 56.9% [95% CI: 40.9%, 71.5%] vs. IIF 1:80: 68.0% [95% CI: 39.5%, 87.4%]; $p=0.5$). FEIA sensitivity was lower than IIF sensitivity (1:80: $p=0.005$; 1:160: $p=0.051$); however, FEIA specificity was higher (seven studies, $n=12,311$, FEIA 93.6% [95% CI: 89.9%, 96.0%] vs. IIF 1:80 72.4% [95% CI: 62.2%, 80.7%]; $p<0.001$; seven studies, $n=3251$, FEIA 93.5% [95% CI: 91.1%, 95.3%] vs. IIF 1:160 81.1% [95% CI: 73.4%, 86.9%]; $p<0.0001$). CLIA sensitivity was similar to IIF (1:80) with higher specificity (four studies, $n=1981$: sensitivity 85.9% [95% CI: 64.7%,

95.3%]; $p=0.86$; specificity 86.1% [95% CI: 78.3%, 91.4%]). More data are needed to make firm inferences for CLIA vs. IIF given the wide prediction region. There were too few studies for the meta-analysis of MIA vs. IIF (MIA sensitivity range 73.7%–86%; specificity 53%–91%).

Conclusions: FEIA and CLIA have good specificity compared to IIF. A positive FEIA or CLIA test is useful to support the diagnosis of a CTD. A negative IIF test is useful to exclude a CTD.

Keywords: antinuclear antibodies; connective tissue disease; immunofluorescence; meta-analysis; solid-phase immunoassay; systematic literature review.

Abbreviations: ACR, American College of Rheumatology; AI, autoimmune; ANA, antinuclear antibody; ARD, autoimmune rheumatic disease; CFS, chronic fatigue syndrome; CI, confidence interval; CLIA, chemiluminescent immunoassay; CTD, connective tissue disease; DC, diseased control; DM, dermatomyositis; DOR, diagnostic odds ratio; ELISA, enzyme-linked immunosorbent assay; EULAR, European League Against Rheumatism; FEIA, fluorescence enzyme immunoassay; FN, false negative; FP, false positive; HC, healthy control; HEp-2, human epithelial type 2 cells; HEp-2000, human epithelial type 2 cells – transfected with Ro60 cDNA; HSROC, hierarchical summary receiver operating characteristic; IA, immunoassay; IIF, indirect immunofluorescence; IM, inflammatory myopathy; Lim SD, limited scleroderma; LR, likelihood ratio; MCTD, mixed connective tissue disease; MeSH, Medical Subject Headings; MIA, multiplex immunoassay; MLE, maximum likelihood estimation; LR–, negative likelihood ratio; LR+, positive likelihood ratio; PBC, primary biliary cholangitis; PM, polymyositis; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RA, rheumatoid arthritis; QUADAS-2, Quality Assessment Tool for Diagnostic Accuracy Studies – version 2; SJS, Sjögren's syndrome; SLE, systemic lupus erythematosus; SSc, systemic sclerosis; SPA, solid-phase assay; TN, true negative; TP, true positive; UCTD, undifferentiated connective tissue disease.

*Corresponding author: Michelle Elaine Orme, ICERA Consulting Ltd, 17 Redbridge Close, Swindon, Wiltshire, UK, Phone: +44 (0) 1793 876761,

E-mail: michelle.orme@iceraconsulting.com

Carmen Andalucia and Sigrid Sjölander: Evidence Generation, Immuno Diagnostics, Thermo Fisher Scientific, Uppsala, Sweden

Xavier Bossuyt: Department of Microbiology, Immunology and Transplantation, Clinical and Diagnostic Immunology, KU Leuven, Leuven, Belgium; and Department of Laboratory Medicine, University Hospitals Leuven, Leuven, Belgium

Introduction

The presence of antinuclear antibodies (ANA) can indicate an autoimmune (AI) disease such as a connective tissue disease (CTD). International guidelines state that the diagnosis of a CTD requires a panel of tests, with the detection of ANA as the first-level screening test [1]. If the ANA screening is positive, then further steps would follow, whereby specific antibody tests are performed to definitely rule in an autoimmune rheumatic disease (ARD). A definitive diagnosis of a CTD including the specific type of CTD would be based on the diagnostic criteria for each CTD classification [2–13]. Therefore, it is important that the ANA test is accurate as this is the first stage in the diagnostic pathway, and the results will determine subsequent follow-up.

There is a broad consensus that the indirect immunofluorescence (IIF) test on human epidermoid laryngeal carcinoma cells (HEp-2 or HEp-2000 cells) is considered the “gold standard” for the detection of ANA [1, 14, 15]. However, as ANA can be present in sera from patients with other rheumatic diseases, in patients with non-rheumatic disorders (e.g. cancer, infection) or in healthy individuals [1, 16–18], the test can have low specificity for CTD. Furthermore, IIF is a labour-intensive technique that requires highly skilled laboratory technicians to interpret the result and is therefore subject to high inter-observer variability [19–21]. Solid-phase immunoassays (IAs) offer an alternative to IIF. IAs have been developed to screen for specific analytes associated with CTD, and fully automated systems can overcome some of the limitations of a manual IIF mentioned earlier.

An enzyme-linked immunosorbent assay (ELISA) is a plate-based assay technique whereby an antigen is immobilized on a solid surface. Autoantibodies bind to the antigen and are complexed with an antibody linked to an enzyme. A generic ELISA detects ANA of a broad specificity by including an extract from HEp-2 cells (which can be complemented by individual autoantigens). Moreover, specific ELISAs are available that react with single autoantigens associated with CTD, such as dsDNA, SS-A/Ro, SS-B/La, Scl-70, Sm and Sm/RNP. The format of the ELISA can be modified to detect the antigen directly via a primary antibody or indirectly via a secondary antibody.

Another method is the solid-phase fluorescence enzyme immunoassay (FEIA) that is designed as a sandwich IA whereby the analyte to be measured is ‘sandwiched’ between an autoantigen coated to the solid phase and a detection antibody that is linked to an enzyme that produces a fluorescence signal. A commercially available automated FEIA test for CTD (EliA CTD Screen, Thermo

Fisher Scientific) is coated with 15 antigens that are associated with CTDs (dsDNA, SSA/Ro 60 kDa, SSA/Ro 52 kDa, SSB/La, U1-RNP (RNP-70, A,C), Sm, Jo-1, Scl-70, centromere B, fibrillarin, RNA Pol III, PM-Scl, Mi-2, Rib-P and PCNA). The fluorescence of the reaction is measured automatically, and the higher the fluorescence intensity, the higher the antibody level in the sample.

A variation on this is the chemiluminescent immunoassay (CLIA), where the enzymes linked to the detection antibody produce a luminescence via a chemical reaction. Automated CLIA tests for CTD provide a qualitative determination of autoantibodies directed against clinically relevant autoantigens (LIAISON ANA Screen, DiaSorin, includes the following autoantigens: SS-A [Ro], RNP/Sm, SS-B [La], Scl-70, Jo-1, CENP-B, mitochondria, dsDNA, HEp-2; QUANTA Flash CTD Screen, Inova, includes 15 autoantigens: dsDNA, Sm/RNP, Ro52, Ro60, SS-B, Scl-70, centromere, Mi-2, Ku, Th/To, RNA Pol III, Pm/Scl, PCNA, Jo-1, Rib-P).

A further method is the multiplex immunoassay (MIA) that provides the simultaneous detection and semi-quantitative assessment of autoantigens using dyed beads. Automated MIA systems with specific antigens are available for screening of CTD (AtheNA Multi-Lyte ANA-II/III Plus, Zeus, includes beads coated with nine analytes [SSA, SSB, Sm, RNP, Scl-70, Jo-1, centromere B, histone, HEp-2]; BioPlex 2200 ANA Screen, Bio-Rad, includes 13 analytes [dsDNA, chromatin, Rib-P, SS-A 60, SS-A 52, SS-B, Sm, RNP/Sm, RNP A, RNP 68, Scl-70, Jo-1, centromere]).

To the best of our knowledge, there has been no systematic assessment of the diagnostic accuracy of different IAs vs. IIF to screen for ANA as an initial step towards diagnosing a CTD. A recent review has been published examining the diagnostic accuracy of two solid-phase assays (SPAs) vs. IIF based on data from seven studies [22]. However, this publication did not use meta-analysis methods to combine the data across studies and the results presented in the paper were reported for the two SPAs combined. A previous publication compared the diagnostic test accuracy of FEIA against IIF as a single test and as a double test strategy [23]. We set out to extend this analysis and assess diagnostic test accuracy for a range of IA techniques vs. IIF for ANA screening as an initial step in the diagnosis of a CTD.

To this end, we had two key objectives. The first was to conduct a comprehensive systematic literature review to identify all published studies evaluating ELISA, FEIA, CLIA or MIA vs. IIF as an initial screening test for CTD, to assess the study quality and to provide an overview of the diagnostic test accuracy data reported in these studies. A second objective was to combine the available

diagnostic test accuracy data in a meta-analysis using a robust statistical method, to provide a direct comparison of the sensitivity and specificity of the different IAs vs. IIF for screening of CTD. Our review provides a better understanding of the available evidence in support of different ANA tests for CTD screening, as well as a formal and robust meta-analysis that allows for the average diagnostic accuracy and variation in test performance to be quantified.

Materials and methods

Systematic literature review process

A structured literature search and systematic literature review was conducted as per the Cochrane Collaboration recommendations for a review of diagnostic test accuracy studies [24]. The search strategy combined search filters for CTD, index tests and diagnostic accuracy test studies using Emtree/Medical Subject Headings (MeSH) terms and free text strings. An electronic search using these filters was conducted using MEDLINE, Embase (see Supplementary Material, Table S-5) and Cochrane databases (from 2000 to March 2018) along with handsearching to identify fully paired, cross-sectional or case-control studies of ANA screening of CTD where the study reported diagnostic test accuracy for an IA of interest and IIF.

All citations retrieved from the electronic search and handsearching were imported into a reference manager (EndNote X8) for screening by two reviewers (MEO, MDO). The initial screening was based on the citation title and abstract, with a second screen using full-text papers to confirm the eligibility of the study for inclusion in the systematic review. The literature search citation flow is reported as per the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement [25].

Inclusion criteria

Study design: Included studies were observational cross-sectional or cohort studies of diagnostic test accuracy. Only fully paired studies were included in the meta-analysis, i.e. studies needed to report results for an IIF and at least one IA using the same cohort of patients.

Population: To be included in the analysis, study populations needed to include a CTD cohort with a range of CTD conditions and a non-CTD diseased control (DC) group (another relevant disease) to reflect the type of patients who may be referred for ANA testing in practice.

The conditions on the CTD spectrum that are associated with the presence of ANA and are included in the CTD group include systemic lupus erythematosus (SLE) incorporating sub-acute cutaneous lupus erythematosus (ScLE); Sjögren's syndrome (SjS); systemic sclerosis (SSc) including limited scleroderma (lim SD); inflammatory myopathies (IM) such as dermatomyositis (DM) and polymyositis (PM); mixed connective tissue disease (MCTD) and undifferentiated connective tissue disease (UCTD). If a study included patients with other rheumatic diseases (e.g. rheumatoid arthritis [RA]) but included these

results in the CTD cohort, then data were adjusted to account for these patients in the DC group instead. Studies that investigated one type of CTD, e.g. SLE patients only, and studies that did not include both a CTD and a control group were excluded as this does not reflect the spectrum of patients referred for ANA testing in practice. If a study included healthy controls (HCs) as part of the DC group, then these patients were excluded from the analysis, wherever this was feasible (see Table S-1). Studies that included only 100% healthy individuals as controls were excluded from the review as test specificity in healthy patients will differ from that in diseased patients and this group does not reflect the spectrum of patients referred for ANA testing in practice. For studies that reported data pre- and post-diagnosis, data for the pre-diagnosis samples were used wherever feasible.

Index tests: All studies needed to report test performance data for both an IIF and an IA method of interest, namely an ELISA, FEIA, CLIA or MIA. Tests that are not available for use in practice or tests that have been discontinued were not included (examples of excluded tests are Bindazyme, Diastat, Varelisa, Synelisa, COBAS Core HEP-2 ANA-EIA). The cut-off for a positive ANA test can vary across studies and, for some studies, results are reported for more than one cut-off. The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [24] asserts that “estimating summary sensitivity and specificity by pooling studies which mix thresholds will produce an estimate that relates to some notional unspecified average of the thresholds that occur in the included studies, which is clinically unhelpful and must be avoided”. To ensure comparisons reflect practice, and to avoid estimating test performance at some nominal ‘average’, the intention was to conduct the analysis using data for IIF at a cut-off of 1:160 as per international recommendations [1], or at 1:80, which is the new European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) entry criterion for SLE classification [26, 27]. Studies that only report results for IIF at other cut-off levels were not included. For the IAs, the cut-off is as per the manufacturer's recommendations for use in practice, the cut-off for FEIA being >1 . Please see our previous paper for a summary of data for IIF and FEIA including other thresholds (1:320 for IIF and 0.7 for FEIA) [23].

Reference standard: The reference standard in this review is the clinical follow-up used to definitely confirm a final diagnosis of a CTD, or definitely rule out a CTD, regardless of whether ANA were detected by the index screening test or not. To be included in the review, a study needed to include a reference standard whereby all patients had their diagnostic status confirmed (i.e. CTD or not CTD). Studies reporting concurrence between tests where the patients' diagnostic status was unknown were not included.

Outcome data: Each study needed to report the number of true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs) for each test. Alternatively, the study needed to report the sensitivity and specificity of each test along with the number of patients in the CTD and DC cohorts such that test count data can be replicated from this information.

Quality assessment

The study quality assessment was adapted from the QUADAS-2 checklist [28] to assess the quality of each study in relation to patient

selection, attrition, flow and timing of the tests, and conduct and interpretation of the index tests and reference standard. As part of the assessment, the reference standard was compared to the most recent clinically accepted diagnostic criteria for CTD classification as follows: SLE: 1997 ACR criteria [2] or 2012 Systemic Lupus International Collaborating Clinics criteria [3]; SjS: 2016 ACR/EULAR criteria [4] or 2012 ACR/EULAR criteria [5]; SSC: 2013 ACR/EULAR criteria [6]; PM/DM: Bohan and Peter [7, 8], Dalakas and Hohlfeld's criteria 2003 [9] or European Neuromuscular Centre criteria 2004 [10]; MCTD/UCTD: Alarcón-Segovia and Villarreal [11], Kasukawa et al. [12] or Sharp and Anderson [13]. The quality of the reference standard was graded A–E. Where the diagnosis/classification of CTD is based on the most recent disease-specific guidelines or classification criteria available at the time of the review as listed earlier, it was graded A. Grade B was used for disease-specific classification criteria that have been superseded by more recent guidelines [29–32]; grade C, where some clinical criteria and most relevant immunological criteria were used (e.g. authors indicated that disease-specific classification criteria were used but did not provide references); grade D, where some relevant clinical criteria were used (e.g. authors indicated that they used some formal criteria but did not provide references for the criteria) and grade E, referring to a reference standard that is not described with sufficient detail in the publication.

Study data summary estimates

The sensitivity of a test is defined as the probability that the index test result will be positive in a patient with CTD. The specificity of a test is defined as the probability that the index test result will be negative in non-CTD DCs. For each study, the sensitivity and specificity of each test were calculated and the 95% confidence interval (CI) around the sensitivity and specificity estimates was calculated using the exact binomial method [33]. The diagnostic odds ratio (DOR) is a summary estimate of how many times higher the odds are of obtaining a positive test result in a diseased rather than a non-diseased person. DOR can be a useful measure when comparing tests if there is no preference for either superior sensitivity or specificity and the focus is on global performance [34]. If the DOR is less than one, then the test is uninformative and is of no clinical value.

In order to summarise all of the available data at all reported test thresholds, a hierarchical summary receiver operating characteristic (HSROC) curve was produced for each index test to provide an overall summary of the diagnostic test accuracy data. A HSROC model [35] is fitted to the study data for each test using the *metandi* package in STATA MP v14.2 [36]. The HSROC model is used to estimate two parameters, test accuracy (lnDOR) and asymmetry (change in DOR relative to change in the test threshold/sensitivity). The estimates from the HSROC model are used to plot a summary ROC curve of sensitivity vs. specificity (expressed as $1 - \text{specificity}$), the 95% confidence region around this summary estimate and a 95% prediction region taking into account unobserved heterogeneity: if a new study was conducted, we would expect the 'true' sensitivity and specificity to lie within the prediction region with a 95% confidence level [24, 36]. The prediction region can be wider than the confidence region as it goes beyond the uncertainty in the available data [34]. The results of this summary analysis are reported in the section Summary of study estimates of diagnostic accuracy by test and Figure 1.

Comparative meta-analysis methods

The meta-analysis was conducted using hierarchical, bivariate, mixed-effect models as recommended in the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [24] and was conducted in STATA MP v14.2 using the *meqrlogit* function [37]. The hierarchical mixed-effect model applies statistical distributions at two levels: at a study level to account for variation within studies (differences between patients) and at a higher level to account for variation between studies. At the study level, the sensitivity and specificity are estimated directly from the TP, TN, FP and FN counts, and assumes that sensitivity and specificity are binomially distributed [38, 39]. Furthermore, the model assumes a correlation between sensitivity and specificity modelled as a single bivariate normal distribution. This correlation models the expected trade-off between sensitivity and specificity (an increase in sensitivity is usually associated with a decrease in specificity). The mixed-effect model includes test-specific sensitivity and specificity (via dummy covariates for the test type [37]) as well as test-specific random-effects, i.e. the model has separate variance estimates for ELISAs, FEIA, MIA, CLIA and IIF to account for the variability in test assays within the test method groups.

In addition to the aforementioned summary estimates, the bivariate meta-analysis estimates the likelihood ratio (LR) which is the probability that a given test result is obtained in the CTD patients compared to the probability of the same results in the controls. The positive likelihood ratio (LR+) describes how many times more likely positive index test results were in the diseased group compared to the non-diseased group. The negative likelihood ratio (LR–) summarises how many times less likely negative index test results were in the diseased group compared to the non-diseased group. In order for the summary estimates to be clinically meaningful, the bivariate meta-analysis has been limited to studies where the test data are reported at a common cut-off such that the results provide an average operating point [24] as well as a 95% CI. The statistical significance of differences between tests is based on the p-value estimated from a two-sided t-test and statistically significant differences are defined as a p-value <0.05.

For the comparative meta-analysis, we conducted separate analyses using data for IIF at thresholds of 1:80 and 1:160.

Results

Summary of studies included in the review

The literature review (up to March 2018) identified 17 studies (from 18 citations [40–57]) that met the inclusion criteria for this review (set out in the section Inclusion criteria) (see PRISMA diagram in Supplementary Material, Figure S-1). A summary of the studies is provided in Table S-1. The 17 studies incorporated 17,201 patients, 13.6% of which had a confirmed diagnosis of a CTD. Three out of the 17 studies [42, 43, 47] had a control group that included some HCs (386 patients, 2.6% of the 14,864 control patients). All 17 studies [40–51, 53–57] reported

Figure 1: HSROC graph with 95% prediction/confidence region.

Top panels: IIF at a cut-off of 1:80 or 1:160 (left) and ELISA all types (right). Middle panels: FEIA at a cut-off of >1 (left) and CLIA (right). Bottom left panel: MIA. The size of the circle corresponds to the size of the study cohort. Prediction region: the 'true' sensitivity and specificity of a new study will lie within this region with a 95% confidence level. Diagnostic odds ratio (DOR) is the ratio of the odds of obtaining a positive result in a diseased patient over the odds of a positive test in a non-diseased person. For the region where $DOR < 1$, the test is uninformative and is of no clinical value. Note that the two types of assay in the CLIA and MIA sub-graphs are CLIA/MIA without HEp-2 (hollow circle) or CLIA/MIA with HEp-2 (coloured circle). Separate HSROC plots of ELISAs with and without HEp-2 are shown in Supplementary Material, Figure S-2.

diagnostic test accuracy data for IIF (by design as all studies were fully paired IA vs. IIF studies), eight different ELISA tests were included in six studies [43, 45–48, 54], 10 studies [40, 41, 44, 49–51, 53, 55–57] reported data for FEIA, four studies [42, 44, 45, 56] for two types of CLIA and three studies [46, 52, 54] for two types of MIA. Table 1 summarises the tests included across the 17 studies identified in the review.

Summary of the quality of the studies included in the review

Overall the quality assessment indicated that the studies were of sufficient quality in terms of patient selection and participant flow (see Supplementary Material, Table S-2), and index test and reference standard conduct (see Supplementary Material, Table S-3). Six studies included a cross-section of all patients referred for ANA testing [48, 53–57], which is more representative of the test population in practice. Some of the publications reported limited information on study methodology such that the quality could not be assessed (judged as an unclear risk of bias). In particular, there was a lack of information regarding the reference standard: for nine out of the 17 studies, the reference standard was graded E [40, 41, 45–47, 50, 53–55].

Summary of study estimates of diagnostic accuracy by test

By design, to be included in the review, all studies must be fully paired and report data for one of the IAs as well as data for IIF at a cut-off of 1:80 or 1:160. Therefore, all 17 studies [40–51, 53–57] contributed diagnostic test accuracy data for IIF (see Supplementary Material, Table S-4). Thirteen studies [40, 42–47, 50, 51, 53, 54, 56, 57] reported data at a cut-off of 1:80, and eight studies [40, 41, 48–51, 53, 55] reported data at a cut-off of 1:160. Figure 1, top left panel, is a plot of the study estimates for IIF sensitivity vs. specificity at these thresholds (one circle per study estimate, where the size of the circle corresponds to the size of the study cohort). Whilst the sensitivity of the IIF test for CTD is good in most studies, there is a large variance in specificity across the studies. The HSROC curve (blue line) shows that an increase in IIF test sensitivity results in a marked decrease in the specificity. The prediction region (grey dashed line) indicates a high level of uncertainty in the estimates for IIF specificity: given the available data, this is the region where the ‘true’ sensitivity and specificity of a new study is expected to lie with a 95% confidence level.

The HSROC plot of the study estimates for ELISA sensitivity vs. specificity is similar to IIF: variance in the test results across the studies lead to a high level of uncertainty in the estimates of diagnostic test accuracy (a large prediction region). Figure 1, top right panel, shows 10 ELISA test results (from six studies [43, 45–48, 54]) with a similar association between increased sensitivity and decreased specificity as IIF. A HSROC plot of ELISA test data for ELISAs with and without a Hep-2 component is shown in Supplementary Material, Figure S-2.

The plots for IIF and ELISA can be contrasted with the plot of the study estimates for FEIA (Figure 1, middle left panel). Of the 17 fully paired studies in this review, 10 studies [40, 41, 44, 49–51, 53, 55–57] reported diagnostic test accuracy data for FEIA at the manufacturer’s recommended cut-off of >1 . As the sensitivity of the FEIA for CTD increases, there is only a small decrease in specificity (Figure 1, middle left panel: HSROC, blue line). The prediction region arising from the FEIA diagnostic test accuracy data is smaller than that for IIF and ELISA, indicating that we can be more certain of our estimates for the sensitivity and specificity of FEIA compared to our estimates for IIF or ELISA.

Of the 17 fully paired studies included in this review, four studies [42, 44, 45, 56] reported diagnostic test accuracy data for CLIA (Figure 1, middle right panel). There is a high level of uncertainty in the estimates of diagnostic test accuracy for CLIA given the large confidence and prediction region, and that the 95% prediction region crosses the line of ‘no effect’ ($\text{DOR}=1$). If a new CLIA study was conducted, we would expect the sensitivity and specificity to lie within the prediction region with a 95% confidence level [24, 36]. Given the size of the prediction region, our estimates lack certainty, and therefore the estimates for CLIA should be viewed with caution.

Of the 17 fully paired studies included in this review, three studies [46, 52, 54] reported diagnostic test accuracy data for MIA (Figure 1, bottom left panel). The prediction region is similar to the prediction region for IIF, though for MIA the uncertainty in the estimates could be driven by a lack of data (HSROC analysis is underpowered) whereas the variance in the IIF and ELISA results could be due to variation in test designs, implementation of the test in practice and operator subjectivity.

Meta-analysis of ELISA vs. IIF

Five studies incorporating 2321 patients reported diagnostic test accuracy data for both IIF at a 1:80 dilution and an ELISA [43, 45–47, 54]. Results from the mixed-effect bivariate model (Table 2) showed no significant difference

Table 1: Summary of ANA tests included in the review by antibodies and associated CTD subtypes.

	Method	IIF	FEIA		CLIA		CLIA	MIA	MIA
	Tests included in the bivariate meta-analysis	Various	EliA CTD		QUANTA-flash CTD		LIAISON ANA	BioPlex 2200	AtheNA
	Number of studies	17 Studies [40–57]	10 Studies [40, 41, 44, 49–51, 53, 55–57]		3 Studies [42, 44, 56]		1 Study [45]	2 Studies [46, 52]	1 Study [54]
Antibody	Associated with								
HEp-2	AARD	✓					✓		✓
dsDNA	SLE ^a		✓		✓		✓		
SSA/Ro	SjS ^a , SLE, PM (Ro52), (SSc)		✓60, 52		✓60, 52		✓	✓60, 52	✓
SSB/La	SjS ^b , SLE		✓		✓		✓		✓
Sm	SLE ^a		✓						✓
Sm/RNP	SLE				✓		✓		
RNP	SLE, MCTD ^a		✓70, A, C					✓68, A	✓
Scl-70	SSc ^a		✓		✓		✓		✓
Centromere B	SSc ^a		✓		✓		✓		✓
Fibrillarin	SSc		✓						
RNA-Pol III	SSc ^a		✓		✓				
PM-Scl	DM/PM, SSc, overlap syndrome		✓		✓				
Jo-1	DM/PM ^a		✓		✓		✓		✓
Mi-2	DM ^a , IM		✓		✓				
Rib-P	SLE		✓		✓			✓	
PCNA	SLE		✓		✓				
Ku	CTD, overlap syndrome				✓				
Th/To	SSc				✓				
Mitochondria	PBC						✓		
Chromatin	SLE							✓	
Histone	SLE, drug-induced lupus								✓
HeLa	AARD								
	Method	ELISA	ELISA	ELISA	ELISA	ELISA	ELISA	ELISA	ELISA
	Tests included in the bivariate meta-analysis	ZEUS EIA	QUANTA Lite ANA	Bio-Rad ANA	ORG 600 ANA Detect	Eurolmmun ANA Screen	Relisa ANA	QUANTA Lite ENA 6	IMTEC ANA Screen
	Number of studies	1 Study [43]	2 Studies [45, 48]	2 Studies [46, 48]	1 Study [45]	1 Study [47]	1 Study [48]	1 Study [54]	1 Study [45]
Antibody	Associated with								
HEp-2	AARD	✓	✓	✓					
dsDNA	SLE ^a	✓	✓	✓	✓	✓	✓		
SSA/Ro	SjS ^a , SLE, PM (Ro52), (SSc)	✓	✓	✓60	✓60, 52	✓	✓	✓	
SSB/La	SjS ^b , SLE	✓	✓	✓	✓	✓	✓	✓	
Sm	SLE ^a	✓		✓		✓	✓	✓	
Sm/RNP	SLE	✓	✓	✓	✓	✓	✓		
RNP	SLE, MCTD ^a			✓				✓	
Scl-70	SSc ^a		✓	✓	✓	✓	✓	✓	
Centromere B	SSc ^a		✓	✓	✓	✓	✓		
Fibrillarin	SSc								
RNA-Pol III	SSc ^a								
PM-Scl	DM/PM, SSc, overlap syndrome				✓				

Table 1 (continued)

	Method	ELISA	ELISA	ELISA	ELISA	ELISA	ELISA	ELISA	ELISA
Jo-1	DM/PM ^a	✓	✓	✓	✓	✓	✓	✓	
Mi-2	DM ^a , IM		✓						
Rib-P	SLE		✓	✓		✓			
PCNA	SLE		✓						
Ku	CTD, overlap syndrome								
Th/To	SSc								
Mitochondria	PBC								
Chromatin	SLE								
Histone	SLE, drug induced lupus		✓		✓	✓			
HeLa	AARD								✓

^aClassification marker; ^bformer classification marker. AARD, ANA-associated rheumatic disease; PBC, primary biliary cholangitis.

Table 2: Meta-analysis results from fully paired studies reporting data for an IA and IIF.

Comparison	Any ELISA vs. IIF at a cut-off of 1:80	FEIA vs. IIF at a cut-off of 1:80 ^a	CLIA vs. IIF at a cut-off of 1:80	ELISA with a HEp-2 component vs. IIF at a cut-off of 1:80	FEIA vs. IIF at a cut-off of 1:160 ^a
Number of fully paired studies	5 Studies [43, 45–47, 54]	7 Studies [40, 44, 50, 51, 53, 56, 57]	4 studies [42, 44, 45, 56]	3 Studies [43, 45, 46]	7 Studies [40, 41, 49–51, 53, 55]
Number of patients	n = 2321	n = 12,311	n = 1981	n = 1782	n = 3251
	ELISA	FEIA	CLIA	ELISA	FEIA
	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Sensitivity	90.3% (80.5%, 95.5%)	78.5% (71.4%, 84.1%)	85.9% (64.7%, 95.3%)	93.6% (87.3%, 96.9%)	72.8% (64.2%, 80.1%)
Specificity	56.9% (40.9%, 71.5%)	93.6% (89.9%, 96.0%)	86.1% (78.3%, 91.4%)	54.6% (36.8%, 71.3%)	93.5% (91.1%, 95.3%)
DOR	12.30 (8.01, 18.91)	53.14 (32.66, 86.46)	37.86 (15.05, 95.19)	17.64 (11.62, 26.76)	38.61 (21.89, 68.09)
LR +	2.09 (1.55, 2.83)	12.23 (7.90, 18.95)	6.19 (4.27, 8.97)	2.06 (1.44, 2.94)	11.22 (7.88, 15.96)
LR –	0.17 (0.10, 0.29)	0.23 (0.17, 0.31)	0.16 (0.06, 0.44)	0.12 (0.07, 0.19)	0.29 (0.22, 0.39)
	IIF (1:80)	IIF (1:80)	IIF (1:80)	IIF (1:80)	IIF (1:160)
	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Sensitivity	86.8% (81.8%, 90.6%)	89.1% (84.4%, 92.5%)	89.2% (82.2%, 93.6%)	87.2% (77.6%, 93.1%)	83.2% (75.4%, 88.9%)
Specificity	68.0% (39.5%, 87.4%)	72.4% (62.2%, 80.7%)	70.9% (60.1%, 79.8%)	69.2% (45.1%, 86.1%)	81.1% (73.4%, 86.9%)
DOR	14.05 (4.93, 40.01)	21.44 (17.12, 26.85)	20.11 (15.07, 26.82)	15.34 (9.26, 25.43)	21.23 (12.11, 37.20)
LR +	2.72 (1.24, 5.93)	3.22 (2.39, 4.34)	3.07 (2.30, 4.09)	2.83 (1.52, 5.28)	4.39 (3.12, 6.19)
LR –	0.19 (0.14, 0.28)	0.15 (0.12, 0.20)	0.15 (0.10, 0.23)	0.18 (0.13, 0.26)	0.21 (0.14, 0.31)
	Any ELISA vs. IIF (1:80)	FEIA vs. IIF (1:80)	CLIA vs. IIF (1:80)	ELISA with a HEp-2 component IIF (1:80)	FEIA vs. IIF (1:160)
Difference in sensitivity (95% CI), p-value	3.5% (–4.9%, 11.8%), 0.41	–10.7% (–18.1%, –3.1%), 0.005 ^b	–3.3% (–18.6%, 12.3%), 0.68	6.4% (–2.4%, 15.1%), 0.15	–10.4% (–20.7%, 0.1%), 0.051
Difference in sensitivity (95% CI), p-value	–11.2% (–39.5%, 19.1%), 0.47	21.2% (11.3%, 30.8%), 0.00002 ^b	15.2% (3.2%, 26.8%), 0.01 ^b	–14.7% (–40.9%, 13.8%), 0.31	12.4% (5.4%, 19.4%), 0.0005 ^b

^aResults are reproduced from Orme et al. [23] with permission; ^bstatistically significant difference between the IA vs. IIF defined as a p-value <0.05.

in sensitivity and specificity between IIF and ELISA. The large 95% CI in specificity for both tests indicates that the estimates are subject to uncertainty. The DOR was comparable for ELISA and IIF. Similar results (Table 2) were obtained for a sensitivity analysis restricting the analysis to three studies reporting diagnostic test accuracy data for both IIF at a 1:80 dilution and an ELISA with a HEp-2 component [43, 45, 46]. There were too few studies to conduct a meta-analysis for ELISA vs. IIF at a cut-off of 1:160.

Meta-analysis of FEIA vs. IIF

Two meta-analyses were conducted using a bivariate mixed-effect model and subsets of studies that reported direct comparisons of FEIA and IIF at a cut-off of 1:160 (seven studies [40, 41, 49–51, 53, 55], 3251 tests) and 1:80 (seven studies [40, 44, 50, 51, 53, 56, 57], 12,311 tests). Table 2 shows the sensitivity, specificity, DOR, LR+ and LR– estimated from the mixed-effect model. The sensitivity of FEIA was statistically significantly lower than the sensitivity of IIF at a cut-off of 1:80 ($p=0.005$) and was lower compared to that of IIF at a cut-off of 1:160 ($p=0.051$). FEIA had a significantly higher specificity than IIF at a cut-off of 1:80 ($p<0.0001$) and 1:160 ($p<0.001$) and the DOR was higher with FEIA compared to IIF.

Meta-analysis of CLIA vs. IIF

A meta-analysis was conducted using a bivariate mixed-effect model and data from four studies that reported data for the CLIA method [42, 44, 45, 56]. Three of these studies report data for the same type of CLIA [42, 44, 56] (see Supplementary Material, Table S-4), with one study reporting data for a different CLIA with a HEp-2 extract [45].

There was no significant difference in the sensitivity of CLIA vs. IIF at a cut-off of 1:80 ($p=0.68$, Table 2). CLIA had a significantly higher specificity than IIF at a cut-off of 1:80 ($p=0.01$). It was noted that the model estimate for the 95% CI for CLIA sensitivity and specificity is large indicating that the estimates are subject to uncertainty: more data are needed to determine whether the large CIs are a fair reflection of the actual variance or are due to the analysis being underpowered. Across the four studies, the reported CLIA sensitivity for CTD was as high as 98.6% (for a CLIA without a HEp-2 component [56]) and as low as 62.9% (for a CLIA with a HEp-2 component [45]), with specificity ranging from 76% [56] to 94% [42] (see Supplementary Material, Table S-4). Given the size of the prediction region, and the lack of a clinical rationale for the variance in the

results, the aforementioned estimates for CLIA should be viewed with caution. There were too few studies to conduct a meta-analysis for CLIA vs. IIF at a cut-off of 1:160.

Meta-analysis of MIA vs. IIF

There were too few studies to conduct a robust meta-analysis using the bivariate mixed-effect model. Across the three studies that did report data [46, 52, 54], the sensitivity of MIA for CTD was as high as 86% [46] and as low as 73.7% [52] with specificity ranging from 53% [54] to 91% [52] (see Supplementary Material, Table S-4).

Pre-test vs. post-test probability for IIF and IAs

Eleven [40–47, 49–51] out of the 17 fully paired studies included in the review have a case-control design, and the ratio of CTD patients to non-CTD patients does not reflect the prevalence of CTD in a clinical setting. The largest prospective cross-sectional study included in this review tested 9856 consecutive patient sera submitted to the clinical laboratory for ANA testing [57]. The prevalence of CTD in the study population was estimated to be 2.7% (267/9856) including 22 cutaneous lupus patients (or 2.5% not including these patients). In 62 patients, the clinician strongly considered the presence of an ANA-associated systemic rheumatic disease (and started treatment), but the patients did not fulfil the diagnostic criteria. If these cases are included as CTD patients, then the prevalence was estimated to be 3.9% or 4.1% including cutaneous lupus. Based on the average operating point estimates from bivariate meta-analysis shown in the previous sections, the pre-test vs. post-test probability of CTD is shown in Figure 2 (after a positive test) and Figure 3 (after a negative test), assuming that the prevalence (pre-test probability) of CTD is in the range of 0–5%. It should be noted that for individual patients the pre-test probability can be higher if typical clinical signs are overt.

Figure 2 indicates that a positive IIF test is more likely to indicate CTD, than a positive ELISA test. Assuming an underlying CTD prevalence of 2.7% and using the average operating point estimates for sensitivity and specificity from Table 2, for every 1000 patients tested, ELISA correctly identifies 24 out of 27 patients with CTD, and 554 out of 973 patients without CTD (57.8% correctly identified). IIF at a cut-off of 1:80 has a similar sensitivity (23 out of 27 with CTD) but more TNs (662 out of 973 without

Figure 2: Post-test probability of CTD as a function of pre-test probability of CTD following a positive test result.

CTD) such that 68.5% of patients are correctly identified by the IIF test. The post-test probability of CTD following a positive IIF (1:80) test is 7.0% (23 TPs out of 335 positive ANA tests) and 5.5% (24 TPs out of 335 positive ANA tests) following a positive ELISA.

Figure 2 indicates that a positive FEIA test is more likely to rule in a CTD than IIF at a cut-off of 1:80 (based on the average operating point estimates from Table 2). For every 1000 patients tested and a background prevalence of 2.7%, FEIA correctly identifies 21 out of 27 patients with CTD, and 911 out of 973 patients without CTD (93.2% correctly identified). The post-test probability of CTD following a positive FEIA test is 25.4% (21 TPs out of 83 positive ANA tests).

Based on the average operating point estimates from Table 2, and given that these results should be viewed with some caution, Figure 2 shows that a positive CLIA test is also more likely to rule in a CTD than IIF at a cut-off

of 1:80. For every 1000 patients tested and a background prevalence of 2.7%, CLIA correctly identifies 23 out of 27 patients with CTD, and 838 patients out of 973 patients without CTD (86.1% correctly identified). The post-test probability of CTD following a positive CLIA test is 14.6% (23 TPs out of 158 positive ANA tests).

Assuming a hypothetical sensitivity of 81.9% and specificity of 74.6% for the MIA method (based on data from three studies [46, 52, 54]), then Figure 2 shows that a positive MIA test is similar to an IIF at a cut-off of 1:80 for ruling in CTD. For every 1000 patients tested and a background prevalence of 2.7%, MIA may correctly identify 22 out of 27 patients with CTD, and 727 out of 973 patients without CTD (74.9% correctly identified). The post-test probability of CTD following a positive MIA test is 8.2% (22 TPs out of 269 positive ANA tests).

Figure 3 indicates that the diagnostic value of a negative FEIA and CLIA test is similar to a negative IIF (1:80) test if the prevalence of CTD is low. At a prevalence rate of 2.7%, the post-test probability of CTD following a negative IIF (1:80) test is 0.54% (four FN tests out of 665 negative tests), 0.47% (three FN tests out of 556 negative tests) following a negative ELISA, 0.63% (six FN tests out of 917 negative tests) following a negative FEIA and 0.45% (four FN tests out of 842 negative tests) following a negative CLIA. Based on the average hypothetical sensitivity and specificity for MIA stated earlier, the post-test probability of CTD following a negative MIA test is estimated to be 0.67% (five FN tests out of 731 negative tests). It should be noted that the differences between the different assays become more pronounced at higher pre-test probabilities.

Discussion

The aim was to provide direct comparisons of the sensitivity and specificity of different IAs vs. IIF for the initial screening of CTD using up-to-date evidence from published diagnostic test accuracy studies. All of the studies included in the review reported diagnostic test accuracy of an ANA test for detecting CTD, that is, the studies reported the number of patients who tested positive or negative for ANA for each assay, and the number of those patients confirmed to have a CTD or not. For FEIA vs. IIF, some studies also reported the number of patients who tested positive or negative for ANA by CTD disease type such that a disease specific sub-group analysis could be conducted. This analysis has been published elsewhere [23]. For the other IAs considered in our review, disease-specific test performance data were reported by very few studies [44,

Figure 3: Post-test probability of CTD as a function of pre-test probability of CTD following a negative test result.

45, 51] such that a meta-analysis by CTD subtype was not possible. The meta-analysis results are generalisable to the initial screening test to rule in or rule out a CTD. Our review and meta-analysis will not be generalisable to the latter stages in the diagnostic pathway and does not report the sensitivity and specificity of ANA tests in the diagnosis of specific diseases such as SSs, SLE, SjS, DM, PM and MCTD.

There were sufficient data to conduct a meta-analysis to assess the diagnostic accuracy of ELISA (various) vs. IIF at a cut-off of 1:80, FEIA (one type) vs. IIF at a cut-off of 1:80 and 1:160 and CLIA (two types) vs. IIF at a cut-off of 1:80 though the latter analysis should be viewed with caution as more data are required to increase the precision of the estimates. The meta-analysis showed no significant difference in sensitivity and specificity between IIF at a cut-off of 1:80 and ELISA for detecting or ruling out CTD ($p=0.8$, $p=0.7$, respectively). Some of the ELISA tests included in the meta-analysis were generic ANA tests that used an extract of the HEp-2 cell line: a meta-analysis limited to data for these tests also showed no significant differences compared to IIF at 1:80. We were unable to conduct a meta-analysis limited to ELISA tests without a HEp-2 component, though we note that the reported sensitivity of an ELISA without HEp-2 is more wide ranging than an ELISA with HEp-2 (see Figure S-2 [right and left panel, respectively]). The 95% CI around the point estimates and the prediction region around the HSROC curve indicate a large variance in results for both IIF and ELISA, particularly for specificity (Figure 1). The variance in test results may be due to differences in the underlying test population across studies, but as this variance was not seen in the FEIA HSROC (and given that the analysis included only fully paired studies), we

assert that this variance is likely to be due to differences in the conduct and interpretation of the IIF tests. IIF and ELISA tests have the lowest positive predictive value: if the pre-test probability of CTD is 2.7%, the post-test probability of CTD after a positive test is estimated to be 7.0% for IIF at a cut-off of 1:80 and 5.5% for ELISA, compared to 25.4% for FEIA, and 14.6% for CLIA (and ~8.2% for MIA based on available data). A positive ANA screening test will be followed up with additional laboratory workup and unnecessary costs that may include a second ANA test and tests for specific antibodies [1]. In one study that followed up 96 patients with laboratory requests for ANA screening, it was observed that a positive HEp-2 test result generated an average of 4.11 follow-up tests [58]. Based on the average FP rate found in our analysis, this will be translated into 1724 unnecessary follow-up tests per 1000 suspected patients tested with ELISA HEp-2 vs. 1280 when IIF is used as the screening test. Moreover, FP results may be misinterpreted by physicians not familiar with systemic CTDs and lead to unnecessary treatments [17, 18, 59].

The manual IIF method and ELISA tests require multiple stages of quality control oversight, and proficient laboratory technicians. Fully automated tests can simplify the process, are less hands-on and provide more consistency across laboratories. The meta-analysis showed that FEIA and CLIA have significantly better specificity than IIF ($p<0.05$), though the estimate for FEIA was supported by more data ($n=12,311$ vs. $n=1981$ for the CLIA analysis) and had narrower 95% CIs (FEIA [95% CI: 89.9%, 96.0%]; CLIA [95% CI: 78.3%, 91.4%]). Based on the HSROC curve for FEIA (Figure 1, middle left panel), we can be 95% confident that the ‘true’ FEIA specificity for CTD lies within the narrow range of values indicated by the prediction region.

The ‘true’ specificity for the other tests is likely to vary more in practice, given the large prediction region.

A good ANA screening assay should be sensitive and exclude CTD in the case of negativity. A recent study in patients with established SLE [60] showed a variance in the frequency of ANA-negative tests across three different IIF (4.9%–22.3%) as well as an ELISA (11.7%) and multiplex bead-based assay (13.6%).

Recently, it has been proposed that combining IIF with FEIA could increase the diagnostic accuracy overall [15, 44, 57, 59, 61, 62]. The current recommendations are to perform an IIF subsequent to a negative IA test (e.g. FEIA, CLIA) or an IA subsequent to a negative IIF test if there is a high clinical suspicion of a CTD [1]. Our analysis supports this recommendation (see Figure 3, post-test probability of CTD after a negative test). A previous assessment of a double-test strategy [23] based on data from four studies [41, 49, 55, 57] showed that concordant IIF and FEIA results correctly classify 96.8% of patients, and where there is a discrepancy in the test results, a positive FEIA/negative IIF result is more likely to occur in a patient with CTD than a negative FEIA/positive IIF result (LR 2.4 vs. 1.4 [23]). The review by Bizzaro [22] which examined the diagnostic accuracy of two SPAs vs. IIF agreed that neither of the two methods alone would identify all patients with CTD and the best diagnostic strategy could be a combination of the two methods.

The comprehensive systematic literature review identified relevant published studies, and the quality review assessed potential bias arising from the study design, conduct and interpretation of results. The recent independent review by Bizzaro [22] identified seven studies, four of which reported diagnostic test data for CLIA and IIF [42, 44, 56, 62] and six for FEIA and IIF [44, 51, 55, 56, 62, 63]. We used four studies for the analysis of CLIA vs. IIF [42, 44, 45, 56] and 10 for FEIA vs. IIF [40, 41, 44, 49–51, 53, 55–57]. Two of the studies reported in the Bizzaro review were not used in our meta-analysis. For one of the studies this was because the IIF cut-off was neither 1:80 nor 1:160 (cut-off was 1:100) [63]. The other study was conducted after the cut-off date for our literature search [62]; however, the publication does not report diagnostic test data for FEIA and CLIA separately so would not be eligible for inclusion in our analysis.

For the analysis of FEIA vs. IIF, we included five studies where the data were from conference abstracts or posters [40, 41, 49, 50, 53] such that full details of the study methodology were not available in a full-text publication. There were too few studies to conduct a comparative meta-analysis of FEIA vs. IIF using full-text publications only. For the five studies where the data were

from conference abstracts/posters [40, 41, 49, 50, 53], the average FEIA sensitivity was 75.4% and specificity 92.1%. However, we noted that for the four full-text publications used in the meta-analysis of FEIA vs. IIF at a cut-off of 1:80 [44, 51, 56, 57], the average FEIA sensitivity and specificity were higher at 78.5% and 93.4%, respectively. Similarly, for the two full-text publications used in the meta-analysis of FEIA vs. IIF at a cut-off of 1:160 [51, 55], the average FEIA sensitivity was similar to the average from the conference abstracts (75.9%) but the average specificity was higher (95.1%). Therefore, whilst we were unable to conduct a thorough assessment of the study quality, inclusion of data from conference abstracts and posters has not overestimated the diagnostic test accuracy estimates for FEIA.

One alternative to the comparative meta-analysis model that we have used is to include an additional independent variable to allow data at different IIF thresholds to be included in the analysis (cf. Leuchten et al. [14]). However, for our dataset this would mean that the models comparing the different IAs vs. IIF would differ as there were not enough data for some of the tests to allow such an analysis to be performed. A minimum of four studies are required for a random-effects bivariate meta-analysis of diagnostic test accuracy data with a binary covariate for the test and separate random-effects by test.

For MIA vs. IIF there were too few studies to conduct a robust meta-analysis and we have provided hypothetical estimates for post-test probability of CTD only. The meta-analysis of CLIA vs. IIF is based on data from four studies; however, the large 95% CIs generated for the CLIA vs. IIF meta-analysis indicate a high level of imprecision in these estimates. There are too few studies to assess whether the variance is driven by the type of CLIA. The study reporting the lowest sensitivity for CLIA (62.9%) used a CLIA test that included a HEp-2 component [45] which is unexpected given the good sensitivity reported for HEp-2 by IIF tests across the 17 studies included in this review. Further diagnostic test accuracy studies are needed to allow for a more robust analysis of CLIA vs. IIF. Whilst using fully paired data is a key advantage for the meta-analysis that we have conducted, it does not account for correlations between tests applied to the same individual.

Most of the studies used a case-control design that may be simpler to conduct compared to a prospective cross-sectional study using unselected patients referred for ANA testing in the clinic which would be more representative of the disease in a clinical setting. The included studies had CTD cohorts with a representative range of CTD subtypes (SLE, SjS, SSc, MCTD and DM/PM). Specificity was calculated from a cohort of (non-CTD) DCs excluding healthy patients wherever feasible and studies without a

representative disease control were excluded. ANA screening tests are used to support diagnosis and ANA levels can change with treatment. In eight of the 17 studies there was no information reported as to whether the sera were sampled before diagnosis or after treatment had been initiated. However, it should be noted that population selection bias would impact all test results within a study, as the analysis included only fully paired studies and the hierarchical meta-analysis grouped test data by study. For 10 studies there was no or a limited description of the reference standard used to confirm the diagnosis of a CTD (see Supplementary Material, Table S-1). There were too few studies to perform sensitivity analyses excluding studies where the patient status is unknown or reference standard is unclear. The diagnosis/classification of CTD is based on the criteria available at the time of the study and it is noted that new criteria for SLE were published in 2019 [26, 27]. Based on the validation cohort, the new SLE criteria appear to have similar sensitivity to the previous criteria [3] for ruling in SLE but better specificity for ruling out SLE. As the 2019 criteria are yet to be implemented in clinical practice, a reference standard grade 'A' reflects the best quality reference standard available at the time of this review. It was also noted that the studies used a mix of diagnostic criteria and classification criteria to define the group of patients used as cases in the study, and that diagnostic criteria may be a better entry option as it includes all available information, rather than a short list of specific criteria.

To the best of our knowledge, this is the first meta-analysis to use a bivariate statistical model to integrate diagnostic test accuracy data and to compare IIF with different IAs in the context of ANA screening as an initial step to diagnosing a CTD. A hierarchical bivariate mixed-effect model is a statistically valid method that produces unbiased estimates of the average sensitivity and specificity for each test method, as well as the expected variance around these estimates by controlling for within- and between-study differences [24]. As automated tests produce more consistent results across different laboratories compared to manual methods, our meta-analysis model included different random-effects estimators for each type of test to account for variances within each method. The use of a hierarchical model structure avoids bias from simply averaging data across studies and the bivariate model allows for correlations between sensitivity and specificity. Furthermore, the use of fully paired data allows for direct comparisons between IA and IIFs to be made [24]. Whilst it is known that IIF and ELISA methods vary, this meta-analysis quantifies the extent to which the sensitivity and specificity are likely to vary in clinical practice. The 95% prediction regions in the HSROC graphs provide a visual

representation of this variance which helps with the interpretation of the performance of the different assays as initial screening tests for CTD.

In conclusion, this meta-analysis demonstrated that there are differences in diagnostic performance between IAs and has quantified the extent of the variation in diagnostic accuracy for IIF and ELISAs which is likely to be due to different assay setups and test outputs, as well as a lack of standardisation for interpretation of results. FEIA and, to a lesser extent, CLIA have a higher specificity and a higher LR+ than IIF whereas ELISAs are expected to have similar accuracy to IIF. A positive test result with FEIA or CLIA is therefore useful to support the diagnosis of a CTD. IIF has a higher sensitivity and a lower LR- than FEIA or CLIA. A negative IIF test, therefore, is useful to exclude a CTD. Consequently, the most favourable strategy could be to combine a highly sensitive test such as IIF with a highly specific test such as FEIA or CLIA.

Acknowledgements: Mark Orme (ICERA Consulting Ltd) aided with the systematic review, drafting and proof reading of the article.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: Authors state no conflict of interest.

Conflict of interest: MEO is the director of ICERA Consulting Ltd, and was paid by Thermo Fisher Scientific to conduct the systematic literature review and analysis. XB has been paid for consultancy related to this manuscript by Thermo Fisher Scientific. CA and SS are employees of Thermo Fisher Scientific.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

References

1. Agmon-Levin N, Damoiseaux J, Kallenberg C, Sack U, Witte T, Herold M, et al. International recommendations for the assessment of autoantibodies to cellular antigens referred to as anti-nuclear antibodies. *Ann Rheum Dis* 2014;73:17–23.
2. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
3. Petri M, Orbai AM, Alarcon GS, Gordon C, Merrill JT, Fortin PR, et al. Derivation and validation of the Systemic Lupus

- International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum* 2012;64:2677–86.
4. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Primary Sjogren's Syndrome: a consensus and data-driven methodology involving three international patient cohorts. *Arthritis Rheumatol* 2017;69:35–45.
 5. Shiboski SC, Shiboski CH, Criswell L, Baer A, Challacombe S, Lanfranchi H, et al. American College of Rheumatology classification criteria for Sjogren's syndrome: a data-driven, expert consensus approach in the Sjogren's International Collaborative Clinical Alliance cohort. *Arthritis Care Res (Hoboken)* 2012;64:475–87.
 6. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum* 2013;65:2737–47.
 7. Bohan A, Peter JB. Polymyositis and dermatomyositis (second of two parts). *N Engl J Med* 1975;292:403–7.
 8. Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). *N Engl J Med* 1975;292:344–7.
 9. Dalakas MC, Hohlfield R. Polymyositis and dermatomyositis. *Lancet* 2003;362:971–82.
 10. Hoogendijk JE, Amato AA, Lecky BR, Choy EH, Lundberg IE, Rose MR, et al. 119th ENMC international workshop: trial design in adult idiopathic inflammatory myopathies, with the exception of inclusion body myositis, 10–12 October 2003, Naarden, The Netherlands. *Neuromuscul Disord* 2004;14:337–45.
 11. Alarcón-Segovia D, Villarreal M. Classification and diagnostic criteria for mixed connective tissue disease. In: Kasukawa R, Sharp G, editors. *Mixed Connective tissue disease and antinuclear antibodies*. Amsterdam: Elsevier, 1987:33–40.
 12. Kasukawa R, Tojo T, Miyawaki S, Yoshida H, Tanimoto K, Nobunaga M, et al. Preliminary diagnostic criteria for classification of mixed connective tissue disease. In: Kasukawa R, Sharp G, editors. *Mixed connective tissue disease and antinuclear antibodies*. Amsterdam: Elsevier, 1987:41–7.
 13. Sharp GC, Anderson PC. Current concepts in the classification of connective tissue diseases. Overlap syndromes and mixed connective tissue disease (MCTD). *J Am Acad Dermatol* 1980;2:269–79.
 14. Leuchten N, Hoyer A, Brinks R, Schoels M, Schneider M, Smolen J, et al. Performance of antinuclear antibodies for classifying systemic lupus erythematosus: a systematic literature review and meta-regression of diagnostic data. *Arthritis Care Res (Hoboken)* 2018;70:428–38.
 15. Perez D, Gilburd B, Azoulay D, Shovman O, Bizzaro N, Shoenfeld Y. Antinuclear antibodies: is the indirect immunofluorescence still the gold standard or should be replaced by solid phase assays? *Autoimmun Rev* 2018;17:548–52.
 16. Abeles AM, Abeles M. The clinical utility of a positive antinuclear antibody test result. *Am J Med* 2013;126:342–8.
 17. Avery TY, van de Cruys M, Austen J, Stals F, Damoiseaux JG. Anti-nuclear antibodies in daily clinical practice: prevalence in primary, secondary, and tertiary care. *J Immunol Res* 2014;2014:401739.
 18. Narain S, Richards HB, Satoh M, Sarmiento M, Davidson R, Shuster J, et al. Diagnostic accuracy for lupus and other systemic autoimmune diseases in the community setting. *Arch Intern Med* 2004;164:2435–41.
 19. Mahler M, Meroni PL, Bossuyt X, Fritzler MJ. Current concepts and future directions for the assessment of autoantibodies to cellular antigens referred to as anti-nuclear antibodies. *J Immunol Res* 2014;2014:315179.
 20. Ricchiuti V, Adams J, Hardy DJ, Katayev A, Fleming JK. Automated processing and evaluation of anti-nuclear antibody indirect immunofluorescence testing. *Front Immunol* 2018;9:927.
 21. Rigon A, Soda P, Zennaro D, Iannello G, Afeltra A. Indirect immunofluorescence in autoimmune diseases: assessment of digital images for diagnostic purpose. *Cytometry B Clin Cytom* 2007;72:472–7.
 22. Bizzaro N. Can solid-phase assays replace immunofluorescence for ANA screening? *Ann Rheum Dis* 2020;79:e32.
 23. Orme ME, Andalucia C, Sjolander S, Bossuyt X. A comparison of a fluorescence enzyme immunoassay versus indirect immunofluorescence for initial screening of connective tissue diseases: systematic literature review and meta-analysis of diagnostic test accuracy studies. *Best Pract Res Clin Rheumatol* 2018;32:521–34.
 24. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0*. London, UK: The Cochrane Collaboration, 2010. <http://srdta.cochrane.org/>.
 25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Br Med J* 2009;339:b2535.
 26. Tedeschi SK, Johnson SR, Boumpas D, Daikh D, Dörner T, Jayne D, et al. Developing and refining new candidate criteria for systemic lupus erythematosus classification: an international collaboration. *Arthritis Care Res (Hoboken)* 2018;70:571–81.
 27. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. *Arthritis Rheumatol* 2019;71:1400–12.
 28. 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. *Ann Intern Med* 2011;155:529–36.
 29. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;25:1271–7.
 30. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* 2002;61:554–8.
 31. Vitali C, Bombardieri S, Moutsopoulos HM, Balestrieri G, Bencivelli W, Bernstein RM, et al. Preliminary criteria for the classification of Sjogren's syndrome. Results of a prospective concerted action supported by the European Community. *Arthritis Rheum* 1993;36:340–7.
 32. ARA DTCC. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 1980;23:581–90.

33. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;26:404–13.
34. Takwoingi Y, Riley RD, Deeks JJ. Meta-analysis of diagnostic accuracy studies in mental health. *Evid Based Ment Health* 2015;18:103–9.
35. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865–84.
36. Harbord RM, Whiting P. Metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009;9:211–29.
37. Takwoingi Y. Meta-analysis of test accuracy studies in Stata: a bivariate model approach. Version 1.1. <http://methods.cochrane.org/sdt/>. April 2016.
38. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239–51.
39. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
40. Alpini C, Valaperta S, Avasse S, Ramoni V, Bonino C, Montecucco C, et al. Role of a new FEIA assay in systemic connective tissue disease diagnosis. *EliA J* 2010:3.
41. Baptista-Fernandes I, Matoso-Ferreira A, Torrão-Mendes A, Faro-Viana J. Performance of a new screening test for connective tissue disease specific antibodies compared to HEp2 screening. *EliA J* 2010:4.
42. Bentow C, Lakos G, Rosenblum R, Bryant C, Seaman A, Mahler M. Clinical performance evaluation of a novel, automated chemiluminescent immunoassay, QUANTA Flash CTD Screen Plus. *Immunol Res* 2014;61:110–6.
43. Bernardini S, Infantino M, Bellincampi L, Nuccetelli M, Afeltra A, Iori R, et al. Screening of antinuclear antibodies: comparison between enzyme immunoassay based on nuclear homogenates, purified or recombinant antigens and immunofluorescence assay. *Clin Chem Lab Med* 2004;42:1155–60.
44. Claessens J, Belmondo T, De Langhe E, Westhovens R, Poesen K, Hue S, et al. Solid phase assays versus automated indirect immunofluorescence for detection of antinuclear antibodies. *Autoimmun Rev* 2018;17:533–40.
45. De Almeida Brito F, Santos SM, Ferreira GA, Pedrosa W, Gradisse J, Costa LC, et al. Diagnostic evaluation of ELISA and chemiluminescent assays as alternative screening tests to indirect immunofluorescence for the detection of antibodies to cellular antigens. *Am J Clin Pathol* 2016;145:323–31.
46. Deng X, Peters B, Ettore MW, Ashworth J, Brunelle LA, Crowson CS, et al. Utility of antinuclear antibody screening by various methods in a clinical laboratory patient cohort. *J Appl Lab Med* 2016;1:36–46.
47. Euphrasia Latha J, Dhason TM, Mohanasundaram K, Kumudhamanoharan M, Rajeswari S. Comparison of performance of ELISA with immunofluorescence and immunoblot for the testing of antinuclear antibodies. *Int J Curr Microbiol App Sci* 2016;5:423–7.
48. Fenger M, Wiik A, Høier-Madsen M, Lykkegaard JJ, Rozenfeld T, Hansen MS, et al. Detection of antinuclear antibodies by solid-phase immunoassays and immunofluorescence analysis. *Clinical Chemistry* 2004;50:2141–7.
49. Korsholm T, Trolborg A, Nielsen BD. Indirect immunofluorescence on HEp-2 cells vs. ELIA CTD screen for the detection of antinuclear antibodies. *Scand J Rheumatol* 2014;43:89.
50. Morozzi G, Fineschi I, Bellisai F, Alpini C, Avasse S, Merlini G, et al. A new strategy to detect ANA: IIF HEp-2 cells at second level after the EliA CTD Screen test. Is the algorithm correct? *ImmunoDiagn J* 2012;2:3–4.
51. Op De Beëck K, Vermeersch P, Verschueren P, Westhovens R, Mariën G, Blockmans D, et al. Detection of antinuclear antibodies by indirect immunofluorescence and by solid phase assay. *Autoimmun Rev* 2011;10:801–8.
52. Op De Beëck K, Vermeersch P, Verschueren P, Westhovens R, Mariën G, Blockmans D, et al. Antinuclear antibody detection by automated multiplex immunoassay in untreated patients at the time of diagnosis. *Autoimmun Rev* 2012;12:137–43.
53. Pereira LM, Garcia-Trujillo JA, Romero-Chala S, Timon M, Galindo J, Camara C. Evaluation of a novel automated CTD screen for connective tissue diseases. *EliA J* 2010:6–7.
54. Pi D, De Bady MH, Nimmo M, White R, Pal J, Wong P, et al. Application of linear discriminant analysis in performance evaluation of extractable nuclear antigen immunoassay systems in the screening and diagnosis of systemic autoimmune rheumatic diseases. *Am J Clin Pathol* 2012;138:596–603.
55. Robier C, Amouzadeh-Ghadikolai O, Stettin M, Reicht G. Comparison of the clinical utility of the EliA CTD Screen to indirect immunofluorescence on Hep-2 cells. *Clin Chem Lab Med* 2016;54:1365–70.
56. van der Pol P, Bakker-Jonges LE, Kuijpers JH, Schreurs MW. Analytical and clinical comparison of two fully automated immunoassay systems for the detection of autoantibodies to extractable nuclear antigens. *Clinica Chimica Acta* 2018;476:154–9.
57. Willems P, De Langhe E, Claessens J, Westhovens R, Van Hoeyveld E, Poesen K, et al. Screening for connective tissue disease-associated antibodies by automated immunoassay. *Clin Chem Lab Med* 2018;56:909–18.
58. Pereira MF, Ventura MR. Investigation of an automated ANA/ENA screening system (EliA) as an alternative to IIF (HEp2) for routine use. *EliA J* 2004:9.
59. Willems P, De Langhe E, Westhovens R, Vanderschueren S, Blockmans D, Bossuyt X. Antinuclear antibody as entry criterion for classification of systemic lupus erythematosus: pitfalls and opportunities. *Ann Rheum Dis* 2019;78:e76.
60. Pisetsky DS, Spencer DM, Lipsky PE, Rovin BH. Assay variation in the detection of antinuclear antibodies in the sera of patients with established SLE. *Ann Rheum Dis* 2018;77:911–3.
61. Bossuyt X, Fieuws S. Detection of antinuclear antibodies: added value of solid phase assay? *Ann Rheum Dis* 2014;73:e10.
62. Bizzaro N, Brusca I, Previtali G, Alessio MG, Daves M, Platzgummer S, et al. The association of solid-phase assays to immunofluorescence increases the diagnostic accuracy for ANA screening in patients with autoimmune rheumatic diseases. *Autoimmun Rev* 2018;17:541–7.
63. Otten HG, Brummelhuis WJ, Fritsch-Stork R, Leavis HL, Wisse BW, van Laar JM, et al. Measurement of antinuclear antibodies and their fine specificities: time for a change in strategy? *Clin Exp Rheumatol* 2017;35:462–70.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cclm-2020-0094>).