High-throughput immunoassays for SARS-CoV-2 – considerable differences in performance when comparing three methods

Oskar Ekelund, MD1,2, Kim Ekblom, MD, PhD3,4, Sofia Somajo, MSc, PhD2, Johanna Pattison-Granberg PhD3, Karl Olsson, BSc2, and Annika Petersson, MSc, PhD3

1Department of Clinical Microbiology, Växjö Central Hospital, Växjö, Sweden; 2Department of Clinical Microbiology, Blekinge County Hospital, Karlskrona, Sweden; 3Department of Clinical Chemistry and Transfusion Medicine, Växjö Central Hospital, Växjö, Sweden; 4Department of Medical Biosciences, Clinical Chemistry, Umeå University, Umeå, Sweden

Corresponding author: Annika Petersson
Department of Clinical Chemistry and Transfusion Medicine
Centrallasaretet Växjö
SE-35185 Växjö
Sweden
phone +46 (0)470 58 7978
cell phone +46 (0) 767 259609
e-mail annika.petersson@kronoberg.se

Running head: Serology methods for SARS-CoV-2

Author ORCID ID: Oskar Ekelund 0000-0001-6621-0244
Kim Ekblom 0000-0003-2844-1310
Sofia Somajo 0000-0003-0433-8079
Johanna Pattison-Granberg 0000-0002-3979-996X
Karl Olsson -
Annika Petersson 0000-0002-4582-3567

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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

Background: The recently launched high-throughput assays for the detection of antibodies against SARS-CoV-2 may change the managing strategies for the COVID-19 pandemic. This study aimed at investigating the performance of three high-throughput assays and one rapid lateral flow test relative to the recommended criteria defined by regulatory authorities.

Methods: A total of 133 samples, including 100 pre-pandemic samples, 20 samples from SARS-CoV-2 RT-PCR positive individuals, and 13 potentially cross-reactive samples were analysed with SARS-CoV-2 IgG (Abbott), Elecsys Anti-SARS-CoV-2 (Roche), LIAISON SARS-CoV-2 S1/S2 IgG (DiaSorin) and 2019-nCOV IgG/IgM Rapid Test (Dynamiker Biotechnology Co).

Results: All assays performed with a high level of specificity; however, only Abbott reached 100% (95% CI 96.3-100). The pre-pandemic samples analysed with Roche, DiaSorin and Dynamiker Biotechnology resulted in two to three false-positive results per method (specificity 96.9-98.0%). Sensitivity differed more between the assays, Roche exhibiting the highest sensitivity (100%, CI 83.9-100). The corresponding figures for Abbott, DiaSorin and Dynamiker Biotechnology were 85.0%, 77.8% and 75.0%, respectively.

Conclusions: The results of the evaluated SARS-CoV-2 assays vary considerably as well as their ability to fulfil the performance criteria proposed by regulatory authorities. Introduction into clinical use in low-prevalent settings, should therefore, be made with caution.

Keywords: COVID-19, SARS-CoV-2, serology, immunology, antibodies
Coronavirus disease 2019 (COVID-19) was first reported in Wuhan, Hubei province, China, in December 2019 [1]. The virus causing this disease has since been designated SARS-CoV-2 [2]. The clinical manifestations of COVID-19 range from mild or no respiratory symptoms to severe viral pneumonia with a need of intensive care and ventilator support. Due to the rapid dissemination of the infection, the outbreak was declared a pandemic by the World Health Organization (WHO) on 11th March 2020 [3].

RT-PCR detection of SARS-CoV-2 in samples primarily obtained from nasopharyngeal and/or pharyngeal swabs is considered gold standard for laboratory confirmation of COVID-19 in patients. However, the presence of viral nucleic acids is transient, and results are dependent on the time of sampling.

Detection of antibodies against the SARS-CoV-2 virus could possibly be used to distinguish patients recovered from COVID-19, and several rapid tests based on immunochromatographic techniques have been developed. These point-of-care tests usually deliver results within 15-30 minutes, however the nature of the tests makes large-scale testing inefficient.

To achieve high-throughput detection of anti-SARS-CoV-2 antibodies in plasma or serum, several manufacturers have developed immunoassays based on recognition of recombinant antigens. Some tests claim to detect IgM or IgG, while others also identify IgA or a mixture of different classes of antibodies. Differences in test design, including choice of antigen, are likely to affect the sensitivity and specificity of the tests [4]. In this study, the performance of three commercially available high-throughput automated SARS-CoV-2 antibody assays, and one rapid immunochromatographic test, was investigated and compared to recommended criteria set by regulatory authorities.
Materials and Methods

Assays

Four commercially available CE marked immunoassays, and their corresponding platforms were used: 1) Abbott SARS-CoV-2 IgG on the ARCHITECT i2000 (Abbott, Illinois, USA); 2) Elecsys Anti-SARS-CoV-2 on the Cobas 8000 e801 (Roche Diagnostic Scandinavia AB, Solna, Sweden); 3) LIAISON SARS-CoV-2 S1/S2 IgG on the LIAISON XL (DiaSorin, Saluggia, Italy); and 4) the lateral flow test 2019-nCOV IgG/IgM Rapid Test (Dynamiker Biotechnology Co., Tianjin, China).

The Abbott SARS-CoV-2 IgG reagent is based on the recognition of IgG antibodies binding to recombinant SARS-CoV-2 nucleoprotein in a chemiluminescent microparticle immunoassay (CMIA). Results are obtained after 29 minutes, and are reported as an index with a cut-off to distinguish between negative and positive results.

The Elecsys Anti-SARS-CoV-2 assay is based on recombinantly produced SARS-CoV-2 nucleocapsid (N) protein and detects antibodies (including IgG). The duration of the analysis is approximately 18 minutes, with results presented as a cut-off index as well as reactive or non-reactive.

The LIAISON SARS-CoV-2 S1/S2 detects IgG antibodies recognising the spike glycoprotein of the coronavirus. As for the previous assays, the antigen is composed of recombinant protein expressed in human cell lines. The analysis takes 35 minutes, and the results are expressed as IgG antibody concentrations in arbitrary units (AU/mL) graded negative, equivocal, or positive.

The manually performed 2019-nCOV IgG/IgM Rapid Test recognises and differentiates IgG from IgM antibodies; however, information on antigen source is scarce. The test result could be obtained after approximately 10 minutes as visual bands across the assay paper strip, any positive result, either for IgG, IgM, or both, was considered positive.
Despite the short reaction time, as compared with the automatic tests, high-throughput analysis of a large number of samples using the manual test procedure is implausible. All automated systems are part of the routine operations in our laboratories and were as such subjected to accepted quality assurance procedures. All tests (including calibration and controls) were performed according to manufacturers’ instructions, using serum samples.

Sample collections

All samples in the study originated from an existing sample collection at the Microbiology department obtained after consent to deposit, store, and use for research and development. Samples were fully anonymised prior to inclusion, and results could hence not be linked to individuals. As a consequence, the study did not require approval from an ethics committee, according to the guidelines of the Swedish Ethical Review Agency. All serum samples were stored at -20°C until analysis.

The specificity of each assay was evaluated using 100 pre-pandemic (2018) samples. To challenge the assays, 13 additional serum samples with possible interferences (antinuclear antibodies (n=2); rheumatoid factor (n=2); anti-cytomegalovirus IgM (n=2); anti-Epstein-Barr virus IgM (n=2) and samples from pregnant donors (n=5)) were analysed.

Assay sensitivities were evaluated using 20 outpatient serum samples from 16 individuals that prior to serum sampling had tested RT-PCR positive for SARS-CoV-2 from nasopharyngeal and/or pharyngeal swabs. RT-PCR had been performed using primers and probes targeting either the envelope (E) and the polymerase (RdRP) genes of SARS-CoV-2 as described by Corman [5], or the nucleocapsid gene (N) of the virus (Abbott RealTime SARS-CoV-2 Assay, Abbott Molecular Inc., Illinois, USA). The interval between onset of COVID-19 symptoms and serum sample collection ranged from 18 to 52 days (median 38 days).
Calculations

Overall per cent agreement, sensitivity (per cent positive agreement), and specificity (per cent negative agreement) were calculated based on a contingency table according to EP12-A2 [6], using Microsoft Excel 2019 (Microsoft Corp. Redmond, WA, USA). The between-test agreement was evaluated using Cohen’s kappa calculated with IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA). For each assay, the positive likelihood ratio (LR+) and positive predictive value (PPV) at different prevalence (P) scenarios were calculated according to equation 1 and 2, respectively.

\[
LR^+ = \frac{\text{sensitivity}}{1 - \text{specificity}} \quad \text{(Eq. 1)}
\]

\[
PPV = \frac{(\text{sensitivity} \times P)}{(\text{sensitivity} \times P) + (1 - \text{specificity}) \times (1 - P)} \quad \text{(Eq. 2)}
\]

Further data analysis, including descriptive statistics, was performed using GraphPad Prism, version 7.04 for Windows (GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com)).

Clinical Performance Requirements

The performance of each assay was compared with published guidelines from three regulatory authorities: the Public Health Agency of Sweden (PHAS); Haute Autorité de Santé, France (HAS); and Centres for Disease Control and Prevention, USA (CDC) [7-9]. Their recommended performance criteria are summarized in Table 1.

Results

Of the 100 pre-pandemic samples, nine tested false positive, two with Roche, five with DiaSorin (including two equivocal results) and two with Dynamiker Biotechnology (figure 1). The negative sample collection resulted in median values of 0.09 COI (0.08-3.16), 0.04 S/CO
(0.01-0.6) and 5.7 AU/mL (1.9-32.6) for Roche, Abbott and DiaSorin, respectively. The panel consisting of 13 potentially cross-reactive pre-pandemic samples mainly gave negative results. However, for Dynamiker Biotechnology, one sample with rheumatoid factor IgM was positive for both SARS-CoV-2 IgM and IgG. Also, one sample obtained during pregnancy showed an equivocal result on the DiaSorin assay. In contrast, all these samples were negative on the Abbott and Roche assays.

Three RT-PCR positive samples gave a weak positive result with Roche, close to the cut-off (1.01 to 1.26, cut-off ≥1.0) (Figure 1). These samples were reported as negative by DiaSorin, Dynamiker Biotechnology, and Abbott. However, for these samples, the Abbott values clearly differed from the rest of the Abbott negative samples by being close to the cut-off (0.81-1.06, cut-off ≥1.4).

The overall agreement was 97.5% for Abbott, 98.3% for Roche, 94.0% for DiaSorin, and 94.2% for Dynamiker Biotechnology. The pairwise inter-assay agreement (Cohen’s kappa) was as follows: Roche and Abbott 0.847 (95% CI, 0.718 to 0.976); Roche and Diasorin 0.654 (95% CI, 0.474 to 0.830); Roche and Dynamiker Biotechnology 0.794 (95% CI, 0.647 to 0.941); Abbott and DiaSorin 0.777 (95% CI, 0.618 to 0.936); Abbott and Dynamiker Biotechnology 0.804 (95% CI, 0.653 to 0.817); and DiaSorin and Dynamiker Biotechnology 0.724 (95% CI, 0.533 to 0.895).

The calculated sensitivity and specificity for each assay is presented in Table 2, together with the corresponding data extracted from each manufacturer’s test kit insert. In the present study Abbott exhibited the highest specificity and Roche the highest sensitivity, both being 100%. Two RT-PCR positive samples and two samples from the negative collection were excluded from the DiaSorin sample collection due to equivocal results. The positive
likelihood ratios (LR+) were $\infty$, 50.0, 25.1 and 37.5 for Abbott, Roche, DiaSorin and Dynamiker Biotechnology, respectively.

Benchmarking the sensitivity and specificity data from the current study against published guidelines showed that the Abbott assay met the specificity criteria set by all three regulatory bodies. The specificities of Roche and Dynamiker Biotechnology were both in accordance with the HAS guidelines, while DiaSorin, in contrast, failed to reach any of the specificity criteria. Neither of the assays from Abbott, DiaSorin and Dynamiker Biotechnology exhibited a sufficient sensitivity to meet the criteria set by PHAS and HAS. In contrast, Roche performed excellent in this regard, with a sensitivity of 100%.

Assay performance recommendations from PHAS are based on a target PPV of $> 90\%$. To demonstrate the seroprevalence required to reach a PPV of $> 90\%$, the PPV based on sensitivity and specificity data from the present study as well as from the manufacturers, was calculated (Figure 2).

**Discussion**

As the pandemic progresses, it will likely become increasingly important not only to establish ongoing COVID-19 but also to confirm past infection of SARS-CoV-2. On a community level, this could be used for assessing the progress of the pandemic and for guiding public health and control policies [10]. On an individual level, a reliable method for proving past COVID-19 could potentially be important for allowing employees to remain at work despite mild respiratory symptoms. It could even serve as a basis for a proposed “immunity passport” [11, 12]. Although the degree of immunity to SARS-CoV-2 is disputed, it is plausible that individuals undergoing COVID-19 will gain partial or temporary protection against new episodes [13]. Serology tests might also be used to identify donors of convalescent plasma, which has been proposed for the treatment of seriously ill COVID-19 patients [14].
However, the rapidly emerging pandemic limits the possibility of state-of-the-art validations according to the EP-12 A2 [6], and there is no international consensus on clinical performance requirements for COVID-19 assays. Neither is there an agreed gold standard method for antibody testing [15]. Using the performance criteria proposed by PHAS, HAS and CDC [7-9], only Abbott managed to reach the specificity recommended by all three authorities while Roche alone fulfilled the sensitivity criteria (PHAS, HAS).

The clinical performance of a qualitative diagnostic test is, with few exceptions, heavily dependent on the prevalence. Both PHAS and CDC accordingly link their respective specificity criteria for SARS-CoV-2 antibody testing to seroprevalence levels. PHAS recommends that a PPV > 90% shall be pursued, regardless of seroprevalence. Based on our results prevalences of approximately 15%, 20% and 25% would be necessary to reach this goal for Roche, Dynamiker Biotechnology and DiaSorin, respectively (Figure 2). In contrast, Abbott would in this regard perform well at any given seroprevalence, owing to its specificity of 100%.

Due to the limited sample size of the study, our results need to be interpreted with caution. Nevertheless, a recent study based on 65 RT-PCR positive and 1154 pre-COVID-19 samples [16] reported similar assay performances, except for the sensitivity of Roche being higher in our study. In contrast, evaluating four different assays using the same outpatient sample collection strengthens this study. Moreover, an aspect not investigated in this study is that assay performance could potentially vary with sample type (serum, plasma, whole blood).

Judging by the results from this study, introducing broad antibody testing for COVID-19 will be cumbersome in a low-prevalence setting. In order to reach the proposed performance criteria, two-tier testing will have to be considered. Alternatively, a high degree of specificity (and thus a high positive predictive value) could be prioritised using modified
cut-off values. Our results suggest that this could be conceivable for the Abbott and Roche
assays, while the more narrowly distributed results for DiaSorin would make it difficult to
find such a specific cut-off with a reasonably preserved degree of sensitivity (Figure 1).
Considering the novelty of the disease, and hence the expedited development of
new assays, possible sources of systematic errors (e.g. cross-reacting antibodies against other
Coronaviridae) need to be carefully considered. In this study, all eleven false-positive
samples originated from unique donors, none of the samples being positive in more than one
assay. Further investigation of possible reasons for the false-positive results is, however,
beyond the scope of this evaluation.
We conclude that the SARS-CoV-2 assays only partly fulfil the performance
criteria proposed by regulatory authorities. Introduction into clinical use in low-prevalent
settings, must therefore, be made with careful consideration and well-informed stakeholders.

Acknowledgements
The expert technical assistance of Ola Forsell and Susanna Bergqvist at the Department of
Clinical Chemistry and Transfusion Medicine, Växjö Central Hospital, and of Sanna
Hjalmarsson, Christina Bojesson and Eline Boesen at the Department of Clinical
Microbiology, Region Kronoberg, is greatly appreciated.

Declaration of interest statement: The authors declare that they have no conflicts of interest
relevant to the manuscript submitted to Infectious Diseases. There is no funding to report.
References


Binder, H. Haslacher, Side by side comparison of three fully automated SARS-CoV-2 antibody assays with a focus on specificity, medRxiv (2020) 2020.06.04.20117911.
Table 1. Summary of recommended criteria for SARS-CoV-2 serology assays issued by authorities in three different countries.

<table>
<thead>
<tr>
<th>Regulatory Authority</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Additional comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Health Agency of Sweden</td>
<td>90.0</td>
<td>99.5</td>
<td>Recommendations based on a seroprevalence of 5%, rendering a target PPV of &gt; 90%</td>
</tr>
<tr>
<td>Haute Autorité de Santé, France</td>
<td>95.0</td>
<td>98.0</td>
<td></td>
</tr>
<tr>
<td>Centers for Disease Control and Prevention, USA</td>
<td>-</td>
<td>99.5</td>
<td>For populations with seroprevalence of ≥ 5%</td>
</tr>
</tbody>
</table>
Table 2. Sensitivity and specificity from this study, and according to the manufacturers’ data. The results are compared to manufacturers’ data from samples collected >14 days (Abbott, Roche) and >15 days (DiaSorin) post PCR confirmation. No information about time of sampling was available for Dynamiker Biotechnology.

<table>
<thead>
<tr>
<th>Study Data</th>
<th>Manufacturer’s Data*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>Abbott</td>
<td>20</td>
</tr>
<tr>
<td>Roche</td>
<td>20</td>
</tr>
<tr>
<td>DiaSorin</td>
<td>18</td>
</tr>
<tr>
<td>Dynamiker</td>
<td>20</td>
</tr>
</tbody>
</table>

* Data retrieved from following versions of product kit inserts: Abbott 06R86 G90418R01, April 2020; Roche V3, 2020-06; DiaSorin 200/007-797, 03, 2020-04; Dynamiker Biotechnology DNK-1419-1.
Figure 1.
Figure 2.
Figure captions

Figure 1. Differences in distribution patterns between Abbott, Roche and DiaSorin immunoassays for detection of SARS-CoV-2 antibodies. Measured values from serology testing of 20 positive (SARS-CoV-2 PCR confirmed) and 100 negative (pre-pandemic 2018) samples. Dotted lines represent cut-off values of A) Abbott: positive result index ≥ 1.4 S/CO, B) Roche: positive cut-off index (COI) ≥ 1.0, and C) DiaSorin: positive cut-off ≥ 15 AU/mL and equivocal 12-15 AU/mL. In C negative samples with signals below detection limit (3.8 AU/ml) were plotted as 1.9 AU/mL.

Figure 2. Estimated PPV for the depicted assays, calculated at seroprevalences of 1%, 5%, 10%, 15%, 20%, 25% and 50%. Calculations were based on sensitivity, specificity and their respective 95% CI limits from the current investigation (solid lines = mean values; grey areas = 95% CI). Dashed lines represent mean sensitivity and specificity data from the manufacturers’ kit inserts. Dotted horizontal lines refer to 90% PPV.