Diagnostic Accuracy of FDA Authorized Serology Tests to Detect SARS-CoV-2 Antibodies: A Systematic Review and Meta-analysis

Rajesh Pandey, a M.D., Anand Gourishankar, b M.B.B.S., MRCP, MAS

Affiliations:
a Division of Neonatology, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA.
b Division of Pediatric Hospital Medicine, Children's National Hospital, The George Washington School of Medicine and Health Sciences, Washington DC, USA.

Address correspondence to
Rajesh Pandey, MD
Assistant Professor,
Department of Pediatrics, Division of Neonatology
McGovern Medical School
University of Texas Health Science Center at Houston
6431 Fannin Street, MSB: 3.230
Houston, TX, 77030
Email: Rajesh.Pandey@uth.tmc.edu

Funding Source: No external funding for this manuscript.

Financial Disclosure: All authors have indicated they have no financial relationships relevant to this article to disclose.

Potential conflict of Interest: The authors have no conflicts of interest relevant to this article to disclose

Clinical Trial Registration: Not applicable

Abbreviations: Coronavirus disease (COVID-19); reverse transcription-polymerase chain reaction (RT-PCR), World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), Food and Drug Administration (FDA), Immunoglobulin G (IgG), Immunoglobulin M (IgM), severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), Area under curve (AUC), Summary receiver operating characteristic (SROC)

Words: 2372
Tables: 2
Figures: 3
Supplement material: 4 files

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Key–points:

Questions: What is the pooled diagnostic accuracy of FDA authorized serology tests to detect SARS-CoV-2 antibodies?

Findings: In this systematic review and meta-analysis of seven reports from FDA authorized serology tests to detect antibodies against SARS-CoV2 antibodies (3336 patients/ samples) pooled sensitivity was 87%, and pooled specificity was almost 100%. There was a wide variation in test performance based on the duration between the onset of symptoms and the tests.

Meaning: FDA authorized tests are highly accurate to detect antibodies against SARS-CoV-2 antibodies if tests are performed under a similar condition, as presented in the original report. There is a wide variation in the test performance based on the time interval between the onset of symptoms to the tests.
Abstract:

Importance: Serology tests are diagnostic and complementary to molecular tests during the COVID-19 pandemic.

Objective: To evaluate the diagnostic accuracy of FDA authorized serology tests for the detection of SARS-CoV-2 infection.

Data sources: A search of MEDLINE, SCOPUS, CINAHL Plus, and EMBASE up to April 4, 2020, was performed to identify studies using the "COVID 19 testing" and "meta-analysis." FDA website was accessed for the list of tests for emergency use authorization (EUA).

Study Selection: Manufacturer reported serology tests published in the FDA website were selected. Two reviewers independently assessed the eligibility of the selected reports.

Data extraction and synthesis: The meta-analysis was performed in accordance with the PRISMA guidelines. A bivariate analysis using the "random-effects model" was applied for pooled summary estimates of sensitivity, specificity, and the summary receiver operating characteristic curves.

Main outcomes and measures: The primary outcome was the diagnostic accuracy of the serology test for detecting SARS-CoV-2 infection. Subgroup analysis of the diagnostic accuracy with lag time between symptom onset and testing were studied.

Results: Seven manufacturer listed reports were included. The pooled sensitivity was 87% (95% CI, 78% - 93%), the pooled specificity was 100% (95% CI, 97% - 100%), and the area under the hierarchical summary receiver operating characteristic curve was 0.97. At ≤ 7 days, sensitivity was 44% (95% CI, 21% - 70%), and for 8-14 days, sensitivity was 84% (95% CI, 67% - 94%).
For blood draws ≥ 15 days after the onset of symptoms, sensitivity was 96% (95% CI, 93% - 98%). Heterogeneity was substantial, and the risk of bias was low in this analysis.

Conclusions and relevance: FDA authorized serology tests demonstrate high diagnostic accuracy for SARS-CoV-2 infection (certainty of evidence: moderate). There is a wide variation in the test accuracy based on the duration between the onset of symptoms and the tests (certainty of evidence: low).
Background:

Coronavirus disease (COVID-19) is a current global threat, and diagnostic tests are critical. COVID-19 was discovered in Hubei Province, China, in December 2019, caused by a pathogen of unknown origin,\(^1\,^2\) which was subsequently named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 has currently caused a pandemic.\(^3\,^4\) The genome sequence of SARS-CoV-2 was added to the GenBank sequence repository on January 10, 2020,\(^2,\,^5\) which enabled laboratories across the world to develop and use reverse transcription-polymerase chain reaction (RT-PCR) test for the diagnosis. Both the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) recommend an RT-PCR for the diagnosis of SARS-CoV-2 infection.\(^6,\,^7\)

SARS-CoV-2 is a new pathogen in humans, so there are no antibodies found in a person not exposed to this virus.\(^7\) The rise in antibody titer between acute and convalescent-phase sera tested in parallel is definite evidence of the viral infection.\(^7\) Seroconversion after SARS-CoV-2 infection appears as early as the fourth day after the onset of symptoms,\(^8\) but can be delayed beyond 15 days.\(^9\) Previous publications have supported the use of serological tests for the detection of SARS-CoV-2 infection.\(^9,\,^10\) WHO has affirmed the use of seroconversion or an increase in antibody titer for the diagnosis of SARS-CoV-2.\(^7\) However, the CDC does not recommend serological tests for diagnosis of SARS-CoV-2. Nonetheless, the Food and Drug Administration (FDA) has authorized emergency use authorizations (EUA) for diagnostic serological tests for COVID-19\(^6,\,^11\) as a valuable tool for the management of this pandemic.
Serological tests offer several advantages including (1) contact tracing; (2) serologic surveillance; (3) identification of those who have already had the virus infection and thus may (if there is protective immunity) be immune; \(^\text{12}\) (4) to identify individuals as a potential source for convalescent serum treatments (currently an experimental treatment or prophylaxis); \(^\text{13}, \text{14}\) (5) to do epidemiological researches; (6) theoretically to diagnose a person in a late stage of disease when RT-PCR may be false negative; \(^\text{15}\) and (7) evaluation and development of a vaccine. \(^\text{15,16}\)

With the novel Coronavirus surge, testing brings challenges with variability in performance, including sensitivity and specificity. \(^\text{15}\) Therefore, the purpose of this review was to evaluate the diagnostic accuracy of the FDA authorized diagnostic serological tests for SARS-CoV-2 infection.

Methods

Literature search

We used the FDA website to search for all FDA authorized, under emergency use authorization (EUA), in-vitro serology diagnostic tests for SARS-CoV-2. We performed a search for systematic reviews up to April 4, 2020, using the term "COVID 19 testing" and “meta-analysis” in the following databases: MEDLINE, CINAHL Plus, EMBASE, and SCOPUS. This search was limited to publication in the "English" language.

Inclusion criteria

We included all FDA authorized diagnostic serological tests (before April 26, 2020) available to use in the United States of America. We collected information from the authorization reports, either the manufactures' instructions/package insert. \(^\text{11}\) Inclusions also had the following criteria: index test must include qualitative detection of anti-SARS-CoV-2
antibodies [Immunoglobulin G (IgG) or Immunoglobulin M (IgM), or both], and target condition must be confirmed by RT-PCR (the reference standard).

**Exclusion criteria**

(1) Serology tests not authorized by the FDA; (2) serology tests not authorized for diagnostic use; (3) molecular tests authorized by the FDA.

**Data extraction and quality assessment**

Two authors (RP and AG) reviewed all eligible reports and extracted the data independently. Disagreements between the researchers were resolved by consensus. We extracted the following data: (1) manufacturer of the test; (2) settings including sample size, disease status at the time of specimen collection, other clinical contexts available in the report; (3) information regarding test-assay methods, specimens used; and (4) data on a true positive, false positive, false negative and true negative. There was a complete agreement on the final data used in the analysis.

We assessed the methodological quality of each study by following the guidelines from the Cochrane Screening and Diagnostic Test Methods Group, a tool adapted from the QUADAS-2. The four domains assessed for risk of bias were patient selection, index test, reference standard, and flow and timing. We assessed the applicability concerns in the first three domains. In each domain, we answered the signaling questions with 'Yes,' ‘No,’ or ‘Unclear.’ For each domain, the risk of bias was judged as ‘Low,’ ‘High,’ or ‘Unclear’ risk. One review author assessed study quality, which a second review-author verified. We summarized the overall quality of evidence using the GRADE methodology recommended for diagnostic tests. The assessment of publication bias was not applicable based on our research design.
**Data synthesis and analysis**

We constructed 2x2 tables for all reports based on dichotomous data from the reference standard (positive or negative). We performed data analysis using methods described in the Cochrane Handbook of Diagnostic Test Accuracy (DTA) Reviews.\(^1\) We created forest plots with 95% confidence intervals (CI) for sensitivity and specificity for each study using Review Manager 5.\(^2\) Summary statistics of four sets of primary data, namely true positive (TP), false positive (FP), false negative (FN), and true negative (TN), were analyzed. Seroconversion, which determines the accuracy of tests, after SARS-CoV-2 infection depends on the time interval between the onset of disease and testing.\(^3\)\(^,\)\(^4\)\(^,\)\(^5\) To address the test accuracy and variation in sensitivity of tests with the lag time, we performed subgroup analysis for tests done at (1) at or before 7 days; (2) between 8-14 days and (3) at or after 15 days after the onset of symptoms. A bivariate random-effects model was used for summary effect size, and Area under curve (AUC) using R software version 3.2.\(^6\)\(^,\)\(^7\) This bivariate model takes account into binomial distribution for directly modeling sensitivity and specificity for within-study variations, and assumes bivariate normal distribution for between-study variation.\(^8\) We used inverse variance when calculating the weights of individual studies. The DerSimonian-Laird estimator was used when calculating between-study variations for statistical assumptions and conservative estimates. The confidence interval was estimated by the Clopper-Pearson method. Heterogeneity was categorized as follows: not important (0% - 40%); moderate (30% - 60%), substantial (50% - 90%); and considerable (75% - 100%).\(^9\) In addition to ROC, the Spearman’s correlation coefficient (Rho) between the sensitivity and false-positive rates was calculated for the presence of “cut-off” point (≥ 0.6).\(^10\) GRADEpro software was used to create a summary of findings.
Publication bias deferred in this analysis as we used manufacturer quoted reports published by the FDA.

**Results**

The systematic literature search yielded no meta-analysis on serological tests for SARS-CoV-2 infection as of April 4, 2020. A total of 48 EUA summaries were identified from the FDA website accessed on April 25, 2020, of which seven reports which met eligibility criteria were included (eFigure 1 in the supplement).

**Characteristics of the included reports**

All seven studies are exclusively obtained from the FDA website. Information was collected from the information for use (IFU) documents. The characteristics of the included reports are summarized in Table 1.

The sample sizes in each report ranged from 70 to 1166 (total n = 3336). Five reports used a qualitative detection of IgM, IgG, or both antibodies against the SARS-CoV-2, while the remaining two used qualitative detection for the detection of IgG only. Demographics such as age, ethnicity, gender were not available. Comorbidities of the patients who tested positive for SARS-CoV-2 infection were unknown. As the cases are rising rapidly, there was no information available regarding the prevalence of the disease.

There was an incomplete and variable reporting of the time interval between the onset of symptoms and the test. Autobio® Anti-SARS-CoV-2 Rapid Test had information on the timing of the test relative to the onset of symptoms while Cellex® qSARS-CoV-2 IgG/IgM Rapid Test did not have such information. Chembio DPP® COVID-19 IgM/IgG System test had information on days of blood samples collected from symptom onset in COVID patients. LIAISON® SARS-CoV-2 S1/S2 IgG had information on the lag time between the onset of symptoms and the tests.
They performed serial tests, up to three times, for negative results. We used only the first of the serial measurements in this review. Mount Sinai Laboratory COVID-19 ELISA IgG Antibody Test had information on the timing of sample collection (7 to 14 days after onset of symptoms and retested at 21 days if the initial test was negative). VITROS® Anti-SARS-CoV-2G had information on all patients. But we did not include in our sub-group analysis as their reported time did not fit into our proposed analysis. VITROS® Anti-SARS-CoV2T had information on 15 out of 30 subjects. The characteristics of included reports are summarized in table 1. We believe the seven studies provided overall good methodological quality and low risk of bias (Figure 1). See supplemental material for details of the risk of bias assessment (eFigure 2 in the supplement).

**Diagnostic performance of serology tests**

The sensitivity of the individual tests ranged from 54% to 94% and specificity ranged from 90% to 100%. We created a coupled forest plot with 95% confidence interval for individual reports with sensitivity and specificity (Figure 2). The total effect size sensitivity (pooled sensitivity) of all seven studies by the random-effects model was 87% (95% CI, 78% to 93%) (eFigure 3 in the supplement). The Higgins’ I² was 86%, and the Cochrane Q statistic was 60; p < 0.001, which suggests substantial heterogeneity. The total effect size specificity (pooled specificity) of all seven studies by the random-effects model was 100% (95% CI, 97% to 100%) (eFigure 4 in the supplement). The Higgins’ I² was 93%, and the Cochrane Q statistic was 26; p < 0.002, which suggests considerable heterogeneity. AUC for the SROC was significant (0.97) (Figure 3). The Spearman’s correlation coefficient between the sensitivity and specificity was negative 0.6, also indicating the presence of an implicit cut-off points. Overall, the diagnostic
accuracy of the test was 96% (95% CI, 95% to 97%). The number needed to diagnose was 1.2
(95% CI, 1.1 to 1.2). Around 12 persons need to be tested to return 10 positive tests.

**Subgroup analysis**

Subgroup analysis detected a notable variation on test sensitivity depending on the timing
of blood drawn after the onset of symptoms (lag time). At ≤ 7 days (51 samples, 4 reports),
sensitivity was 44% (95% CI, 21% to 70%), for 8-14 days (152 samples, 5 reports), sensitivity
was 84% (95% CI, 67% to 94%). For blood draws ≥ 15 days after the onset of symptoms (329
samples, 3 reports), sensitivity was 96% (95% CI, 93% to 98%) (Table 2).

**Discussion**

Pooled sensitivity of 87% and specificity of 100% support that the FDA authorized
serology tests are useful in ruling out SARS-CoV-2 infection if the tests are done under similar
conditions as conducted in the reports. We noticed variations in testing time from the onset of
symptoms as well as serial measurements on some reports. Although heterogeneity of specificity
was higher (93%) compared to those of sensitivity (86%), SROC curve with less scattering of
test values and moderate Spearman’s correlation coefficient of negative 0.59 validates these
studies. The diagnostic accuracy and the number needed to diagnose reflect the strength of the
serology tests.

After the onset of infection with SARS-CoV-2, seroconversion occurs variably, ranging
from days to weeks. Zhao J et al. showed empirical evidence for the role of antibody titers
with a lag time similar to our study findings. In that study, among 173 patients, the median
duration for IgM and IgG antibodies seroconversion was 12 days and 14 days, respectively.
Moreover, less than 40% had antibody detected within the first week of illness. In another study
of 285 patients, median seroconversion was 13 days for IgG, with 100% positive within 19 days.
IgM antibodies are detectable earlier than IgG antibody (208 plasma sample: IgM median 5 (IQR, 3-6 days) versus IgG median 14(IQR, 10-18 days). We found an increase in test sensitivity with an increase in time between the illness and the test [sensitivity 96% for tests done ≥ 15 days from symptom onset (Table 2)]. This high sensitivity finding from only three reports warrants validation from more extensive studies. Negative tests, on appropriate clinical settings, should be interpreted with caution, mostly if the tests are done early in the courses of the disease.

Cross-reactivity of SARS-CoV-2 antibodies with other proteins in the blood is under investigation. Reports of false-positive test results are published. Cross-reactivity between SARS-CoV-2 antibodies with other known antibodies were tested in all reports. Chembio DPP® COVID-19 IgM/IgG System reported cross-reactivity against human coronavirus HKU1 and human coronavirus 229E. LIAISON® SARS-CoV-2 S1/S2 IgG reported cross-reactivity against the Hepatitis B Virus, Anti-Influenza A, and rheumatoid factors. These cross-reactivity limitations should be considered while interpreting positive tests.

**Limitations**

Information on the study population is not available to critically appraise individual reports. We could not verify the spectrum bias. Per protocol, we limited our search to FDA authorized serology tests for detecting antibodies against SARS-CoV-2. None of the serology tests have been independently verified for accuracy by the FDA. Five reports used detection of IgG or IgM or both, and the remaining two used detection of IgG only for the diagnosis of SARS-CoV-2. We accepted the threshold detection of antibodies as defined by individual reports.
Heterogeneity could not be explained by threshold as the serology tests were qualitative.

Heterogeneity was substantial. We could not perform meta-regression or sensitivity analysis as limited information was available. Higher heterogeneity in this review can arise from chance, patient selection, and the type of diagnostic test.

We found a significant area under the curve (AUC), which represents the global summary performance of the test. The symmetric shoulder of SROC represents the variability in the studies and the trade-off between sensitivity and specificity.

Conclusions

WHO and CDC recommend RT-PCR for the diagnosis of SARS-CoV-2 infection. Both CDC and WHO agree that serology tests play a pivotal role in the COVID-19 pandemic considering cost, diagnostic accuracy, and relatively quick results. The overall sensitivity and specificity of the FDA authorized serological tests for detecting SARS-CoV-2 infection are high (certainty of evidence: moderate). There is a wide variation in the test accuracy based on the duration between the onset of symptoms and the tests (certainty of evidence: low).

Funding: None

Acknowledgment: Drs. Lisa M. Scheid and Eric W. Reynolds for their review and comments on the draft, and Dr. Audrey Wanger for useful discussions.
References:


26. McMaster University, 2015 (developed by Evidence Prime.


Table 1: Characteristics of the reports included in the systematic review

<table>
<thead>
<tr>
<th>Tests</th>
<th>Sample source</th>
<th>Antibody detected</th>
<th>TYPE</th>
<th>Target protein</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autobio</td>
<td>Serum/plasma</td>
<td>IgM and IgG</td>
<td>RDT</td>
<td>Spike protein</td>
<td>357</td>
<td>48</td>
<td>3</td>
<td>309</td>
<td>717</td>
</tr>
<tr>
<td>Cellex</td>
<td>Plasma/serum/whole blood</td>
<td>IgM &amp; IgG</td>
<td>RDT</td>
<td>Nucleocapsid</td>
<td>120</td>
<td>8</td>
<td>11</td>
<td>239</td>
<td>378</td>
</tr>
<tr>
<td>Chembio</td>
<td>Plasma/serum/whole blood</td>
<td>IgM &amp; IgG</td>
<td>RDT</td>
<td>Nucleocapsid</td>
<td>29</td>
<td>2</td>
<td>4</td>
<td>35</td>
<td>70</td>
</tr>
<tr>
<td>LIAISON</td>
<td>Serum/plasma</td>
<td>IgG</td>
<td>RDT</td>
<td>Spike protein</td>
<td>41</td>
<td>35</td>
<td>8</td>
<td>1082</td>
<td>1166</td>
</tr>
<tr>
<td>MSL</td>
<td>Serum/plasma</td>
<td>IgG</td>
<td>ELISA</td>
<td>Spike protein</td>
<td>37</td>
<td>3</td>
<td>0</td>
<td>74</td>
<td>114</td>
</tr>
<tr>
<td>VITROS – CoV2G</td>
<td>Serum/plasma</td>
<td>IgG</td>
<td>Modified ELISA</td>
<td>Spike protein</td>
<td>42</td>
<td>6</td>
<td>0</td>
<td>407</td>
<td>455</td>
</tr>
<tr>
<td>VITROS–COV2T</td>
<td>Serum/plasma</td>
<td>IgM and IgG</td>
<td>Modified ELISA</td>
<td>Spike protein</td>
<td>30</td>
<td>6</td>
<td>0</td>
<td>400</td>
<td>436</td>
</tr>
</tbody>
</table>

a. See text for description
b. RDT – Rapid Diagnostic test, ELISA – Enzyme Linked Immunosorbent Assay
c. TP – True Positive, FN – False Negative, FP – False Positive, TN – True Negative
d. Autobio® Anti-SARS-CoV-2 Rapid Test: Included 717 patients (405 PCR positive and 312 PCR negative for SARS-CoV-2 infection), done January to Mid-March 2020. They reported information on the time of testing after onset of symptoms.
e. Cellex® qSARS-CoV-2 IgG/IgM Rapid Test: included 378 (98 samples collected from individuals who exhibited mild or no clinical symptoms, 30 samples were collected from hospitalized, with severe symptoms, 250 negative samples collected prior to September 2019. The day of sample collection relative to the onset of illness was unknown.
f. Chembio DPP® COVID-19 IgM/IgG System: They had 31 (RT PCR confirmed positive) symptomatic subjects, in New York or East Asia during COVID-19 pandemic, 41 symptomatic subjects who were confirmed negative by PCR. They reported information on the time of testing after onset of symptoms. Reported cross-reactivity to Human Coronavirus HKU1 and Human Coronavirus 229E.
g. LIAISON® SARS-CoV-2 S1/S2 IgG: 76 European patients (44 patients hospitalized with moderate symptoms and 32 severe symptoms), 1090 presumed SARS-CoV-2 negative samples from a European laboratory routine (n=90) and European blood donors (n=1000) were collected prior to COVID-19. They performed a serial antibody measurement based on testing time after days of diagnosis. Cross reactive to: Hepatitis B Virus, Anti-Influenza A and rheumatoid factors.
h. Mount Sinai Laboratory COVID-19 ELISA IgG Antibody Test: Total 114 samples reported on table (58 samples collected between 7 and 14 days after onset of symptoms. Fifty-five presumed true negative samples collected prior to the onset of the emergence of COVID-19 pandemic (that adds to 113)
i. VITROS® Anti-SARS-CoV-2G: 48 samples collected from patients confirmed to be SARS-CoV-2 positive by PCR were tested. Of the 48 PCR positive samples 42 were reactive in the VITROS Anti-SARS-CoV-2 IgG assay and 6 were non-reactive 407 PCR negative control. They reported information on the time of testing after onset of symptoms in test positive subjects. They were not included in our subgroup analysis because the time frame did not fit in to our proposed protocol.
j. VITROS® Anti-SARS-CoV2T: 36 samples collected from patients confirmed to be SARS-CoV-2 positive by PCR were tested, 400 presumed SARS-CoV-2 antibodies negative healthy blood donor. They reported information on the time of testing after onset of symptoms in 15/30 test positive subjects.
### Table 2: Summary of Findings: Primary outcomes and subgroup analysis

#### Diagnostic Accuracy of FDA Authorized Serology Tests to Detect SARS-CoV-2 Antibodies

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 10% Typically seen in</td>
<td>Prevalence 20% Typically seen in</td>
<td>Prevalence 60% Typically seen in</td>
</tr>
<tr>
<td>True positives</td>
<td>8720 (7750 to 9310)</td>
<td>17440 (15500 to 18620)</td>
<td>52320 (46500 to 55860)</td>
</tr>
<tr>
<td>False negatives</td>
<td>1280 (690 to 2250)</td>
<td>2560 (1380 to 4500)</td>
<td>7680 (4140 to 13500)</td>
</tr>
<tr>
<td>True negatives</td>
<td>89550 (86940 to 89910)</td>
<td>79600 (77280 to 79920)</td>
<td>39800 (38640 to 39960)</td>
</tr>
<tr>
<td>False positives</td>
<td>450 (90 to 3060)</td>
<td>400 (80 to 2720)</td>
<td>200 (40 to 1360)</td>
</tr>
</tbody>
</table>

CI: Confidence interval

#### Diagnostic Accuracy of FDA Authorized Serology Tests to Detect SARS-CoV-2 Antibodies based on the lag period between on onset of symptoms and testing.

**Accuracy of test result (blood draw ≤ 7 days after onset of symptoms)**

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>44000 (21000 to 70000)</td>
<td>51</td>
<td>☀✪✪✪ LOW</td>
</tr>
<tr>
<td>False negatives</td>
<td>56000 (30000 to 79000)</td>
<td>(-4)</td>
<td>✽✪ ✽ ✽ LOW</td>
</tr>
</tbody>
</table>
**Accuracy of test result (blood draw between 8-14 days of onset of symptoms)**

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True positives</strong></td>
<td>84000 (67000 to 94000)</td>
<td>152 (5)</td>
<td>⬤⬤⬤◯◯ LOW</td>
</tr>
<tr>
<td><strong>False negatives</strong></td>
<td>16000 (6000 to 33000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Accuracy of test result (blood draw ≥15 days of onset of symptoms)**

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True positives</strong></td>
<td>96000 (93000 to 98000)</td>
<td>329 (3)</td>
<td>⬤⬤⬤◯◯ LOW</td>
</tr>
<tr>
<td><strong>False negatives</strong></td>
<td>40000 (2000 to 7000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI: Confidence interval
Figure 1. Forest plot: Sensitivity and specificity of serological antibody tests for diagnosing SARS-CoV-2 infection.

Figure 2. Summary receiver operating characteristic (SROC) plot of serological antibody tests for SARS-CoV-2 infection. Study estimates of sensitivity and specificity are shown with the SROC curve.

Figure 3. Risk of bias and applicability concerns summary: as judged by the review authors.
<table>
<thead>
<tr>
<th>Test</th>
<th>Risk of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient Selection</td>
<td>Index Test</td>
</tr>
<tr>
<td>Autobio® Anti-SARS-CoV-2 Rapid Test</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Cellex® qSARS-CoV-2 IgG/IgM Rapid Test</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Chembio DPP® COVID-19 IgM/IgG System</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>LIAISON® SARS-CoV-2 S1/S2 IgG</td>
<td>?</td>
<td>-</td>
</tr>
<tr>
<td>Mount Sinai Laboratory COVID-19 ELISA IgG Antibody</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>VITROS® Anti-SARS-CoV-2G</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>VITROS® Anti-SARS-CoV2T</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

- **High**
- **Unclear**
- **Low**
<table>
<thead>
<tr>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autobio Anti-SARS-CoV-2 Rapid Test</td>
<td>357</td>
<td>3</td>
<td>48</td>
<td>309</td>
<td>0.88 [0.85, 0.91]</td>
<td>0.99 [0.97, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Celllex qSARS-CoV-2 IgG/IgM Rapid Test</td>
<td>120</td>
<td>11</td>
<td>8</td>
<td>239</td>
<td>0.94 [0.88, 0.97]</td>
<td>0.96 [0.92, 0.98]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPP COVID-19 IgM/IgG System</td>
<td>29</td>
<td>4</td>
<td>2</td>
<td>35</td>
<td>0.94 [0.79, 0.99]</td>
<td>0.90 [0.76, 0.97]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIAISON SARS-CoV-2 S1/S2 IgG</td>
<td>41</td>
<td>8</td>
<td>35</td>
<td>1082</td>
<td>0.54 [0.42, 0.65]</td>
<td>0.99 [0.99, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mount Sinai Laboratory COVID-19 ELISA IgG Antibody</td>
<td>37</td>
<td>0</td>
<td>3</td>
<td>74</td>
<td>0.93 [0.80, 0.98]</td>
<td>1.00 [0.95, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VITROS Anti-SARS-CoV2T</td>
<td>30</td>
<td>0</td>
<td>6</td>
<td>400</td>
<td>0.83 [0.67, 0.94]</td>
<td>1.00 [0.99, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VITROS CoV2G</td>
<td>42</td>
<td>0</td>
<td>6</td>
<td>407</td>
<td>0.86 [0.75, 0.95]</td>
<td>1.00 [0.99, 1.00]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SROC (bivariate model) for Diagnostic Test Accuracy
<table>
<thead>
<tr>
<th>Tests</th>
<th>COD</th>
<th>Sample source</th>
<th>Antibody detected</th>
<th>TYPE</th>
<th>Target protein</th>
<th>Time to result</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Total</th>
<th>Cross reactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autobio®</td>
<td>China</td>
<td>Serum/plasma</td>
<td>IgM and IgG</td>
<td>RDT</td>
<td>Spike protein</td>
<td>10-30 minutes</td>
<td>357</td>
<td>48</td>
<td>3</td>
<td>309</td>
<td>717</td>
<td>None</td>
</tr>
<tr>
<td>Cellex®</td>
<td>USA/China</td>
<td>Plasma/serum/whole blood</td>
<td>IgM &amp; IgG</td>
<td>RDT</td>
<td>Nucleocapsid</td>
<td>10-30 minutes</td>
<td>120</td>
<td>8</td>
<td>11</td>
<td>239</td>
<td>378</td>
<td>None</td>
</tr>
<tr>
<td>Chembio®</td>
<td>USA</td>
<td>Plasma/serum/whole blood</td>
<td>IgM &amp; IgG</td>
<td>RDT</td>
<td>Nucleocapsid</td>
<td>10-30 minutes</td>
<td>29</td>
<td>2</td>
<td>4</td>
<td>35</td>
<td>70</td>
<td>2</td>
</tr>
<tr>
<td>LIAISON®</td>
<td>USA</td>
<td>Serum/plasma</td>
<td>IgG</td>
<td>RDT</td>
<td>Spike protein</td>
<td>10-30 minutes</td>
<td>41</td>
<td>35</td>
<td>8</td>
<td>1082</td>
<td>1166</td>
<td>3/168 positive</td>
</tr>
<tr>
<td>MSL®</td>
<td>USA</td>
<td>Serum/plasma</td>
<td>IgG</td>
<td>ELISA</td>
<td>Spike protein</td>
<td>1-5 hours</td>
<td>37</td>
<td>3</td>
<td>0</td>
<td>74</td>
<td>114</td>
<td>None</td>
</tr>
<tr>
<td>VITROS – CoV2G®</td>
<td>USA</td>
<td>Serum/plasma</td>
<td>IgG</td>
<td>Modified ELISA</td>
<td>Spike protein</td>
<td>48 minutes</td>
<td>42</td>
<td>6</td>
<td>0</td>
<td>407</td>
<td>455</td>
<td>None</td>
</tr>
<tr>
<td>VITROS- COV2T®</td>
<td>USA</td>
<td>Serum/plasma</td>
<td>IgM and IgG</td>
<td>Modified ELISA</td>
<td>Spike protein</td>
<td>48 minutes</td>
<td>30</td>
<td>6</td>
<td>0</td>
<td>400</td>
<td>436</td>
<td>None</td>
</tr>
</tbody>
</table>

a. See text for description
b. COD – Country of development
c. RDT – Rapid Diagnostic test, ELISA – Enzyme Linked Immunosorbent Assay
d. TP – True Positive, FN – False Negative, FP - False Positive, TN – True Negative
e. Autobio® Anti-SARS-CoV-2 Rapid Test: Included 717 patients (405 PCR positive and 312 PCR negative for SARS-CoV-2 infection), done January to Mid-March 2020. They reported information on the time of testing after onset of symptoms.
f. Cellex® qSARS-CoV-2 IgG/IgM Rapid Test: included 378 (98 samples collected from individuals, who exhibited mild or no clinical symptoms, 30 samples were collected from hospitalized, with severe symptoms, 250 negative samples collected prior to September 2019. The day of sample collection relative to the onset of illness was unknown.
g. Chembio DPP® COVID-19 IgM/IgG System: They had 31 (RT PCR confirmed positive) symptomatic subjects, in New York or East Asia during COVID-19 pandemic, 41 symptomatic subjects who were confirmed negative by PCR. They reported information on the time of testing after onset of symptoms.
h. LIAISON® SARS-CoV-2 S1/S2 IgG: 76 European patients (44 patients hospitalized with moderate symptoms and 32 severe symptoms), 1090 presumed SARS-CoV-2 negative samples from a European laboratory routine (n=90) and European blood donors (n=1000) were collected prior to COVID-19. They performed a serial antibody measurement based on testing time after days of diagnosis
i. Mount Sinai Laboratory COVID-19 ELISA IgG Antibody Test: Total 114 samples reported on table (58 samples (collected between 7 and 14 days after onset of symptoms. Fifty-five presumed true negative samples collected prior to the onset of the emergence of COVID-19 pandemic (that adds to 113)
j. VITROS® Anti-SARS-CoV-2: 48 samples collected from patients confirmed to be SARS-CoV-2 positive by PCR were tested. Of the 48 PCR positive samples 42 were reactive in the VITROS Anti-SARS-CoV-2 IgG assay and 6 were non-reactive 407 PCR negative control. They reported information on the time of testing after onset of symptoms in test positive subjects. They were not included in our subgroup analysis because the time frame did not fit in to our proposed protocol.
k. VITROS® Anti-SARS-CoV2T: 36 samples collected from patients confirmed to be SARS-CoV-2 positive by PCR were tested, 400 presumed SARS-CoV-2 antibodies negative healthy blood donor. They reported information on the time of testing after onset of symptoms in 15/30 test positive subjects.
l. Cross reactive to: Human coronavirus HKU1 and Human coronavirus 229E
m. Cross reactive to: Hepatitis B Virus, Anti-Influenza A and rheumatoid factors
Table 2: Summary of Findings

Diagnostic Accuracy of FDA Authorized Serology Tests to Detect SARS-CoV-2 Antibodies

Number of studies (subjects): 7 (3336)
True positive: 710
True Negative: 2546
New test: Anti SARS-CoV-2 antibody (IgM and/or IgG)
Reference test: Reverse Transcription Polymerase Chain Reaction
Pooled sensitivity: 0.87 (95% CI: 0.78 to 0.93) | Pooled specificity: 0.99 (95% CI: 0.97 to 1.00)

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 10% Typically seen in</td>
<td>Prevalence 20% Typically seen in</td>
<td>Prevalence 60% Typically seen in</td>
</tr>
<tr>
<td>True positives</td>
<td>8720 (7750 to 9310)</td>
<td>17440 (15500 to 18620)</td>
<td>52320 (46500 to 55860)</td>
</tr>
<tr>
<td>False negatives</td>
<td>1280 (690 to 2250)</td>
<td>2560 (1380 to 4500)</td>
<td>7680 (4140 to 13500)</td>
</tr>
<tr>
<td>True negatives</td>
<td>89550 (86940 to 92910)</td>
<td>79600 (77280 to 79920)</td>
<td>39800 (38640 to 39960)</td>
</tr>
<tr>
<td>False positives</td>
<td>450 (90 to 3060)</td>
<td>400 (80 to 2720)</td>
<td>200 (40 to 1360)</td>
</tr>
</tbody>
</table>

Diagnostic Accuracy of FDA Authorized Serology Tests to Detect SARS-CoV-2 Antibodies based on the lag period between on onset of symptoms and testing.

Accuracy of test result (blood draw ≤ 7 days after onset of symptoms)
Risk of Bias: Serious
Indirectness: Not serious
Inconsistency: Serious
Imprecision: Not serious
Publication bias: Not applicable
Pooled sensitivity: 0.44 (95% CI: 0.21 to 0.70)

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 100% Typically seen in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True positives</td>
<td>44000 (21000 to 70000)</td>
<td>51 (4)</td>
<td>⬤⬤⬤⬤</td>
</tr>
<tr>
<td>False negatives</td>
<td>56000 (30000 to 79000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Accuracy of test result (blood draw between 8-14 days of onset of symptoms)
Risk of Bias: Serious
Indirectness: Not serious
Inconsistency: Serious
Imprecision: Not serious
Publication bias: Not applicable
Pooled sensitivity: 0.84 (95% CI: 0.67 to 0.94)

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>84000 (67000 to 94000)</td>
<td>152 (5)</td>
<td>⨁⨁◯◯ LOW</td>
</tr>
<tr>
<td>False negatives</td>
<td>16000 (6000 to 33000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Accuracy of test result (blood draw ≥15 days of onset of symptoms)
Risk of Bias: Serious
Indirectness: Not serious
Inconsistency: Serious
Imprecision: Not serious
Publication bias: Not applicable
Pooled sensitivity: 0.96 (95% CI: 0.93 to 0.98)

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 100,000 patients tested (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Certainty of the Evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives</td>
<td>96000 (93000 to 98000)</td>
<td>329 (3)</td>
<td>⨁⨁◯◯ LOW</td>
</tr>
<tr>
<td>False negatives</td>
<td>4000 (2000 to 7000)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI: Confidence interval