

Effectiveness of tests to detect the presence of SARS-CoV-2 virus, and antibodies to SARS-CoV-2, to inform COVID-19 diagnosis: a rapid systematic review

David Jarrom ¹, Lauren Elston,¹ Jennifer Washington,¹ Matthew Prettyjohns,¹ Kimberley Cann,^{1,2} Susan Myles,¹ Peter Groves¹

10.1136/bmjebm-2020-111511

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjebm-2020-111511>).

¹Health Technology Wales, Velindre NHS Trust, Cardiff, UK
²Local Public Health Team, Cwm Taf Morgannwg University Health Board, Abercynon, UK

Correspondence to:
Dr David Jarrom, Health Technology Wales, Velindre NHS Trust, Cardiff, UK; david.jarrom@wales.nhs.uk



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Jarrom D, Elston L, Washington J, *et al.* *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2020-111511

Abstract

Objectives We undertook a rapid systematic review with the aim of identifying evidence that could be used to answer the following research questions: (1) What is the clinical effectiveness of tests that detect the presence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to inform COVID-19 diagnosis? (2) What is the clinical effectiveness of tests that detect the presence of antibodies to the SARS-CoV-2 virus to inform COVID-19 diagnosis?

Design and setting Systematic review and meta-analysis of studies of diagnostic test accuracy. We systematically searched for all published evidence on the effectiveness of tests for the presence of SARS-CoV-2 virus, or antibodies to SARS-CoV-2, up to 4 May 2020, and assessed relevant studies for risks of bias using the QUADAS-2 framework. **Main outcome measures** Measures of diagnostic accuracy (sensitivity, specificity, positive/negative predictive value) were the main outcomes of interest. We also included studies that reported influence of testing on subsequent patient management, and that reported virus/antibody detection rates where these facilitated comparisons of testing in different settings, different populations or using different sampling methods.

Results 38 studies on SARS-CoV-2 virus testing and 25 studies on SARS-CoV-2 antibody testing were identified. We identified high or unclear risks of bias in the majority of studies, most commonly as a result of unclear methods of patient selection and test conduct, or because of the use of a reference standard that may not definitively diagnose COVID-19. The majority were in hospital settings, in patients with confirmed or suspected COVID-19 infection. Pooled analysis of 16 studies (3818 patients) estimated a sensitivity of 87.8% (95% CI 81.5% to 92.2%) for an initial reverse-transcriptase PCR test. For antibody tests, 10 studies reported diagnostic accuracy outcomes: sensitivity ranged from 18.4% to 96.1% and specificity 88.9% to 100%. However, the lack of a true reference standard for SARS-CoV-2 diagnosis makes it challenging to assess the true diagnostic accuracy of these tests. Eighteen studies reporting different sampling methods suggest that for virus tests, the type of sample obtained/type of tissue sampled could influence test accuracy. Finally, we searched for, but did not identify, any evidence

Summary box

What is already known about this subject?

- Tests for the presence of the SARS-CoV-2 virus, and antibodies to the virus, are being deployed rapidly and at scale as part of the global response to COVID-19.
- At the outset of this work (March 2020), no high-quality evidence reviews on the effectiveness of SARS-CoV-2 virus or antibody tests were available.
- High-quality evidence reviews are required to help decision-makers deploy and interpret these tests effectively.

What are the new findings?

- Here, we synthesise evidence on the diagnostic accuracy of all known tests for SARS-CoV-2, as well as tests for antibodies to SARS-CoV-2.
- We also systematically summarise evidence on the influence of tissue sample site on virus test detection rates and the influence of test timing relative to disease course on antibody detection. The results suggest that both these factors could influence test results.
- We conclude that evidence on SARS-CoV-2 virus and antibody tests is nascent and significant uncertainties remain in the evidence base regarding their clinical and public health application. We also note that potential risks of bias exist within many of the available studies.

on how any test influences subsequent patient management.

Conclusions Evidence is rapidly emerging on the effectiveness of tests for COVID-19 diagnosis and management, but important uncertainties about their effectiveness and most appropriate application remain. Estimates of diagnostic accuracy should be interpreted bearing in mind

Summary box

How might it impact on clinical practice in the foreseeable future?

- ▶ In a rapidly developing pandemic, the widespread use of testing is an essential element in the development of effective public health strategies, but it is important to acknowledge the gaps and limitations that exist in the current evidence base and that, where possible, these should be addressed in future studies.
- ▶ In particular, more evidence is needed on the performance of point-of-care or near-patient tests compared with their laboratory equivalents, and results of testing in people with no or minimal symptoms in community-based settings need further analysis.

the absence of a definitive reference standard to diagnose or rule out COVID-19 infection. More evidence is needed about the effectiveness of testing outside of hospital settings and in mild or asymptomatic cases. Implementation of public health strategies centred on COVID-19 testing provides opportunities to explore these important areas of research.

Introduction

In December 2019, a novel coronavirus was discovered in Wuhan, China, which has since spread rapidly across the world. This virus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and the disease that it causes, COVID-19. Early on in the pandemic, the World Health Organization (WHO) stated that testing for the virus should be considered for symptomatic patients on the basis of the suspicion and likelihood of COVID-19, as well as in those who are asymptomatic or minimally symptomatic but who have been in contact with confirmed cases.¹ More recently, WHO highlighted the importance of testing for disease surveillance, to limit the spread of the disease and to manage COVID-19 risk during attempts to restore normal economic and social functioning.² Furthermore, the Organization for Economic Co-operation and Development have identified the potential importance of testing when combined with effective contact tracing in suppressing local outbreaks of COVID-19 as well as in determining individuals who have been previously infected who may safely re-integrate into work and healthcare environments.³

Tests for COVID-19 fall into two broad groups: tests that detect the presence of SARS-CoV-2 virus and tests that detect the presence of antibodies to SARS-CoV-2. Tests for the presence of virus usually use methods that recognise and amplify SARS-CoV-2 viral nucleic acid, such as reverse-transcriptase polymerase chain reaction (RT-PCR) or isothermal amplification. SARS CoV-2 virus testing is usually done in a specialised laboratory setting using respiratory samples, such as nasopharyngeal swabs, but near-patient tests have also been developed. SARS CoV-2 antibody testing (also called serology testing) is done on blood or serum samples and tests have been developed both for analysis in a laboratory and a near-patient setting. Since antibodies are produced as part of the body's immune response to infection, serology tests may be useful to identify ongoing, recovering (convalescent) or previous SARS-CoV-2 infection.

The validation and application of the different tests for COVID-19, whether for individual clinical decision-making or population-based public health strategies, is dependent on the accuracy and performance of these tests. The purpose of this review is to identify, appraise and summarise the published evidence on the diagnostic performance and effectiveness of SARS-CoV-2 virus and antibody tests in the diagnosis and management of current or previous COVID-19. The review also explores the influence of a range of factors on test outcomes, such as the timing of testing relative to first diagnosis/symptom onset, sampling methods, and whether testing is laboratory based or done at point of care.

Methods

We systematically searched for evidence to answer the following questions:

1. What is the clinical effectiveness of tests that detect the presence of the SARS-CoV-2 virus to inform COVID-19 diagnosis?
2. What is the clinical effectiveness of tests that detect the presence of antibodies to the SARS-CoV-2 virus to inform COVID-19 diagnosis?

Searching and screening for both questions was undertaken based on one search strategy. Initial scoping-level evidence searches were conducted using online databases set up to aggregate COVID-19-specific evidence.⁴⁻⁶

Based on the results of these, a specific search strategy (online supplementary appendix 1; developed and run by JW) was used to capture published evidence on SARS-CoV-2 diagnostics. The databases searched were Medline, Embase, Cochrane Library, International Network of Agencies for Health Technology Assessment (INAHTA) database and Open Grey, to include all evidence published up to 4 May 2020. The sources included in the Health Technology Wales COVID-19 Evidence Digest⁷ were hand-searched for relevant evidence and key stakeholders in Wales contacted for any published or unpublished data of relevance to this review. Because this was a rapid review, the protocol was not prospectively published.

Articles were included that studied any test to detect the presence of SARS-CoV-2, or antibodies to SARS-CoV-2, in people suspected of having recent or ongoing infection, and reported detection rates, influence of test result on changes in patient management or diagnostic accuracy. For the latter outcome, we included studies that used any suitable reference standard method of diagnosis (we excluded studies that used CT scan results alone as a reference standard). The detailed criteria used to select evidence are provided in online supplemental appendix 1. The following data were extracted from all studies deemed relevant: study design; number of centres and their location(s); dates of enrolment; inclusion/exclusion criteria; number of patients included; age and sex of included patients; test type; test target; test supplier or manufacturer; reference standard; outcome data for each relevant outcome reported. We used the QUADAS-2 tool to assess risk of bias and applicability of relevant articles.⁸ Two authors (DJ and LE) screened studies, extracted data and carried out QUADAS-2 assessments; results were checked by a third author (KC) and any disagreements were resolved by consensus.

Meta-analysis of diagnostic accuracy outcomes (sensitivity and/or specificity) was conducted only for suitable studies that reported numbers of true and false-positive and false-negative results validated against a suitable reference standard. Pooled estimates were calculated for diagnostic accuracy outcomes using a random-effects bivariate binomial model in MetaDTA V.1.25.⁹

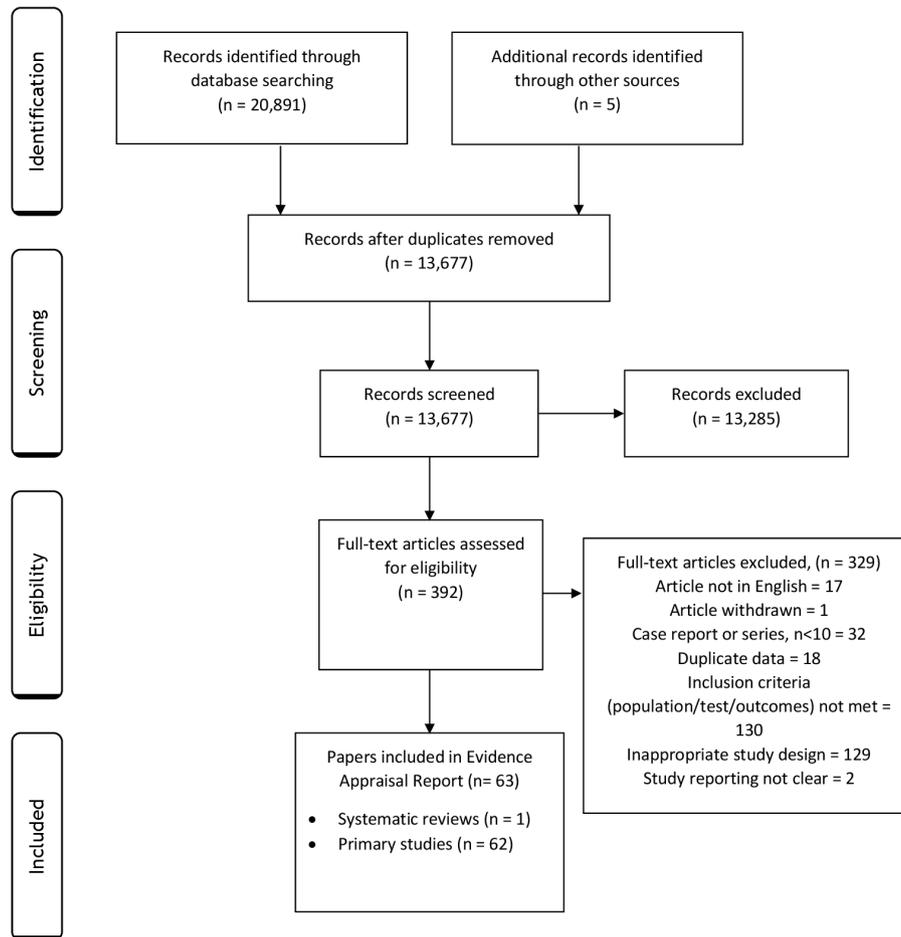


Figure 1 Summary of study selection.

Results

Figure 1 summarises articles included and excluded at each stage, and reasons that studies were excluded. A total of 13 677 unique articles were screened for eligibility, of which 13 285 were excluded after reading the title and abstract because they did not meet our inclusion criteria. The full text of the remaining 392 articles were read and checked for eligibility, and a further 329 were excluded. Of the remaining 63 relevant articles, 38 studied virus tests and 25 studied antibody tests. Tables 1 and 2 summarise the design and characteristics of studies reporting diagnostic accuracy outcomes for virus and antibody tests, respectively. Characteristics of studies that reported other outcomes are reported in online supplemental appendix 2.

All the articles that reported on virus detection were based on the detection of amplified viral SARS-CoV-2 nucleic acid sequences. Most studies used laboratory-based RT-PCR tests conducted using standard in-house or commercially available reagents and equipment, although in some cases, assay details were not reported. The RT-PCR primer used (ie, which part of the viral RNA is targeted and amplified) varied between studies, although again in some cases, primer details were not reported. In addition to RT-PCR, we identified five studies reporting the diagnostic performance of isothermal amplification assays.

The antibody tests studied used a range of different assay methods to detect one or more antibody type (different immunoglobulin classes and/or antibody targeted). In seven of the studies, tests were laboratory based (ELISA).¹⁰⁻¹⁶ We identified 17 studies using assays (lateral-flow immunoassay (LFIA); chemiluminescent

immunoassay (CLIA); colloidal gold immunochromatographic assay (GICA)) that could be suitable for point-of-care use,^{13 17-32} but the tests were not applied at point of care, or it was not clearly reported that the test had been applied at point of care, in 14 of these studies. In two studies, the type of assay was unclear.^{33 34}

The reliability and applicability of each study's conduct and reporting was assessed using the QUADAS-2 tool.⁸ Virus tests and antibody tests were assessed separately and summary judgements are shown in figure 2; signalling questions used and judgements per study are shown in online supplemental appendix 3.

For virus tests, most studies were judged to be of high or unclear risk of bias regarding patient selection, either because patients were selected for the study in a way that could have introduced bias (11% of studies) or because the method of patient selection was unclear (56% of studies). Risk of bias regarding how the index test was conducted or interpreted was judged to be high or unclear for 14% and 44% of studies, respectively, either because aspects of how the tests were conducted were unclear or because tests were not conducted in a uniform manner. For the 12 studies that included a reference standard, we judged the risk of bias to be unclear in 42% and to be high in 8%, largely because not all tests were compared against a uniform reference standard or some details of the reference standard were uncertain.

For antibody tests, the method of patient selection was judged to be unclear for 36% of studies and high for a further 52%. There was an unclear and high risk of bias regarding how the index test was conducted or interpreted in 72% and 12% of studies, respectively. For the 19 studies that included a reference standard, 42%

Table 1 Characteristics of included studies reporting diagnostic accuracy outcomes for virus tests

Reference	Study design	Population/samples	Test (supplier) (target); sample site
Kim <i>et al</i> (2020) ³⁵	Systematic review and meta-analysis. 68 studies were included; 19 studies (n=1502) reported on RT-PCR. Initial search in Medline and Embase from 1 December 2019 to 16 March 2020. The search was updated to 3 April 2020	Studies on COVID-19 that reported the diagnostic sensitivity and/or specificity of chest CT scans and/or RT-PCR assays	Index tests: initial RT-PCR test (target varied among studies); nasopharyngeal swab, throat swab or sputum. Reference standard: repeated RT-PCR tests. RT-PCR results were extracted within 14 days of symptom onset
Baek <i>et al</i> (2020) ³⁶	Design/validation study (samples used collected retrospectively). Korea, number of centres not clear	Patients with COVID-19, n=14. No demographic details reported	RT-LAMP assay (developed in-house); (N primer); nasal swabs
Fang <i>et al</i> (2020) ⁶⁸	Retrospective case series. Single centre (China). 19 January 2020 to 4 February 2020	People with eventual confirmed diagnosis of COVID-19 infection who had an RT-PCR test and CT scan within 3 days or less. Eventual confirmed diagnosis is defined as through repeated RT-PCR testing of negative patients, until a positive test is received. n=51. Median age 4.5 years (IQR 39 to 55 years). 29 men; 22 women	Index test: initial RT-PCR (Shanghai ZI Bio-Tech) (primer not specified); throat or sputum samples. Reference standard: eventual confirmed diagnosis through RT-PCR
Fang <i>et al</i> (2020) ⁴³	Retrospective case series. Single centre (China). January 2020 to February 2020 (specific dates not specified)	People with COVID-19, n=32 (8 ICU patients; 24 non-ICU patients). Age range 35 to 54 years. Sex not reported	RT-PCR (NR) (primer not specified); from nasal swabs, blood, faecal, urine, saliva and tear samples
Harrington <i>et al</i> (2020) ³⁷	Prospective case series. Five centres (USA)	Symptomatic patients meeting current criteria for diagnosis of COVID-19, n=524. Demographic details not reported	Isothermal amplification (ID NOW COVID-19 assay (Abbott)) (RdRp); nasal swabs. Reference standard was RT-PCR (Abbott RealTime SARS-CoV-2 assay performed on the Abbott m2000 system)
He <i>et al</i> (2020) ⁶⁹	Retrospective case series. Single centre (China). 10 January 2020 to 28 February 2020	Hospitalised patients with suspected COVID-19 who underwent high-resolution chest CT and real-time RT-PCR. n=82. Median age 52 years (range 8 to 74 years). 49 males	RT-PCR (BGI Genomics); (primer NR); nasopharyngeal swab, oropharyngeal swab, endotracheal aspirate or bronchoalveolar lavage. Reference standard: eventual confirmed diagnosis through RT-PCR
Lee <i>et al</i> (2020) ⁷⁰	Retrospective case series. Single centre (Singapore). Up to 29 February 2020 (start date not reported)	Patients admitted to hospital with suspected COVID-19 infection. n=70. Demographics not reported	RT-PCR (in-house or A*STAR Fortitude Kit (Accelerate Technologies)); (N, Orf1ab); nasopharyngeal swabs
Lu <i>et al</i> (2020) ³⁸	Design/validation study. China. Dates of sampling not reported	Patients with suspected COVID-19 admitted to hospital and quarantined. n=56. Demographics not reported	Index test: RT-LAMP (in-house assay) (N); throat swabs. Reference standard: RT-PCR (LifeRiver Bio)
Shen <i>et al</i> (2020) ⁶²	Retrospective case series. Single centre (China). 22 January to 18 February 2020	Subjects judged at high risk of SARS-CoV-2 infection. n=5630. Median age 51 years (IQR 36–63). Male 2631 (46.7%)	RT-PCR (Shanghai Huirui Biotechnology); (Orf1ab, N); throat swabs
Yan <i>et al</i> (2020) ³⁹	Development/validation study. Centre NR. Dates NR	Patients with pneumonia and suspected SARS-CoV-2 infection. n=130 specimens. Characteristics NR	RT-LAMP (Loopamp RNA amplification kit; Loopamp Real-time Turbidimeter, both Eiken Chemical, Tokyo, Japan, used to perform and monitor the RT-LAMP reaction) (Orf1ab and spike). Reference standard: RT-PCR. Sampling from swabs (not specified) and bronchoalveolar lavage fluid

Continued

Reference	Study design	Population/samples	Test (supplier) (target); sample site
Zhen <i>et al</i> (2020) ⁴⁰	Prospective comparative evaluation. Single centre (USA)	Symptomatic patients suspected of having COVID-19. n=108 samples (unclear if from unique patients). Patient characteristics not reported	Three index tests, all 'sample to answer' platforms: (1) RT-PCR (Cepheid Xpert Xpress SARS-CoV-2 assay, performed on GeneXpert instrument system); (N2, E); (2) isothermal amplification (Abbott ID NOW COVID-19 assay); (RdRp); (3) DNA hybridisation and electrochemical detection (GenMark ePlex SARS-CoV-2 assay) (target NR). All compared with RT-PCR (Hologic Panther Fusion SARS-CoV-2 assay); (Orf1ab), Nasopharyngeal specimens
Zhang <i>et al</i> (2020) ⁷¹	Retrospective case series. China, two centres. Collection from 29 December 2019 to 16 February 2020	People with laboratory-confirmed COVID-19 (via RT-PCR). n=290. Median age 57 years (22–88 years). 155 (53.4%) male	RT-PCR (Shanghai Bio-germ Medical Technology) (Orf1ab, N primers); pharyngeal swab samples

NR, details not reported; RT-LAMP, reverse transcription loop-mediated isothermal amplification.

were judged to have an unclear risk of bias and 16% a high risk of bias.

We identified one existing published meta-analysis estimating the sensitivity of an initial RT-PCR test, using the results of repeated RT-PCR tests as the reference standard.³⁵ This included studies published up to 3 April 2020, and aligned closely with our inclusion criteria regarding virus tests, although the authors included studies of any population size, whereas we elected to exclude studies that included less than 10 patients, meaning we omitted seven studies (total 46 patients) that were included in the earlier meta-analysis. We used data from this analysis and studies published subsequently to determine that the overall sensitivity of RT-PCR is 87.8% (95% CI 81.5% to 92.2%), based on 16 studies of 3818 patients. Because these studies only included cases where COVID-19 was confirmed to be present, specificity cannot be legitimately estimated.

Five studies (972 patients or samples in total) reported the diagnostic accuracy of isothermal amplification assays in the diagnosis of patients with suspected COVID-19, using test results from RT-PCR as a reference standard.^{36–40} Because the use of a single RT-PCR test as a reference standard may not be representative of true disease presence, we deemed it inappropriate to use the results of these studies to derive a single pooled estimate of sensitivity and specificity. Reported diagnostic sensitivity and specificity estimates range from 74.7% to 100% and 87.7% to 100%, respectively. Table 3 provides a detailed breakdown of results.

Ten studies on antibody testing (757 participants included; number not clear for two studies) reported sensitivity and specificity^{11 12 17 18 20–24 33} or sufficient information to allow these to be calculated. Two additional studies reported specificity only.^{10 19} Where a reference standard was included, this was usually RT-PCR (initial and repeats until a positive confirmation); one study that used either RT-PCR or clinical diagnosis to determine final disease status. Furthermore, studies used a range of different antibody types and targets. Because of these limitations, we concluded that pooling data across studies was not appropriate.

The reported sensitivity in these studies ranged from 18.4% to 96.1%. Notably, the lowest reported sensitivity was using a point-of-care test,¹⁷ although sensitivity figures below 50% were also reported for one laboratory test.²⁰ Specificity was reported in 12 studies (682 participants included; number not clear for two studies) and ranged from 88.9% to 100%. Full outcomes from these studies are shown in table 4.

The positive predictive value (PPV) and negative predictive value (NPV) of RT-PCR was estimated at different prevalence levels. We used our pooled sensitivity estimate of 87.8%, and because we were unable to calculate specificity using the evidence found, we used a previously published estimate of 98.0% for specificity.⁴¹ Prevalence estimates were based on data from Public Health England (PHE).⁴² A prevalence rate of 3.0% was estimated based on PHE data up to 6 August 2020, which showed that there have been 308 134 confirmed cases of COVID-19 from 10 236 970 tests. At this prevalence level, RT-PCR testing was estimated to have a PPV of 57.7% and NPV of 99.6%. To estimate the utility of the test at times of high prevalence, PPV and NPV were also estimated using PHE data up to 1 May 2020 (the date at which the daily number of cases was at its highest point). On this date, a prevalence rate of 24.6% was estimated based on 177 454 confirmed cases of COVID-19 from 721 124 tests. At this prevalence level, RT-PCR testing was estimated to have a PPV of 93.5% and NPV of 96.1%.

We identified 18 studies^{30 43–59} that compared RT-PCR for SARS-CoV-2 results from samples taken from different parts of

Table 2 Characteristics of included studies reporting diagnostic accuracy outcomes for antibody tests

Reference	Study design	Population/samples	Test (supplier) (target); sample site
Cassaniti <i>et al</i> (2020) ¹⁷	Cohort study, Single centre (Italy). Collection date NR	3 cohorts: (1) healthy volunteers with negative RT-PCR for COVID-19; (2) hospitalised patients with positive COVID-19 RT-PCR; (3) patients with suspected COVID-19 at their first access at emergency room. n=110 (30 healthy volunteers; 30 patients with COVID-19; 50 patients with suspected COVID-19). Baseline characteristics reported separately for each cohort	VivaDiag COVID-19-19 IgM/IgG rapid point-of-care lateral flow immunoassay (VivaChek) (target NR); serum or blood samples. Serum samples were obtained at median 7 days (IQR 4–11 days) after positive result for hospitalised patients. Reference/comparator: RT-PCR (RdRp and E primers); respiratory samples
Dohla <i>et al</i> (2020) ¹⁸	Single centre (Germany). Dates NR	People within a community setting (high-prevalence area), presenting with COVID-19 symptoms (n=39) and people diagnosed with COVID-19 (n=10). Median age 46 years (IQR 28–72 years). 29/49 female (49.0%)	IgG/IgM point-of-care test (NR) (target NR); fingertip prick blood or serum. Reference standard: repeated RT-PCR (Altona Diagnostics) (target NR); throat swabs. Serum samples for previously diagnosed individuals were also analysed using the antibody test
Hoffman <i>et al</i> (2020) ¹⁹	Validation study. Centre NR. Study dates NR	Patients with confirmed COVID-19 or convalescents (n=29). Controls: healthy volunteers without any known history of COVID-19 (n=24); blood donor sera from healthy adults (n=80) and babies (n=20) collected during 2018	IgG/IgM Rapid Test Cassette (Zhejiang Orient Gene Biotech Company) (target NR); blood/serum samples. Reference standard: RT-PCR
Jin <i>et al</i> (2020) ²⁰	Retrospective study, Single centre (China)	People with a laboratory-confirmed SARS-CoV-2 infection in hospital, and at least one viral serological test (n=43). Median age 47.0 years (IQR 34.0–59.0 years). 39.5% male. Control group: patients with suspected SARS-CoV-2 infection who were excluded and quarantined at home (n=33). Median age 31.0 years (IQR 25.5–37.5 years). 66.7% male. Suspected infected patients were discharged from hospital when they received two negative PCRs, performed in a 24-hour interval	IgM and IgG chemiluminescence assay (CLIA)(Shenzhen YHLO Biotech) (targets N protein and spike protein). Reference standard: confirmed diagnosis from RT-PCR (target not specified); sampling not clearly reported but includes oral swabs, anal swabs and sputum. Duration between first symptoms and serological test (CLIA) was 18 days (IQR 11–23 days) in the COVID-19 group. 3.0 days (2.0–8.0 days)
Li <i>et al</i> (2020) ²¹	Prospective development study, Single centre (China). 12 February 2020 to 20 February 2020	People with suspected (RT-PCR negative) or confirmed (RT-PCR positive) COVID-19. n=278 (89 confirmed; 189 probable). n=273 controls were included. Baseline characteristics NR	IgM and IgG colloidal gold assay (NR) (targets serum antibodies against N protein); serum specimens. RT-PCR assumed to be the reference standard (described as a 'control' by the authors); primer/reference and sampling methods not known
Li <i>et al</i> (2020) ²²	Prospective development study, 8 centres (China). Dates NR	People with suspected COVID-19. n=525 specimens (397 clinical positive; 128 clinical negative). Characteristics NR	IgM/IgG rapid point-of-care lateral flow immunoassay (Jiangsu Medomics Medical Technologies) (targets antibodies against spike protein); blood (including serum and plasma). Reference standard: RT-PCR; respiratory specimens
Liu <i>et al</i> (2020) ¹⁰	Prospective study, Single centre (China). 18 January to 26 February 2020	Hospitalised patients diagnosed with COVID-19. All patients were laboratory confirmed by RT-PCR. n=314 (214 patients; 100 healthy controls). Baseline characteristics NR	IgM ELISA IgG ELISA (NR) (targets antibodies against N and spike) serum. Median time of sample collection was 15 days (range 0 to 55)

Continued

Table 2 Continued

Reference	Study design	Population/samples	Test (supplier) (target); sample site
Shen <i>et al</i> (2020) ²³	Prospective cohort study. Single centre (China). 20 January and 2 February 2020	People with suspected COVID-19. Suspected COVID-19 was defined as a pneumonia that had related epidemiological history and fulfilled two of the following: fever and/or respiratory symptoms; imaging indicative of pneumonia; low/normal white cell count or low lymphocyte count. n=150. Median age: PCR-positive group 38 years (IQR 46–56 years); PCR-negative group 32 years (IQR 20–42.5). Sex: PCR-positive group 60.8% male; PCR-negative group 56.6% male. Control: healthy donors (n=26)	IgM/IgG colloidal gold immunochromatography antibody kit (Shanghai Outdo Biotech Company) (target NR); blood. Reference standard: RT-PCR (target NR); nasopharyngeal and oropharyngeal swabs. At least two different samples were obtained from each patient for RT-PCR. If the result was inconclusive, repeated sample collection was required. A patient with at least one positive RT-PCR was confirmed as positive. Patients with two consecutive negative results were defined as PCR negative, but would only be diagnosed as non-COVID-19 if the symptoms could be explained by another condition or infection resolved following the corresponding treatments. Any other PCR-negative result was treated as inconclusive
Spicuzza <i>et al</i> (2020) ²⁴		People with confirmed COVID-19 (n=23) or suspected COVID-19 (n=7). Confirmed COVID-19 was defined as consistent radiological/clinical findings, with positive RT-PCR. Suspected COVID-19 was defined as suggestive radiological/clinical findings but negative RT-PCR. Control: asymptomatic controls with negative RT-PCR (n=7). n=37. Mean age: confirmed COVID-19 57±17 years; suspected COVID-19 67±15 years. Controls age NR. Sex characteristics NR	IgG/IgM POC Antibody Rapid Test Kit (Beijing Diagreat Biotechnologies Company) (spike); blood/serum/plasma. Reference standard: RT-PCR (NR) (target NR); nasopharyngeal swab or bronchial aspirate
Xiang <i>et al</i> (2020) ¹¹	Single centre (China). 19 January to 2 March 2020	People with suspected (n=24) or confirmed (n=85) COVID-19. Diagnosis of laboratory-confirmed COVID-19 was defined as positive nucleic acid tests for SARS-CoV-2 by RT-PCR. Diagnosis of suspected COVID-19 was based on negative RT-PCR, but satisfying 1 of the epidemiological history criteria and 2 of the clinical criteria. Control: samples from healthy blood donors or from hospitalised patients with other diseases (n=60)	IgM/IgG ELISA (Livzon) (NR); serum. Serum samples were obtained at different time periods after symptom onset. Reference standard: RT-PCR (ORF1ab and N); nasopharyngeal and/or oropharyngeal swabs
Xu <i>et al</i> (2020) ³³	Retrospective study. Single centre (China). 20 January 2020 to 17 February 2020	Patients with suspected COVID-19. n=284 participants: 186 COVID-19 patients with RT-PCR-positive result; 19 COVID-19 cases diagnosed by clinical symptoms; 79 controls with other diseases (negative RT-PCR). Baseline characteristics NR	IgM and IgG fully automated assay (NR) (target NR); serum samples. Comparator: RT-PCR. Reference standard: Diagnosis through positive RT-PCR or clinical symptoms
Zhao <i>et al</i> (2020) ¹²	Retrospective study. Single centre (China). 11 January 2020 to 9 February 2020	People with COVID-19. All enrolled cases were confirmed to be infected with SARS-CoV-2 by RT-PCR. n=173 patients (535 samples). Median age 48 years (IQR 35 to 61). 51.4% female	Index tests: IgM ELISA (Beijing Wantai Biological Pharmacy Enterprise) (spike protein). IgG ELISA (N). Total antibody (Ab) ELISA (spike protein); plasma samples. Comparator: RT-PCR result. Reference standard: confirmed COVID-19 through positive RT-PCR

NR, details not reported.

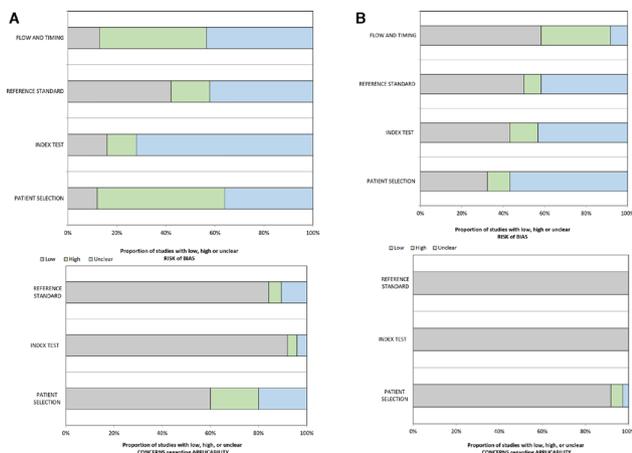


Figure 2 QUADAS-2 risk of bias judgements. Summary of risk of bias and applicability assessments for (A) virus tests and (B) antibody tests.

the body. These results are presented in [table 5](#). Most samples were taken from the upper respiratory tract (online supplemental appendix 4 summarises detection rates for individual sites in the upper respiratory tract, where reported). Other sample sites were saliva, sputum and stool/rectal swab. The detection rates varied across sample sites but the heterogeneous nature of the studies makes meaningful comparison difficult. Detection rates

were consistently low with urine or tears/conjunctiva sampling. Detection rates in blood samples were mixed, with some studies reporting very low detection rates, while others reported rates that were comparable with samples from other sites in the same population.

The majority of studies tested people with relatively severe disease and a high suspicion of COVID-19 infection. Considering other populations, two studies^{60 61} tested UK healthcare workers and three studies⁶²⁻⁶⁴ tested people outside of hospital (or as outpatients). One further study⁶⁵ routinely tested pregnant women. All these studies only reported detection rates: results are summarised in online supplemental appendix 4.

Ten studies provided data on antibody detection (seroprevalence) at different points in time after the onset of confirmed COVID-19 disease.^{10 12-15 26-28 31 34} Detailed results are shown in online supplemental appendix 4.

Discussion

This review summarises the available published evidence of the effectiveness of tests that are used in the diagnosis of current or previous COVID-19 infection up to 4 May 2020. Despite this work taking place relatively early in the COVID-19 pandemic, 38 published studies were identified that reported on the effectiveness of tests for detecting the presence of SARS CoV-2 virus and 25 studies were identified that reported on testing for the presence of

Table 3 Diagnostic accuracy outcomes for virus tests

Reference	Assay and target	Number of patients/ samples	Index test; comparator (if applicable)
Pooled sensitivity			
Updated analysis including results from Kim <i>et al</i> (2020) systematic review and meta-analysis and subsequent publications ³⁵	NR	n=3818; 16 studies	RT-PCR: 87.8% (95% CI 81.5% to 92.2%)
Sensitivity for studies that could not be pooled			
Baek <i>et al</i> (2020) ³⁶	N	n=154 samples	RT-LAMP: 100%
Harrington <i>et al</i> (2020) ³⁷	RdRp	n=524 patients	Isothermal amplification (Abbott ID-NOW assay): 74.7% (95% CI 67.8% to 80.8%)
Lu <i>et al</i> (2020) ³⁸	N	n=56 patients	RT-LAMP: 94.4% (95% CI 81.3% to 99.3%)
Yan <i>et al</i> (2020) ³⁹	Orf1ab and spike	n=130 samples	RT-LAMP: 100% (95% CI 92.3% to 100%)
Zhen <i>et al</i> (2020) ⁴⁰	N2, E	n=108 samples	RT-PCR (Cepheid Xpert Xpress SARS-CoV-2 assay): 98.3% (95% CI 90.7% to 99.9%)
	RdRp	n=108 samples	Isothermal amplification (Abbott ID NOW COVID-19 assay): 87.7% (95% CI 76.3% to 94.9%)
	NR	n=108 samples	DNA hybridisation and electrochemical detection (GenMark ePlex): 98.3% (95% CI 90.7% to 99.9%)
Specificity			
Baek <i>et al</i> (2020) ³⁶	N	n=154 samples	RT-LAMP: 98.7%
Lu <i>et al</i> (2020) ³⁸	N	n=56 patients	RT-LAMP: 90.0% (95% CI 68.3% to 98.8%)
Harrington <i>et al</i> (2020) ³⁷	RdRp	n=524 patients	Isothermal amplification (Abbott ID-NOW assay): 99.4% (95% CI 97.8% to 99.9%)
Yan <i>et al</i> (2020) ³⁹	Orf1ab and spike	n=130 specimens	RT-LAMP: 100% (95% CI 93.7% to 100%)
Zhen <i>et al</i> (2020) ⁴⁰	N2, E	n=108 samples	RT-PCR (Cepheid Xpert Xpress SARS-CoV-2 assay): 98.3% (95% CI 92.3% to 100%)
	RdRp	n=108 samples	Isothermal amplification (Abbott ID NOW COVID-19 assay): 87.7% (95% CI 92.3% to 100%)
	NR	n=108 samples	DNA hybridisation and electrochemical detection (GenMark ePlex): 100% (95% CI 92.3% to 100%)

RT-LAMP, reverse transcription loop-mediated isothermal amplification; NR, details not reported.

Table 4 Diagnostic accuracy outcomes for antibody tests

Reference	Assay and target	Number of patients/samples	Index test; comparator (if applicable)
<i>Sensitivity</i>			
Cassaniti <i>et al</i> (2020) ¹⁷	LFIA, VivaChek POC	n=50 patients (suspected cases only)	IgM/IgG: 18.4%
Dohla <i>et al</i> (2020) ¹⁸	IgM/IgG POC test	n=49	IgM/IgG: 36.4% (95% CI 17.2 to 59.3)
Jin <i>et al</i> (2020) ²⁰	CLIA (N and spike proteins)	n=27	IgM: 48.1% (13/27) IgG: 88.9% (24/27)
Li <i>et al</i> (2020) ²¹	Colloidal gold	Population not clear	IgM: 78.7% IgG: 73.0% IgM/IgG: 87.6%
Li <i>et al</i> (2020) ²²	LFIA, Jiangsu Medomics POC	n=525 specimens	IgM/IgG: 88.66%
Spicuzza <i>et al</i> (2020) ²⁴	LFIA (spike)	n=37	IgG/IgM: 82.6%
Shen <i>et al</i> (2020) ²³	Colloidal gold (NR)	n=150	IgM/IgG: 71.1% (95% CI 0.609 to 0.797)
Xiang <i>et al</i> (2020) ¹¹	ELISA (NR)	n=66	IgM: 77.3% (51/66) IgG: 83.3% (55/66)
Xu <i>et al</i> (2020) ³³	Fully automated assay (NR)	n=205 patients	IgM: 70.24% (144/205) IgG: 96.10% (197/205)
Zhao <i>et al</i> (2020) ¹²	ELISA (spike for IgM and Ab; N for IgG)	n=173 samples	IgM: 82.7% (143/173) IgG: 64.7% (112/173) Ab: 93.1% (161/173) RT-PCR: 67.1%* (112/?)
<i>Specificity</i>			
Cassaniti <i>et al</i> (2020) ¹⁷	LFIA, VivaChek POC	n=50 (suspected cases only)	IgM/IgG: 91.7%
Li <i>et al</i> (2020) ²¹	Colloidal gold	Population not clear	IgM: 98.2% IgG: 99.3% IgM/IgG: 98.2%
Li <i>et al</i> (2020) ²²	LFIA, Jiangsu Medomics POC	n=525 specimens	IgM/IgG: 90.63%
Liu <i>et al</i> (2020) ¹⁰	ELISA (spike)	n=100 healthy controls	IgM: 100% (0/100) IgG: 100% (0/100) IgM and/or IgG: 100% (0/100)
Xu <i>et al</i> (2020) ³³	Fully automated assay (NR)	n=79 patients	IgM: 96.20% (76/79) IgG: 92.41% (73/79)
Zhao <i>et al</i> (2020) ¹²	ELISA (spike for IgM and Ab; N for IgG)	Not clear	Total Ab: 99.1% (211/213) IgM: 98.6% (210/213) IgA: 99.0% (195/197)
Jin <i>et al</i> (2020) ²⁰	CLIA (N and spike proteins)	n=33	IgM: 100% (33/33) IgG: 90.9% (30/33)
Xiang <i>et al</i> (2020) ¹¹	ELISA (NR)	n=60	IgM: 100% (60/60) IgG: 95.0% (57/60)
Dohla <i>et al</i> (2020) ¹⁸	IgM/IgG POC test	n=49	IgM/IgG: 88.9% (95% CI 70.8 to 97.7)
Spicuzza <i>et al</i> (2020) ²⁴	LFIA (spike)	n=37	IgG/IgM 92.9%
Hoffman <i>et al</i> (2020) ¹⁹	LFIA (NR)	n=124 (controls)	IgM: 100% (0/124) IgG: 99.2% (1/124)
Shen <i>et al</i> (2020) ²³	Colloidal gold (NR)	n=150	IgM/IgG: 96.2% (95% CI 0.859 to 0.993)

CLIA, chemiluminescent immunoassay; LFIA, lateral flow immunoassay; NR, details not reported.

antibodies. Analysis of these studies using the QUADAS-2 framework revealed high or unclear risks of bias in the majority, most commonly as a result of unclear methods of patient selection and test conduct, or because of the use of a reference standard that may not definitively diagnose COVID-19. Nonetheless, the available evidence provides information on which to begin to judge the possible clinical effectiveness of COVID-19 testing, although significant uncertainties remain in the evidence base regarding their clinical and public health application.

In the course of our work, the first meta-analysis of diagnostic accuracy for SARS-CoV-2 virus tests was published by Kim *et al*.³⁵ This included pooled analysis of 19 studies (1502 patients) and

used the results of repeated laboratory-based RT-PCR as the reference standard. This aligned closely with our own inclusion criteria for virus tests, although Kim *et al* included studies of any population size, we excluded studies with less than 10 patients, meaning we omitted 7 studies (46 patients). However, by including more recently published studies in our analysis the number of patients more than doubles from 1502 to 3818 patients. The analysis by Kim *et al* estimated that the sensitivity of an initial RT-PCR test is 89% (95% CI 81% to 94%). Our addition of data from more recent studies leads us to conclude a sensitivity of 87.8% (95% CI 81.5% to 92.2%). An analysis estimating the PPV and NPV of RT-PCR showed that the NPV is likely to be high while PPV

Table 5 Virus test detection rates in studies comparing different sample sites

Study	BLF	Pharyngeal*	Throat wash	Lingual	Saliva	Sputum	Plasma/blood	Urine	Faeces and/or rectal swabs	Tears/ conjunctival swab	Fibrobronchoscope brush biopsy
Azzi <i>et al</i> (2020) ⁴⁴	n/a	25/25 (100%)	n/a	n/a	25/25 (100%)	n/a	n/a	n/a	n/a	n/a	n/a
Chan <i>et al</i> (2020), RdRp/ Hel ⁴⁵	n/a	30/34 (88.2%)	n/a	n/a	59/72 (81.9%)	13/14 (92.9%)	10/87 (11.5%)	0/33 (0.0%)	7/33 (21.2%)	n/a	n/a
Chan <i>et al</i> (2020), RdRp-P2 ⁴⁵	n/a	22/34 (64.7%)	n/a	n/a	38/72 (52.8%)	13/14 (92.9%)	0/87 (0.0%)	0/33 (0.0%)	4/33 (12.1%)	n/a	n/a
Chen <i>et al</i> (2020) ⁴⁶	n/a	42/42 (100%)	n/a	n/a	n/a	n/a	n/a	0/10 (0%)	28/42 (66.7%)	n/a	n/a
Guo <i>et al</i> (2020) ⁴⁷	n/a	1/24 (4.2%)	7/24 (29.2%)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Fang <i>et al</i> (2020) ⁴³	n/a	32/32 (100%)	n/a	n/a	25/32 (78%)	n/a	23/32 (72%)	0	NR	5/32 (16%)	n/a
Huang <i>et al</i> (2020) ⁴⁸	n/a	10/16	n/a	n/a	n/a	16/16 (100%)	1/16	1/16	11/16	1/15	n/a
Lin <i>et al</i> (2020) ⁴⁹	n/a	23/52 (44.2%)	n/a	n/a	n/a	40/52 (76.9%)	n/a	n/a	n/a	n/a	n/a
Liu <i>et al</i> (2020) ⁵⁰	4/5 (80%)	1843/4818 (38.25%)	n/a	n/a	n/a	28/57 (49.12%)	n/a	n/a	n/a	n/a	n/a
Wang <i>et al</i> (2020) ⁵²	14/15 (93%)	131/406 (32%)	n/a	n/a	n/a	75/104 (72%)	3/307 (1%)	0/72 (0%)	44/153 (29%)	n/a	6/13 (46%)
Williams <i>et al</i> (2020) ⁵⁴	n/a	39/622 (6.3%)	n/a	n/a	33/522 (6.3%)	n/a	n/a	n/a	n/a	n/a	n/a
Wu <i>et al</i> (2020) ⁶⁰	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	41/74 (55%)	n/a	n/a
Xia <i>et al</i> (2020) ⁶¹	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1/30 (3.3%)	n/a
Xie <i>et al</i> (2020) ⁴³	n/a	9/19	n/a	n/a	n/a	n/a	0/19	0/19	8/19	n/a	n/a
Ye <i>et al</i> (2020) ⁵⁷	n/a	40/91 (44.0%)	n/a	n/a	33/91 (36.3%)	n/a	n/a	n/a	n/a	n/a	n/a
Zhang <i>et al</i> (2020) ⁵⁸	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	5/14 (35.7%)	n/a	n/a
Zheng <i>et al</i> (2020) ⁵⁹	n/a	n/a	n/a	n/a	96/96 (100%)†	n/a	39/95 (41%)	1/67 (1%)	55/93 (59%)	n/a	n/a

*Includes nasopharyngeal swabs, nasopharyngeal aspirate, nose and throat swabs.

†Sputum samples were collected from the respiratory tract of patients with sputum, and saliva after deep cough was collected from patients without sputum.

BLF, bronchoalveolar lavage fluid; n/a, not included in study; NR, sampling included in study but outcome not reported.

may be low at times where the prevalence in the tested population is low. The likely prevalence in the tested population should therefore be a key consideration for decision-makers when interpreting test results and deciding on testing strategies. Despite our finding of a high NPV for RT-PCR, uncertainty may remain with a negative test result, especially in the context of high clinical suspicion, and the possibility of a false-negative result also needs to be considered. Possible causes for false-negative tests include laboratory error, sampling error, and variability in viral shedding with the lack or negligible presence of virus nucleic acid in the tissue sampled at the time of sampling. Determining the specificity of SARS-CoV-2 nucleic acid testing is particularly challenging because of the inclusion in the published studies of patients considered to be suffering from COVID-19 as well as the lack of a reference standard that validates the absence of disease. The assessment of overall diagnostic accuracy in laboratory testing for the presence of the SARS-CoV-2 virus is hampered by the absence of a definitive reference standard and by a wide range of target primers, methods and types of sampling used in the published studies. In addition, there is very limited published information on the diagnostic accuracy of point-of-care or near-patient tests.

Of the 25 studies that assessed antibody tests, 10 reported diagnostic accuracy in terms of both sensitivity and specificity, almost all using RT-PCR (initial or repeat testing) as the reference standard.^{11 12 17 18 20-22 24 33 62} Accepting the limitations already discussed around the absence of a diagnostic reference standard, the overall sensitivity reported in these studies varied widely, from 18.4% to 96.1%, although the specificity was more consistent and ranged from 88.9% to 100%. The clinical implications of these data are that considerable uncertainty remains about the implications of a negative antibody test with a significant possibility of false negativity, while the presence of a positive antibody test carries with it a high likelihood of previous COVID-19 infection. There is very limited information available on the accuracy of point-of-care antibody tests.

Our study has some limitations, primarily due to the nature of the evidence found by our searches. The rapid nature of this work (to help inform decision-makers at the outset of the COVID-19 pandemic in the UK) meant some steps in a full systematic review were not completed: there was minimal consultation with decision-makers on the inclusion and exclusion criteria for the review, and we did not publish our protocol in advance of commencing the review. Other limitations relate to the nature of the evidence we found, and that this work was completed during the early stages of the COVID-19 pandemic. The lack of a recognised reference standard meant we considered studies for inclusion that used any appropriate method to verify test results. While initial suspicion of COVID-19 may be based on clinical assessment combined with radiological results, WHO advice is that laboratory-based nucleic acid testing (such as RT-PCR) should be used to confirm cases with further confirmation by nucleic acid sequencing when necessary or feasible.⁶⁶ The only suitable studies that allowed the diagnostic accuracy of RT-PCR to be assessed compared initial test results with repeated RT-PCR testing in the same individuals: this allowed us to estimate the sensitivity of an initial RT-PCR test, using final (positive) results of the repeated test as the reference standard. Use of this reference standard, which only validates the presence of disease and not its absence, means specificity cannot be determined. We estimated the PPV and NPV of RT-PCR and at different prevalence rates; estimates from PHE were judged to provide the best current evidence for prevalence. However, it should be noted that there are limitations with this approach. Most notably, it is based on the total number of tests rather than the

number of people tested. As such, the estimates may underestimate prevalence as many people will have been tested more than once. Crucially, because we could not calculate specificity from the evidence found by our own systematic review, we relied on a previously published estimate of 98.0%. PPV is highly sensitive to this estimate, emphasising the need for further reliable published estimates of the sensitivity of RT-PCR to the interpretation of this test, particularly in low prevalence populations. Furthermore, the evidence included in this pooled analysis and other individual studies we identified used a range of target primers, methods and type of sampling.

We observed similar limitations with evidence on other tests. We found studies reporting diagnostic accuracy of antibody tests and of loop-mediated isothermal amplification (LAMP) as a method of virus detection. However, the reference standard used was RT-PCR (initial and repeat tests), except for one study that used either RT-PCR or clinical diagnosis to determine final disease status. As already concluded, a true assessment of the accuracy of RT-PCR test results is very challenging, and using RT-PCR for validation means the same limitations apply to the results of any antibody or LAMP tests studied in this way. These tests also varied considerably in their conduct and protocols used. These limitations led us to conclude that it was inappropriate to conduct pooled analysis of diagnostic accuracy, meaning limited conclusions about antibody and LAMP tests can be drawn based on the data currently available. Lastly, for all types of test, there are a wide range of different commercially available testing products and kits, as well as some that use protocols developed in-house by academic and public health testing laboratories. Where available, we have detailed the exact test used for each data source (tables 1 and 2, and online supplemental appendix 2), but our evidence synthesis does not take into account similarities or differences between specific test kits or protocols, and the results should be interpreted with this in mind.

Alternative approaches to validating COVID-19 test results could use genomic sequencing, testing for multiple primer targets, confirmatory testing for other respiratory viruses, long-term follow-up, or clinical signs and symptoms. Each would have potential advantages and disadvantages. For example, genomic sequencing would determine the exact strain of virus present and could detect cases of infection where primer target regions were not conserved, resulting in a false-negative RT-PCR result. However, sequencing could only be used to 'rule in' the presence of the virus and not to rule it out. Furthermore, it is time and resource intensive and so would be highly unlikely to be undertaken for all samples in routine practice. Future studies might need to use a combination of these factors as a composite reference standard that could validate results from both positive and negative diagnoses, allowing sensitivity and specificity of the test to be estimated with greater certainty.

In applying the results of the published studies on testing for COVID-19 to influencing the development of evidence-based testing strategies, it should be noted that the majority of published studies reported on the results of COVID-19 virus or antibody testing were done in a hospital setting and in symptomatic patients with confirmed or suspected COVID-19 infection. Data on testing in other settings are comparatively limited. Only three studies⁶²⁻⁶⁴ were identified that used RT-PCR to detect SARS-CoV-2 in the general population in the context of mild influenza-like symptoms while only two studies were found^{60 61} that reported on the testing of UK healthcare workers. Furthermore, only one study¹⁸ was identified that reported on the results of antibody testing outside of a hospital setting. In a rapidly developing pandemic,

the widespread use of testing is an essential element in the development of effective public health strategies, but it is important to acknowledge the gaps that exist in the current evidence base and that, where possible, these should be addressed in future studies.

In regard to future research, more data are required to substantiate the effectiveness of tests to detect the presence of SARS-CoV-2 virus or antibodies to SARS-CoV-2 in different populations, and more evidence is needed to compare the effectiveness of laboratory-based testing and point-of-care testing strategies. Further clarity is required about the optimal timing of tests relative to symptom onset. The results of testing in people with no or minimal symptoms in a community-based setting needs further analysis and the impact of these data on public health measures needs to be fully analysed. Evidence should be prospectively collected during the implementation of public health strategies that combine testing with tracing and isolating individuals who have been in contact with COVID-19 sufferers. The UK National Institute for Health and Care Excellence have recently published an evidence standards framework which describes a three-stage approach to collecting the best possible data and evidence in the short and long term which is applicable both to established and developing COVID-19 tests.⁶⁷ The framework assumes that the tests' analytical performance is already established, that developers are complying with existing quality systems for manufacturers (ISO 13485) and laboratories (ISO 15198 or 17025), and recommends that a good diagnostic accuracy study should be followed by demonstrating the clinical significance as well as the economic impact of applying the index test.

Contributors DJ, MP, SM and PG conceived the project and prepared the review inclusion and exclusion criteria. JW prepared and ran search strategies with input from DJ. DJ and LE screened evidence, extracted data from relevant studies and carried out risk of bias assessments; data were independently checked by KC. DJ and MP analysed data to generate outcomes. DJ, LE, JW and PG wrote the manuscript with input and editing from all other authors.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iD

David Jarrom <http://orcid.org/0000-0002-6008-4758>

References

1 World Health Organization. Laboratory testing strategy recommendations for COVID-19: interim guidance, 2020. Available: <https://apps.who.int/iris/handle/10665/331509> [Accessed 11 May 2020].

2 World Health Organization. Surveillance strategies for COVID-19 human infection, 2020. Available: <https://www.who.int/publications/i/item/surveillance-strategies-for-covid-19-human-infection> [Accessed 11 May 2020].

3 Organisation for Economic Co-operation and Development. Testing for COVID-19: a way to lift confinement restrictions, 2020. Available: <http://www.oecd.org/coronavirus/policy-responses/testing-for-covid-19-a-way-to-lift-confinement-restrictions-89756248/> [Accessed 11 May 2020].

4 EPPICentre. COVID-19: a living systematic map of the evidence. The NIHR policy research programme reviews facility, 2020. Available: <http://eppi.ioe.ac.uk/cms/Projects/DepartmentofHealthandSocialCare/Publishedreviews/COVID-19LivingSystematicMapoftheEvidence/tabid/3765/Default.aspx> [Accessed 11 May 2020].

5 National Center for Biotechnology Information, U.S. National Library of Medicine. LitCovid - diagnosis, 2020. Available: <https://www.ncbi.nlm.nih.gov/research/coronavirus/> [Accessed 11 May 2020].

6 World Health Organization. Global research on coronavirus disease (COVID-19), 2020. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov> [Accessed 11 May 2020].

7 Health Technology Wales. Coronavirus (COVID-19): evidence digest, 2020. Available: <https://www.healthtechnology.wales/covid-19/covid-19-evidence-digest/> [Accessed 11 May 2020].

8 Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.

9 Freeman SC, Kerby CR, Patel A, *et al*. Development of an interactive web-based tool to conduct and interrogate meta-analysis of diagnostic test accuracy studies: MetaDTA. *BMC Med Res Methodol* 2019;19:81.

10 Liu W, Liu L, Kou G, *et al*. Evaluation of nucleocapsid and spike protein-based enzyme-linked immunosorbent assays for detecting antibodies against SARS-CoV-2. *J Clin Microbiol* 2020;58:e00461–20.

11 Xiang F, Wang X, He X, *et al*. Antibody detection and dynamic characteristics in patients with COVID-19. *Clin Infect Dis* 2020:ciaa461.

12 Zhao J, Yuan Q, Wang H, *et al*. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin Infect Dis* 2020:ciaa344.

13 Gao H-X, Li Y-N, Xu Z-G, *et al*. Detection of serum immunoglobulin M and immunoglobulin G antibodies in 2019 novel coronavirus infected patients from different stages. *Chin Med J* 2020;133:1479–80.

14 Guo L, Ren L, Yang S, *et al*. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). *Clin Infect Dis* 2020;71:778–85.

15 Sun B, Feng Y, Mo X, *et al*. Kinetics of SARS-CoV-2 specific IgM and IgG responses in COVID-19 patients. *Emerg Microbes Infect* 2020;9:940–8.

16 Zeng Z, Chen L, Pan Y, *et al*. Re: Profile of specific antibodies to SARS-CoV-2: the first report. *J Infect* 2020;81:e80–1.

17 Cassaniti I, Novazzi F, Giardina F, *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. doi:10.1002/jmv.25800. [Epub ahead of print: 30 Mar 2020].

18 Döhla M, Boesecke C, Schulte B, *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.

19 Hoffman T, Nissen K, Krambrich J, *et al*. Evaluation of a COVID-19 IgM and IgG rapid test; an efficient tool for assessment of past exposure to SARS-CoV-2. *Infect Ecol Epidemiol* 2020;10:1754538.

20 Jin Y, Wang M, Zuo Z, *et al*. Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019. *Int J Infect Dis* 2020;94:49–52.

21 Li H, Li Y, Zhang Z, *et al*. Establishment and clinical performance evaluation of 2019 novel coronavirus antibody colloidal gold detection method. *Chinese Journal of Infectious Diseases* 2020;38:E017.

22 Li Z, Yi Y, Luo X, *et al*. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol* 2020. doi:10.1002/jmv.25727. [Epub ahead of print: 27 Feb 2020].

23 Shen B, Zheng Y, Zhang X, *et al*. Clinical evaluation of a rapid colloidal gold immunochromatography assay for SARS-Cov-2 IgM/IgG. *Am J Transl Res* 2020;12:1348–54.

- 24 Spicuzza L, Montineri A, Manuele R, *et al.* Reliability and usefulness of a rapid IgM-IgG antibody test for the diagnosis of SARS-CoV-2 infection: a preliminary report. *J Infect* 2020;81:e53-4.
- 25 Lee Y-L, Liao C-H, Liu P-Y, *et al.* Dynamics of anti-SARS-Cov-2 IgM and IgG antibodies among COVID-19 patients. *J Infect* 2020;81:e55-8.
- 26 Lippi G, Salvagno GL, Pegoraro M, *et al.* Assessment of immune response to SARS-CoV-2 with fully automated MAGLUMI 2019-nCoV IgG and IgM chemiluminescence immunoassays. *Clin Chem Lab Med* 2020;58:1156-9.
- 27 Long Q-X, Liu B-Z, Deng H-J, *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med* 2020;26:845-8.
- 28 Pan Y, Li X, Yang G, *et al.* Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. *J Infect* 2020;81:e28-32.
- 29 Wu X, Fu B, Chen L, *et al.* Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China. *J Med Virol* 2020. doi:10.1002/jmv.25904. [Epub ahead of print: 20 Apr 2020].
- 30 Xie C, Jiang L, Huang G, *et al.* Comparison of different samples for 2019 novel coronavirus detection by nucleic acid amplification tests. *Int J Infect Dis* 2020;93:264-7.
- 31 Yong G, Yi Y, Tuantuan L, *et al.* Evaluation of the auxiliary diagnostic value of antibody assays for the detection of novel coronavirus (SARS-CoV-2). *J Med Virol* 2020. doi:10.1002/jmv.25919. [Epub ahead of print: 22 Apr 2020].
- 32 Yongchen Z, Shen H, Wang X, *et al.* Different longitudinal patterns of nucleic acid and serology testing results based on disease severity of COVID-19 patients. *Emerg Microbes Infect* 2020;9:833-6.
- 33 Xu W, Li J, He X, *et al.* The diagnostic value of joint detection of serum IgM and IgG antibodies to 2019-nCoV in 2019-nCoV infection. *Chinese Journal of Laboratory Medicine* 2020;43:E012.
- 34 Zhang G, Nie S, Zhang Z, *et al.* Longitudinal change of severe acute respiratory syndrome coronavirus 2 antibodies in patients with coronavirus disease 2019. *J Infect Dis* 2020;222:183-8.
- 35 Kim H, Hong H, Yoon SH. Diagnostic performance of CT and reverse transcriptase polymerase chain reaction for coronavirus disease 2019: a meta-analysis. *Radiology* 2020;296:E145-55.
- 36 Baek YH, Um J, Antigua KJC, *et al.* Development of a reverse transcription-loop-mediated isothermal amplification as a rapid early-detection method for novel SARS-CoV-2. *Emerg Microbes Infect* 2020;9:998-1007.
- 37 Harrington A, Cox B, Snowdon J, *et al.* Comparison of Abbott ID Now and Abbott m2000 methods for the detection of SARS-CoV-2 from nasopharyngeal and nasal swabs from symptomatic patients. *J Clin Microbiol* 2020;58:e00798-20.
- 38 Lu R, Wu X, Wan Z, *et al.* A novel reverse transcription loop-mediated isothermal amplification method for rapid detection of SARS-CoV-2. *Int J Mol Sci* 2020;21:2826.
- 39 Yan C, Cui J, Huang L, *et al.* Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. *Clin Microbiol Infect* 2020. [Epub ahead of print: 8 Apr 2020].
- 40 Zhen W, Smith E, Manji R, *et al.* Clinical evaluation of three sample-to-answer platforms for detection of SARS-CoV-2. *J Clin Microbiol* 2020;58:e00783-20.
- 41 Corman VM, Landt O, Kaiser M, *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 2020;25:2000045.
- 42 Public Health England. COVID-19: track coronavirus cases. 2020, 2020. Available: <https://www.gov.uk/government/publications/covid-19-track-coronavirus-cases> [Accessed 6 Aug 2020].
- 43 Fang Z, Zhang Y, Hang C, *et al.* Comparisons of viral shedding time of SARS-CoV-2 of different samples in ICU and non-ICU patients. *J Infect* 2020;81:147-78.
- 44 Azzi L, Carcano G, Gianfagna F, *et al.* Saliva is a reliable tool to detect SARS-CoV-2. *J Infect* 2020;81:e45-50.
- 45 Chan JF-W, Yip CC-Y, To KK-W, *et al.* Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay validated *in vitro* and with clinical specimens. *J Clin Microbiol* 2020;58:e00310-20.
- 46 Chen Y, Chen L, Deng Q, *et al.* The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *J Med Virol* 2020;92:833-40.
- 47 Guo W-L, Jiang Q, Ye F, *et al.* Effect of throat washings on detection of 2019 novel coronavirus. *Clin Infect Dis*;382:ciaa416.
- 48 Huang Y, Chen S, Yang Z, *et al.* SARS-CoV-2 viral load in clinical samples from critically ill patients. *Am J Respir Crit Care Med* 2020;201:1435-8.
- 49 Lin C, Xiang J, Yan M, *et al.* Comparison of throat swabs and sputum specimens for viral nucleic acid detection in 52 cases of novel coronavirus (SARS-Cov-2)-infected pneumonia (COVID-19). *Clin Chem Lab Med* 2020;58:1089-94.
- 50 Liu R, Han H, Liu F, *et al.* Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clin Chim Acta* 2020;505:172-5.
- 51 Péré H, Podglajen I, Wack M, *et al.* Nasal swab sampling for SARS-CoV-2: a convenient alternative in times of nasopharyngeal swab shortage. *J Clin Microbiol* 2020;58:e00721-20.
- 52 Wang W, Xu Y, Gao R, *et al.* Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* 2020;323:1843-4.
- 53 Wang X, Tan L, Wang X, *et al.* Comparison of nasopharyngeal and oropharyngeal swabs for SARS-CoV-2 detection in 353 patients received tests with both specimens simultaneously. *Int J Infect Dis* 2020;94:107-9.
- 54 Williams E, Bond K, Zhang B, *et al.* Saliva as a noninvasive specimen for detection of SARS-CoV-2. *J Clin Microbiol* 2020;58:e00776-20.
- 55 Wu Y, Guo C, Tang L, *et al.* Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *Lancet Gastroenterol Hepatol* 2020;5:434-5.
- 56 Xia J, Tong J, Liu M, *et al.* Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection. *J Med Virol* 2020;92:589-94.
- 57 Ye G, Li Y, Lu M, *et al.* Experience of different upper respiratory tract sampling strategies for detection of COVID-19. *J Hosp Infect* 2020;105:1-2.
- 58 Zhang J, Wang S, Xue Y. Fecal specimen diagnosis 2019 novel coronavirus-infected pneumonia. *J Med Virol* 2020;92:680-2.
- 59 Zheng S, Fan J, Yu F, *et al.* Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang Province, China, January-March 2020: retrospective cohort study. *BMJ* 2020;369:m1443.
- 60 Hunter E, Price DA, Murphy E, *et al.* First experience of COVID-19 screening of health-care workers in England. *Lancet* 2020;395:e77-8.
- 61 Keeley AJ, Evans C, Colton H, *et al.* Roll-out of SARS-CoV-2 testing for healthcare workers at a large NHS Foundation Trust in the United Kingdom, March 2020. *Eurosurveillance* 2020;25:2000433.
- 62 Shen N, Zhu Y, Wang X, *et al.* Characteristics and diagnosis rate of 5630 subjects receiving SARS-CoV-2 nucleic acid tests from Wuhan, China. *JCI Insight* 2020;5:e137662.
- 63 Kong W-H, Li Y, Peng M-W, *et al.* SARS-CoV-2 detection in patients with influenza-like illness. *Nat Microbiol* 2020;5:675-8.
- 64 Spellberg B, Haddix M, Lee R, *et al.* Community prevalence of SARS-CoV-2 among patients with influenzalike illnesses presenting to a Los Angeles medical center in March 2020. *JAMA* 2020;323:1966-7.
- 65 Sutton D, Fuchs K, D'Alton M, *et al.* Universal screening for SARS-CoV-2 in women admitted for delivery. *N Engl J Med* 2020;382:2163-4.
- 66 World Health Organization. Global surveillance for human infection with novel coronavirus (2019-nCoV), 2020. Available: <https://apps.who.int/iris/bitstream/handle/10665/330857/WHO-2019-nCoV-SurveillanceGuidance-2020.3-eng.pdf> [Accessed 11 May 2020].
- 67 National Institute for Health and Care Excellence. Evidence standards framework for SARS-CoV-2 and anti-SARS-CoV-2 antibody diagnostic tests, 2020. Available: <https://www.nice.org.uk/Media/Default/About/what-we-do/covid-19/Diagnostic-tests-for-COVID-19-evidence-standards-framework.pdf> [Accessed 11 May 2020].
- 68 Fang Y, Zhang H, Xie J, *et al.* Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* 2020;296:E115-7.
- 69 He J-L, Luo L, Luo Z-D, *et al.* Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. *Respir Med* 2020;168:105980.
- 70 Lee TH, Lin RJ, Lin RTP, *et al.* Testing for SARS-CoV-2: can we stop at two? *Clin Infect Dis* 2020:ciaa459.
- 71 Zhang J-J, Cao Y-Y, Dong X, *et al.* Distinct characteristics of COVID-19 patients with initial rRT-PCR-positive and rRT-PCR-negative results for SARS-CoV-2. *Allergy* 2020;75:1809-12.