Technoleg Iechyd Cymru
Health Technology Wales

The clinical effectiveness of tests to detect the presence of SARS-CoV-2 virus, and antibodies to SARS-CoV-2, to inform COVID-19 diagnosis

This report has been produced to assist the Welsh Government and Health and Social Care in Wales respond to the Coronavirus disease 2019 (COVID-19) pandemic. It is based on the most recent available evidence at the time of publication (date of publication 23 April 2020, to include all evidence published up to 14 April 2020) but will be updated frequently.

Executive summary

Tests for the presence of SARS CoV-2 virus

- Health Technology Wales (HTW) Researchers searched for, appraised and summarised all published evidence on the diagnostic performance, effectiveness or economic impact of tests used to detect the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus to inform COVID-19 diagnosis.
- We identified 22 studies reporting diagnostic accuracy, detection rates and the time taken to obtain test results. The majority of studies (20/22) tested hospitalised, symptomatic patients with a strong clinical suspicion of COVID-19. Studies in people with milder symptoms are comparatively limited in number (two studies were identified).
- All the tests studied used laboratory-based polymerase chain reaction (PCR) protocols, with one exception; a single study of loop-mediated isothermal amplification assay. The majority of studies investigated tests at an early stage in their development, before any wider deployment and commercialisation. We did not identify any evidence on the effectiveness of any specific commercially available tests for the presence of SARS-CoV-2.
- The majority of the studies did not include methods of confirmatory/differential diagnosis to validate the test results obtained (e.g. the proportion of likely false positive and negative results). The lack of a generally accepted reference standard to compare reverse transcription PCR (RT-PCR) results against makes it challenging to assess the true diagnostic accuracy of these tests as method of diagnosing COVID-19.
- Tentatively, it appears the type of sample obtained, the part of the body sampled, and the timing of test relative to symptom onset could be influential on test results and accuracy, but we did not identify evidence with enough certainty to guide how these factors could be used to optimise testing.
- There are important gaps in the available evidence on the effectiveness of tests for the presence of SARS-CoV-2. We did not identify any studies of virus testing in asymptomatic patients, or in any specific populations such as healthcare workers. Furthermore, no evidence is available for any point-of-care or near-patient tests. We did not identify any evidence on the economic impact of any test, or how any test influences subsequent patient management.

Tests for the presence of SARS CoV-2 antibodies

- Health Technology Wales Researchers searched for, appraised and summarised all published evidence on the diagnostic performance, effectiveness or economic impact of tests used to detect antibodies to the SARS-CoV-2 virus to inform COVID-19 diagnosis.
- We identified 9 studies reporting diagnostic accuracy or detection rates of SARS-CoV-2 antibody tests. These used a range of different assay methods to detect a range of different antibody targets. Most tests were laboratory-based but we identified two studies evaluating point of care tests.
- The majority of studies investigated tests at an early stage in their development, before any wider deployment and commercialisation. Published evidence is available on the effectiveness of two commercially available (point of care) tests. All studies tested hospitalised, symptomatic patients with a strong clinical suspicion of COVID-19 (or in a few cases, healthy volunteers).
- Six studies reported estimates of test sensitivity and specificity. Sensitivity reported in the studies ranged from 18.4% to 96.1%. Specificity was more consistent across studies and ranged from 90.9% to 100%. Test results were, in most cases, validated by comparing them to the results of RT-PCR tests: as noted on page 1, a true assessment of the accuracy of RT-PCR test results is very challenging, and using these RT-PCR for validation mean the same issues apply to the results of antibody tests studied in this way.
- Several studies reported how the timing of testing relative to symptom onset influences test results, which could potentially be used to guide appropriate timing of antibody testing, although more evidence is required to allow firm conclusions on this to be reached.
- At present, key gaps exist in the available evidence on antibody tests as a method of informing COVID-19 diagnosis. We did not identify any studies of antibody testing in people outside of hospital, such as those with milder symptoms or in other settings such as community or home-based testing. We also did not identify any evidence on use of the tests in specific populations, such as healthcare workers. Finally, we did not identify any evidence on the time taken to obtain test results, the economic impact of any test, or how any test influences subsequent patient management.

1. Purpose of the evidence appraisal report

This report aims to identify and summarise evidence that addresses the following questions:

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

HTW Evidence Appraisal Reports are based on rapid systematic literature searches, with the aim of identifying the best published clinical and economic evidence on health technologies. Researchers critically appraise and summarise this evidence. The methods used to identify, assess and summarise evidence are described in Section 5.

Updated literature searches for this report will be performed regularly and any new evidence materially influencing findings will be included in an updated report. Please see Appendix 1 for the revision history of the document.

2. Introduction/Background

In December 2019, a novel coronavirus was discovered in Wuhan, China and has since spread rapidly across the world. This novel coronavirus was named SARS-CoV-2 and causes a disease called COVID-19.

Tests for COVID-19 fall into two broad groups:

- Tests that detect the presence of the SARS-CoV-2 virus. These tests can be used to diagnose people with ongoing COVID-19 infection. We will refer to these as 'virus tests'.

- Tests that detect the presence of antibodies to SARS-CoV-2. Antibodies are produced after SARS-CoV-2 infection as part of the body's immune response. These tests can be used to diagnose COVID-19 cases after infection. We will refer to these as 'antibody tests'.

Tests can be carried out in a laboratory or at point of care in a range of settings.

The purpose of this review will be to identify, appraise and summarise evidence on the diagnostic performance and effectiveness of these tests. This has initially involved reviewing all evidence published since the beginning of the COVID-19 outbreak. HTW will now carry out routine surveillance for new evidence and produce frequent updates to this report as new evidence emerges.

3. Virus tests

3.1. Clinical effectiveness of virus tests

We identified one systematic review that searched for evidence on potential rapid diagnostics, vaccines or therapeutics for SARS-CoV-2 published between 1 December 2019 and 6 February 2020 (Pang et al. 2020). Characteristics are outlined in Appendix 5, Table 1. Only one study was identified, which explored development of RT-PCR assays (Corman et al. 2020). However, this study reported no outcomes of interest so was excluded from this review. The review also included studies on the related previous SARS coronavirus and Middle East respiratory syndrome (MERS) coronavirus, but these studies were excluded based on our selection criteria.

We identified a further 22 sources reporting primary data on the evaluation of tests for SARS-CoV-2 virus detection. The design and characteristics of each study is summarised in Appendix 5, Table 1. Key outcomes are summarised in **Table** 1

All of the virus tests we identified were molecular, i.e. based on detection of amplified viral SARS-CoV-2 nucleic acid sequences. We did not identify any evidence on the effectiveness of tests that use immunological assays to directly detect SARS-CoV-2, i.e. the detection of the presence of viral antigens. The majority of tests were laboratory-based RT-PCR tests, conducted using standard in-house or commercially available PCR reagents and equipment (in some cases assay details were not reported). The RT-PCR primer used (i.e. which part of the viral RNA is targeted and amplified) varied between studies, although in some cases primer details were not reported. We did not identify any evidence on commercially available point-of-care or near-patient PCR tests.

In addition to RT-PCR, we identified one study reporting the diagnostic performance of a loop-mediated isothermal amplification assay (LAMP) to detect viral nucleic acids. The authors state that this has the potential to be used at point of care, but in the study concerned the test was laboratory-based.

The lack of a generally accepted reference standard to compare RT-PCR results against makes it challenging to assess the true diagnostic accuracy of these tests as a method of diagnosing COVID-19. Several studies reported detection rate (proportion of test results that were positive) without reporting any validation of the results. In other studies, serial tests for virus carried out at different time points were either compared to the eventual confirmed molecular diagnosis (any patient that eventually returned a positive PCR result was treated as positive) or PCR was compared to clinical diagnosis such as chest imaging. Some studies also compared different PCR methods, or different methods of sampling. We also identified one study that assessed the diagnostic performance of LAMP (Yan et al. 2020a) using the results from RT-PCR as a reference standard. Key results are described in the following sections and in **Table** 1; studies are described in more detail in Table 1 of Appendix 5.

3.1.1. Diagnostic accuracy

One study reported the diagnostic accuracy of LAMP in the diagnosis of 130 patients with suspected COVID-19, using equivalent test results from RT-PCR as a reference standard. There was complete concordance between LAMP and RT-PCR: 58 patients tested positive and 72 tested negative with each test method. Sensitivity and specificity for LAMP were therefore 100% (95% CI 92.3% - 100%) and 100% (95% CI 93.7% - 100%) respectively.

3.1.2. Detection rates

Five studies reported the proportion of initial positive test results in people admitted to hospital and ultimately diagnosed as having COVID-19 (either on the basis of laboratory or clinical findings). Four studies included people (n = 409 total) with signs or symptoms of coronavirus and who eventually tested positive for SARS-CoV-2 by RT-PCR. The initial detection rate (on first RT-PCR test) varied from 71% to 90.6%. In some cases multiple tests were required before a positive result was obtained; the maximum reported number of negative tests before a subsequent positive test was six (1 patient out of a sample of 290). A fifth study included patients clinically diagnosed with COVID-19 based on a chest computed tomography (CT) image that demonstrated viral pneumonia. Out of 610 included cases, 168 (27.5%) had a positive initial RT-PCR test; a further 48 (7.9%) were positive on second test.

In addition to the studies above, which studied people with relatively severe disease and high suspicion of COVID-19 infection, we identified two studies (Kong et al. 2020, Spellberg et al. 2020) that used RT-PCR to detect SARS-CoV-2 in people with milder, influenza-like symptoms. These

analysed samples from a total of 731 patients and reported SARS-CoV-2 detection rates of 1.4% (one study, 640 patients in Wuhan, China) and 5.3% (one study, 131 patients in California).

3.1.3. Time to diagnosis

Two studies of laboratory-based RT-PCR tests for SARS-CoV-2 reported the time taken to obtain a diagnosis. One study (Amrane et al. 2020) reported a mean time to result from the time a sample arrived at the laboratory of 175 minutes (range 150 to 195 minutes). This was based solely on the first 22 tests but the authors noted that the time to obtain subsequent results did not exceed 3 hours. A second study (Won et al. 2020) reported an estimated whole procedure time (including collection of sample) of 230 minutes. In the single study that used LAMP to diagnose SARS-CoV-2, mean procedure time was 26.3 minutes (Yan et al. 2020a).

3.1.4. Comparisons to other methods of diagnosis

Three studies compared laboratory diagnosis of COVID-19 using RT-PCR to clinical diagnosis based on chest CT scan. Two of the studies included confirmed positive cases: in the study by Fang et al (2020a) the disease detection rates using RT-PCR and CT scan were 36/51 (71%) and 50/51 (98%) respectively; in the study by Long et al (2020) the disease detection rates using RT-PCR and CT scan were 30/36 (84.6%) and 35/36 (97.2%) respectively. A third study (Ai et al. 2020) included 1,014 patients with suspected COVID-19 but did not report a confirmed final diagnosis. Disease detection rates using RT-PCR and CT scan were 601/1014 (59%) and 888/1014 (88%) respectively

3.1.5. Other comparisons

One study (Chan et al. 2020) reported SARS-CoV-2 detection rates for RT-PCR assays using two different sets of primers. The detection rate for RT-PCR using the RdRp/Hel primer was 119/273 (43.6%). The corresponding detection rate with the RdRp-P2 primer was 77/273 (28.2%).

We identified ten studies that compared RT-PCR of SARS-CoV-2 with samples from different parts of the body. We have collated these in Table 2. Chan et al. (2020) also reported detection rates from different sample sites for two assays: RT-PCR RdRp/Hel assay (their developed assay and RdRp-P2 assay (standard assay used in many laboratories). RdRp/Hel had a higher rate of detection in respiratory tract samples than non-respiratory samples, with 102/120 (85%) respiratory specimens and 17/153 (11.1%) of non-respiratory specimens. This was significantly higher than the detection rate of the RdRp-P2 assay (73/120 [60.8%], $p<0.001$ and 4/153 [2.6%], $p = 0.005$, respectively). Further breakdown of detection by sample site can be seen in Table 2.

Table 1. SARS-CoV-2 viral tests: outcomes of interest

Outcome	Reference	Index test target	Number of patients/samples	Index test; Comparator (if applicable)	Comments
Detection rate	Ai et al. (2020)	Not specified	n = 1,014 patients	RT-PCR: 601/1014 (59%; 95% CI 56% to 62%); CT scan: 888/1014 (88%, 95% CI 86% to 90%)	
	Amrane et al. (2020)	E and spike assays	n = 280 patients	0/280 (0%)	A multiplex molecular assay for other respiratory pathogens detected non-SARS-CoV-2 viral infection in 137/280 (48.9%) patients.
	Chan et al. (2020)	RdRp/Hel	n = 273 samples	RT-PCR (RdRp/Hel): 119/273 (43.6%); RT-PCR (RdRp-P2): 77/273 (28.2%) p < 0.001	Results on first testing. Reference standard: eventual confirmed diagnosis with RT-PCR (RdRp-P2)
	(Kong et al. 2020)	Orf1ab, N	n = 640 samples	RT-PCR: 9/640 (1.4%)	Tests conducted on outpatients with influenza-like symptoms. Some samples were collected before the first recording cases of COVID-19 were reported.
	(Liu et al. 2020a)	Orf1ab, N	n = 4,880 patients	RT-PCR (Orf1ab AND N assay): 1875/4880 (38.42%)	Based on positive detection of in both primer assays. Individual assay detection rates were 39.80% for the N assay and 40.98% for Orf1ab.
	(Spellberg et al. 2020)	NR	n = 131 samples	RT-PCR: 7/131 (5.3%)	Tests conducted on patients presenting with mild influenza-like symptoms; no suspicion of COVID-19.
	Wang et al. (2020)	NR	n = 1,070 samples	273/1070 (25.5%)	Includes samples obtained from various sites.
	Xie et al. (2020)	NR	n = 19 patients	RT-PCR: 9/19 (47.4%)	

Outcome	Reference	Index test target	Number of patients/samples	Index test; Comparator (if applicable)	Comments
	Ye et al. (2020)	NR	n = 91 patients	47/91 (51.6%)	
Detection rate/Sensitivity	Fang et al. (2020a)[1]	NR	n = 51 patients	RT-PCR: 36/51 (71%, 95% CI 56% to 83%); CT scan: 50/51 (98%, 95% CI 90% to 100%) p < 0.001	Based on first RT-PCR testing. (12/51 received a positive second test; 2/51 received a positive third test; 1/51 received a positive fourth test.) Eventual positive from RT-PCR was the reference standard.
	Fang et al. (2020b)[1]	NR	n = 32 patients	RT-PCR: 29/32 (90.6%)	Based on first RT-PCR testing result. Eventual positive from RT-PCR was the reference standard.
	Li et al. (2020b)[2]	NR	n = 610 patients	RT-PCR: 168/610 (27.5%)	Based on initial RT-PCR testing. 48/610 (7.9%) received a positive on second test. Reference standard: Clinical diagnosis of COVID-19 (CT scan)
	Long et al. (2020)[1]	NR	n = 36 patients	RT-PCR: 30/36 (84.6%); CT scan: 35/36 (97.2%)	Based on initial RT-PCR testing. 3/36 had a positive result at second testing and the remaining 3/36 had a positive third test. (Reference standard: eventual confirmed positive RT-PCR)
	(Zhang et al. 2020b)[1]	Orf1ab, N	n = 290	RT-PCR: 249/290 (85.9%)	Based on first RT-PCR testing result. Patients testing negative were re-tested and only those who eventually tested postivie were included in the results. Cumulative proportion of patients tested positive after each round of testing: 2nd test: 270/290 (93.1%)

Outcome	Reference	Index test target	Number of patients/samples	Index test; Comparator (if applicable)	Comments
					3rd test: 283/290 (97.6%) 4th test: 287/290 (99.0%) 5th test: 289/290 (99.7%) 6th test: 290/290 (100%) Eventual positive from RT-PCR was the reference standard.
Sensitivity	Yan et al. (2020b)	Orf1ab and spike	n = 130 specimens	100% (95% CI 92.3% to 100%)	RT-PCR (primer NR) was the reference standard.
Specificity	Yan et al. (2020b)	Orf1ab and spike	n = 130 specimens	100% (95% CI 93.7% to 100%)	RT-PCR (primer NR) was the reference standard.
Mean time to test result	Amrane et al. (2020)	E and spike assays	n = 22 patients	175 minutes (range 150 to 195 minutes)	Based on first 22 tests. Subsequent results did not exceed 3 hours.
Procedure time	Won et al. (2020)	NR	n = 12 healthy volunteers	230 minutes	Includes collection of sample
Procedure time, mean (±SD)	Yan et al. (2020b)	Orf1ab and spike	n = 130 specimens	26.28 minutes ± 4.48 minutes	

[1]All COVID-19 diagnoses assumed to be positive by the study authors based on positive RT-PCR results (after multiple tests in some cases) [21]All diagnoses assumed to be positive by the study authors based on positive chest imaging.

CI: confidence interval; CT: computed tomography; SD: standard deviation; RT-PCR: reverse transciption polymerase chain reaction; NR: details not reported

Table 2. SARS-CoV-2 virus tests: detection by sample site

Study, RT-PCR assay target	BLF	Pharyngeal*	Throat wash	Lingual	Saliva	Sputum	Plasma/blood	Urine	Faeces and/or rectal swabs	Tears	Fibrobronchoscope brush biopsy
Chan et al. (2020), RdRp/Hel	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 et al. (2020), RdRp-P2	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 et al. 2020b)	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 et al. 2020b)	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 et al. (2020b)	n/a	32/32 (100%)	n/a	n/a	25/32 (78%)	n/a	23/32 (72%)	0/32 (0.0%)	NR	5/32 (16%)	n/a
Liu et al. (2020a), ORF1ab & N	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 et al. (2020)	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%)
Wu et al. (2020)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	41/74 (55%)	n/a	n/a
Xia et al. (2020)	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 et al. (2020)	n/a	9/19	n/a	n/a	n/a	n/a	0/19	0/19	8/19	n/a	n/a
Ye et al. (2020)	n/a	40/91 (44.0%)	n/a	33/91 (36.3%)	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Zhang et al. (2020a)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	5/14 (35.7%)	n/a	n/a

RT-PCR; reverse transcriptase polymerase chain reaction; BLF: Bronchoalveolar lavage fluid; n/a: not included in study; NR: sampling included in study but outcome not reported;

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

4. Antibody tests

Clinical effectiveness

We identified nine primary studies evaluating the detection of antibodies against SARS-CoV-2. Two of these were published in Chinese with English-language abstracts but have been included based on information reported in the abstracts.

Details of each study's design and characteristics are summarised in Appendix 5, Table 2. The tests studied used a range of different assay methods to detect one or more antibody type (different immunoglobulin classes and/or antibody targeting). Most tests were laboratory-based and used standard reagents and equipment used to conduct antibody testing. We identified two studies (Cassanati 2020, Li 2020c) assessing point-of-care tests, both of which were lateral flow immunoassays targeting IgM/IgG (VivaChek, Jiangsu Medomics). Where a reference standard was included, this was RT-PCR except for one study that used either RT-PCR or clinical diagnosis to determine final disease status. As noted in Section 3, using RT-PCR to diagnose COVID-19 also results in a proportion of tests that are falsely negative or positive, and this should be taken into account when interpreting the diagnostic accuracy figures reported for antibody tests. Study outcomes are summarised in Table 3 and the following section.

4.1.1. Diagnostic accuracy

Six studies (980 patients included; number not clear for one study) reported sensitivity and specificity, or sufficient information to allow these to be calculated. As noted above, the range of different antibody types and targets used means that pooling data across studies would not be appropriate. Sensitivity reported in the studies ranged from 18.4% to 96.1%. Notably, the lowest reported sensitivity was for a point-of-care test (Cassanati, 2020), although sensitivity figures below 50% were also reported for one laboratory test (Jin, 2020). Specificity was more consistent across studies and ranged from 90.9% to 100%.

4.1.2. Other comparisons

One study (Liu et al. 2020b) tested two different immunoassays in the same population (n = 214): one targeting antibodies for the SARS-CoV-2 N protein and one targeting the spike protein. This study only reported detection rates and did not verify test results against a reference standard. Detection rates were comparable for assays against the two targets (detection of Immunoglobulin M (IgM) and/or Immunoglobulin G (IgG): 172/214 (80.4%) for N protein assay and 176/214 (82.2%) for spike protein assay; see Table 3 for results for individual immunoglobulins).

(Li et al. 2020c) measured sensitivity and specificity in a sample of 525 patients using inactivated venous blood. However, they also compared results with fingerstick blood, venous blood and plasma in a smaller sample (seven COVID-19 patients and three healthy volunteers were recruited). Test results were consistent across the different blood samples: 3 of the 7 COVID-19 patients were IgM-only positive and 4 patients were both IgM and IgG positive; all healthy volunteers tested negative.

Table 3. SARS-CoV-2 immunological tests: outcomes of interest

Outcome	Reference	Index test assay	Number of patients/samples	Result	Comments
Detection rate	Cassaniti et al. (2020)	LFIA , VivaChek POC	n = 110 patients	Healthy volunteers 0/30 (0%); COVID-19 patients 19/30 (63.3%); Suspected cases 0/50 (0%)	Based on being fully positive for IgM and IgG together (weakly positive not included). Authors considered sensitivity of the rapid LFIA to be sub-optimal based on the results with known COVID-19 patients (data not reported). Suggested reasons were low antibody titers or delayed immune response.
	Guo et al. (2020a)	IgM, IgG or IgA ELISA	n = 208 specimens	IgM: 188/208 (90.4%); IgA: 194/208 (93.3%); IgG: 162/208 (77.9%)	Samples were obtained from acute, middle or late stages of infection. This includes confiirmed and probable cases of COVID-19.
	Li et al. (2020a)	IgM or IgG colloidal gold	n = 189	IgM: 113/189 (59.8%); IgG: 100/189 (52.9%); IgM/IgG: 125/189 (66.1%)	Population was probably cases of COVID-19 (PCR negative test but clinical manifestations).
	Jin et al. (2020)	CLIA (N and spike proteins)	n = 34	IgM: 19/34 (55.9%) IgG: 32/34 (94.1%)	Detection rate of antibody tests after 2 negative PCR tests (in a 24 hour interval).
Detection rate/Sensitivity	Gao et al. (2020)	CLIA/ELISA/GICA	n = 14 patients	IgM CLIA: 6/14 (42.9%);	Data for the late stage of infection (18-24 days post-disease onset)

				IgM ELISA: 6/14 (42.9%); IgM GICA: 9/14 (64.3%)	
	Gao et al. (2020)	CLIA/ELISA/GICA	n = 14 patients	IgG CLIA: 9/14 (64.3%); IgG ELISA: 12/14 (85.7%); IgG GICA: 11/14 (78.6%)	Data for the late stage of infection (18-24 days post-disease onset)
	Liu et al. (2020b)	ELISA (target: N-protein)	n = 214 patients	IgM: 146/214 (68.2); IgG: 150/214 (70.1%); IgM and/or IgG: 172/214 (80.4%)	Samples were acquired at different times post disease onset (median 15 days, range 0 to 55).
	Liu et al. (2020b)	ELISA (target: spike protein)	n = 214 patients	IgM: 165/214 (77.1%); IgG: 159/214 (74.3%) IgM and/or IgG: 176/214 (82.2%)	Samples were acquired at different times post disease onset (median 15 days, range 0 to 55).
Sensitivity	Cassaniti et al. (2020)	LFIA , VivaChek POC	n = 50 (suspected cases only)	IgM/IgG: 18.4%	Diagnostic accuracy considered both 'positive' and 'weakly positive' test results as positive.
	Li et al. (2020a)	colloidal gold	Population not clear	IgM: 78.7%; IgG: 73.0% IgM/IgG: 87.6%	Limited based on abstract so calculations etc not clear.

	Li et al. (2020c)	LFIA, Jiangsu Medomics POC	n = 525 specimens	IgM/IgG: 88.66%	A positive result was whether results were IgM positive, IgG positive or IgM and IgG positive.
	Xu et al. (2020)	Fully-automated assay (NR)	n = 205 patients	IgM: 70.24%(144/205) IgG: 96.10%(197/205)	Cohort included COVID-19 diagnosed by positive RT-PCR (n = 186) and COVID-19 diagnosed by clinical manifestations (n = 19).
	Zhao et al. (2020)	ELISA (spike for IgM and Ab; N for IgG)	n = 173	IgM: 82.7% (143/173); IgG: 64.7% (112/173); Ab: 93.1% (161/173); RT-PCR: 67.1%*	Includes samples acquired at different time points post-disease onset. *112 patients tested positive using RT-PCR at various time points. The total population on which the sensitivity figure is calculated is not clear; potentially due to some patients not receiving PCR tests at certain time points.
	Jin et al. (2020)	CLIA (N and spike proteins)	n = 27	IgM: 48.1% (13/27) IgG: 88.9% (24/27)	Sensitivity was calculated using a subgroup of the full COVID-19 cohort (n = 43); patient who had a serological test prior to getting a negative RT-PCR (reference standard).
Specificity	Cassaniti et al. (2020)	LFIA , VivaChek POC	n = 50 (suspected cases only)	IgM/IgG: 91.7%	Diagnostic accuracy included both positive and weakly positive results as positive.
	Li et al. (2020a)	colloidal gold	Population not clear	IgM: 98.2%; IgG: 99.3%; IgM/IgG: 98.2%	Limited based on abstract so calculations etc not clear.
	Li et al. (2020c)	LFIA, Jiangsu Medomics POC	n = 525 specimens	IgM/IgG: 90.63%.	A positive result was whether results were IgM positive, IgG positive or IgM and IgG positive.
	Liu et al. (2020b)	ELISA (spike)	n = 100 healthy controls	IgM: 100% (0/100); IgG: 100% (0/100)	

				IgM and/or IgG: 100% (0/100)	
	Xu et al. (2020)	Fully-automated assay (NR)	n = 79 patients	IgM: 96.20% (76/79) IgG: 92.41%(73/79)	Based on 'control' cohort with other diseases (but negative for COVID-19)
	Zhao et al. (2020)	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)	Specificity was based on a cohort of healthy individuals who were tested with the assays prior to the SARS-CoV-2 outbreak.
	Jin et al. (2020)	CLIA (N and spike proteins)	n = 33	IgM: 100% (33/33) IgG: 90.9% (30/33)	Based on a 'control' cohort of patients with suspected COVID-19, but were discharged from hospital based on 2 negative PCR tests in a 24 hour period.
NPV	Cassaniti et al. (2020)	LFIA , VivaChek POC	n = 50 (suspected cases only)	IgM/IgG: 26.2%,	Diagnostic accuracy included both positive and weakly positive results as positive.
	Xu et al. (2020)	Fully-automated assay (NR)	n = 79	IgM/IgG: 91.03% (71/78); RT-PCR: 80.61% (79/98)	Based on 'control' cohort with other diseases (but negative for COVID-19). It is not clear how the IgM/IgG caculation was derived, in terms of whether it used double positive results only (IgM and IgG) or included patients that were positive for one antibody test (IgM and/or IgG).
	Jin et al. (2020)	CLIA (N and spike proteins)	n = 60	IgM: 100% (13/13) IgG: 88.9% (24/27)	Based on a control group (n = 33) and a subgroup of the COVID-19 cohort where patients had received an antibody test before testing negative on RT-PCR (n = 27).
PPV	Cassaniti et al. (2020)	LFIA , VivaChek POC	n = 50 (suspected cases only)	IgM/IgG: 87.5%	Diagnostic accuracy included both positive and weakly positive results as positive.
	Xu et al. (2020)	Fully-automated assay (NR)	n = 205 patients	IgM/IgG: 95.63%(197/206);	Cohort included COVID-19 diagnosed by positive RT-PCR (n = 186) and COVID-19 diagnosed by clinical manifestations (n = 19).

				RT-PCR: 100% (186/186)	It is not clear how the IgM/IgG caculation was derived, in terms of whether it used double positive results only (IgM and IgG) or included patients that were positive for one antibody test (IgM and/or IgG).
	Jin et al. (2020)	CLIA (N and spike proteins)	n = 60	IgM: 70.2% (33/47) IgG: 90.9% (30/33)	Based on a control group (n = 33) and a subgroup of the COVID-19 cohort where patients had received an antibody test before testing negative on RT-PCR (n = 27).
CLIA: Chemiluminescent immunoassay; ELISA: enzyme-linked immunosorbent assay; GICA: gold immunochromatography assay; IgA: Immunoglobulin A; IgG: Immunoglobulin G; IgM: Immunoglobulin M; LFIA: lateral flow immunoassay; RT-PCR: reverse transcription polymerase chain reaction; NR: details not reported					

5. Conclusions

This is the first version of a living evidence review on the effectiveness of tests to inform COVID-19 diagnosis. We intend to carry out ongoing surveillance of the evidence, and this report will be updated frequently as new evidence emerges.

We searched for and appraised all available evidence on the effectiveness of tests for the presence of the SARS-CoV-2 virus, or antibodies to the virus, up to 14 April 2020. As of this date, we identified 22 published studies reporting on the effectiveness of tests for the presence of virus, and 9 studies testing for presence of antibodies. In some cases, evidence was reported as correspondence or short communications (exemplifying the rapid pace of research on COVID-19) which limited the reporting of detail on how some tests were conducted. Two studies were also available only in Chinese, with an English abstract: these included sufficient outcome data to be included here, but again this limits the details available about these studies.

The majority of evidence is from China and almost all studies reported on the use of laboratory-based PCR tests in the hospital setting, in symptomatic patients with confirmed or suspected COVID-19 infection. Data on testing in other settings is comparatively limited: two studies (Kong et al. 2020, Spellberg et al. 2020) used RT-PCR to detect SARS-CoV-2 in people with milder, influenza-like symptoms. These only reported SARS-CoV-2 detection rates and not any other outcomes. No studies were found that used antibody tests outside of hospital settings.

Of the 22 studies of virus tests, five studies attempted to validate detection rates, (i.e. assess the proportion of positive tests that could be considered true positive, and the proportion that were false negative). However, the lack of a generally accepted reference standard to compare reverse transcription PCR (RT-PCR) results against makes it challenging to assess the true diagnostic accuracy of these tests as method of diagnosing COVID-19. False negative results can be attributed to a range of causes, including laboratory error, sampling error, or lack of/negligible presence of virus in the tissue sampled at the time of sampling. False positive results are less likely but also possible, due to, for example detection of viral genome in cases that do not result in infection. Tentatively, evidence identified so far indicates that the type of sample obtained, the part of the body sampled, and the timing of test relative to symptom onset could be influential on test results and accuracy, and this will be explored in more detail in future versions of this evidence review.

Of the nine studies assessing antibody tests, six reported a measure of diagnostic accuracy. Where a reference standard was included, this was RT-PCR except for one study that used either RT-PCR or clinical diagnosis to determine final disease status. As noted above, using RT-PCR to diagnose COVID-19 also results in a proportion of tests that are falsely negative or positive, and this should be taken into account when interpreting the diagnostic accuracy figures reported for antibody tests. With this caveat, sensitivity reported in the studies ranged from 18.4% to 96.1%. Specificity was more consistent across studies and ranged from 90.9% to 100%.

To conclude, more data is required on the effectiveness of tests to detect the presence of SARS-CoV-2 virus, and antibodies to SARS-CoV-2, to inform their use in COVID-19 diagnosis and management. For both types of tests, there is a particular lack of evidence on point-of-care tests (and how these compare to laboratory tests), and the use of tests outside of hospital settings and/or in mild/asymptomatic cases. Some of the evidence identified suggests that for virus tests, the type of sample obtained, and the part of the body sampled could influence test accuracy, whilst for both virus and antibody tests, the timing of test relative to symptom onset is likely to be influential. These factors will be explored in more depth in a future version of this evidence review.

6. Evidence search methods

We searched for evidence that could be used to answer the following review questions:

1. What is the clinical effectiveness and/or economic impact of tests that detect the presence of the SARS-CoV-2 virus to inform COVID-19 diagnosis?
2. What is the clinical effectiveness and/or economic impact 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, but the results for each question were reported separately. Initial scoping-level evidence searches were conducted using the following databases, set up to aggregate COVID-19-specific evidence:

- WHO Global research on coronavirus disease (COVID-19) database
- COVID-19: a living systematic map of the evidence, produced by The NIHR Policy Research Programme Reviews Facility
- LitCovid, Diagnostic set

Based on the results of these, we developed a specific search strategy to capture published evidence on SARS-CoV-2 diagnostics. A copy of this search strategy is available on request. We also hand-searched the sources included in the HTW COVID-19 Evidence Digest for relevant evidence, and contacted key stakeholders in Wales for any published or unpublished data of relevance to this review.

The criteria used to select evidence for the appraisal are outlined in Appendix 3. We followed the recommendations made in the Interim Guidance from the Cochrane Rapid Reviews Methods Group with regards to study selection, data extraction and evidence synthesis. Because of the timescales involved in producing this report we did not conduct formal risk of bias assessments. Appendix 4 summarises the selection of articles for inclusion in the review.

7. Contributors

This topic was proposed by Welsh Government to assist with their response to the COVID-19 outbreak.

The HTW staff involved in writing this report were:

- D Jarrom: preparation of scope, screening of evidence, author of clinical effectiveness summary and conclusions, data verification
- L Elston: screening of evidence, extraction of data from relevant studies
- J Washington: preparation and running of search strategies
- K Cann: internal quality assurance
- M Prettyjohns: preparation of scope, identification of external reviewers, health economics oversight
- P Groves: review of draft report, identification of external reviewers
- S McAllister: project management of report production, coordination of external review
- S Myles: project oversight, review of draft report, identification of external reviewers

Experts who contributed to this appraisal:

- A Freedman, Reader and Honorary Consultant in Infectious Diseases, Cardiff University School of Medicine/Cardiff and Vale University Health Board
- C Fegan, R&D Director, Cardiff and Vale University Health Board
- E Campbell, Scottish Health Technologies Group
- I Weeks, Dean of Clinical Innovation, Cardiff and Vale University Health Board
- S Jolles, Consultant Clinical Immunologist, Cardiff and Vale University Health Board

8. References

Ai T, Yang Z, Hou H, et al. (2020). Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology. 0(0): 200642. doi: https://doi.org/10.1148/radiol.2020200642

Amrane S, Tissot-Dupont H, Doudier B, et al. (2020). Rapid viral diagnosis and ambulatory management of suspected COVID-19 cases presenting at the infectious diseases referral hospital in Marseille, France, - January 31st to March 1st, 2020: A respiratory virus snapshot. Travel Med Infect Dis. doi: https://doi.org/10.1016/j.tmaid.2020.101632

Cassaniti I, Novazzi F, Giardina F, et al. (2020). Performance of VivaDiagTM COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department. Journal of Medical Virology. n/a(n/a). doi: https://doi.org/10.1002/jmv.25800

Chan JF-W, Yip CC-Y, To KK-W, et al. (2020). Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-polymerase chain reaction assay validated in vitro and with clinical specimens. Journal of Clinical Microbiology. JCM.00310-20. doi: https://doi.org/10.1128/jcm.00310-20

Chen C, Gao G, Xu Y, et al. (2020a). SARS-CoV-2-Positive Sputum and Feces After Conversion of Pharyngeal Samples in Patients With COVID-19. Ann Intern Med. doi: https://doi.org/10.7326/M20-0991

Chen Y, Chen L, Deng Q, et al. (2020b). The Presence of SARS-CoV-2 RNA in Feces of COVID-19 Patients. Journal of Medical Virology. 03: 03. doi: https://dx.doi.org/10.1002/jmv.25825

Corman VM, Landt O, Kaiser M, et al. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. doi: 10.2807/1560-7917.ES.2020.25.3.2000045

Fang Y, Zhang H, Xie J, et al. (2020a). Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology. 200432-. doi: https://doi.org/10.1148/radiol.2020200432

Fang Z, Zhang Y, Hang C, et al. (2020b). Comparisons of nucleic acid conversion time of SARS-CoV-2 of different samples in ICU and non-ICU patients. Journal of Infection. doi: https://doi.org/10.1016/j.jinf.2020.03.013

Gao H-X, Li Y-N, Xu Z-G, et al. (2020). Detection of serum immunoglobulin M and immunoglobulin G antibodies in 2019-novel coronavirus infected cases from different stages. Chinese medical journal. doi: https://doi.org/10.1097/CM9.0000000000000820

Guo L, Ren L, Yang S, et al. (2020a). Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). Clin Infect Dis. doi: https://doi.org/10.1093/cid/ciaa310

Guo W-L, Jiang Q, Ye F, et al. (2020b). Effect of throat washings on detection of 2019 novel coronavirus. Clin Infect Dis. doi: https://doi.org/10.1093/cid/ciaa416

Jin Y, Wang M, Zuo Z, et al. (2020). Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019. Int J Infect Dis. doi: 10.1016/j.ijid.2020.03.065

Kong W-H, Li Y, Peng M-W, et al. (2020). SARS-CoV-2 detection in patients with influenza-like illness. Nat Microbiol. doi: https://doi.org/10.1038/s41564-020-0713-1

Li H, Li Y, Zhang Z, et al. (2020a). Establishment and clinical performance evaluation of 2019 novel coronavirus antibody colloidal gold detection method. Chinese Journal of Infectious Diseases. 38(0): E017-E.

Li Y, Yao L, Li J, et al. (2020b). Stability Issues of RT-PCR Testing of SARS-CoV-2 for Hospitalized Patients Clinically Diagnosed with COVID-19. J Med Virol. doi: https://doi.org/10.1002/jmv.25786

Li Z, Yi Y, Luo X, et al. (2020c). Development and Clinical Application of A Rapid IgM-IgG Combined Antibody Test for SARS-CoV-2 Infection Diagnosis. Journal of Medical Virology. n/a(n/a). doi: https://doi.org/10.1002/jmv.25727

Liu R, Han H, Liu F, et al. (2020a). 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. Clinica Chimica Acta. 505: 172-5. doi: https://doi.org/10.1016/j.cca.2020.03.009

Liu W, Liu L, Kou G, et al. (2020b). Evaluation of Nucleocapsid and Spike Protein-based ELISAs for detecting antibodies against SARS-CoV-2. Journal of Clinical Microbiology. JCM.00461-20. doi: https://doi.org/10.1128/jcm.00461-20

Long C, Xu H, Shen Q, et al. (2020). Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? European Journal of Radiology. 108961-. doi: https://doi.org/10.1016/j.ejrad.2020.108961

Pang J, Wang MX, Ang IYH, et al. (2020). Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV): A Systematic Review. J Clin Med. doi: 10.3390/jcm9030623

Spellberg B, Haddix M, Lee R, et al. (2020). Community Prevalence of SARS-CoV-2 Among Patients With Influenzalike Illnesses Presenting to a Los Angeles Medical Center in March 2020. JAMA. doi: https://doi.org/10.1001/jama.2020.4958

Wang W, Xu Y, Gao R, et al. (2020). Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. doi: https://doi.org/10.1001/jama.2020.3786

Won J, Lee S, Park M, et al. (2020). Development of a Laboratory-safe and Low-cost Detection Protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19). Experimental Neurobiology. 0(0). doi: https://doi.org/10.5607/en20009

Wu Y, Guo C, Tang L, et al. (2020). Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. The Lancet Gastroenterology & Hepatology. doi: https://doi.org/10.1016/S2468-1253(20)30083-2

Xia J, Tong J, Liu M, et al. (2020). Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection. J Med Virol. doi: https://doi.org/10.1002/jmv.25725

Xie C, Jiang L, Huang G, et al. (2020). Comparison of different samples for 2019 novel coronavirus detection by nucleic acid amplification tests. Int J Infect Dis. doi: https://doi.org/10.1016/j.ijid.2020.02.050

Xu W, Li J, He X, et al. (2020). The diagnostic value of joint detection of serum IgM and IgG antibodies to 2019-nCoV in 2019-nCoV infection. Chinese Journal of Laboratory Medicine. 43(0): E012-E.

Yan C, Cui J, Huang L, et al. (2020a). Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. Clinical Microbiology & Infection. 07: 07. doi: https://dx.doi.org/10.1016/j.cmi.2020.04.001

Yan C, Cui J, Huang L, et al. (2020b). Rapid and visual detection of 2019 novel coronavirus (SARS-CoV-2) by a reverse transcription loop-mediated isothermal amplification assay. Clinical Microbiology and Infection. doi: https://doi.org/10.1016/j.cmi.2020.04.001

Ye G, Li Y, Lu M, et al. (2020). Experience of different upper respiratory tract sampling strategies for detection of COVID-19. Journal of Hospital Infection. doi: https://doi.org/10.1016/j.jhin.2020.03.012

Zhang J, Wang S, Xue Y. (2020a). Fecal specimen diagnosis 2019 novel coronavirus–infected pneumonia. Journal of Medical Virology. n/a(n/a). doi: https://doi.org/10.1002/jmv.25742

Zhang JJ, Cao YY, Dong X, et al. (2020b). Distinct characteristics of COVID-19 patients with initial rRT-PCR positive and negative results for SARS-CoV-2. Allergy. 13: 13. doi: https://dx.doi.org/10.1111/all.14316

Zhao J, Yuan Q, Wang H, et al. (2020). Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin Infect Dis. doi: https://doi.org/10.1093/cid/ciaa344

Appendix 1. Document revision history

Date of revision	*Reasons for changes*
23 April 2020	Original version, incorporating all evidence up to 14 April 2020

Appendix 2. List of abbreviations

Abbreviation	Definition
CLIA	Chemiluminescent immunoassay
COVID-19	Coronavirus disease 2019
CT	Computed tomography
ELISA	Enzyme-linked immunosorbent assay
GICA	Gold immunochromatography assay
HTW	Health Technology Wales
ICU	Intensive care unit
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IQR	Inter-quartile range
LAMP	Loop-mediated isothermal amplification assay
LFIA	Lateral flow immunoassay
MERS	Middle East respiratory syndrome
NPV	Negative predictive value
NR	Not reported
PCR	Polymerase chain reaction
PPV	Positive predictive value
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SARS	Severe acute respiratory syndrome
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2

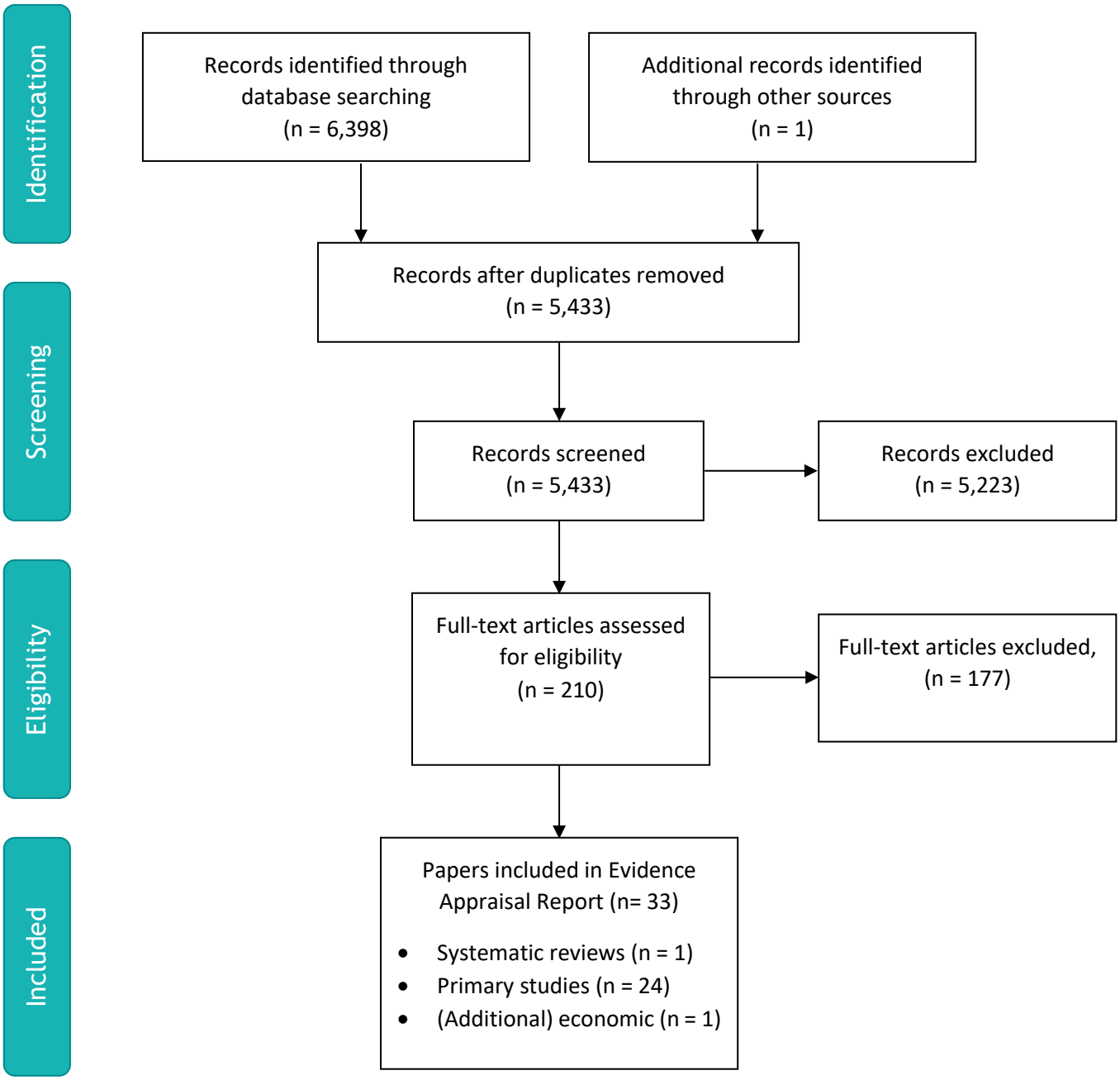
Appendix 3. Study selection criteria

Research Question	What is the clinical effectiveness and/or economic impact of tests that detect the presence of the SARS-CoV-2 virus to inform COVID-19 diagnosis? What is the clinical effectiveness and/or economic impact of tests that detect the presence of antibodies to the SARS-CoV-2 virus to inform COVID-19 diagnosis?

	Inclusion criteria	Exclusion criteria
Population	People with suspected ongoing or recent SARS-CoV-2 infection	
Intervention	Any test that is designed to detect the presence of SARS-CoV-2, or antibodies to SARS-CoV-2, in people suspected of recent or ongoing infection.	We will not include evidence on the accuracy of diagnosing COVID-19 based on clinical information alone, e.g. signs and symptoms, chest imaging. We will however include studies if they compare these methods to virus or antibody detection. We will not include tools used for mass non-contact screening such as fever screening at airports or other transit hubs.
Comparisons	Where available, we will report comparisons of: • different tests or test protocols with each other • virus or antibody tests in comparison to clinical diagnosis	
Outcome measures	• Diagnostic performance (rates of true/false positive/negative results). We will report or calculate measures of diagnostic accuracy (sensitivity, specificity, positive/negative predictive value) where data is available to do so. We will consider any 'gold standard' method used to confirm test results, but will report different methods of calculating these separately. • Virus/antibody detection rates • Time to test result • Influence on/changes in patient management	
Study design	We will prioritise evidence according to its reliability and certainty using established methodology for rapid evidence reviews. We will only include evidence from "lower priority" evidence where outcomes are not reported by a "higher priority" source. We will include data from published sources and also any unpublished data provided by test developers where available, but priority will be given to published, peer-reviewed sources of evidence.	

	We will also search for economic evaluations or original research that can form the basis of an economic assessment. Where possible, we will obtain costs directly from test developers and use this information to carry out assessments of the economic impact of introducing the tests.
Search limits	We will only include evidence published in English or that has an English translation available. We will search for evidence published from December 2019 onwards (the date when the first SARS-CoV-2 infections in humans were identified).
Other factors	We will report evidence on virus and antibody tests separately. Where available, we will also compare or analyse outcomes separately for the factors listed below: • Point-of-care and laboratory testing methods • Quantitative or qualitative reporting of test results • Different sites or methods of tissue sampling • Any variations in test performance in different populations – a range of different genetic, ethnicity and demographic factors will be considered • Tests conducted in different clinical or community settings • Self-administered tests versus those administered and/or interpreted by a healthcare professional

Appendix 4. PRISMA flow diagram outlining selection of evidence (14 April update)



Appendix 5. Study Characteristics

Table 1. Study characteristics: molecular tests

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Secondary evidence					
Pang et al. (2020)	Systematic review	Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and the 2019 novel coronavirus (SARS-CoV-2). For SARS-CoV-2, the authors searched for all in vitro, animal or human studies published in English between 1 December 2019 and 6 February 2020.	Rapid diagnostics, vaccines or therapeutics.	• Sensitivity and/or specificity for rapid diagnostic tests of point-of-care tests. • Impact of drug therapy • Vaccine efficacy	No studies or outcomes relevant to our review were included.
Primary evidence					
Ai et al. (2020)	Retrospective case series Single centre (China) 6 January 2020 to 6 February 2020 RT-PCR results were extracted from the patient's electronic medical record	Patients with suspected novel coronavirus who underwent both chest CT imaging and RT-PCR. n = 1,014 Mean age 51 (±15 years) 46% male	Index test: initial real-time RT-PCR using TaqMan One-Step RT-PCR kits [Shanghai Huirui Biotechnology Co., Ltd or Shanghai BioGerm Medical Biotechnology Co., Ltd] (primer target not specified); throat swab Comparator index test: CT scan	• Detection rate (number of postitive tests) • 'Missed' cases from negative RT-PCR (probable/highly likely) • Test conversion (changes from negative to positive, or positive to negative).	Written informed consent waived. Comparative tests were not necessarily performed at the same time. The CT scan performed closest to the RT-PCR was used (≤7 days; 35 patients excluded due to longer time interval).

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
			Reference test: confirmation of diagnosis with RT-PCR up to and including 3 days after first RT-PCR test		
Amrane et al. (2020)	Prospective case series Single centre (France) 31 January 2020 to 1 March 2020	Patients with suspected COVID-19 n = 280 Mean age 21 years (ranging from 1 to 84 years) Male:Female ratio 1:1.2	Index test: RT-PCR [NR] (E and Spike primers); nasopharyngeal samples A concurrent point-of-care molecular assay was performed to detect other respiratory pathogens.	• Detection rate (number of postitive tests) • Time to result • Differential diagnoses	Definition of 'possible COVID-19' changed throughout the course of the study.
Chan et al. (2020)	Design/Validation study (retrospective) Single centre (China)	Patients with laboratory confirmed COVID-19 (from RT-PCR RdRp-P2 assay). n = 15 (n = 273 specimens) Median age 63 years (range 37 to 75 years) 8 males, 7 females	Index tests: RT-PCR using QuantiNova Probe RT-PCR Kit [Qiagen] (RdRp/Hel, Spike and N primers); respiratory samples (nasopharyngeal aspirates/swabs, throat swabs, saliva, and sputum) and non-respiratory samples (plasma, urine, and feces / rectal swabs) Comparator: RT-PCR (RdRP-P2) [current standard]	• Analytic sensitivity (limit of detection [LOD], copies per reaction) • Detection rate (number of positive tests)	Permission from patients not clear. Collection period of samples not specified.

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Chen et al. (2020a)	Retrospective case series Single centre (China) 20 January 2020 to 27 February 2020	Patients with a diagnosis of COVID-19 and paired RT-qPCR testing of pharyngeal swabs with either sputum or feces samples. A diagnosis of COVID-19 required at least 2 RT-qPCR-positive pharyngeal swabs. n = 22 (545 specimens) 18/22 patients were aged 15 to 65 years old; 4/22 were children. 14/22 (64%) male.	RT-qPCR [NR] (Orf1ab and N primers); pharyngeal, sputum and faecal samples.	• Detection in faecal & sputum samples after conversion of pharyngeal samples.	Letter format, so limited detail.
Chen et al. (2020b)	Retrospective case series Single centre (China) 20 January 2020 to 9 February 2020	Hospital admissions who tested positive for SARS-CoV-2 RNA in pharyngeal swab specimens by RT-PCR 42 patients (multiple specimens from each, total number not reported) Median age 51 years (IQR 42-62 years) 27 (64% female)	RT-PCR [NR] (primer NR) of pharyngeal swab, stool and urine specimens	• Detection rate in pharyngeal swab, stool and urine specimens at multiple time points	Each patient was sampled and tested multiple times, but the total number of samples and sampling interval varied. The minimum time between first and last test was 8 days; the maximum was 24 days.
Fang et al. (2020a)	Retrospective case series Single centre (China)	People with eventual confirmed diagnosis of COVID-19 infection who had an RT-	Index test: Initial RT-PCR [Shanghai ZJ Bio-Tech Co., Ltd] (primer	• Detection rate	Patient consent waived.

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
	19 January 2020 to 4 February 2020	PCR test and CT scan within 3 days or less. Evential confirmed diagnosis is defined as through repeated RT-PCR testing of negative patients, until a positive test is received. n = 51 Median age 45 years (IQR 39 to 55 years) 29 men:22 women	not specified); throat or sputum samples. Comparator: CT scan Reference standard: eventual confirmed diagnosis through RT-PCR		
Fang et al. (2020b)	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 old Sex not reported	RT-PCR [NR] (primer not specified); from nasal swabs, blood, faecal, urine, saliva and tears samples	• Positive rate (Detection rate/Sensitivity) • Conversion time (positive test to negative test)	Letter so limited detail. It is not clear how the population was diagnosed (clinical diagnosis or laboratory confirmed) Reporting/language not always clear. Patient selection not clear (ie. not clear if serial selection, convenience selection)

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Guo et al. (2020b)	Retrospective case series Single centre (China)	People who were hospitalised, and diagnosed with COVID-19 according to the Chinese Management Criteria for COVID-19. 11 patients, 24 samples 9/11 male. Age range 26 to 83 years	RT-PCR [NR] (N, Orflab); pharyngeal swabs and throat washings. Samples were taken simultaneously on 24 occasions.	• Detection rate from different sampling methods	Patient selection not clear (ie. not clear if serial selection, convenience selection)
Kong et al. (2020)	Retrospective case series Two centres, China (Wuhan) 6 October 2019 to 21 January 2020	Hospital outpatients with influenza-like illness (sudden onset of fever and cough or sore throat). Samples were collected as part of routine influenza surveillance. 640 patients (all sampled on a single occasion) Mean/median age: 22.7/8 years (range 9 months to 87 years). 315 males/325 females	Quantitative PCR [Biogerm] (N, Orflab); throat swabs	• Detection rate	Data collection began before the start of the COVID-19 outbreak.
Li et al. (2020b)	Retrospective case series Single centre (China) 2 February 2020 to 17 February 2020	People who were hospitalised, with clinically diagnosed COVID-19. Patients were clinically diagnosed with COVID-19 based on CT scans indicative of viral pneumonia.	RT-PCR [NR] (primer not specified); pharyngeal swabs. Confirmed COVID-19 was defined according to the positive (do not contain suspicious positive) RT PCR test result for	• Detection rate/sensitivity • Conversion	Short communication so limited data Informed consent waived

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
		n = 610 Median age 52.7 years (range 20 to 88) 55.8% male	pharyngeal swab specimens.		
Liu et al. (2020a)	Retrospective case series Single centre (China) 22 January 2020 to 14 February 2020	People tested for SARS-CoV-2 who were suspected or at high risk of infection because of, 1) typical respiratory infection symptoms such as fever, cough and hard breath , or 2) close contact with a SARS-CoV-2 patients. n = 4,880 Median age 50 years (IQR = 27) 2251(46.13%) male	RT-PCR [Shanghai Huirui Biotechnology Co.,Ltd.] (Orf1ab, N primers); respiratory specimens When two targets (ORF1ab, NP) tested positive by specific real-time RT-PCR, the case would be considered to be laboratory-confirmed.	• Detection rate	Informed consent waived
Long et al. (2020)	Retrospective study Single centre (China) 20 January 2020 to 8 February 2020.	Patients with a fever of >38℃ and COVID-19 pneumonia suspicion who underwent both thin-section CT of the chest and RT-PCR examinations. Exclusion criteria: Patients transferred to another hospital or lost to follow-up. n = 87 (n = 36 diagnosed with COVID-19, n = 51 with non-COVID-19 pneumonia [controls])	Index test: initial RT-PCR [NR] (primer not reported); sampling not reported Comparator: CT scan Reference standard: eventual confirmed diagnosis with RT-PCR.	• Detection rate	Consent exempted

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
		The gold standard for a final diagnosis was positivity of first or repeated RT-PCR tests. No significant different between groups for sex and age.			
Spellberg et al. (2020)	Prospective study One centre, United States 12 March 2020 to 16 March 2020	Patients presenting to the emergency department or urgent care with mild influenza-like illness. Patients were excluded if they had specific risk factors for SARS-CoV-2 (eg,travel exposure; known contact with a traveller; severe respiratory tract infection) N = 131 (assumed to be one test per patient, but not clearly reported)	RT-PCR [Quest Diagnostics] (primer NR) using nasopharyngeal swabs	• Detection rate	Letter so limited detail. Convenience sample used (only samples collected during normal laboratory working hourse were tested for SARS-CoV-2
Wang et al. (2020)	Retrospective study 3 centres (China) 1 January 2020 to 17 February 2020	In-patients with coronavirus disease 2019 (COVID-19) diagnosed based on symptoms and radiology and confirmed by SARS-CoV-2 detection. n = 205 (1,070 specimens) Mean age 44 years (range 5 to 67 years) 68% male	RT-PCR [NR] (primer NR); Pharyngeal swabs, faeces, urine, and nasal samples, bronchoalveolar lavage fluid and fibrobronchoscope brush biopsy (severe patients)	• Detection rate	Letter so limited detail. Informed consent waived.

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Won et al. (2020)	Prospective protocol development One centre (South Korea)	Asymptomatic volunteers n = 12	RT-PCR [Applied Biosystems] (primer not clear); self-collected pharyngeal swab.	• Procedure time • Cost	
Wu et al. (2020)	Case series Single centre (China) 16 January and 15 March 2020	Patients with COVID-19. Patients with suspected SARS-CoV-2 were confirmed after two sequential positive respiratory tract sample results. n = 74 Baseline characteristics NR	RT-PCR [NR] (primer NR); sampling NR	• Detection rate (from faecal samples).	Correspondence so limited reporting.
Xia et al. (2020)	Prospective case series Single centre (China) 26 January 2020 to 9 February 2020	People with confirmed COVID-19. The diagnostic criteria were (a) real-time RT-PCR assay of respiratory or blood specimens yielded positive results for the novel coronavirus nucleic acid and (b) CT lung imaging findings were consistent with viral pneumonia. n = 30 Mean age 54.50 ± 14.17 years Male:female ratio of 7:3	RT-PCR [Shanghai Berger Medical Technology Co Ltd] (primer NR); sputum and tear swab.	• Detection rate (in tear samples).	

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Xie et al. (2020)	Case series Two centres (China) Collection dates NR	People with suspected COVID-19. n = 19 Age range 8 to 62 8 male, 11 female	RT-PCR [GeneoDx (GZ-TRM2, China), Maccura (Sichuan, China) and Liferiver (W-RR-0479-02, China) assay kits] (primer not specified); throat, stool, urine and blood samples	• Detection rate	Short communication.
Ye et al. (2020)	Cohort study Two centres (China) Collection dates NR	People with suspected COVID-19. n = 91 Baseline characteristics NR.	RT-PCR [NR] (primer not specified); throat and lingual samples	• Detection rate.	"Practice points" short article
Yan et al. (2020b)	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 Co., Ltd., Tokyo, Japan, used to perform and monitor the RT-LAMP reaction] (Orf1ab and spike) Reference standard: RT-PCR Sampling from swabs (not specified) and bronchoaliveolar lavage fluid.	• Sensitivity and specificity • Procedure time	Potential to be performed as a point of care test but it is not clear whether this is how the test was carried out in the study.

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Zhang et al. (2020a)	Retrospective case series Single centre (China) Collection from 27 January 2020 to 9 February 2020.	People with laboratory confirmed COVID-19 (via RT-PCR) n = 14 Median age 41 years (18–87 years) 7 (50%) female	RT-PCR [NR] (primer not specified); pharyngeal and faecal samples	• Detection rate (of faecal samples)	
(Zhang et al. 2020b)	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 Co Ltd] (Orf1ab, N primers); pharyngeal swab samples	• Detection rate; number of tests before a positive test result	
CT: computed tomography; ICU: intensive care unit; IQR: inter-quartile range; NR: not reported; RT-PCR: reverse transcription polymerase chain reaction;					

Table 2. Study characteristics: immunological tests

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Primary evidence					
Cassaniti et al. (2020)	Cohort study Single centre (Italy) Collection date NR.	3 cohorts: Health volunteers with negative RT-PCR for COVID-19 Hospitalised patients with positive COVID-19 RT-PCR Patients with suspected COVID-19 at their first access at emergency room. n = 110 (30 healthy volunteers; 30 COVID-19 patients; 50 patients with suspected COVID-19) Baseline characteristics reported separately for each cohort.	VivaDiag COVID-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 to 11 days) after positive result for hospitalised patients. Reference/Comparator: RT-PCR (RdRp and E primers); respiratory samples	• Detection rate • Sensitivity and specificity (suspected cohort only) • NPV and PPV (suspected cohort only)	Letter so limited reporting.
Gao et al. (2020)	Case series Single centre (China) 21 January 2020 to 24 February 2020	People with confirmed cases of COVID-19 (confirmed by RT-PCR) n = 22 (37 samples) Median age 40 years (range 4 to 72 years) 8 females, 14 males	IgM and IgG chemiluminescent immunoassays (CLIA), gold immunochromatographic assays (GICA), and enzyme-linked immunosorbent assays (ELISA) [all Beier Bioengineering Company] (targets serum antibodies	• Detection rate • Antibody detection over time	Letter so limited reporting. IgG and IgM tests were separate.

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
			against spike and N); serum samples. Samples were obtained at early (1-7 dpo), middle (8 to 14 dpo) and late (14-24 dpo) stages of infection.		
Guo et al. (2020a)	Prospective study Single centre Dates NR.	Two cohorts People with confirmed COVID-19 (n = 82) People with probable COVID-19 (RT-PCR negative but clinical manifestations) (n = 58) n = 140 (208 specimens)	IgM, IgG, IgA ELISA [in-house protocol] (targets serum antibodies against N gene) Samples were obtained at early (1-7 dpo), middle (8 to 14 dpo) and late (>14 dpo) stages of infection.	• Detection rate • Antibody detection over time.	Consent was waived.
Jin et al. (2020)	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 to 59.0 years) 39.5% male Control group: patient with suspected SARS-CoV-2 infection who were	IgM and IgG chemiluminescence assay (CLIA) [Shenzhen YHLO Biotech Co., Ltd] (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	• Sensitivity and specificity • PPV and NPV • Antibody detection over time. • Detection rate after negative RT-PCR	Characteristics may not be fairly represented across the COVID-19 and control group; in particular the time between symptom onset and first serological test. Therefore validity of the results should be taken with caution. Diagnostic accuracy outcomes were calculated using the control group and a sub

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
		excluded and quarantined at home (n = 33) Median age 31.0 years (IQR 25.5 to 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.	test (CLIA) was 18 days (IQR 11 to 23 days) in the COVID-19 group, 3.0 days (2.0 to 8.0 days).		group of the COVID-19 cohort where patients had received an antibody test before testing negative on RT-PCR (n = 27). Median duration of symptom onset to serological test in this subgroup was 16 days (IQR 9 to 20 days).
Li et al. (2020a)	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/target and sampling methods not known.	• Detection rate • Sensitivity and specificity	Abstract only so limited reporting. Definitions of 'suspected' cases is not clear. Collection time of sample not clear. Type of immunoassay not clear (colloidal gold technique)
Li et al. (2020c)	Prospective development study 8 centres (China) Dates NR	People with suspected COVID-19.	IgM/IgG rapid point-of-care lateral flow immunoassay [Jiangsu Medomics Medical Technologies] (targets antibodies against spike	• Detection rate • Sensitivity and specificity	

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
		n = 525 specimens (397 clinical positive; 128 clinical negative) Characteristics NR	protein); blood (including serum and plasma). Reference standard: RT-PCR; respiratory specimens		
Liu et al. (2020b)	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)	• Detection rates • Antibody detection over time.	Written informed consent waived.
Xu et al. (2020)	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.	• Sensitivity and specificity • NPV and PPV • Coincidence rate	Abstract only so limited reporting. For example time of sampling was not clear.

Reference	Study design	Population/samples	Test [supplier] (target); sample site	Outcomes	Comments
Zhao et al. (2020)	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 Co.,Ltd] (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	• Specificity (based on testing health individuals before SARS-CoV-2 outbreak) • Sensitivity • Median time to seroconversion • Antibody detection over time.	
CLIA: Chemiluminescent immunoassay; dpo: days post onset; ELISA: enzyme-linked immunosorbent assay; GICA: gold immunochromatography assay; IgA: Immunoglobulin A; IgG: Immunoglobulin G; IgM: Immunoglobulin M; LFIA: lateral flow immunoassay: RT-PCR: reverse transcription polymerase chain reaction: NR: details not reported					