Head-to-head comparison of four antigen-based rapid detection tests for the diagnosis of SARS-CoV-2 in respiratory samples

Thomas Weitzel\textsuperscript{a,b,#}, Paulette Legarraga\textsuperscript{a}, Mirentxu Iruretagoyena\textsuperscript{a}, Gabriel Pizarro\textsuperscript{a}, Valeska Vollrath\textsuperscript{a}, Rafael Araos\textsuperscript{b,c,d}, José M. Munita\textsuperscript{b,c,d}, Lorena Porte\textsuperscript{a,#}

\textsuperscript{a}Laboratorio Clínico, Clínica Alemana, Universidad del Desarrollo, Santiago, Chile
\textsuperscript{b}Instituto de Ciencias e Innovación en Medicina (ICIM), Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, Santiago, Chile
\textsuperscript{c}Servicio de Infectología, Clínica Alemana de Santiago, Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, Santiago, Chile
\textsuperscript{d}Millennium Initiative for Collaborative Research On Bacterial Resistance (MICROB-R), Santiago, Chile

Running head: Comparison of antigen tests for SARS-CoV-2

Address correspondence to Lorena Porte, lporte@alemana.cl; or Thomas Weitzel, thomas.weitzel@gmail.com

Paulette Legarraga and Mirentxu Iruretagoyena contributed equally to this work.
Abstract

In the context of the Covid-19 pandemic, the development and validation of rapid and easy-to-perform diagnostic methods are of high priority. We compared the performance of four rapid antigen detection tests for SARS-CoV-2 in respiratory samples. Immunochromatographic SARS-CoV-2 assays from RapiGEN, Liming bio, Savant, and Bioeasy were evaluated using universal transport medium containing naso-oropharyngeal swabs from suspected Covid-19 cases. The diagnostic accuracy was determined in comparison to SARS-CoV-2 RT-PCR. A total of 111 samples were included; 80 were RT-PCR positive. Median patients’ age was 40 years, 55% were female, and 88% presented within the first week after symptom onset. The evaluation of the Liming bio assay was discontinued due to insufficient performance. The overall sensitivity values of RapiGEN, Liming bio, and Bioeasy tests were 62.0% (CI95% 51.0–71.9), 16.7% (CI95% 10.0–26.5), and 85.0% (CI95% 75.6–91.2), respectively, with specificities of 100%. Sensitivity was significantly higher in samples with high viral loads (RapiGEN, 84.9%; Bioeasy, 100%). The study highlighted the significant heterogeneity of test performance among evaluated assays, which might have been influenced by the use of a non-validated sample material. The high sensitivity of some tests demonstrated that rapid antigen detection has the potential to serve as an alternative diagnostic method, especially in patients presenting with high viral loads in early phases of infection. This is particularly important in situations with limited access to RT-PCR or prolonged turnaround time. Further comparative evaluations are necessary to select products with high performance among the growing market of diagnostic tests for SARS-CoV-2.

Key words: Coronavirus; SARS-CoV-2; Covid-19; diagnosis; rapid diagnostic test; antigen detection
Introduction

Since its emergence in 2019, the SARS-CoV-2 pandemic has caused tremendous public health challenges worldwide (1). Early detection of cases is essential to help curtail this unprecedented pandemic; thus, rapid and easy-to-perform diagnostic tools that can be used to test large numbers of samples in a short period of time are crucial (2). To date, the recommended diagnostic method for SARS-CoV-2 infection (known as Covid-19) is real-time reverse-transcription polymerase chain reaction (RT-PCR), which was introduced in January 2020 (3), and is now performed using WHO or CDC protocols (4, 5), as well as various commercial assays (6).

The gap between the number of samples and the capacity to perform RT-PCR in a timely manner is considered a mayor limitation of public health containment strategies (7). Therefore, there is a critical demand for alternative assays, especially rapid diagnostic tests (RDTs), which are timely, easy to perform, and can serve for point-of-care testing (POCT) or community-based testing (8). RDTs for antibody detection have been developed, but due to the delay in humoral immune response, they have limited use for early diagnosis and low sensitivity for community-based screening (9, 10). Antigen detection tests, on the other hand, have the advantage of detecting the presence of the virus itself and might therefore be a better tool for early cases, but require sufficient viral loads and high-quality sampling (11). Although SARS-CoV-2-specific antigen testing has only recently been developed (12); the market pressure generated by this unprecedented pandemic has resulted in many novel assays that are now commercially available (13). Unfortunately, scientific literature supporting their accuracy is scarce and real-world performance of these assays is uncertain; their validation and comparison are therefore of high
priority (12, 14). Here we present a head-to-head comparison and evaluation of four novel antigen-based RDTs for the detection of SARS-CoV-2 in respiratory specimens from suspected Covid-19 cases.

**Material and Methods**

We conducted a head-to-head study of the diagnostic accuracy of four rapid SARS-CoV-2 antigen detection tests compared to RT-PCR. Samples derived from patients with respiratory symptoms and/or fever, attending Clínica Alemana, a private medical center in Santiago, Chile, between March 16 and April 26, 2020. Specimens were obtained by trained personnel in a specially dedicated “Respiratory Emergency Room” and consisted of naso-oropharyngeal (NOP) swabs, which were placed in tubes with 3 mL universal transport medium (UTM-RT® System, Copan Diagnostics, Murrieta, CA, USA). Samples were initially examined for SARS-CoV-2 RNA by COVID-19 Genesig® Real-Time PCR assay (Primerdesign Ltd., Chander’s Ford, UK) after RNA extraction with the Magna Pure Compact system (Roche Molecular Systems Inc., Pleasanton, Ca, USA) or using a manual protocol with the High Pure Viral Nucleic Acid kit (Roche Molecular Systems Inc., Mannheim, Germany). The Primerdesign RT-PCR received FDA Emergency Use Authorization (EUA) and is within the WHO Emergency Use Listing (EUL) tests eligible for procurement. The test kit includes a positive control template of the target gene and a RNA internal extraction control. It targets the RNA-dependent RNA polymerase (RdRp) with a detection limit of 0.58 copies/µL, according to the manufacturer. Samples showing an exponential amplification curve and a cycle threshold (Ct) value ≤40 were considered as positive.
PCR-characterized samples (UTM with swabs) were kept at -80° C and tested on April 28 and 29 by the following lateral flow antigen-detection kits: 1) “Biocredit COVID-19 Ag One Step SARS-CoV-2 Antigen Test” (RapiGEN Inc., Anyang-si, Gyeonggi-do, Republic of Korea), 2) “COVID-19 Antigen Rapid Test Device StrongStep® COVID-19 Antigen Test” (Liming Bio-Products Co., Jiangsu, China; 3) “Huaketai New Coronavirus (SARS-CoV-2) N Protein Detection Kit (Fluorescence immunochromatography)” (Savant Biotechnology Co., Beijing, China), and 4) “Diagnostic Kit for 2019-Novel Coronavirus (2019-nCoV) Ag Test (Fluorescence Immunochromatographic Assay)” (Bioeasy Biotechnology Co., Shenzhen, China). All kits have a cassette format and display test and control lines, permitting a rapid use without positive and negative control specimens (Table 1). Tests must be read after a specific incubation period (5 to 15 minutes). The first two assays use colloid gold conjugated antibodies, resulting in visible colored bands, whereas the latter two kits are based on fluorescein-marked antibodies. For the Savant assay we used a UV flashlight provided by the manufacturer for visual readout, while the Bioeasy kit was automatically read by the immunofluorescence analyzer EASY-11 (Bioeasy Biotechnology Co.). Importantly, our evaluation protocol included a deviation from the manufacturer’s instructions. Instead of using the provided test solution, we used the equivalent volume of UTM, as described in other studies (15-17). This method allowed us to compare all four assays using the same specimens and to rapidly evaluate a large number of samples previously characterized by RT-PCR.

Positive and negative samples were selected by convenience among the 5,276 respiratory specimens processed for SARS-CoV-2 in the clinical laboratory during the study period. Due to the shortage of available test kits, a 2:1 distribution of positive to negative samples was chosen.
Seventeen of the positive specimens had been used in a previous evaluation of the Bioeasy RDT by our group (15). Assays were tested in parallel from the same sample and by the same trained technician, who was blinded to the RT-PCR results. All test procedures, except the reading of the cassette, were performed under a BSL2 cabinet. Assays with visual output were read by two independent observers, who conferred with a third observer in case of disagreement. Results of the RDTs were compared to those of RT-PCR as reference method; for samples with discordant result, tests were repeated. Demographic and clinical data were obtained from the mandatory Covid-19 notification forms and analyzed in an anonymized manner. Samples with high viral loads (defined as Ct value <25) were compared to those with low viral load (Ct values >25) (16). Statistical analyses were performed with OpenEpi (version 3.01) and GraphPad Prism (version 8.4.2).

All materials and personnel for this evaluation except the test from Savant Biotechnology Co. were purchased using funds for routine diagnostics of the Clinical Laboratory of Clínica Alemana; the Savant RDT was provided free-of-charge by the manufacturer. The study was approved by the respective institutional review board (Comité Etico Científico, Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, Santiago, Chile) and need for informed consent was waived.

**Results**

The study included a total of 111 samples from symptomatic patients; 55% were female, with a median age of 40 years, which was similar to the population of all patients tested for Covid-19 in our laboratory during the study period (57% female, median age 38 years). Eighty of the tested specimens were RT-PCR positive for SARS-CoV-2, which represented 22% of all
positives during the study period, while 31 samples were RT-PCR negative. The median duration from symptom onset to sampling was 2 days (IQR 1-5 days); 88% of specimens (96/109; missing data, n=2) were taken during the first week of symptoms. Ct values ranged from 10.7 to 37.7 (mean, 22.5).

While all four tests were user-friendly, test performances showed significant differences (Table 2). The evaluation of the Liming bio kit was stopped after 19 samples, due to its poor results with a sensitivity of 0% (0/9), specificity of 90% (9/10), and a Kappa coefficient of -0.1. Sensitivities of the other three assays ranged from 16.7% for the Savant test to 85% for Bioeasy (Table 2). Similarly, Bioeasy had the highest accuracy (89.2%) and Kappa coefficient (0.8). All three assays had a specificity of 100% and proved robust; invalid results occurred only with RapiGen (n=2) and Savant (n=2) due to insufficient liquid migration.

In addition, we correlated the viral loads of samples with true positive and false negative results. As shown in Figure 1, mean Ct values of false negative samples were significantly higher than in true positives samples in all three tests. RapiGen and Bioeasy mainly missed to detect samples with high viral load (Ct values >25); this threshold was less clear for the Savant assay. Accordingly, sensitivity values of RapiGen and Bioeasy differed significantly among subgroups with high viral loads and low viral loads (Ct >25) (Figure 2). Indeed, the two assays identified 84.9% and 100% of specimens with high viral load, but only 15.4% and 53.8% of those with low viral load, respectively. In case of the Savant assay, although the difference in sensitivity among samples with high and low viral loads remained, it was far less striking (21.2% vs. 7.7%, respectively) (Figure 2). RapiGen and Bioeasy had a concordance of 82%, while agreement between these two tests and Savant was 67% and 50%, respectively. As visualized in Figure 3, all samples identified as positive by RapiGen or Savant were also detected by Bioeasy.
Moreover, the latter platform identified various additional positive samples, especially among those with lower viral loads (Ct 22-30) and it was the only test to detect specimens with very low viral loads (Ct >30).

All four tests were incubated on a flat surface at room temperature and delivered results in short time (5 to 15 minutes). The visual readout of RapiGen was clear, regardless of the intensity of the band, due to its dual colored format in which the control band was red and positive results appeared as black lines. The visual interpretation of the fluorescence Savant assay was difficult under daylight conditions. This test might perform differently with its reader, which was not available in Chile. Bioeasy had a user-friendly readout performed by a desktop instrument that interpreted the cassettes in <1 minute. The instrument included options for saving patient results, QR coding, and printing, and is suitable to be connected to a laboratory information system.

Discussion

The evaluated assays are among the 22 antigen detection tests for SARS-CoV-2 available at the time of writing, of which 14 are commercialized with CE (Conformité Européene) marking (13). However, it is important to note that the CE licencing process is based on self-reporting of manufacturers, does not grant high performance, and can be misused (18, 19). Indeed, the challenge and problems presented by this procedure has recently been addressed by the European Commission in light of the evolving Covid-19 pandemic (12). The emerging marketing and use of novel PCR-independent tests in the absence of robust performance data has been criticized and it has the potential to cause more damage than benefit (20, 21). Independent evaluations of such diagnostic assays are therefore urgently needed, with larger comparative
studies for antibody tests only recently available or under way (22, 23). Antigen-based testing is still in its infancy and evaluations are scarce (12). Up to now, we only found four publications available (three peer-reviewed), evaluating two assays (15-17, 24), while comparative studies are lacking.

Sensitivity is the most important performance parameter of SARS-CoV-2 antigen detection (12). However, it is also considered the main limiting factor due to experiences with influenza and RSV lateral flow detection kits (6, 25). Newer biosensor-based methods to detect SARS-CoV-2 antigen have shown promising results and could offer a highly sensitive alternative for rapid diagnosis in the future (26, 27).

Our comparison detected significant performance heterogeneity among the evaluated kits. The Bioeasy assay had the highest sensitivity, which was in accordance with our previous evaluation (15). The here reported sensitivity (85%) was almost identical to the value reported by the manufacturer (85.5%). A study from China (with participation of the manufacturer) found an overall sensitivity of 68%; however, for samples with Ct values ≤30, the sensitivity increased to 98% (24). This excellent identification rate in specimens with high viral loads (Ct <25) was confirmed in this report and our recently published study (15). The RapiGen test also showed an acceptable sensitivity (84.9%) with high viral load samples, but was much less sensitive (15.4%) when the viral burden was low. This test had a visual readout, which might have contributed to lower sensitivity. Another assay with visual bands (Respi-Strip CORIS) was recently evaluated in two European studies. Overall sensitivity ranged from 50% to 57.6%; however, detection rates improved for samples with high viral loads (Ct <25) reaching sensitivities of 73.9% to 82.2% (16, 17). The other two assays evaluated by us showed an insufficient performance. Savant had an overall sensitivity of 16.7%, which did not increase significantly in high viral load samples
(Figure 2), while the Liming bio assay did not detect any of 10 positives and was not further evaluated.

All four tests had a cassette format; two had visual readout and two were based on immunofluorescence, one of which required an automated reader. In our experience, all systems were easy to use, robust, and gave a qualitative result in short time (10-20 minutes). The fluorescence reader eased interpretation, but it required a higher standard including technical support. The user-friendliness of all tests demonstrated their potential for decentralized screening. However, the application as POCT is hampered by the inherent biological hazard of respiratory specimens, which require processing in a biosafety cabinet (28). This problem could be overcome if extraction buffers or solutions with virus inactivating properties are used.

The enormous performance gaps detected in our study highlight the urgent need of comparative studies of commercialized antigen tests (12). A possible explanation of performance variations might be related to differences in protein targets. However, details on the detection system (target antigen and detecting antibody) are only reported by a minority of manufacturers (29). The methodology of using specimens stored in UTM for evaluation purposes is another critical point with advantages and disadvantages (15), which should be systematically validated, e.g. with spiked samples.

Performance data are critical for local decision making and also for global agencies in the procurement of simpler, scalable diagnostic tests. Although these tests are less sensitive than RT-PCR, they might be useful during outbreak situations, when timely results are important, but access to molecular testing is limited (14). As shown in this and other reports, antigen tests are more reliable in samples with high viral loads, which usually occur during the first days of clinical disease (30-32). In our population, for example, 96% of high viral load specimens
derived from patients in their first week of symptoms. Accordingly, the sensitivities for first week samples reached 91%-95% (Bioeasy) in this and previous studies (15). Interestingly, a similar value (94%) was reported for SARS-CoV antigen detection during the first 5 days of disease in a publication from 2004 (33).

The potential to detect early infections might be crucial for the design of new RDT-based algorithms, which are particularly important in weaker health systems and low resource settings such as Latin America and Africa. Possible frontline applications include community-based testing, e.g. in drive-through test stations, and at-home self-testing (34). However, in light of the imperfection of tests, large scale strategies need to be well designed to avoid negative effects (35). Another application might be as an adjunct to RT-PCR to achieve rapid preliminary results, e.g. for healthcare workers.

A limitation regarding the evaluation of specificity was the low number of negative samples and that the study was performed during a season with reduced circulation of other respiratory viruses.

In conclusion, our comparative study highlighted a significant heterogeneity of test performance. The high sensitivity of some tests demonstrated that antigen detection has the potential to serve as an alternative or adjunct diagnostic method to RT-PCR, especially in patients presenting with high viral loads in early phases of infection. This might be of particular importance in situations with limited access to RT-PCR or prolonged turnaround time. Further comparative evaluations are necessary to select for tests with high performance among the growing market of diagnostics for Covid-19.

Acknowledgements
We thank Savant Biotechnology Co. for providing the test kits free of charge for evaluation.

Contributors

LP and TW conceived the study. LP, PL, MG, and GP curated the data. LP, TW, PL, and MI analysed the data. LP and GP performed the investigation. LP and TW administered the project. LP, PL, and VV supervised the study. LP, TW, PL, MI, JMM, and RA validated the data. TW wrote the first draft. All authors contributed in reviewing and editing later drafts, and approved the final version.

Declaration of interests

All authors declare no competing interests.

Funding

This work did not receive funding.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Test No 1</th>
<th>Test No 2</th>
<th>Test No 3</th>
<th>Test No 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>RapiGen</td>
<td>Liming bio</td>
<td>Savant</td>
<td>Bioeasy</td>
</tr>
<tr>
<td>Commercial name</td>
<td>Biocredit One Step SARS-CoV-2 Antigen Test</td>
<td>StrongStep® COVID-19 Antigen Test</td>
<td>Huaketai New Coronavirus (SARS-CoV-2) N Protein Detection Kit (FIA)</td>
<td>Diagnostic Kit for 2019-Novel Coronavirus (2019-nCoV) Ag Test (FIA)</td>
</tr>
<tr>
<td>Catalogue No</td>
<td>G61RHA20</td>
<td>500200</td>
<td>BCT-HKT-050</td>
<td>YRLF04401025</td>
</tr>
<tr>
<td>Lot No</td>
<td>H073001SD</td>
<td>2003014</td>
<td>20031501</td>
<td>2002N408</td>
</tr>
<tr>
<td>Certification ²</td>
<td>CE-IVD</td>
<td>CE-IVD</td>
<td>CE-IVD</td>
<td>CE-IVD</td>
</tr>
<tr>
<td>Antigen detected</td>
<td>Not specified</td>
<td>Not specified</td>
<td>N protein</td>
<td>Not specified¹</td>
</tr>
<tr>
<td>Specimen</td>
<td>NP/OP swab</td>
<td>NP/OP swab</td>
<td>Throat swab</td>
<td>NP/OP swab, sputum</td>
</tr>
<tr>
<td>Extraction/Dilution</td>
<td>Swab immersed in assay diluent and strongly squeezed</td>
<td>Swab immersed in extraction buffer and strongly squeezed</td>
<td>Swab immersed in conservation solution immediately before testing</td>
<td>Swab immersed in extraction buffer and strongly squeezed</td>
</tr>
<tr>
<td>Volume applied into cassette</td>
<td>3-4 drops (~90-150 µL)</td>
<td>3 drops (~100 µL)</td>
<td>60 µL</td>
<td>2 drops (100 µL)</td>
</tr>
<tr>
<td>Study sample</td>
<td>150 µL of UTM</td>
<td>150 µL of UTM</td>
<td>60 µL of UTM</td>
<td>100 µL of UTM</td>
</tr>
<tr>
<td>Incubation (at room temperature)</td>
<td>5-8 minutes</td>
<td>15-20 minutes</td>
<td>15 minutes ±1 minute</td>
<td>10 minutes ±0 minutes</td>
</tr>
</tbody>
</table>

FIA, fluorescence immune assay; NP, nasopharyngeal; OP, oropharyngeal
¹ N protein (24)
² Information from (13)
Table 2. Performance of four antigen detection tests for SARS-CoV-2 compared to RT-PCR

<table>
<thead>
<tr>
<th>Antigen detection test</th>
<th>RT-PCR</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Positive %</td>
<td>CI95%</td>
<td>Negative %</td>
<td>CI95%</td>
</tr>
<tr>
<td>RapiGen (n = 109)¹</td>
<td>Positive</td>
<td>49</td>
<td>62.0</td>
<td>100</td>
<td>88.7-100</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>30</td>
<td>51.0-71.9</td>
<td>88.7-100</td>
<td>100</td>
</tr>
<tr>
<td>Liming bio (n = 19)²</td>
<td>Positive</td>
<td>0</td>
<td>0.0-29.9</td>
<td>90.0</td>
<td>59.6-98.2</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>9</td>
<td>0.0-29.9</td>
<td>59.6-98.2</td>
<td></td>
</tr>
<tr>
<td>Savant (n = 109)¹</td>
<td>Positive</td>
<td>13</td>
<td>16.7</td>
<td>100</td>
<td>89.0-100</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>65</td>
<td>10.0-26.5</td>
<td>89.0-100</td>
<td></td>
</tr>
<tr>
<td>Bioeasy (n = 111)</td>
<td>Positive</td>
<td>68</td>
<td>85.0</td>
<td>100</td>
<td>89.0-100</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>12</td>
<td>75.6-91.2</td>
<td>89.0-100</td>
<td></td>
</tr>
</tbody>
</table>

¹ Two invalid results were excluded
² Testing was suspended after 19 samples due to poor test performance
References

   https://www.who.int/blueprint/priority-diseases/key-action/Coronavirus_Roadmap_V9.pdf?ua=1
   (accessed 19 May 2020)


Figure 1. Comparison of viral loads (Ct values) among true positive and false negative results of three rapid antigen assays. Red lines represent mean values; p values calculated by two-tailed t-Test.
Figure 2. Sensitivity (%) of three rapid antigen tests in subgroups of samples with high viral load (Ct < 25) and low viral load (Ct > 25). Sensitivity values are represented by dots and squares, while bars demonstrate 95% confidence intervals.
Figure 3. Concordance of four rapid antigen tests among 80 RT-PCR positive samples. Each line represents an individual sample. Samples are listed from high Ct values (top) to low Ct values (bottom).