Systematic Reviews

A systematic and meta-analysis review on the diagnostic accuracy of antibodies in the serological diagnosis of COVID-19

Check for updates

Arthur Vengesai[1,2]* , Herald Midzi[1], Maritha Kasambala[1], Hamlet Mutandadzi[3], Tariro L. Mduluza-Jokonya[1,2], Simbarashe Rusakaniko[3], Francisca Mutapi[4], Thajasvarie Naicker[2] and Takafira Mduluza[1,2]

Abstract

Background: Serological testing based on different antibody types are an alternative method being used to diagnose SARS-CoV-2 and has the potential of having higher diagnostic accuracy compared to the current gold standard rRT-PCR. Therefore, the objective of this review was to evaluate the diagnostic accuracy of IgG and IgM based point-of-care (POC) lateral flow immunoassay (LFIA), chemiluminescence enzyme immunoassay (CLIA), fluorescence enzyme-linked immunoassay (FIA) and ELISA systems that detect SARS-CoV-2 antigens.

Method: A systematic literature search was carried out in PubMed, Medline complete and MedRxiv. Studies evaluating the diagnostic accuracy of serological assays for SARS-CoV-2 were eligible. Study selection and data-extraction were performed by two authors independently. QUADAS-2 checklist tool was used to assess the quality of the studies. The bivariate model and the hierarchical summary receiver operating characteristic curve model were performed to evaluate the diagnostic accuracy of the serological tests. Subgroup meta-analysis was performed to explore the heterogeneity.

Results: The pooled sensitivity for IgG (n = 17), IgM (n = 16) and IgG-IgM (n = 24) based LFIA tests were 0.5856, 0.4637 and 0.6886, respectively compared to rRT-PCR method. The pooled sensitivity for IgG (n = 9) and IgM (n = 10) based CLIA tests were 0.9311 and 0.8516, respectively compared to rRT-PCR. The pooled sensitivity the IgG (n = 10), IgM (n = 11) and IgG-IgM (n = 5) based ELISA tests were 0.8292, 0.8388 and 0.8531 respectively compared to rRT-PCR. All tests displayed high specificities ranging from 0.9693 to 0.9991. Amongst the evaluated tests, IgG based CLIA expressed the highest sensitivity signifying its accurate detection of the largest proportion of infections identified by rRT-PCR. ELISA and CLIA tests performed better in terms of sensitivity compared to LFIA. IgG based tests performed better compared to IgM except for the ELISA.

(Continued on next page)

* Correspondence: arthurvengesai@gmail.com
[1]Department of Biochemistry, University of Zimbabwe, P.O. Box MP 167, Mt Pleasant, Harare, Zimbabwe
[2]Optics and Imaging, Doris Duke Medical Research Institute, College of Health Sciences, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: We report that IgG-IgM based ELISA tests have the best overall diagnostic test accuracy. Moreover, irrespective of the method, a combined IgG/IgM test seems to be a better choice in terms of sensitivity than measuring either antibody type independently. Given the poor performances of the current LFIA devices, there is a need for more research on the development of highly sensitivity and specific POC LFIA that are adequate for most individual patient applications and attractive for large sero-prevalence studies.

Systematic review registration: PROSPERO CRD42020179112

Keywords: Serology, COVID-19, SARS-CoV2, rRT-PCR, IgG, IgM, Specificity, Sensitivity

Introduction

Coronavirus disease 2019 (COVID-19) is a major contagious pandemic of respiratory disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is also known as novel (new) coronavirus 2019-nCoV [1–3]. The first COVID-19 cases were identified in December 2019 from Wuhan, Hubei Province, China [4]. On November 18 2020, according to the European Centre for Disease Prevention and Control, COVID-19 Situation update, there were 55,743,951 confirmed cases and 1,339,436 deaths reported worldwide [5]. Although the COVID-19 clinical features are not yet fully known and understood, clinicians have reported clinical manifestations that range from asymptomatic cases to patients with mild and severe respiratory illness, with or without pneumonia, fever, cough and shortness of breath. Older people (>65 years) and people of all ages with severe chronic medical conditions such as lung disease, heart disease and diabetes seem to have a higher risk of succumbing to severe COVID-19 illness [6].

Early and accurate diagnostic testing for COVID-19 is critical for tracking the SARS-CoV-2, understanding the virus epidemiology, informing case management, suppressing transmission and for quarantine purposes [7, 8]. The standard diagnostic confirmatory test for COVID-19 is based on the detection of nucleic acids of SARS-CoV-2 by nucleic acid amplification tests, such as real-time reverse-transcriptase polymerase chain reaction (rRT-PCR). The test identifies viral nucleic acids when present in sufficient quantity in sputum, throat swabs and secretions of the lower respiratory tract. In some patients, SARS-CoV-2 RNA detection in blood and oral fluid specimens has been reported; however, limited data is available on adequacy of SARS-CoV-2 detection in these specimens [9]. The rRT-PCR test is time consuming as it takes between 4 to 6 h for completion. It requires expensive specialist equipment, skilled laboratory personal for sample preparation and testing and PCR reagents, creating diagnostic delays and limiting use in real-life situations when rapid diagnosis is required for fast intervention decisions. Therefore, less expensive and easy implementable tests are required for SARS-CoV-2 detection. Another limitation of using rRT-PCR involves the use of swabs from the upper respiratory tract which can be falsely diagnosed as

negative due to the poor quality of the sample or acquiring the sample at an incorrect timeframe; notably, viral load in upper respiratory tract secretions peak in the first week of symptoms but may decline below the limit of detection in patients presenting late with symptoms [8, 10–12]. Missing the time-window of viral replication may also provide false negative results. Moreover, after a variable period of time, one expects the rRT-PCR result to become negative due to cessation of viral shedding [13].

False-negative rRT-PCR results are common during diagnosis of SARS-CoV-2. The Fever Clinic of the Beijing Haidian Hospital collected data from January 2020 which indicated that only two out of ten negative cases diagnosed by rRT-PCR test were confirmed to be true positive for COVID-19. This yielded an approximately 20% false-negative rate of rRT-PCR [12]. Zhang et al. also showed that the current strategy for the detection of viral RNA in oral swabs used for SARS-CoV-2 diagnosis is not 100% accurate. The presence of the virus has been detected in anal swabs or blood samples of patients whilst their oral swabs diagnosis reports a negative result. This observation implies that a patient cannot be discharged based purely on oral swab samples being negative [14].

A false negative diagnosis may have grave consequences, especially at this stage of the COVID-19 pandemic by allowing SARS-CoV-2 infected patients to spread the infection thereby hampering the efforts to contain the spread of the virus [8]. Additional screening methods that can detect the presence of infection despite lower viral titres are highly beneficial to ensure timely diagnosis of all COVID-19 patients. Detection of serum specific anti-SARS-CoV-2 antibodies, both immunoglobulin G (IgG) and M (IgM), which are produced rapidly after the infection provides an alternative highly sensitive and accurate solution and compensates for the limitations of rRT-PCR. The serological methods could also be a more practical alternative to chest CT [8, 15, 16]. Immunoglobulin G antibodies permit the use of serological tools to better understand the overall rate of COVID-19 infections including the rate of asymptomatic infections [8].

Vengesai *et al. Systematic Reviews*   (2021) 10:155

Page 3 of 23

However, the dynamics of blood or serum antibodies in the cases of COVID-19 are not well evaluated. Currently, the serological dynamics of COVID-19 patients remain limited. Also, before diagnostic assays are widely deployed, their performance characteristics need to be evaluated. Therefore, the objective of this review was to evaluate the diagnostic accuracy of IgG and IgM (together or separately) based point-of-care (POC) lateral flow immunoassay (LFIA), chemiluminescence enzyme immunoassay (CLIA), fluorescence enzyme-linked immunoassay (FIA) and enzyme-linked immunosorbent assay (ELISA) to detect antigens against SARS-CoV-2.

Review questions
The primary research question of this systematic review was 'What is the diagnostic accuracy of antibody serology tests for COVID-19 using the bivariate model and the hierarchical summary receiver operating characteristic curve (HSROC) model?'

Materials and methods
We conducted a systematic review and meta-analysis in accordance with the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic test accuracy (PRISMA-DTA) [17] (Additional file 1). We used the Cochrane recommendations to report systematic reviews and meta-analyses of studies on diagnostic accuracy [18]. We also used protocols from published systematic and meta-analysis reviews to develop our protocol [19–22]. The developed systematic review protocol was registered in the International Prospective Register of Systematic Reviews registration number CRD42020179112.

Eligibility criteria
Cross-sectional studies would be the ideal study type to answer our review questions. However, as we anticipated that serological diagnosis cross-sectional studies for COVID-19 would be very sparse, we decided to include case-control studies. The inclusion criteria comprised studies in which the study population ($n \geq 10$) were subjected to COVID-19 rRT-PCR testing or genetic sequencing as the reference standards and either one or all of the following serological tests; point-of-care (POC), lateral flow immunoassay (LFIA), chemiluminescence enzyme immunoassay (CLIA), fluorescence immunoassays (FIA) and enzyme-linked immunoassay (ELISA). Studies that used chest CT images, epidemiological history, well-defined clinical features accompanied by rRT-PCR as a reference standard were included. The diagnostic accuracy of the tests was defined as the primary outcome. Original studies were included without restriction based on language or geographical location. We included studies between 1 January 2020, and 27 April 2020. Animal studies, in vitro-based studies and survey studies investigating seroprevalence were excluded from the study.

Information sources and search strategies
The following databases were searched for studies: MEDLINE Complete (EBSCO), PubMed, and MedRxiv (a preprint server for health sciences which distributes complete but unpublished manuscripts). We performed the search strategy on studies dated until 29 April 2020. The data bases were searched using predefined keywords: COVID-19 and serologic test and their synonyms. Appendix 1 illustrates the search strategy for PubMed, which was adapted for the other data bases. Additional studies were identified by contacting experts in the field and by searching reference lists from primary studies, review articles and textbook chapters.

Study selection and data extraction
Two authors (AV and HM) assessed the titles identified by the search, excluding those obviously irrelevant to the serological diagnosis of COVID-19. Letters, review articles and articles clearly irrelevant based on examination of the abstract and other notes and duplicates were excluded next. The eligibility of the remaining potentially relevant articles was judged on full-text publications.

Data extraction was conducted independently by two authors (AV and HM) to avoid bias and discrepancies and was resolved by discussion. Where an agreement could not be reached, a third author was consulted. Where it remained unclear whether a study is eligible for inclusion, it was then excluded. Whilst extracting data, authors also had to decide whether a study was a case-control or a cross-sectional study. The following data were extracted.

- Study authors and publication year
- Study design
- Case definition (inclusion/exclusion criteria)
- Participant demographics
- Reference standard (including criteria for positive test)
- Index tests [cutoff values (prespecified or not) and whether the test was a commercial of in-house test]
- Geographical location of data collection
- Index/reference time interval
- Distribution of severity of disease in those with target condition
- Other diagnoses in those without target condition
- Numbers of TPs, FNs, FPs and FNs

Other considerations and exclusion criteria
Studies, from which a 2×2 table containing true positives, false positives, false negatives and true negatives

could not be drawn, were excluded. Furthermore, studies that were too unspecific in their reporting to ensure that they fulfilled the above criteria, were excluded.

Assessment of methodological quality

The QUADAS-2 tool was used to assess the methodological quality of all studies included in this systematic review [23]. QUADAS-2 consists of four key domains: patient selection, index test, reference standard and flow and timing. We assessed all domains for risk of bias (ROB) potential and the first three domains for applicability concerns. Risk of bias was judged as 'low,' 'high' or 'unclear'. Details are shown in Appendix 2. Two review authors (AV and HM) independently completed QUADAS-2. The divergences were resolved by consensus amongst the researchers.

Statistical analysis and data synthesis

Diagnostic accuracy

For each study, we constructed 2×2 tables for true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). Where only sensitivity and specificity estimates were reported, we derived the two-by-two table from the reported data. We constructed forest plots displaying sensitivity and specificity of the index tests from contingency tables assuming that the reference method was 100% sensitive and specific. Data were entered into the Review Manager (RevMan) software for Windows v.5.3 (Cochrane Collaboration, Copenhagen, Denmark) and forest plots were created with 95% confidence interval (CI) for sensitivity and specificity for each study.

Studies were submitted to meta-analysis when three conditions were met: sample size was greater than 20; sensitivity and specificity were available for the index and control group was included in the analysis. We used the two recommended random-effects hierarchical methods: the bivariate model and the hierarchical summary receiver operating characteristic (HSROC) model for performing the meta-analysis. The focus of the bivariate model is estimation of a summary point (summary sensitivity and specificity). HSROC model is on estimating an SROC curve [24]. The summary estimates of sensitivity and specificity and 95% CI and the HSROC were calculated using OpenMeta-Analyst for windows 10 (open-source, cross-platform for advanced meta-analysis).

Investigations of heterogeneity

We investigated heterogeneity by adding antigen type as the covariate. The following approach was taken: Firstly, the variation in accuracy between IgG or IgM or IgG-IgM based LFIA, CLIA and ELISA serological testing was analysed (Table 2). Then, the effect of the antigen type was investigated using subgroup meta-analysis in OpenMeta-analyst. I^2 values close to 0% were considered as having no heterogeneity between studies; values close to 25 %, low heterogeneity; values close to 50%, moderate heterogeneity and values close to 75%, high heterogeneity between studies [25].

Assessment of publication bias

In this review, we did not assess for reporting bias. The studies included in our meta-analysis showed a lot of heterogeneity; therefore, assessments for reporting bias may not yield conclusive results. This was adopted from the approach used by Ochodo et al. [26].

Test sensitivity by time since onset of symptoms

We stratified data by days since COVID-19 symptom onset to specimen collection. Then, we constructed forest plots (95 % CI) displaying test sensitivity by time since onset of symptoms using the RevMan software for Windows v.5.3 (Cochrane Collaboration, Copenhagen, Denmark).

Results

Study inclusion

The results of the search and selection process are presented in Fig. 1. A total of 202 articles were identified. Amongst these, 40 were MedRxiv preprints and 162 were fully published articles from MEDLINE Complete (EBSCO) and PubMed. Two articles were identified from other sources for example manual search. After abstract/title exclusion and removing duplicates, 74 articles were submitted to full-text screening and 31 of these were included for the systematic review. Most articles were excluded because they did not present sufficient data hence it was not possible to extract data to construct 2 × 2 table and 1 article was excluded because it was only available in Chinese. A total of 29 articles describing the results of 99 independent studies/data sets (19, 23 and 57 investigating LFIA, CLIA and ELISA respectively) were eligible for the meta-analysis.

Characteristics of the studies

The general characteristics of the included articles are presented in Table 1. All the published articles ($n = 14$) included in the review were published in 2020 because COVID-19 is an emerging disease. The 17 unpublished articles were MedRxiv preprints which have been submitted to different journals for publication. Twenty five articles included in the review had a case-control design, comparing a group of well-defined cases with a group of healthy controls or controls with diseases or COVID-19 rRT-PCR negative patients, and only six studies were cross sectional studies. One study had no control

Vengesai et al. Systematic Reviews   (2021) 10:155

Page 5 of 23

Fig. 1 PRISMA flow diagram for selection of articles for meta-analysis
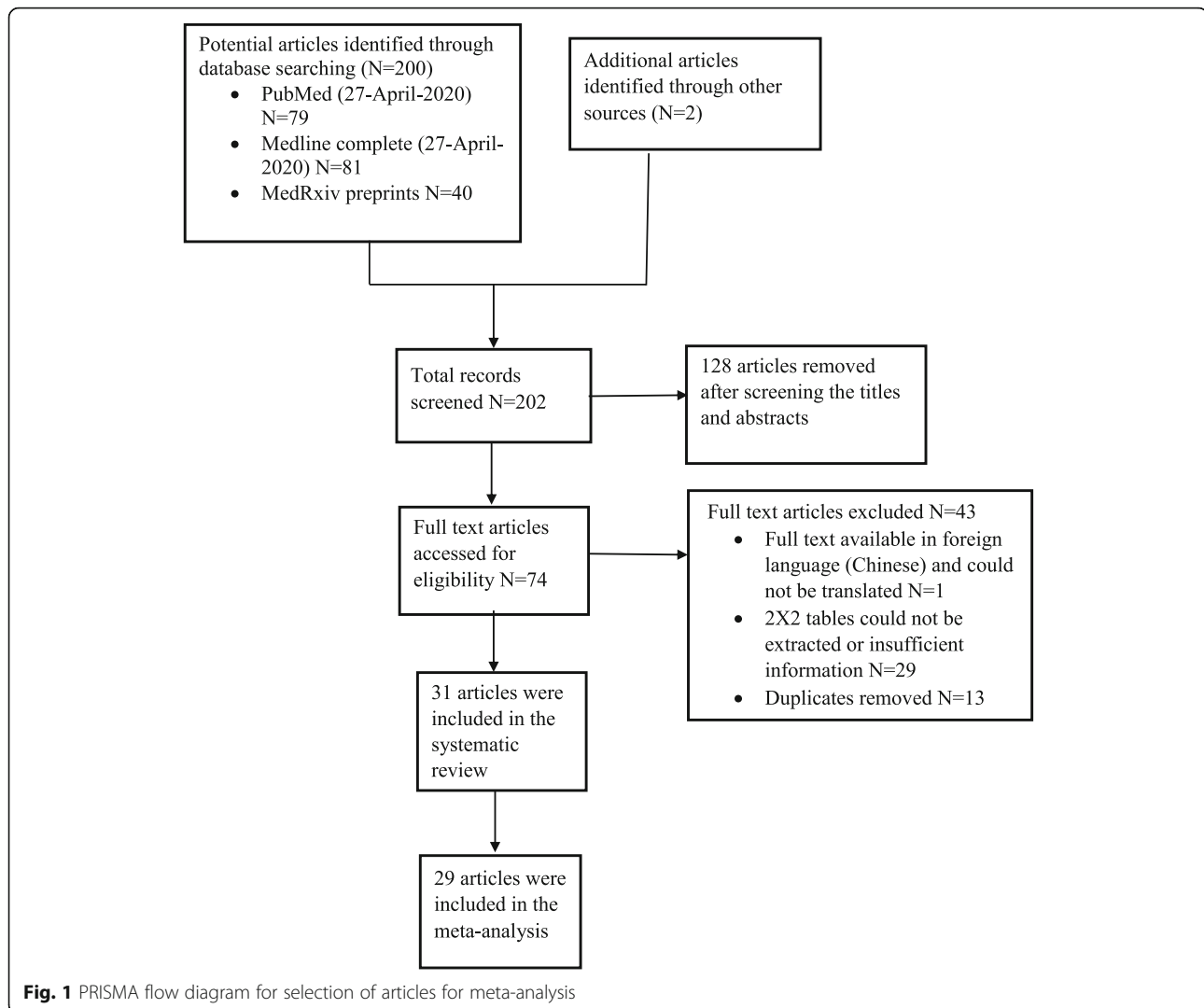
group and was excluded in the meta-analysis [27]. Most of the studies (*n* = 22) were conducted in China where the COVID-19 pandemic began and 3 studies were conducted in Italy whilst, USA whilst UK, Denmark, Germany, Spain and Japan each conducted one study.

Most articles (*n* = 26) included in the review clearly stated that the gold standard nucleic acid tests (rRT-PCR or deep sequencing) were used as the reference standard. However, five articles used a combination of epidemiological risk, clinical features, chest CT images and rRT-PCR. In one article, the reference standard used was not stated but all the patients in the study were COVID-19 patients [38].

Point-of-care (POC) lateral flow immunoassays (LFIA) were used in 14 articles, CLIA were used in 9 articles and ELISA were used in 13 articles. We did not identify articles using FIA that met our inclusion criteria. One study did not specify the serological assay used and it was excluded from the review [31]. One study used a LIPS which is performed in solution, thus maintaining the native antigen conformation [54]. Most of the serological assay test kits were commercial (*n* = 21) and 12 were in-house. Three SARS-CoV-2 antigens, Spike protein (S), nucleocapsid protein (N) and envelope protein (E) were used together or separately in studies included in the review. The spike protein and nucleocapsid were used as the antigen in 9 articles and 6 articles respectively. Five articles used both S and N as the antigens separately. In 3 articles, S and N antigens (S-N) were used together as the antigen. In 1 article, N and E antigens (N-E) were used together as the antigen. In 7 articles, the name of antigen used was not given.

Vengesai *et al. Systematic Reviews*   (2021) 10:155

Page 6 of 23

Table 1 The general characteristics of the studies included in the review

Study ID	Country	Antibody type	Antigen type	Commercial	Reference standard	Index test	Control group/comparison group
Kai-Wang To [27] [P, CS]	China	IgG and IgM	S and N	Inhouse	rRT-PCR	ELISA	No control
Cassaniti [28] [P, CS]	Italy	IgG, IgM and IgG-IgM	S	Commercial	rRT-PCR	POC LFIA	Patients with fever and respiratory syndrome/RT-PCR negative
Duchuan Lin [29]	China	IgG, IgM and IgG-IgM	N	Inhouse	Epidemiological risk/clinical features/rRT-PCR	CLIA	Healthy individuals and tuberculosis patients
Jie Xiang [30]	China	IgG, IgM and IgG-IgM		Commercial	rRT-PCR	ELISA and POC LFIA	Healthy individuals
Li Guo [13] [P]	China	IgM	N	Inhouse	Deep sequencing and rRT-PCR	ELISA	Adult patients with acute lower respiratory tract infections (ALRTIs)
Rui Liu [31] [CS]	China	IgM	N	Inhouse	rRT-PCR		COVID-19 rRT-PCR negative patients
Wanbing Liu [32] [P]	China	IgG, IgM and IgG-IgM	S and N	Commercial	rRT-PCR	ELISA	Healthy individuals
Xuefei Cai [33]	China	IgG, IgM and IgG-IgM	S	Inhouse	rRT-PCR	Peptide-based magnetic CLIA	Mixed diseases and healthy controls
Yu bao Pan [34] [CS]	China	IgG, IgM and IgG-IgM		Commercial	rRT-PCR	POC LFIA	COVID-19 rRT-PCR negative patients
Yujiao Jin [35] [P]	China	IgG, IgM and IgG-IgM	S-N	Commercial	rRT-PCR	CLIA	Patients with suspected SARS-CoV-2 infection but with negative rRT-PCR results
Zhao [36] [P]	China	IgG and IgM	S	Commercial	Chest CT images/epidemiological history/clinical diagnosis/rRT-PCR	ELISA	Healthy individuals
Zhengtu Li [37] [P]	China	IgG, IgM and IgG-IgM	S	Commercial	rRT-PCR	POC LFIA	Healthy individuals
Rongqing Zhao [38]	China	IgG-IgM	S	Inhouse	Not clear but all cases were confirmed COVID-19 patients	ELISA	Healthy individuals (samples collected before and during the COVID-19 pandemic)
Pingping Zhang [39]	China	IgG-IgM	S	Inhouse	rRT-PCR	POC LFIA	COVID-19 rRT-PCR negative patients
Paradiso [40]	Italy	IgG-IgM	S	Commercial	rRT-PCR	POC LFIA	Patients with COVID-19 disease orienting-symptoms but rRT-PCR negative
Huan Ma [41]	China	IgA, IgG, IgM, IgG-IgM and Ab	S and N	Inhouse	rRT-PCR	CLIA	Healthy individuals, COVID-19 suspected individuals and mixed disease group
Qian [15]	China	IgG and IgM	S-N	Commercial	rRT-PCR	CLIA	Healthy individuals and hospitalised individuals
Ling Zhong [42] [P]	China	IgG and IgM		Inhouse	rRT-PCR	CLIA and ELISA	Healthy individuals
Jiajia Xie [43] [P, CS]	China	IgG and IgM	E-N	Commercial	Chest CT images/epidemiological history/clinical diagnosis/rRT-PCR	CLIA	Clinically confirmed COVID-19 rRT-PCR negative patients
Infantino [44] [P]	Italy	IgG and IgM	S-N	Commercial	rRT-PCR	CLIA	Mixed diseases patients and blood donors pre-COVID-19

Vengesai *et al. Systematic Reviews*   (2021) 10:155

Page 7 of 23

Table 1 The general characteristics of the studies included in the review *(Continued)*

Study ID	Country	Antibody type	Antigen type	Commercial	Reference standard	Index test	Control group/comparison group
Adams [45]	UK	IgG, IgM and IgG-IgM	S	Inhouse (ELISA) and Commercial (LFIA)	rRT-PCR	ELISA and POC LFIA	Healthy blood and ICU cerebral organ donors before the COVID-19 pandemic
Lassaunière [46]	Denmark	IgA, IgG and Ab	S	Commercial	rRT-PCR	ELISA and POC LFIA	Healthy individuals and mixed diseases patients (including acute respiratory tract infections caused by other corona viruses and non-corona viruses
Qiang Wang [47] [P]	China	IgG, IgM and IgG-IgM		Commercial	Chest CT images/ epidemiological history/ clinical diagnosis/rRT-PCR	ELISA and POC LFIA	COVID-19 clinical negative mixed diseases patients
Fei Xiang [30] [P]	China	IgG and IgM	N	Commercial	rRT-PCR	ELISA	Healthy blood donors or from patients with other disease hospitalised
Bin Lou [48]	China	IgG, IgM and Ab	S and N	Commercial	rRT-PCR	ELISA, CLIA and POC LFIA	Healthy Individuals
Lei Liu [49]	China	IgG-IgM	N	Commercial	rRT-PCR	ELISA	Randomly selected ordinary patients and healthy blood donors
Imai [50]	Japan	IgG, IgM and IgG-IgM		Commercial	rRT-PCR	POC LFIA	Non-COVID-19 patients (from April to October 2019
Pérez-García [51]	Spain	IgG, IgM and IgG-IgM		Commercial	rRT-PCR	POC LFIA	Healthy individuals (samples collected before the COVID-19 pandemic)
Zhenhua Chen [52] [P]	China	IgG	N	Inhouse	rRT-PCR	POC LFIA	Clinically suspicious for the presence of anti-SARS-CoV-2
Dohla [53] [P, CS]	German	IgG, IgM and IgG-IgM		Commercial	RT-qPCR	POC LFIA	COVID-19 RT-qPCR negative patients
Burbelo [54]	USA	Ab	S and N	Inhouse	RT-PCR	LIPS	Subjects with COVID-19-like symptoms or household contacts of persons with COVID-19 (not tested by PCR), and blood donors who donated samples before 2018.

Studies with P superscripts were published articles and without P superscripts were MedRxiv preprints
Studies with CS superscripts are cross sectional studies and without CS superscripts are case control studies
IgG-IgM means that either one of them or both were detected in serum
Ab means total antibodies

Vengesai *et al. Systematic Reviews*   (2021) 10:155

Page 8 of 23

Methodological quality of included studies

The methodological quality of the included studies for the IgG or IgM or IgG-IgM based LFIA, CLIA and ELISA summarised across all studies are shown in Figs. 2b, 3b and 4b. Figures 2a, 3a and 4a show for the risk of bias and applicability concerns summary results for the LFIA, CLIA and ELISA individual studies respectively. None of the studies included in
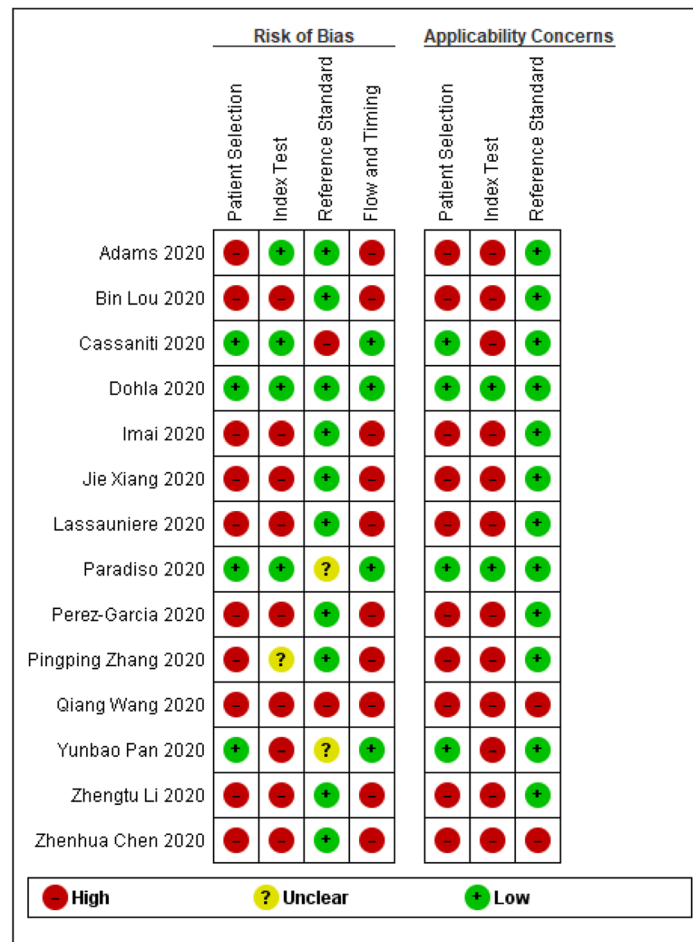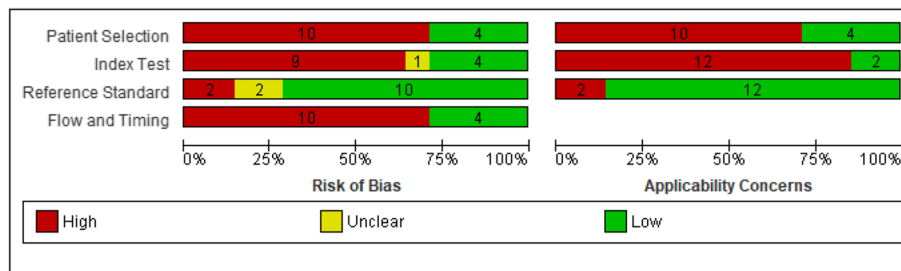
Fig. 2 LFIA methodological quality summary table and graph. **a** Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study. **b** QUADAS-2 bias assessment and QUADAS-2 applicability assessment item presented as percentages across all included studies. On the left, risk of bias graph and on the right applicability concerns graph. **c** Risk of bias and applicability concerns summary: review authors. Low, low risk of bias; high, high risk of bias; unclear, bias is unclear

Fig. 3 CLIA methodological quality summary table and graph. **a** Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study. **b** QUADAS-2 bias assessment and QUADAS-2 applicability assessment item presented as percentages across all included studies. On the left, risk of bias graph and on the right applicability concerns graph. **c** Risk of bias and applicability concerns summary: review authors. Low, low risk of bias; high, high risk of bias; unclear, bias is unclear
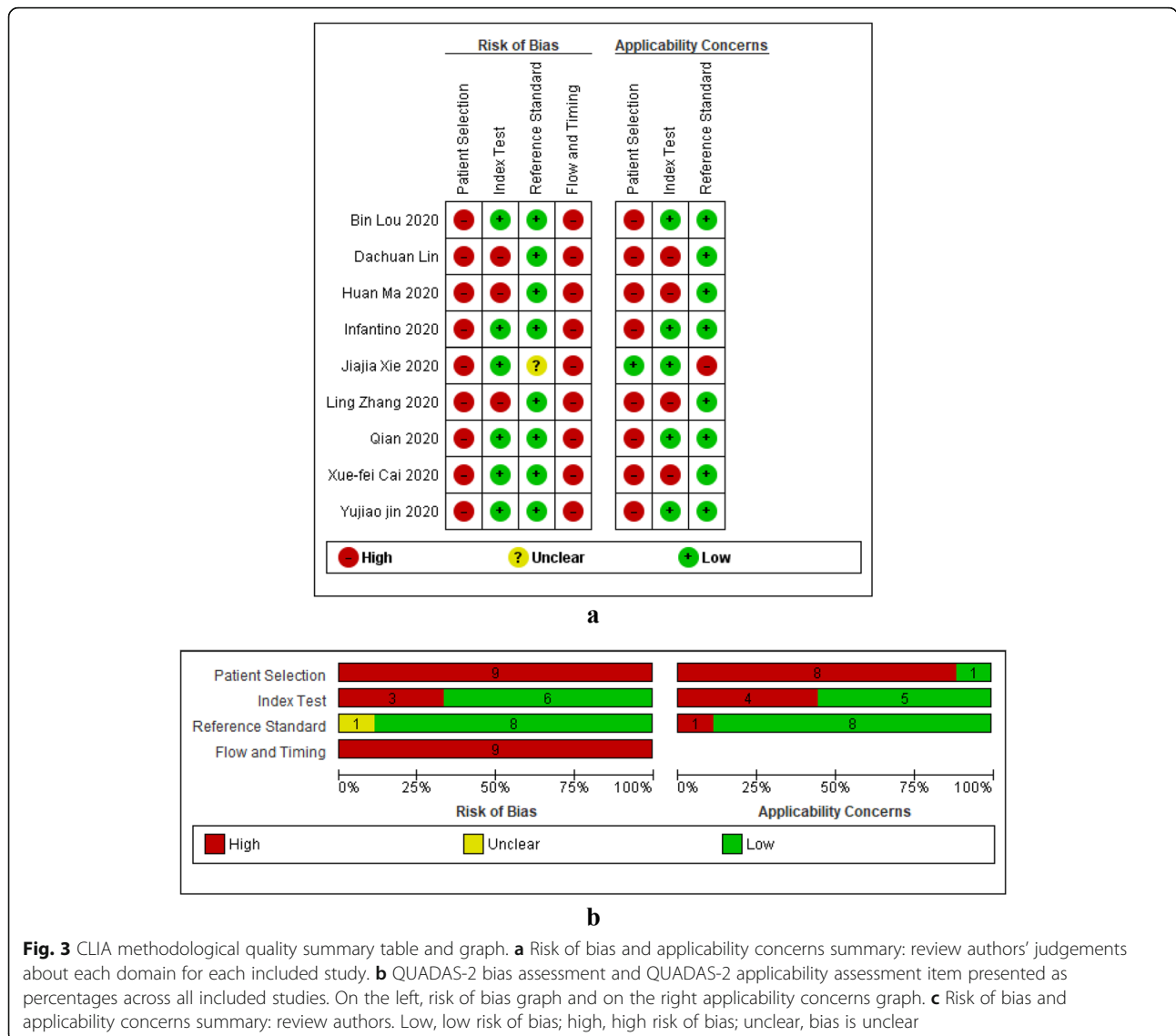
this review had low risk of bias in all four QUADAS-2 domains. Generally, case control studies were of high risk of bias and high concern in the patients and timing and flow domains and cross sectional studies were of low risk of bias and low concern in all domains.

Patient selection domain
Generally, most studies included were at risk of bias and had high concerns regarding applicability. Studies were mostly case control studies and they did not include a consecutive or random series of participants implying that the patients that were included are not representative for clinical use. All thirteen ELISA studies were at high risk of bias and had high concerns regarding applicability. For CLIA, all the 9 studies included had high risk of bias and only 1 cross

sectional study had low applicability concerns. Generally, LFIA had more studies (*n* = 4) with low risk of bias and applicability concerns in the patient selection domain because there were 4 LFIA cross-sectional studies.

Index test domain
The LFIA studies had a high risk of bias (9/14) and high applicability concerns (12/14) in the index test domain. The high risk of bias was due to no blinding between the index test and the reference test. The high applicability concerns were due to tests using serum or plasma instead of whole blood which would make the test less amenable to use at the point of care. The CLIA and ELISA studies generally had a low risk of bias (6/9 and 8/13 respectively). This was because most studies were automated and had a pre-

Vengesai *et al. Systematic Reviews*   (2021) 10:155

Page 10 of 23
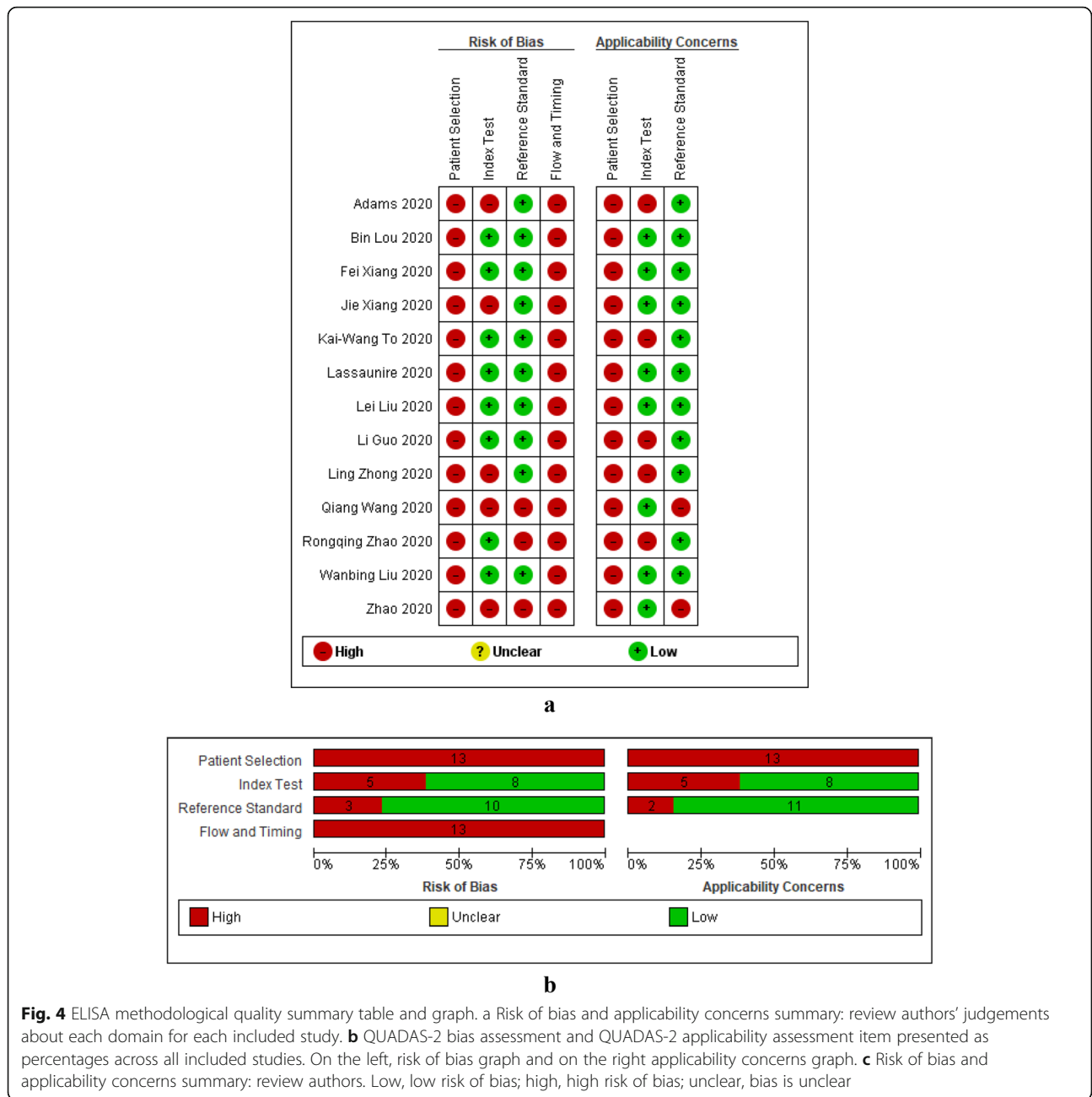
Fig. 4 ELISA methodological quality summary table and graph. **a** Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study. **b** QUADAS-2 bias assessment and QUADAS-2 applicability assessment item presented as percentages across all included studies. On the left, risk of bias graph and on the right applicability concerns graph. **c** Risk of bias and applicability concerns summary: review authors. Low, low risk of bias; high, high risk of bias; unclear, bias is unclear

specified threshold (cut-off value to decide whether a test is positive or negative). The studies that had high risk of bias did not have a pre-specified threshold. Likewise, CLIA and ELISA studies generally had low applicability concerns in the index test domain (5/9 and 8/13 respectively) because they used commercial index tests.

Reference standard domain

Like the index test domain, studies generally had a low risk of bias (10/14, 8/9 and 10/13 for LFIA, CLIA and ELISA respectively) in the reference standard domain.

Generally, the studies were of low applicability concern, 10/14, 8/9 and 11/13 for LFIA, CLIA and ELISA respectively.

Flow and timing domain

All the CLIA ($n = 9$) and ELISA ($n = 13$) studies were at high risk of bias in the flow and timing domain. These studies were all case control studies. Most of the LFIA studies were also at a high risk of bias; however, 4 cross sectional LFIA studies were at low risk of bias.

a

IgG LFIA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Adams 2020	0	0	0	0	Not estimable	Not estimable
Adams 2020	17	0	23	142	0.42 [0.27, 0.59]	1.00 [0.97, 1.00]
Adams 2020	20	2	13	58	0.61 [0.42, 0.77]	0.97 [0.88, 1.00]
Adams 2020	15	0	17	60	0.47 [0.29, 0.65]	1.00 [0.94, 1.00]
Adams 2020	25	1	13	59	0.66 [0.49, 0.80]	0.98 [0.91, 1.00]
Adams 2020	17	0	14	60	0.55 [0.36, 0.73]	1.00 [0.94, 1.00]
Adams 2020	15	1	16	59	0.48 [0.30, 0.67]	0.98 [0.91, 1.00]
Adams 2020	0	0	0	0	Not estimable	Not estimable
Adams 2020	0	0	0	0	Not estimable	Not estimable
Adams 2020	21	2	12	58	0.64 [0.45, 0.80]	0.97 [0.88, 1.00]
Bin Lou 2020	61	1	11	208	0.85 [0.74, 0.92]	1.00 [0.97, 1.00]
Cassaniti 2020	5	0	33	12	0.13 [0.04, 0.28]	1.00 [0.74, 1.00]
Dohla 2020	8	3	14	24	0.36 [0.17, 0.59]	0.89 [0.71, 0.98]
Imai 2020	20	0	119	48	0.14 [0.09, 0.21]	1.00 [0.93, 1.00]
Jie Xiang 2020	74	0	17	35	0.81 [0.72, 0.89]	1.00 [0.90, 1.00]
Perez-Garcia 2020	23	0	32	45	0.42 [0.29, 0.56]	1.00 [0.92, 1.00]
Qiang Wang 2020	14	22	0	50	1.00 [0.77, 1.00]	0.69 [0.57, 0.80]
Yunbao Pan 2020	47	13	39	9	0.55 [0.44, 0.65]	0.41 [0.21, 0.64]
Zhengtu Li 2020	280	1	117	127	0.71 [0.66, 0.75]	0.99 [0.96, 1.00]
Zhenhua Chen 2020	7	1	0	11	1.00 [0.59, 1.00]	0.92 [0.62, 1.00]

b

IgM LFIA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Adams 2020	18	1	13	59	0.58 [0.39, 0.75]	0.98 [0.91, 1.00]
Adams 2020	2	0	36	60	0.05 [0.01, 0.18]	1.00 [0.94, 1.00]
Adams 2020	3	0	30	60	0.09 [0.02, 0.24]	1.00 [0.94, 1.00]
Adams 2020	17	1	16	59	0.52 [0.34, 0.69]	0.98 [0.91, 1.00]
Adams 2020	14	1	17	59	0.45 [0.27, 0.64]	0.98 [0.91, 1.00]
Adams 2020	15	4	25	138	0.38 [0.23, 0.54]	0.97 [0.93, 0.99]
Adams 2020	14	0	18	60	0.44 [0.26, 0.62]	1.00 [0.94, 1.00]
Bin Lou 2020	71	95	9	205	0.89 [0.80, 0.95]	0.68 [0.63, 0.74]
Cassaniti 2020	6	1	32	11	0.16 [0.06, 0.31]	0.92 [0.62, 1.00]
Dohla 2020	8	3	14	24	0.36 [0.17, 0.59]	0.89 [0.71, 0.98]
Imai 2020	60	1	79	47	0.43 [0.35, 0.52]	0.98 [0.89, 1.00]
Jie Xiang 2020	52	0	39	35	0.57 [0.46, 0.67]	1.00 [0.90, 1.00]
Perez-Garcia 2020	12	0	43	45	0.22 [0.12, 0.35]	1.00 [0.92, 1.00]
Qiang Wang 2020	14	22	0	50	1.00 [0.77, 1.00]	0.69 [0.57, 0.80]
Yunbao Pan 2020	48	8	38	14	0.56 [0.45, 0.67]	0.64 [0.41, 0.83]
Zhengtu Li 2020	328	10	69	118	0.83 [0.79, 0.86]	0.92 [0.86, 0.96]

c

IgG-IgM LFIA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Adams 2020	23	3	10	57	0.70 [0.51, 0.84]	0.95 [0.86, 0.99]
Adams 2020	20	1	11	59	0.65 [0.45, 0.81]	0.98 [0.91, 1.00]
Adams 2020	18	0	15	60	0.55 [0.36, 0.72]	1.00 [0.94, 1.00]
Adams 2020	18	0	14	60	0.56 [0.38, 0.74]	1.00 [0.94, 1.00]
Adams 2020	22	4	18	138	0.55 [0.38, 0.71]	0.97 [0.93, 0.99]
Adams 2020	23	1	15	91	0.61 [0.43, 0.76]	0.99 [0.94, 1.00]
Adams 2020	19	2	12	58	0.61 [0.42, 0.78]	0.97 [0.88, 1.00]
Adams 2020	25	1	13	59	0.66 [0.49, 0.80]	0.98 [0.91, 1.00]
Adams 2020	21	2	12	58	0.64 [0.45, 0.80]	0.97 [0.88, 1.00]
Cassaniti 2020	7	1	31	11	0.18 [0.08, 0.34]	0.92 [0.62, 1.00]
Dohla 2020	8	3	14	24	0.36 [0.17, 0.59]	0.89 [0.71, 0.98]
Imai 2020	60	1	79	47	0.43 [0.35, 0.52]	0.98 [0.89, 1.00]
Jie Xiang 2020	75	0	16	35	0.82 [0.73, 0.90]	1.00 [0.90, 1.00]
Lassauniere 2020	28	0	2	32	0.93 [0.78, 0.99]	1.00 [0.89, 1.00]
Lassauniere 2020	27	0	3	32	0.90 [0.73, 0.98]	1.00 [0.89, 1.00]
Lassauniere 2020	25	0	5	17	0.83 [0.65, 0.94]	1.00 [0.80, 1.00]
Lassauniere 2020	4	3	1	12	0.80 [0.28, 0.99]	0.80 [0.52, 0.96]
Lassauniere 2020	27	0	3	32	0.90 [0.73, 0.98]	1.00 [0.89, 1.00]
Paradiso 2020	21	13	49	107	0.30 [0.20, 0.42]	0.89 [0.82, 0.94]
Perez-Garcia 2020	26	0	29	45	0.47 [0.34, 0.61]	1.00 [0.92, 1.00]
Pingping Zhang 2020	106	4	16	656	0.87 [0.80, 0.92]	0.99 [0.98, 1.00]
Qiang Wang 2020	14	22	0	50	1.00 [0.77, 1.00]	0.69 [0.57, 0.80]
Yunbao Pan 2020	59	14	27	8	0.69 [0.58, 0.78]	0.36 [0.17, 0.59]
Zhengtu Li 2020	352	12	45	116	0.89 [0.85, 0.92]	0.91 [0.84, 0.95]

Fig. 5 Forest plot of sensitivity, specificity and heterogeneity of serological LFIA diagnosis of COVID-19. **a** Forest plot for the IgG LFIA. **b** Forest plot for the IgM based LFIA. **c** Forest plot for the IgG-IgM based LFIA

Quantitative synthesis and meta-analysis

Firstly, we considered performance of the LFIA devices using rRT-PCR-confirmed cases as the reference standard. The forest plots in Fig. 5 show the sensitivity, specificity range and heterogeneity for the three IgG or IgM or IgG-IgM based LFIA detecting COVID-19 across the included studies. Overall, the sensitivity varied widely across studies in contrast to the specificity which did not vary much 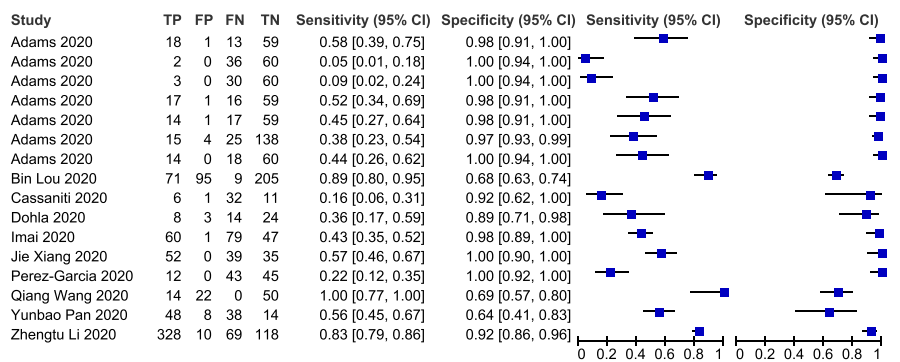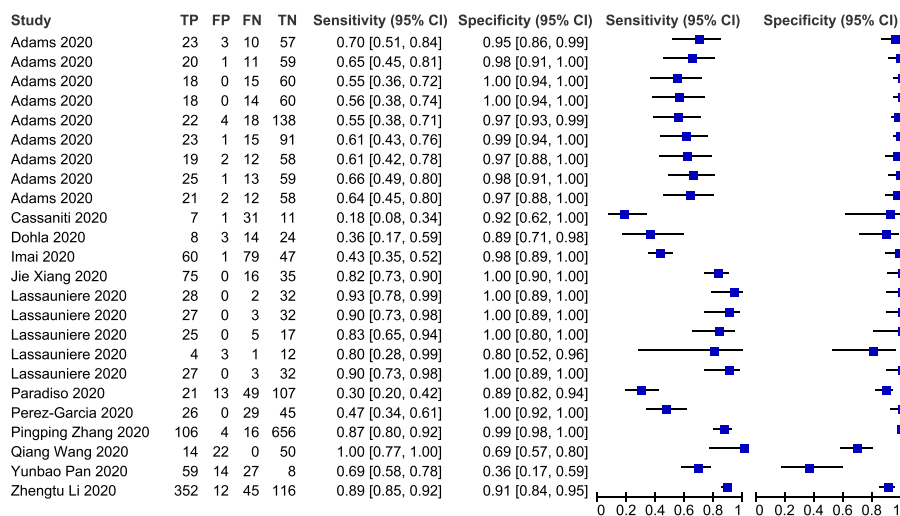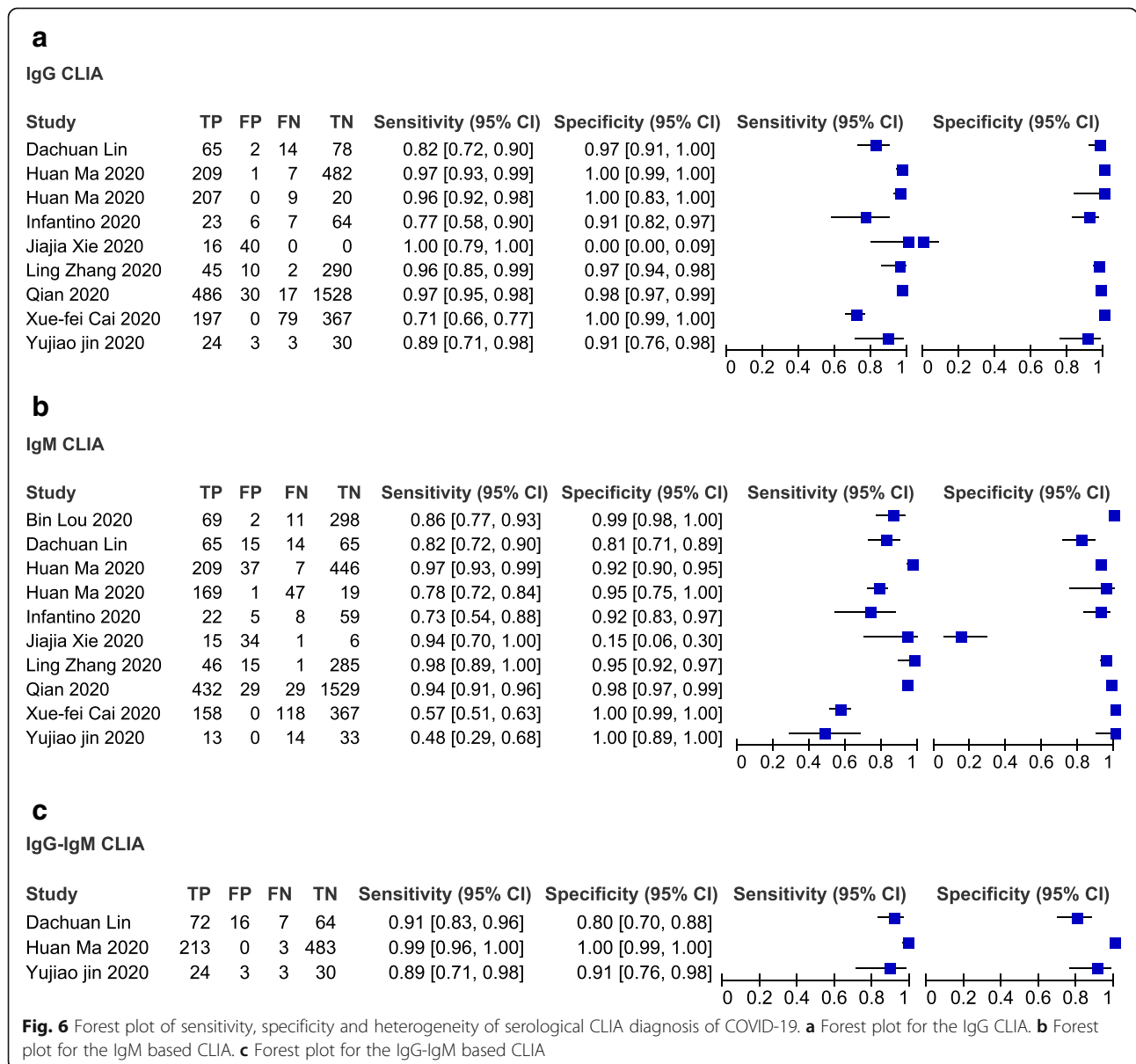except for 2 studies, Yunbao Pan, 2020, and Qiang Wang, 2020, which had the lowest and second lowest specificities respectively. Amongst the IgG based LFIA tests (n = 17), the sensitivity estimates ranged from 0.14 (95% CI 0.09-0.21) (Imai, 2020) to 1.00 (95% CI 0.77-1.00) (Qiang Wang, 2020) and specificity estimates ranged
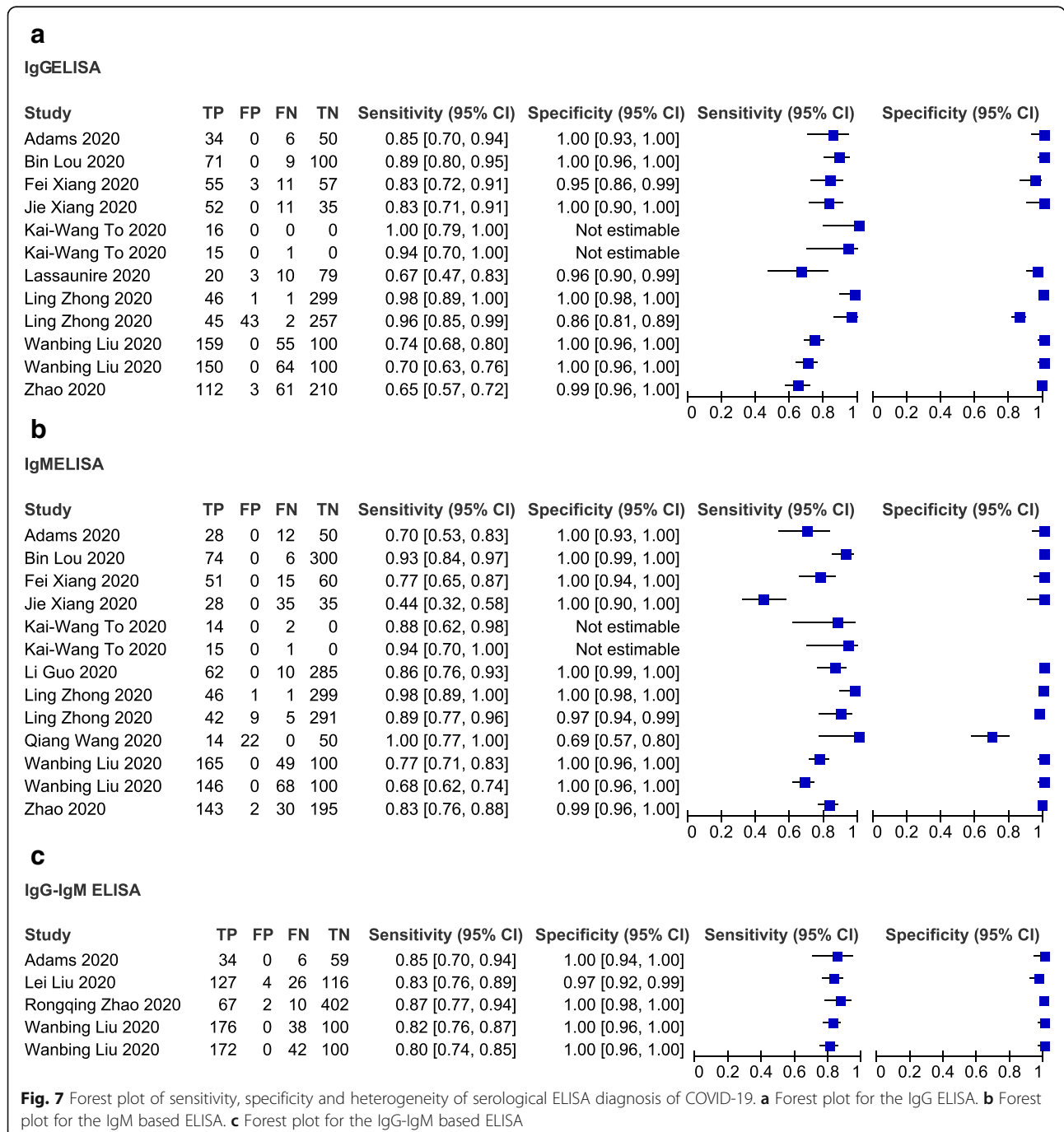
from 0.41 (95% CI 0.21-0.64) (Yunbao Pan, 2020) to 1.00 (95% CI 0.97-1.00) (Bin Lou, 2020) (Fig. 5a). For the IgM based LFIA tests (n = 16), the sensitivity estimates ranged from 0.05 (95% CI 0.01-0.18) Adams (assay 4 to 1.00) (95% CI 0.77-1.00) (Qiang Wang, 2020) and specificity estimates ranged from 0.64 (95% CI 0.41-0.83) (Yunbao Pan, 2020) to 1.00 (95% CI 0.94-1.00) (Adams assays 4 and 5) (Fig. 5b). For the IgG-IgM based LFIA tests (n = 24), the sensitivity estimates ranged from 0.18 (95% CI 0.08-0.34) (Cassaniti, 2020) to 1.00 (95% CI 0.77-1.00) (Qiang Wang, 2020), with most of the studies having sensitivities over 0.55 and specificity estimates ranged from 0.36 (95% CI 0.17-0.59) (Yunbao Pan, 2020) to 1.00 (95% CI 0.94-1.00) (Adams assays 2 and 3) (Fig. 5c).

a

IgG CLIA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dachuan Lin	65	2	14	78	0.82 [0.72, 0.90]	0.97 [0.91, 1.00]
Huan Ma 2020	209	1	7	482	0.97 [0.93, 0.99]	1.00 [0.99, 1.00]
Huan Ma 2020	207	0	9	20	0.96 [0.92, 0.98]	1.00 [0.83, 1.00]
Infantino 2020	23	6	7	64	0.77 [0.58, 0.90]	0.91 [0.82, 0.97]
Jiajia Xie 2020	16	40	0	0	1.00 [0.79, 1.00]	0.00 [0.00, 0.09]
Ling Zhang 2020	45	10	2	290	0.96 [0.85, 0.99]	0.97 [0.94, 0.98]
Qian 2020	486	30	17	1528	0.97 [0.95, 0.98]	0.98 [0.97, 0.99]
Xue-fei Cai 2020	197	0	79	367	0.71 [0.66, 0.77]	1.00 [0.99, 1.00]
Yujiao jin 2020	24	3	3	30	0.89 [0.71, 0.98]	0.91 [0.76, 0.98]

b

IgM CLIA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bin Lou 2020	69	2	11	298	0.86 [0.77, 0.93]	0.99 [0.98, 1.00]
Dachuan Lin	65	15	14	65	0.82 [0.72, 0.90]	0.81 [0.71, 0.89]
Huan Ma 2020	209	37	7	446	0.97 [0.93, 0.99]	0.92 [0.90, 0.95]
Huan Ma 2020	169	1	47	19	0.78 [0.72, 0.84]	0.95 [0.75, 1.00]
Infantino 2020	22	5	8	59	0.73 [0.54, 0.88]	0.92 [0.83, 0.97]
Jiajia Xie 2020	15	34	1	6	0.94 [0.70, 1.00]	0.15 [0.06, 0.30]
Ling Zhang 2020	46	15	1	285	0.98 [0.89, 1.00]	0.95 [0.92, 0.97]
Qian 2020	432	29	29	1529	0.94 [0.91, 0.96]	0.98 [0.97, 0.99]
Xue-fei Cai 2020	158	0	118	367	0.57 [0.51, 0.63]	1.00 [0.99, 1.00]
Yujiao jin 2020	13	0	14	33	0.48 [0.29, 0.68]	1.00 [0.89, 1.00]

c

IgG-IgM CLIA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Dachuan Lin	72	16	7	64	0.91 [0.83, 0.96]	0.80 [0.70, 0.88]
Huan Ma 2020	213	0	3	483	0.99 [0.96, 1.00]	1.00 [0.99, 1.00]
Yujiao jin 2020	24	3	3	30	0.89 [0.71, 0.98]	0.91 [0.76, 0.98]

Fig. 6 Forest plot of sensitivity, specificity and heterogeneity of serological CLIA diagnosis of COVID-19. **a** Forest plot for the IgG CLIA. **b** Forest plot for the IgM based CLIA. **c** Forest plot for the IgG-IgM based CLIA

We then considered performance of the different IgG or IgM or IgG-IgM based CLIA test using rRT-PCR-confirmed cases as the reference standard (Fig. 6a, b and c). Considering any positive result (IgM positive, IgG positive or both), CLIA serological tests achieved sensitivity ranging from 0.48 (95% CI 0.29-0.68%) (Yujiao Jin, 2020) to 1.00 (95% CI 0.79-1.00) with most studies being between 0.80 and 1. The specificity was over 0.80 in most tests except for 2 tests, one IgG based test and one
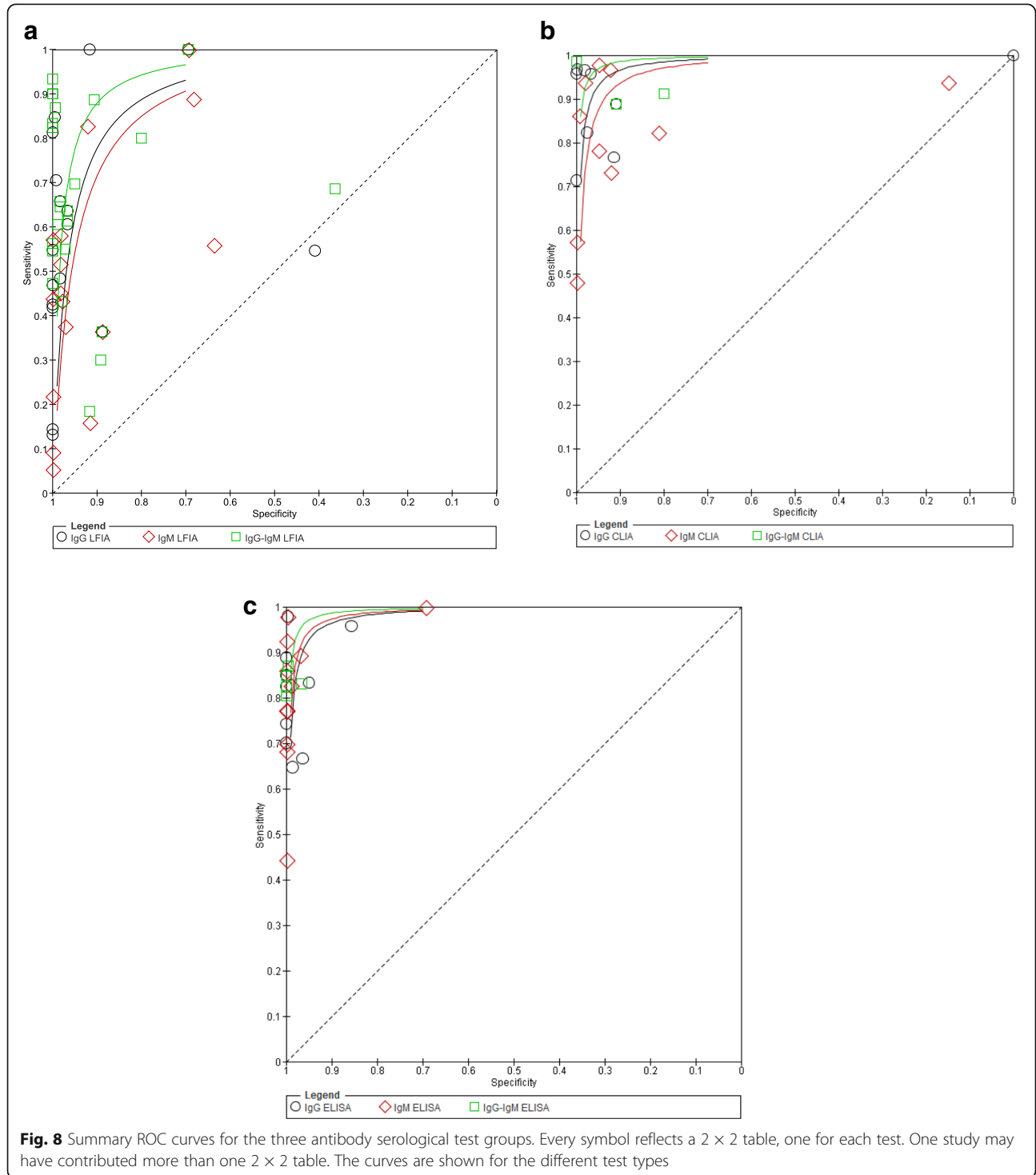
IgM based test which had the lowest 0.00 (95% CI 0.00-0.009) and second lowest 0.15 (95% CI 0.06-0.30) specificities respectively.

Lastly, we evaluated the performance of the different IgG or IgM or IgG-IgM based ELISA tests using rRT-PCR-confirmed cases as the reference standard (Fig. 7a, b and c). The sensitivities and specificities were generally high, ranging from 0.80 to 1.00 and 0.95 to 1.00 in most studies. For all the IgG based

a

IgGELISA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Adams 2020	34	0	6	50	0.85 [0.70, 0.94]	1.00 [0.93, 1.00]
Bin Lou 2020	71	0	9	100	0.89 [0.80, 0.95]	1.00 [0.96, 1.00]
Fei Xiang 2020	55	3	11	57	0.83 [0.72, 0.91]	0.95 [0.86, 0.99]
Jie Xiang 2020	52	0	11	35	0.83 [0.71, 0.91]	1.00 [0.90, 1.00]
Kai-Wang To 2020	16	0	0	0	1.00 [0.79, 1.00]	Not estimable
Kai-Wang To 2020	15	0	1	0	0.94 [0.70, 1.00]	Not estimable
Lassaunire 2020	20	3	10	79	0.67 [0.47, 0.83]	0.96 [0.90, 0.99]
Ling Zhong 2020	46	1	1	299	0.98 [0.89, 1.00]	1.00 [0.98, 1.00]
Ling Zhong 2020	45	43	2	257	0.96 [0.85, 0.99]	0.86 [0.81, 0.89]
Wanbing Liu 2020	159	0	55	100	0.74 [0.68, 0.80]	1.00 [0.96, 1.00]
Wanbing Liu 2020	150	0	64	100	0.70 [0.63, 0.76]	1.00 [0.96, 1.00]
Zhao 2020	112	3	61	210	0.65 [0.57, 0.72]	0.99 [0.96, 1.00]

b

IgMELISA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Adams 2020	28	0	12	50	0.70 [0.53, 0.83]	1.00 [0.93, 1.00]
Bin Lou 2020	74	0	6	300	0.93 [0.84, 0.97]	1.00 [0.99, 1.00]
Fei Xiang 2020	51	0	15	60	0.77 [0.65, 0.87]	1.00 [0.94, 1.00]
Jie Xiang 2020	28	0	35	35	0.44 [0.32, 0.58]	1.00 [0.90, 1.00]
Kai-Wang To 2020	14	0	2	0	0.88 [0.62, 0.98]	Not estimable
Kai-Wang To 2020	15	0	1	0	0.94 [0.70, 1.00]	Not estimable
Li Guo 2020	62	0	10	285	0.86 [0.76, 0.93]	1.00 [0.99, 1.00]
Ling Zhong 2020	46	1	1	299	0.98 [0.89, 1.00]	1.00 [0.98, 1.00]
Ling Zhong 2020	42	9	5	291	0.89 [0.77, 0.96]	0.97 [0.94, 0.99]
Qiang Wang 2020	14	22	0	50	1.00 [0.77, 1.00]	0.69 [0.57, 0.80]
Wanbing Liu 2020	165	0	49	100	0.77 [0.71, 0.83]	1.00 [0.96, 1.00]
Wanbing Liu 2020	146	0	68	100	0.68 [0.62, 0.74]	1.00 [0.96, 1.00]
Zhao 2020	143	2	30	195	0.83 [0.76, 0.88]	0.99 [0.96, 1.00]

c

IgG-IgM ELISA

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Adams 2020	34	0	6	59	0.85 [0.70, 0.94]	1.00 [0.94, 1.00]
Lei Liu 2020	127	4	26	116	0.83 [0.76, 0.89]	0.97 [0.92, 0.99]
Rongqing Zhao 2020	67	2	10	402	0.87 [0.77, 0.94]	1.00 [0.98, 1.00]
Wanbing Liu 2020	176	0	38	100	0.82 [0.76, 0.87]	1.00 [0.96, 1.00]
Wanbing Liu 2020	172	0	42	100	0.80 [0.74, 0.85]	1.00 [0.96, 1.00]

Fig. 7 Forest plot of sensitivity, specificity and heterogeneity of serological ELISA diagnosis of COVID-19. **a** Forest plot for the IgG ELISA. **b** Forest plot for the IgM based ELISA. **c** Forest plot for the IgG-IgM based ELISA

ELISA tests (*n* = 10), the sensitivity estimates ranged from 0.65 (95% CI 0.57-0.72) (Zhao, 2020) to 1.00 (95% CI 0.79-1.00) (Kai-Wang To, 2020) and specificity estimates from 0.86 (95% CI 0.51-0.89) to 1.00 (95% CI 0.98-1.00) (Ling Zhong, 2020) (Fig. 7a). In the IgM based tests (*n* = 11), the sensitivity and specificity in the individual studies ranged from 0.44

(95% CI 0.32-0.58) (Jie Xiang, 2020) to 1.00 (95% CI 0.77–1.00) (Qiang Wang, 2020) and 0.69 (95% CI 0.57-0.80) (Qiang Wang, 2020) to 1.00 (95% CI 0.99–1.00) (Ling Zhong, 2020), respectively (Fig. 7b). The sensitivity across the 5 studies included in the IgG-IgM based ELISA tests ranged from 0.80 (95% CI 0.74-0.85) (Wanbing Liu, 2020) to 0.87 (95% CI

Fig. 8 Summary ROC curves for the three antibody serological test groups. Every symbol reflects a 2 × 2 table, one for each test. One study may have contributed more than one 2 × 2 table. The curves are shown for the different test types

0.77-0.94) (Rongqing Zhao, 2020). On the other hand, specificity across the 5 studies ranged from 0.97 (95% CI 0.92-0.99) (Lei Liu, 2020) to 1.00 (95% CI 0.98-1.00) (Rongqing Zhao, 2020) (Fig. 7c).

We also constructed the SROC curves for all the three antibody based serological tests (Fig. 8). However, we did not calculate the area under the ROC (AUROC). From the SROC, we visually assessed heterogeneity between the different tests. Diagonal line indicated useless tests and the best tests were clustered further up to the top left hand corner.

The bivariate model and the hierarchical summary receiver operating characteristic curve (HSROC) model were performed to evaluate the diagnostic accuracy of the serological tests. The outputs of the meta-analysis (bivariate and HSROC parameter estimates, as well as the summary values of sensitivity and specificity) are presented in Table 2 and Fig. 9. The pooled sensitivity for the IgG, IgM and IgG-IgM based LFIA tests were 0.5856, 0.4637 and 0.6886 respectively compared to rRT-PCR. The pooled sensitivity for the IgG and IgM based CLIA tests were 0.9311 and 0.8516 respectively compared to rRT-PCR. The pooled sensitivity for the IgG, IgM and IgG-IgM based ELISA tests were 0.8292, 0.0.8388 and 0.8531 respectively compared to rRT-PCR. All the tests had high specificities ranging from 0.9693 to 0.9991 compared to rRT-PCR. The estimated SROC curves for bivariate models are not presented.

HSROCs were also used to visually access the overall performance of the diagnostic tests, to access the overall diagnostic accuracy of the tests and to compare the diagnostic accuracy of the different tests used for diagnosing COVID-19 in the review (Fig. 9). The overall diagnostic test accuracy was measured by the proximity of the curve to the top left corner which represents high sensitivity and specificity. The closer the curve was to the upper left hand corner,

the better the diagnostic accuracy [55]. From Fig. 9, it can be observed that ELISA and CLIA have better diagnostic accuracy compared to LFIA and IgG-IgM based ELISA tests have the best overall diagnostic test accuracy. Of importance, it is noteworthy that in the study the evidence base was too weak to definitively state that one class of test was more accurate than the other class of tests.

We identified one study (Burbelo, 2020) reporting total antibody (Ab) based luciferase immunoprecipitation assay system (LIPS) using N and S antigens with sensitivities and specificities of 0.91 (95% CI 0.77-0.99) and 1.00 (0.80-1.00) and 1.00 (0.92-1.00) and 1.00 (0.92-1.00) respectively. We also identified studies reporting other Ab based serological assays and IgA based serological assays but results are not reported in this review.
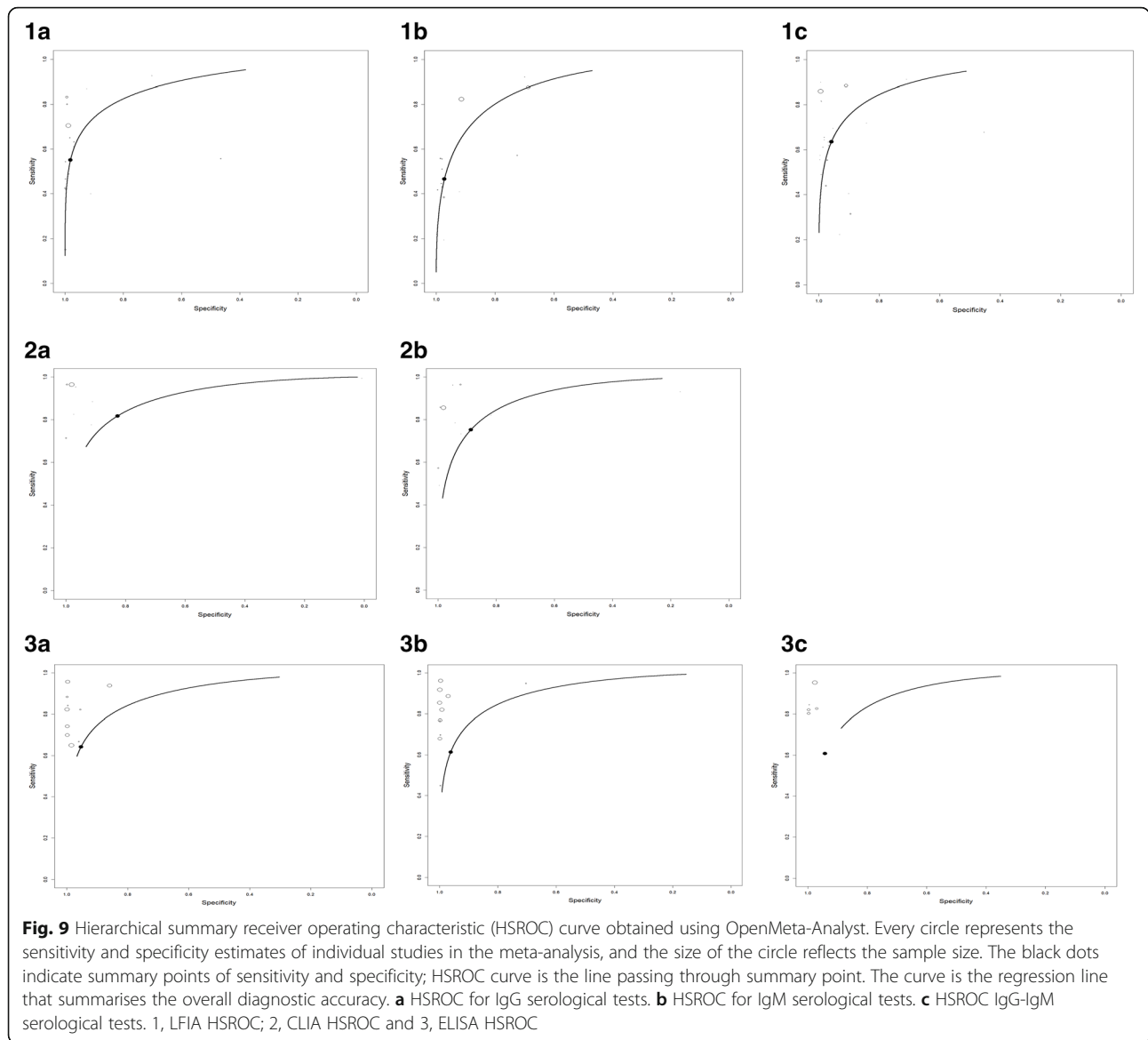
Heterogeneity investigations
Generally high overall I^2 values above 85%, which indicate high heterogeneities, were observed for both the sensitivities and specificities when we performed antigen subgroup meta-analysis with the exception of IgG-IgM based ELISA. IgG-IgM based ELISA had an overall sensitivity I^2 value of 52. 12% which is considered moderate heterogeneity and overall specificity I^2 value of 0% which is considered to be low heterogeneity. However, it should be noted that only 5 studies were included for this subgroup meta-analysis. Overall I^2 values for sensitivities and specificities heterogeneities for the antigen type subgroup meta-analysis are shown in Table 3. We did not investigate heterogeneity for LFIA because most studies included in the review did not specify the type of antigen they used in their serological tests.

Detailed results of heterogeneity for the different antigen type sensitivities and specificities for each test type and antibody type combination are presented in Additional file 2.

Table 2 Summary estimates of test accuracy

Test type	Antibody type	Number of studies/tests	Sensitivity (95% CI)	Specificity (95% CI)	Correlation
LFIA	IgG	17	58.56 (43.97-71.79)	98.96 (95.61-99.76)	−0.4454
CLIA	IgG	9	93.11 (93.09-93.12)	97.57 (97.57-97.58)	−0.511
ELISA	IgG	10	82.92 (74.16-89.15)	99.48 (96.75-99.92)	−0.1709
LFIA	IgM	16	46.37 (30.16-0.6339)	97.34 (92.75-99.05)	−0.7925
CLIA	IgM	10	85.16 (73.56-0.9221)	96.93 (85.5-99.41)	−0.7074
ELISA	IgM	11	83.88 (0.7307-0.909)	99.91 (97.78-100)	−0.7247
LFIA	IgG-IgM	24	68.86 (58.78-77.42)	97.57 (94.66-98.92)	0.1011
CLIA	IgG-IgM	3			
ELISA	IgG-IgM	5	85.31 (78.51-90.23)	99.01 (92.87-99.87)	−0.6771

Fig. 9 Hierarchical summary receiver operating characteristic (HSROC) curve obtained using OpenMeta-Analyst. Every circle represents the sensitivity and specificity estimates of individual studies in the meta-analysis, and the size of the circle reflects the sample size. The black dots indicate summary points of sensitivity and specificity; HSROC curve is the line passing through summary point. The curve is the regression line that summarises the overall diagnostic accuracy. **a** HSROC for IgG serological tests. **b** HSROC for IgM serological tests. **c** HSROC IgG-IgM serological tests. 1, LFIA HSROC; 2, CLIA HSROC and 3, ELISA HSROC

Test sensitivity by time since onset of symptoms

Figure 10 shows forest plots for antibody positive rates for IgG (25 tests), IgM (22 tests) and IgG-IgM (30 tests), stratified by days since initial symptom onset to specimen collection. The sensitivity of the serological tests generally increased with increased time from symptoms onset. Regardless of test method (ELISA or CLIA or LFIA), the sensitivities for IgG and IgM based tests were generally low in the first week (1-7 days) of symptom onset followed by the second (7-14 days) and the sensitivities were generally highest in the third week or later (>14 days) for each test. Data on specificity stratified by specimen collection since symptom onset was not available for all the studies.

Discussion

COVID-19 is a major healthcare challenge globally. One key aspect of limiting SARS-CoV-2 spread is to ensure early and accurate diagnosis of the viral infection. In this study, we performed a meta-analysis to evaluate the diagnostic accuracy of IgG and IgM based serological assays offered to detect antigens against rRT-PCR positive SARS-CoV-2 patients. The meta-analysis showed that all serological assays yielded high specificities ranging from 0.9693 (95% CI 0.855-0.9941) to 0.9991 (95% CI 0.9778-1) in comparison to rRT-PCR. Meta-analysis by Wang et al., Bastos et al. and Deeks et al. observed similar pooled specificities ranging from 0.95 (95% CI 0.91–

Table 3 Overall antibody type subgroup meta-analysis heterogeneity

Test type	Antibody type	Heterogeneity (I^2)	
		Sensitivity	Specificity
LFIA	IgG		
	IgM		
	IgG-IgM		
CLIA	IgG	93.56%	86.5%
	IgM	93.42%	95.17%
	IgG-IgM		
ELISA	IgG	78.07%	84.97%
	IgM	85.47%	90.08%
	IgG-IgM	52.12%	0%

0.98) to 99.6 (97.3 to 99.9) [56–58]. The sensitivities of all serological assays varied widely across studies. COVID-19 serological assays rely on antibodies binding to SARS-CoV-2 nucleocapsid protein, spike protein or spike protein fragments (i.e. receptor binding domain). Some tests or test devices even use a combination of the N and S proteins and protein fragments. This may have resulted in the inconsistencies observed in the different serological assays [59]. Overall IgG-IgM based ELISA had superior diagnostic accuracy with sensitivity and specificity of 0.8531 (95% CI 0.7851-0.9023) and 0.9991 (95% CI 0.9778-1). IgG based CLIA had the highest sensitivity 0.9311 (95% CI 0.9309-0.9312). The pooled sensitivity results are in agreement with other meta-analysis that showed POC LFIA had lower sensitivities than CLIA and ELISA within each antibody class [56–58]. However, it must be noted that a meta-analysis by Deeks et al. found that for IgG and IgG-IgM, concentration gradient immunoassay POC LFIA had higher pooled sensitivity than ELISA [56].

Similar to other meta-analysis, IgM based serological assays had the lowest sensitivities compared IgG based serological tests in each respective test method [56–58]. Low antibody concentrations and especially timing of the IgM based serological assays could potentially explain the low ability of the IgM based tests to identify people infected with SARS-CoV-2. Notably, immediately after a person is infected, antibodies may not have been developed yet or too late IgM antibodies may have decreased or disappeared [60].
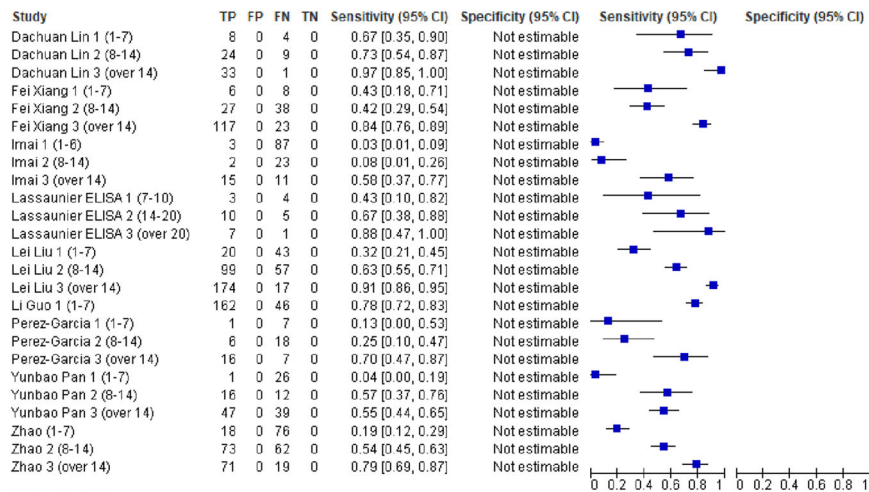
Generally combined IgG-IgM based serological assays seem to be a better choice in terms of sensitivity than measuring either antibody type alone, though in this review we did not estimate the pooled sensitivity of the IgG-IgM based CLIA due to the limited number of studies. Subgroup meta-analysis using Open Meta-Analyst showed that IgM based serological

assays that use the S antigen are more sensitive than IgM based serological assays that used the N antigen-based tests probably due to higher sensitivity and earlier immune response to the S antigen [32]. However, subgroup meta-analysis showed that IgG based serological assays that use N antigen are more sensitive than IgG based serological assays that use S antigen.
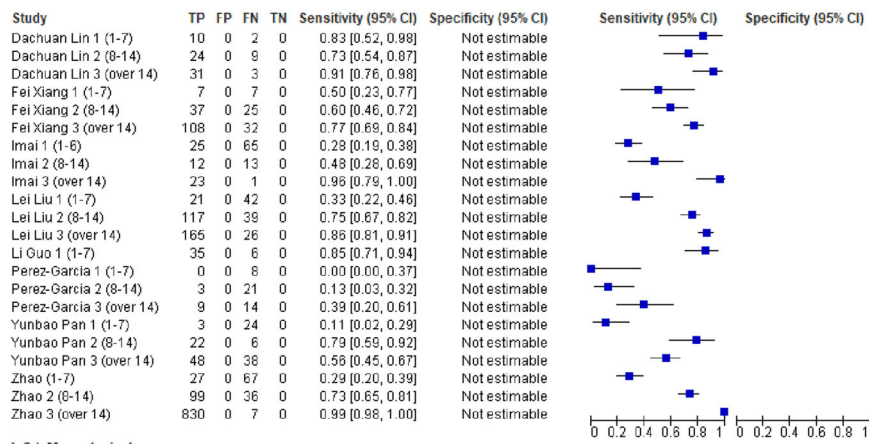
The sensitivity and specificity estimates from this review were used to indicate the consequences of testing when the test is being used in clinical practice. Prevalence estimates were calculated using metaDTA [61] to predict how many patients in practice you would expect to have true positive, false positive, true negative and false negative results for a given prevalence based on the meta-analysis results giving the results some clinical context. Using IgG-IgM based ELISA test which had the best overall diagnostic accuracy as an example, one can see that with 1000 patients and a COVID-19 prevalence of 50%, we would expect 432 (95% CI 394-486) patients to test positive for SARS-CoV-2, of which 427 (95% CI 393-451) will be true positives (are diseased and test positive) and 5 (95% CI 1-35) will be false positives (are not diseased but test positive). We also noted 568 (95% CI 514-606) patients test negative for COVID-19, of which 495 (95% CI 465-499) were true negatives (are not diseased and test negative) and 73 (95% CI 49-107) were false negatives (are diseased but test negative).

A prevalence of 1% and 1000 patients would mean that 19 (95% CI 9-79) patients test positive for SARS-CoV-2, 9 (95% CI 8-9) of these will be true positives and 10 (95% CI 1-70) false positives and 981 (95% CI 921-991) patients will test negative for SARS-COV-2, 980 (95% CI 920-989) of these will be true negatives and 1 (95% CI 1-2) patient will be false negative. This means that the false positive will be treated for COVID-19 whilst they have another disease, thus delaying their final diagnosis and

IgG serology assays

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Dachuan Lin 1 (1-7)	8	0	4	0	0.67 [0.35, 0.90]	Not estimable		
Dachuan Lin 2 (8-14)	24	0	9	0	0.73 [0.54, 0.87]	Not estimable		
Dachuan Lin 3 (over 14)	33	0	1	0	0.97 [0.85, 1.00]	Not estimable		
Fei Xiang 1 (1-7)	6	0	8	0	0.43 [0.18, 0.71]	Not estimable		
Fei Xiang 2 (8-14)	27	0	38	0	0.42 [0.29, 0.54]	Not estimable		
Fei Xiang 3 (over 14)	117	0	23	0	0.84 [0.76, 0.89]	Not estimable		
Imai 1 (1-6)	3	0	87	0	0.03 [0.01, 0.09]	Not estimable		
Imai 2 (8-14)	2	0	23	0	0.08 [0.01, 0.26]	Not estimable		
Imai 3 (over 14)	15	0	11	0	0.58 [0.37, 0.77]	Not estimable		
Lassaunier ELISA 1 (7-10)	3	0	4	0	0.43 [0.10, 0.82]	Not estimable		
Lassaunier ELISA 2 (14-20)	10	0	5	0	0.67 [0.38, 0.88]	Not estimable		
Lassaunier ELISA 3 (over 20)	7	0	1	0	0.88 [0.47, 1.00]	Not estimable		
Lei Liu 1 (1-7)	20	0	43	0	0.32 [0.21, 0.45]	Not estimable		
Lei Liu 2 (8-14)	99	0	57	0	0.63 [0.55, 0.71]	Not estimable		
Lei Liu 3 (over 14)	174	0	17	0	0.91 [0.86, 0.95]	Not estimable		
Li Guo 1 (1-7)	162	0	46	0	0.78 [0.72, 0.83]	Not estimable		
Perez-Garcia 1 (1-7)	1	0	7	0	0.13 [0.00, 0.53]	Not estimable		
Perez-Garcia 2 (8-14)	6	0	18	0	0.25 [0.10, 0.47]	Not estimable		
Perez-Garcia 3 (over 14)	16	0	7	0	0.70 [0.47, 0.87]	Not estimable		
Yunbao Pan 1 (1-7)	1	0	26	0	0.04 [0.00, 0.19]	Not estimable		
Yunbao Pan 2 (8-14)	16	0	12	0	0.57 [0.37, 0.76]	Not estimable		
Yunbao Pan 3 (over 14)	47	0	39	0	0.55 [0.44, 0.65]	Not estimable		
Zhao (1-7)	18	0	76	0	0.19 [0.12, 0.29]	Not estimable		
Zhao 2 (8-14)	73	0	62	0	0.54 [0.45, 0.63]	Not estimable		
Zhao 3 (over 14)	71	0	19	0	0.79 [0.69, 0.87]	Not estimable		

IgM serological assays

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Dachuan Lin 1 (1-7)	10	0	2	0	0.83 [0.52, 0.98]	Not estimable		
Dachuan Lin 2 (8-14)	24	0	9	0	0.73 [0.54, 0.87]	Not estimable		
Dachuan Lin 3 (over 14)	31	0	3	0	0.91 [0.76, 0.98]	Not estimable		
Fei Xiang 1 (1-7)	7	0	7	0	0.50 [0.23, 0.77]	Not estimable		
Fei Xiang 2 (8-14)	37	0	25	0	0.60 [0.46, 0.72]	Not estimable		
Fei Xiang 3 (over 14)	108	0	32	0	0.77 [0.69, 0.84]	Not estimable		
Imai 1 (1-6)	25	0	65	0	0.28 [0.19, 0.38]	Not estimable		
Imai 2 (8-14)	12	0	13	0	0.48 [0.28, 0.69]	Not estimable		
Imai 3 (over 14)	23	0	1	0	0.96 [0.79, 1.00]	Not estimable		
Lei Liu 1 (1-7)	21	0	42	0	0.33 [0.22, 0.46]	Not estimable		
Lei Liu 2 (8-14)	117	0	39	0	0.75 [0.67, 0.82]	Not estimable		
Lei Liu 3 (over 14)	165	0	26	0	0.86 [0.81, 0.91]	Not estimable		
Li Guo 1 (1-7)	35	0	6	0	0.85 [0.71, 0.94]	Not estimable		
Perez-Garcia 1 (1-7)	0	0	8	0	0.00 [0.00, 0.37]	Not estimable		
Perez-Garcia 2 (8-14)	3	0	21	0	0.13 [0.03, 0.32]	Not estimable		
Perez-Garcia 3 (over 14)	9	0	14	0	0.39 [0.20, 0.61]	Not estimable		
Yunbao Pan 1 (1-7)	3	0	24	0	0.11 [0.02, 0.29]	Not estimable		
Yunbao Pan 2 (8-14)	22	0	6	0	0.79 [0.59, 0.92]	Not estimable		
Yunbao Pan 3 (over 14)	48	0	38	0	0.56 [0.45, 0.67]	Not estimable		
Zhao (1-7)	27	0	67	0	0.29 [0.20, 0.39]	Not estimable		
Zhao 2 (8-14)	99	0	36	0	0.73 [0.65, 0.81]	Not estimable		
Zhao 3 (over 14)	830	0	7	0	0.99 [0.98, 1.00]	Not estimable		

IgG-IgM serological assays

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Artron 2 (14-20)	13	0	3	0	0.81 [0.54, 0.96]	Not estimable		
Artron 1 (7-10)	5	0	2	0	0.71 [0.29, 0.96]	Not estimable		
Artron 3 (over 20)	8	0	8	0	0.50 [0.25, 0.75]	Not estimable		
Auto Bio 1 (7-10)	6	0	1	0	0.86 [0.42, 1.00]	Not estimable		
Auto Bio 2 (14-20)	14	0	1	0	0.93 [0.68, 1.00]	Not estimable		
Auto Bio 3 (over 20)	8	0	8	0	0.50 [0.25, 0.75]	Not estimable		
CTK Biotech 1 (7-10)	5	0	2	0	0.71 [0.29, 0.96]	Not estimable		
CTK Biotech 2 (14-20)	14	0	1	0	0.93 [0.68, 1.00]	Not estimable		
CTK Biotech 3 (over 20)	8	0	8	0	0.50 [0.25, 0.75]	Not estimable		
Dachuan Lin 1 (1-7)	10	0	2	0	0.83 [0.52, 0.98]	Not estimable		
Dachuan Lin 2 (8-14)	29	0	4	0	0.88 [0.72, 0.97]	Not estimable		
Dachuan Lin 3 (over 14)	33	0	1	0	0.97 [0.85, 1.00]	Not estimable		
Dynamiker 2 (14-20)	14	0	1	0	0.93 [0.68, 1.00]	Not estimable		
Dynamiker 1 (7-10)	5	0	2	0	0.71 [0.29, 0.96]	Not estimable		
Dynamiker 3 (over 20)	8	0	8	0	0.50 [0.25, 0.75]	Not estimable		
Imai 1 (1-6)	25	0	65	0	0.28 [0.19, 0.38]	Not estimable		
Imai 2 (8-14)	12	0	13	0	0.48 [0.28, 0.69]	Not estimable		
Imai 3 (over 14)	23	0	1	0	0.96 [0.79, 1.00]	Not estimable		
Lei Liu 1 (1-7)	34	0	39	0	0.47 [0.35, 0.59]	Not estimable		
Lei Liu 2 (8-14)	133	0	23	0	0.85 [0.79, 0.90]	Not estimable		
Lei Liu 3 (over 14)	185	0	6	0	0.97 [0.93, 0.99]	Not estimable		
Perez-Garcia 1 (1-7)	0	0	8	0	0.00 [0.00, 0.37]	Not estimable		
Perez-Garcia 2 (8-14)	3	0	21	0	0.13 [0.03, 0.32]	Not estimable		
Perez-Garcia 3 (over 14)	9	0	14	0	0.39 [0.20, 0.61]	Not estimable		
Yunbao Pan 1 (1-7)	3	0	24	0	0.11 [0.02, 0.29]	Not estimable		
Yunbao Pan 2 (8-14)	26	0	2	0	0.93 [0.76, 0.99]	Not estimable		
Yunbao Pan 3 (over 14)	59	0	27	0	0.69 [0.58, 0.78]	Not estimable		
Zhao (1-7)	0	0	0	0	Not estimable	Not estimable		
Zhao 2 (8-14)	0	0	0	0	Not estimable	Not estimable		
Zhao 3 (over 14)	0	0	0	0	Not estimable	Not estimable		

Fig. 10 Forest plot of studies evaluating tests for detection of IgG, IgM and IgG-IgM according to days since COVID-19 symptom onset to specimen collection. In brackets () are the number of days since symptom onset to specimen collection. Artron, Auto Bio CTK Biotech CTK Biotech are test names all reported in a study by Lassaunire et al

subsequent treatment. In the false-negatives, COVID-19 diagnosis will be missed or delayed and the patients will not be quarantined and they will thus spread the SARS-CoV-2 to other patients in the hospital/clinic. Nonetheless, despite calculations like these providing insight in the consequences of testing, they should be taken with caution.

Using the inferences for LIFA using the IgG-IgM based test [sensitivity 0.6886 (0.5878-0.7742) and specificity 0.9757 (0.9466-0.9892)]. We see that for 1000 patients and a COVID-19 prevalence of 50%, we would expect 356 (95% CI 299-414) patients to test positive for SARS-CoV-2, 324 (95% CI 294-387) of which will be true positives and 12 (95% CI 5-27) will be false positives. We would also expect 644 (95% CI 586-701) patients to test negative for the COVID-19, 488 (95% CI 473-495) of which will be true negatives and 156 (95% CI 113-206) are false negatives. A prevalence of 1% and 1000 patients would mean 31 (95% CI 16-62) patients test positive for SARS-CoV-2, 7 (95% CI 6-8) of these will be true positives and 24 (95% CI 10-54) false positives and 969 (95% CI 938-984) patients to test negative for COVID, 966 (95% CI 936-980) of these will be true negatives and 3 (95% CI 2-4) are false negatives. Compared to IgG-IgM based ELISA, the IgG-IgM based LFIA has higher rates of false positives and false negatives.

Limitation

Serological tests for SARS-CoV-2 have accuracy issues that warrant attention. They measure specific antibody responses which may take some weeks to develop after disease onset reducing the sensitivity of the assay. If blood samples were collected during the early stage of the infection, they may produce false negative results. They do not directly detect the presence of the virus. Further, antibodies may be present when SARS-CoV-2 is no longer present giving false positive case diagnosis. Moreover, since the identity of the N protein of SARS-CoV-2 and SARS-CoV reached up to 91.2%, there is probability of a cross reaction between the N protein of SARS-CoV-2 and antibodies against other human coronaviruses. Other molecules including interferon, rheumatoid factor and non-specific IgM may cause false positive results [42].

Most studies included in the meta-analysis were case-control studies. These may be easy to perform in a laboratory setting than cross-sectional designs, but their results are less representative for clinical practice. The performance of diagnostic tests very much depends on the population in which the test is being used. Future studies should therefore be prospective cross-sectional studies including a consecutive sample of presenting patients.

Index tests need to be evaluated to determine their sensitivity and specificity, ideally by comparison with a standard confirmatory test. An important limitation with the rRT-PCR, the standard confirmatory test for COVID-19 is the risk of false-negative results [62]. Two reviews of the accuracy of rRT-PCR COVID-19 tests reported false negative rates of between 2% and 29%, based on negative rRT-PCR tests which were positive on repeat rRT-PCR testing [63, 64]. False negative results of rRT-PCR tests can lead decreased specificity of the serological tests (index tests). The rRT-PCR negative results picked up as positive tests by the serological tests will be treated as false positives thereby lowering the specificity of the serological tests. In order to reduce false-negative results, Bastos et al. recommended that the standard confirmatory test should consist of RT-PCR performed on at least two consecutive specimens and when possible it must include viral cultures [57].

Conclusion

Given the poor performance of the current LFIA devices, we recommend more research to develop highly sensitivity and specific POC LFIA that are adequate for most individual patient applications and attractive for large seroprevalence studies. The use of CLIA and ELISA for diagnosis has high sensitivity and is comparable to using rRT-PCR. They may be calibrated to be specific for detecting and quantifying SARSCoV-2 IgM and IgG. More serological data should be collected to elucidate the clinical and epidemiological utility of IgG and IgM serological measurements to detect symptomatic and asymptomatic cases of COVID-19.

Appendix 1
Search strategy in PubMed

Search	Query	Items found
#20	Search (((((COVID-19[Title/Abstract]) OR SARS-CoV-2[Title/Abstract]) OR 2019-nCoV[Title/Abstract]) OR Wuhan Coronavirus[Title/Abstract])) AND (((((((((((((Serologic test) OR Serologic method) OR Serological test) OR Serological method) OR Serodiagnosis) OR Immunodiagnosis) OR Immunological test) OR Immunological method) OR Antibody detection) OR Antigen detection) OR IgM) OR IgG) OR Immunochromatography)	78
#19	Search (((((((((((Serologic test) OR Serologic method) OR Serological test) OR Serological method) OR Serodiagnosis) OR Immunodiagnosis) OR Immunological test) OR Immunological method) OR	822733

Search strategy in PubMed (Continued)

Search	Query	Items found
	Antibody detection) OR Antigen detection) OR IgM) OR IgG) OR Immunochromatography	
#18	Search Immunochromatography	69624
#17	Search IgG	143941
#16	Search IgM	66041
#15	Search Antigen detection	82456
#14	Search Antibody detection	99123
#13	Search Immunological method	39637
#12	Search Immunological test	456120
#11	Search Immunodiagnosis	450622
#10	Search Serodiagnosis	184574
#9	Search Serological method	24322
#8	Search Serological test	191599
#7	Search Serologic method	16939
#6	Search Serologic test	184758
#5	Search ((((COVID-19[Title/Abstract]) OR SARS-CoV-2[Title/Abstract]) OR 2019-nCoV[Title/Abstract]) OR Wuhan Coronavirus[Title/Abstract]	6734
#4	Search Wuhan Coronavirus[Title/Abstract]	14
#3	Search 2019-nCoV[Title/Abstract]	547
#2	Search SARS-CoV-2[Title/Abstract]	1898
#1	Search COVID-19[Title/Abstract]	5985

Appendix 2
Rating of QUADAS-2 items
1. Patient selection
1a. Risk of bias, signalling questions

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

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

- Is there concern that the included patients do not match the review question?
  - All case-control studies are by default rated 'high concern'. All cross-sectional studies are by default 'low concern', except when the used case definition was not very clear.

2. Index test
2a. Risk of bias, signalling questions

- Were the index test results interpreted without knowledge of the results of the reference standard?
- If a threshold was used, was it pre-specified? By selecting the cut-off value with the highest sensitivity and/or specificity, researchers artificially optimise the accuracy of their tests, which may lead to an overestimation of sensitivity and specificity. POC LFIA threshold were pre-specified and the signal was the development of a line or colour.
- Overall judgement:
  - If a study was not reporting results from automated assays and it did not explicitly mention blinding then tests were automatically rated as high risk.
  - If there was blinding and the second question was answered with 'yes', overall judgement was 'low'.

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

- Are there concerns that the index test, its application or interpretation deviate from the review question?
  - All in-house tests were automatically rated as 'high concern'.
  - If serum or plasma samples not blood were used for POC LFIA then tests were automatically rated high concern

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

3. Reference standard
3a. Risk of bias, signalling questions

- Is the reference standard likely to correctly classify the target condition?

- ▪ Assumption: This will likely be the case for case-control studies that use the 'correct' case definitions
- ▪ This is also likely for cross-sectional studies which use the 'correct' case definitions.
- Were the reference standard results interpreted without knowledge of the results of the index test?
  - ▪ Assumption: This will likely be the case for most case-control studies, but only if serology was not part of the case definition.
  - ▪ For cross-sectional studies, this should be explicitly stated.
- Overall judgement risk of bias:
  - ▪ Case-control studies with clear case definitions were rated as having a 'low' risk of bias.
  - ▪ case-control studies with unclear/wrong case definitions rated as 'unclear' Or 'high risk' of bias respectively
  - ▪ Cross-sectional studies with a clear case definition and the second question answered with 'yes': low risk of bias.
  - ▪ Otherwise 'unclear'.

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

- ▪ If serology is included in the case definition, there is an incorporation bias and thus a high risk of bias.
- ▪ If a case-control study uses clear criteria and does not include serology in these criteria: 'low' concern of bias.

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

- Was there an appropriate interval between index test(s) and reference standard?
  - ▪ We expected that studies with a cross-sectional design conducted most tests on a date sufficiently close to the final diagnosis. If we had reason to suspect that the patient status changed between the time of testing and the time of diagnosis, we rated this as 'no'.
  - ▪ For case-control studies, this was always rated as 'no', because serology was always determined after the case definitions were defined, sometimes with a long delay.
- Did patients receive the same reference standard?
- Were all patients included in the analysis?
  - ▪ This was rated 'no' for all case-control studies.
- Overall judgement:
  - ▪ Case-control studies were always rated as having a high risk of bias.

- ▪ For cross-sectional studies, we perceived a low risk of bias if all three questions were answered with 'yes'; a high risk of bias was perceived if at least one of them was answered 'no'. All other possibilities were rated as 'unclear' [65].

Abbreviations
COVID-19: Coronavirus disease 2019; SARS: Severe acute respiratory syndrome; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; STARD: Standards for the reporting of diagnostic accuracy studies; QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies-2; rRT-PCR: Real-time reverse-transcriptase polymerase chain reaction; PRISMA-DTA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic test accuracy; DTA: Diagnostic test accuracy; RT-qPCR: Quantitative real-time polymerase chain reaction; POC: Point-of-care; LFIA: Lateral flow immunoassays; CLIA: Chemiluminescence enzyme immunoassay; FIA: Fluorescence; EISA: Enzyme-linked immunoassay; LIPS: Luciferase immunoprecipitation assay systems; S: SARS-CoV-2 spike protein; N: SARS-CoV-2 nucleocapsid protein; E: SARS-CoV-2 envelope protein; HSROC: Hierarchical summary receiver operating characteristic; TP: True positive; FN: False negative; FP: False positive; TN: True negative

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13643-021-01689-3.

Additional file 1. PRISMA-DTA Checklist

Additional file 2. Antigen type subgroup-meta analysis results

---

Declarations

Author details
[1]Department of Biochemistry, University of Zimbabwe, P.O. Box MP 167, Mt Pleasant, Harare, Zimbabwe. [2]Optics and Imaging, Doris Duke Medical Research Institute, College of Health Sciences, University of KwaZulu-Natal, Durban, KwaZulu-Natal, South Africa. [3]College of Health Sciences, University of Zimbabwe, Box A178 Mazowe Street Avondale, Harare, Zimbabwe. [4]Institute for Immunology and Infection Research and Centre for Immunity,

Vengesai *et al. Systematic Reviews*   (2021) 10:155

Page 22 of 23

Infection and Evolution, School of Biological Sciences, Ashworth Laboratories, University of Edinburgh, King's Buildings, Charlotte Auerbach Rd, Edinburgh EH9 3JT, UK.

References

1. Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. Acta Paediatr. 2020;109(6):1088–95 Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/apa.15270. Cited 2020 May 27.

2. Ren L-L, Wang Y-M, Wu Z-Q, Xiang Z-C, Guo L, Xu T, et al. Identification of a novel coronavirus causing severe pneumonia in human. Chin Med J. 2020; 133(9):1015–24 Available from: http://journals.lww.com/10.1097/CM9.0000000000000722. Cited 2020 May 28.

3. Situation Summary | CDC. Available from: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/summary.html. Cited 2020 May 27

4. NIH clinical trial of investigational vaccine for COVID-19 begins | National Institutes of Health (NIH). Available from: https://www.nih.gov/news-events/news-releases/nih-clinical-trial-investigational-vaccine-covid-19-begins. Cited 2020 May 27

5. COVID-19 situation update worldwide, as of 11 November 2020. Available from: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases. Cited 2020 Nov 12

6. Coronavirus. Available from: https://www.who.int/health-topics/coronavirus#tab=tab_1. Cited 2020 Nov 12

7. Coronavirus disease 2019 (COVID-19) Situation Report-63 HIGHLIGHTS. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

8. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Clinical infectious diseases Clinical Infectious Diseases ® 2020;XX(XX):1-8. Available from: https://academic.oup.com/cid/advance-article-abstract/doi/10.1093/cid/ciaa310/5810754. Cited 2020 May 28

9. Diagnostic testing for SARS-CoV-2. Available from: https://www.who.int/publications/i/item/diagnostic-testing-for-sars-cov-2. Cited 2020 Nov 18

10. To KKW, Tsang OTY, Leung WS, Tam AR, Wu TC, Lung DC, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. Lancet Infect Dis. 2020;20(5):565–74.

11. Wikramaratna P, Paton RS, Ghafari M, Lourenco J. Estimating false-negative detection rate of SARS-CoV-2 by RT-PCR. medRxiv. 2020:2020.04.05.20053355 Available from: https://doi.org/10.1101/2020.04.05.20053355. Cited 2020 May 28.

12. Li D, Wang D, Dong J, Wang N, Huang H, Xu H, et al. False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-learning-based ct diagnosis and insights from two cases. Korean J Radiol. 2020;21(4):505–8. https://doi.org/10.3348/kjr.2020.0146.

13. Guo L, Ren L, Yang S, Xiao M, Chang D, Yang F, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). Clin Infect Dis. 2020; Available from: http://www.ncbi.nlm.nih.gov/pubmed/32198501. Cited 2020 Apr 2.

14. Zhang W, Du RH, Li B, Zheng XS, Lou YX, Hu B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. Emerg Microbes Infect. 2020;9(1):386–9 Available from: https://www.tandfonline.com/doi/full/10.1080/22221751.2020.1729071. Cited 2020 May 28.

15. Qian C, Zhou M, Cheng F, Lin X, Gong Y, Xie X, et al. Development and multicenter performance evaluation of the first fully automated SARS-CoV-2 IgM and IgG immunoassays. medRxiv. 2020:2020.04.16.20067231.

16. The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks in: Clinical Chemistry and Laboratory Medicine (CCLM) - Ahead of print. Available from: https://www.degruyter.com/view/journals/cclm/ahead-of-print/article-10.1515-cclm-2020-0240/article-10.1515-cclm-2020-0240.xml. Cited 2020 May 28

17. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies The PRISMA-DTA Statement. JAMA. 2018;319(4):388–96. https://doi.org/10.1001/jama.2017.19163.

18. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Cochrane handbook for systematic reviews of diagnostic test accuracy Chapter 10 Analysing and presenting results. Available from: http://srdta.cochrane.org/. Cited 2020 May 31

19. Zeller H, Van Bortel W. A systematic literature review on the diagnosis accuracy of serological tests for Lyme borreliosis; 2016.

20. Steingart KR, Henry M, Laal S, Hopewell PC, Ramsay A, Menzies D, et al. A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Postgrad Med J. 2007;83:705–12 BMJ Publishing Group Ltd.

21. Volpe Chaves CE, do Valle Leone de Oliveira SM, Venturini J, Grande AJ, Sylvestre TF, Poncio Mendes R, et al. Accuracy of serological tests for diagnosis of chronic pulmonary aspergillosis: a systematic review and meta-analysis. PLoS One. 2020;15(3):e0222738 Available from: https://dx.plos.org/10.1371/journal.pone.0222738. Agarwal R, editor. Cited 2020 May 31.

22. Abba K, Deeks JJ, Olliaro PL, Naing C-M, Jackson SM, Takwoingi Y, et al. Rapid diagnostic tests for diagnosing uncomplicated P. falciparum malaria in endemic countries. Cochrane Database Syst Rev. 2011;(7).

23. Whiting PF, Rutjes AWS, 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 American College of Physicians.

24. Takwoingi Y, Riley RD, Deeks JJ. Meta-analysis of diagnostic accuracy studies in mental health. Evid Based Ment Health. 2015;18(4):103–9. https://doi.org/10.1136/eb-2015-102228.

25. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J. 2003;327:557–60 British Medical Journal Publishing Group.

26. Ochodo EA, Gopalakrishna G, Spek B, Reitsma JB, van Lieshout L, Polman K, et al. Circulating antigen tests and urine reagent strips for diagnosis of active schistosomiasis in endemic areas. Cochrane Database Syst Rev. 2015; 2015 John Wiley and Sons Ltd.

27. Kai-Wang To K, Tak-Yin Tsang O, Leung W-S, Raymond Tam A, Wu T-C, Christopher Lung D, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. 2020; Available from: www.thelancet.com/infection. Cited 2020 May 31

28. Cassaniti I, Novazzi F, Giardina F, Salinaro F, Sachs M, Perlini S, et al. Performance of VivaDiag COVID-19 IgM/IgG rapid test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department. J Med Virol. 2020:jmv.25800 Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25800. Cited 2020 May 28.

29. Lin D, Liu L, Zhang M, Hu Y, Yang Q, Guo J, et al. Evaluations of serological test in the diagnosis of 2019 novel coronavirus (SARS-CoV-2) infections during the COVID-19 outbreak. medRxiv. 2020:2020.03.27.20045153 Available from: https://www.medrxiv.org/content/10.1101/2020.03.27.20045153v1. Cited 2020 May 31.

30. Xiang J, Yan M, Li H, Liu T, Lin C, Huang S, et al. Evaluation of enzyme-linked immunoassay and colloidal gold-immunochromatographic assay kit for detection of novel coronavirus (SARS-Cov-2) causing an outbreak of pneumonia (COVID-19). medRxiv. 2020:2020.02.27.20028787 Available from: https://doi.org/10.1101/2020.02.27.20028787. Cited 2020 May 31.

31. Liu R, Liu X, Han H, Shereen MA, Niu Z, Li D, et al. The comparative superiority of IgM-IgG antibody test to real-time reverse transcriptase PCR detection for SARS-CoV-2 infection diagnosis. medRxiv. 2020:2020.03.28.20045765.

32. Liu W, Liu L, Kou G, Zheng Y, Ding Y, Ni W, et al. Evaluation of nucleocapsid and Spike protein-based ELISAs for detecting antibodies against SARS-CoV-2. J Clin Microbiol. 2020;58(6).

33. Chen J, Hu J, Long Q, Deng H, Fan K, Liao P, et al. A peptide-based magnetic chemiluminescence enzyme immunoassay for 1 Xue-fei. medrxiv.org. 2020:2020.02.22.20026617 Available from: https://doi.org/10.1101/2020.02.22.20026617. Cited 2020 May 31.

34. Pan Y, Li X, Yang G, Fan J, Tang Y, Zhao J, et al. Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. medRxiv. 2020:2020.03.13.20035428.

35. Jin Y, Wang M, Zuo Z, Fan C, Ye F, Cai Z, et al. Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019. Int J Infect Dis. 2020;94:49–52 Available from: https://doi.org/10.1016/j.ijid.2020.03.065. Cited 2020 May 31.

36. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin Infect Dis. 2020; Available from: http://www.ncbi.nlm.nih.gov/pubmed/32221519. Cited 2020 Apr 2.

37. Li Z, Yi Y, Luo X, Xiong N, Liu Y, Li S, et al. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. J Med Virol. 2020; Available from: http://www.ncbi.nlm.nih.gov/pubmed/32104917. Cited 2020 Apr 2.

38. Zhao R, Li M, Song H, Chen J, Ren W, Feng Y, et al. Serological diagnostic kit of SARS-CoV-2 antibodies using CHO-expressed full-length SARS-CoV-2 S1 proteins. medRxiv. 2020:2020.03.26.20042184 Available from: https://www.medrxiv.org/content/10.1101/2020.03.26.20042184v1. Cited 2020 May 31.

39. Zhang P, Gao Q, Wang T, Ke Y, Mo F, Jia R, et al. Evaluation of recombinant nucleocapsid and spike proteins for serological diagnosis of novel coronavirus disease 2019 (COVID-19). medRxiv. 2020:2020.03.17.20036954 Available from: https://www.medrxiv.org/content/10.1101/2020.03.17.20036954v1. Cited 2020 May 31.

40. Paradiso AV, De Summa S, Loconsole D, Procacci V, Sallustio A, Centrone F, et al. Clinical meanings of rapid serological assay in patients tested for SARS-Co2 RT-PCR. medRxiv. 2020:2020.04.03.20052183.

41. Ma H, Zeng W, He H, Zhao D, Yang Y, Jiang D, et al. COVID-19 diagnosis and study of serum SARS-CoV-2 specific IgA, IgM and IgG by a quantitative and sensitive immunoassay. medRxiv. 2020:2020.04.17.20064907.

42. Zhong L, Chuan J, Gong B, Shuai P, Zhou Y, Zhang Y, et al. Detection of serum IgM and IgG for COVID-19 diagnosis. Sci China Life Sci. 2020;63(5):777–80. https://doi.org/10.1007/s11427-020-1688-9 Science in China Press.

43. Xie J, Ding C, Li J, Wang Y, Guo H, Lu Z, et al. Characteristics of patients with coronavirus disease (COVID-19) confirmed using an IgM-IgG antibody test. J Med Virol. 2020:jmv.25930 Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25930. Cited 2020 May 31.

44. Infantino M, Grossi V, Lari B, Bambi R, Perri A, Manneschi M, et al. Diagnostic accuracy of an automated chemiluminescent immunoassay for anti-SARS-CoV-2 IgM and IgG antibodies: an Italian experience. J Med Virol. 2020:jmv.25932 Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25932. Cited 2020 May 31.

45. Crook D. Evaluation of antibody testing for SARS-CoV-2 using ELISA and lateral flow immunoassays. Available from: https://doi.org/10.1101/2020.04.15.20066407. Cited 2020 May 31

46. Lassaunière R, Frische A, Harboe ZB, Nielsen AC, Fomsgaard A, Krogfelt KA, et al. Evaluation of nine commercial SARS-CoV-2 immunoassays. medRxiv. 2020:2020.04.09.20056325.

47. Wang Q, Du Q, Guo B, Mu D, Lu X, Ma Q, et al. A method to prevent SARS-CoV-2 IgM false positives in gold immunochromatography and enzyme-linked immunosorbent assays. J Clin Microbiol. 2020;58(6).

48. Lou B, Li T, Zheng S, Su Y, Li Z, Liu W, et al. Serology characteristics of SARS-CoV-2 infection since the exposure and post symptoms onset. medRxiv. 2020:2020.03.23.20041707.

49. Liu L, Liu W, Wang S, Zheng S. A preliminary study on serological assay for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 238 admitted hospital patients. medRxiv. 2020:2020.03.06.20031856.

50. Imai K, Tabata S, Ikeda M, Noguchi S, Kitagawa Y, Matuoka M, et al. Clinical evaluation of an immunochromatographic IgM/IgG antibody assay and chest computed tomography for the diagnosis of COVID-19. medRxiv. 2020:2020.04.22.20075564.

51. Garcia FP, Tanoira RP, Cabrera JPR, Serrano TA, Herruz PG, Gonzalez JC. Rapid diagnosis of SARS-CoV-2 infection by detecting IgG and IgM antibodies with an immunochromatographic device: a prospective single-center study. medRxiv. 2020:2020.04.11.20062158.

52. Chen Z, Zhang Z, Zhai X, Li Y, Lin L, Zhao H, et al. Rapid and sensitive detection of anti-SARS-CoV-2 IgG, using lanthanide-doped nanoparticles-based lateral flow immunoassay. Anal Chem. 2020.

53. Döhla M, Boesecke C, Schulte B, Diegmann C, Sib E, Richter E, et al. Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity. Public Health. 2020;182:170–2. https://doi.org/10.1016/j.puhe.2020.04.009.

54. Burbelo PD, Riedo FX, Morishima C, Rawlings S, Smith D, Das S, et al. Detection of nucleocapsid antibody to SARS-CoV-2 is more sensitive than antibody to Spike protein in COVID-19 patients. medRxiv. 2020:2020.04.20.20071423.

55. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg. 2005 Jan 1;79(1):16–20. https://doi.org/10.1016/j.athoracsur.2004.09.040.

56. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane Database Syst Rev. 2020;2020 Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD013652/full. John Wiley and Sons Ltd. Cited 2021 Mar 19.

57. Lisboa Bastos M, Tavaziva G, Abidi SK, Campbell JR, Haraoui LP, Johnston JC, et al. Diagnostic accuracy of serological tests for covid-19: Systematic review and meta-analysis. BMJ. 2020;370:2516 Available from: https://connect.medrxiv.org/relate/content/181. Cited 2021 Mar 19.

58. Wang H, Ai J, Loeffelholz MJ, Tang YW, Zhang W. Meta-analysis of diagnostic performance of serology tests for COVID-19: impact of assay design and post-symptom-onset intervals. Emerg Microbes Infect. 2020;9(1):2200–11 Available from: https://www.tandfonline.com/doi/full/10.1080/22221751.2020.1826362. Cited 2021 Mar 19.

59. Zhang X, Wu X, Wang D, Lu M, Hou X, Wang H, et al. Proteome-wide analysis of differentially-expressed SARS-CoV-2 antibodies in early COVID-19 infection. medRxiv. 2020:2020.04.14.20064535 Available from: https://www.medrxiv.org/content/10.1101/2020.04.14.20064535v2?%253fcollection=. Cited 2020 May 31.

60. Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. medRxiv. 2020:2020.04.22.20074914 Available from: http://medrxiv.org/content/early/2020/04/25/2020.04.22.20074914.abstract. Cited 2020 May 31.

61. Freeman SC, Kerby CR, Patel A, Cooper NJ, Quinn T, Sutton AJ. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. BMC Med Res Methodol. 2019;19(1).

62. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn. 2020;20:453–4 Available from: https://www.tandfonline.com/doi/abs/10.1080/14737159.2020.1757437. Taylor and Francis Ltd. Cited 2020 Nov 18.

63. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. BMJ. 2020;369 Available from: https://calculator.testingwisely.com/playground. BMJ Publishing Group. Cited 2020 Nov 18.

64. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, del Campo R, Ciapponi A, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. medRxiv. 2020:2020.04.16.20066787 Available from: https://doi.org/10.1101/2020.04.16.20066787. Cited 2020 Nov 18.

65. Ecdc. A systematic literature review on the diagnostic accuracy of serological tests for Lyme borreliosis. Available from: www.ecdc.europa.eu. Cited 2020 Nov 19

Publisher's Note