Evaluating the role of a fully automated SARS-CoV-2 antigen ECLIA immunoassay in the management of the SARS COV 2 pandemic on general population

https://doi.org/10.1515/dx-2022-0067
Received June 23, 2022; accepted September 15, 2022; published online October 3, 2022

Abstract

Objectives: Chemiluminescence immunoassay (CLIA) automated assays (fourth-generation antigen test) for SARS-CoV-2 detection are promising because of their analytical productivity, but have lower sensitivity and specificity than rt-PCR assays. The authors of this paper evaluated a recent immunoassay implemented on Siemens Atellica IM, investigating how much this could affect the actual feasibility of this diagnostic during the pandemic.

Methods: From the three-day routine 134 positive and 241 negative swab samples by rt-PCR test were evaluated, selected as 1/3 positive – 2/3 negative.

Results: Using rt-PCR as gold standard, the specificity of immunoassay was 96.7%, while sensitivity was 68.0%. Sensitivity is inversely proportional to the viral load: 100% for cycles threshold (CT) values from 14 to 29, 95% until 30 CT, then 85, 74, 72, 68%, for 31–35 CT respectively.

Conclusions: Our study confirms the reliability of the fourth-generation antigen assay in recognizing negative samples. Conversely, sensitivity appears to be less reliable (68.0%) than reported in the literature. This could be due to a non-randomized study group: many swab samples were taken from patients with expected low viral load (hospitalized for COVID for more than 10–12 days or asymptomatic patients for epidemiological surveillance). The strong correlation of sensitivity and viral load could prove significant to track the infectiousness of infected people, as previous studies reported that a viral load of at least 10E6 copies of RNA/mL, corresponding to 25 CT, is the threshold of transmission of the disease.

Keywords: fourth-generation antigen test; immunoassays; SARS-CoV-2; viral load.

Introduction

Diagnosis of the Covid 19 disease is based on the isolation of the virus genetic material (RNA) in the patient’s biological samples. From the onset of the pandemic, Nucleic Acid Amplification Molecular Testing (NAAT) has been the pillar of this diagnostic. According to the directions of the World Health Organization (WHO), the Center for Disease Control and Prevention in Atlanta and the Superior Institute of Health, the diagnostic gold standard is the detection of viral RNA by reverse transcriptase polymerase chain reaction (rt-PCR) on material collected from the upper (nasal and pharyngeal swabs) or lower respiratory tract (bronchoalveolar lavage) [1].

Unfortunately, the molecular test requires dedicated instruments, and the test execution procedure needs both specialized personnel and several hours of working process. Moreover, bio-containment measures in analytical working areas are mandatory [2].

In peak phases of the pandemic, when the number of test requests increase significantly, an organization based solely on NAAT testing can lead to diagnostic delays, worsening the outcome for the patient and compromising the ability to contain the pandemic [3].

This is why antigenic tests (Ag-RDTs) have become widespread, firstly rapid qualitative (RATs) and, more recently, quantitative (QATs). These tests can identify proteins produced by the replicating viruses and present in the secretions of the respiratory tract.
In particular, antigenic tests identify the viral nucleocapsid proteins (N) involved in the formation of viral RNA; these proteins are produced by the virus during the SARS-CoV-2 infection [4].

As opposed to molecular tests, the fundamental feature of antigenic tests is the faster turnaround, and the greater cost-effectiveness, both in terms of the cost of reagents and instruments, and of specialized personnel.

There are numerous antigen tests assays on the market [5]: 85% are mainly lateral flow immunochromatographic assays (LFIA), which can be used immediately after the collection of swabs (point of care test, POCT). Only 15% are intended for implementation on automated immunochemistry tools, based on immunochemiluminescence techniques (CLIA).

The accuracy of antigen tests may differ, but, generally, antigen tests have lower Sensitivity (Sn) than rt-PCR tests [6].

In certain contexts, however, antigen tests can play a determinant role [7] especially when the use of molecular tests does not guarantee an adequate response time. The CLIA automated assays are particularly promising for their high analytical productivity. In fact, although this type of assay does not allow an immediate response for the single case, the productivity of these analytical instruments allows to conduct tens of tests per hour. Furthermore, their use allow to reduce the workload of nurses and health workers at the swab collection center. In fact, nurses are often responsible not only for collecting the swabs, but also for the manual execution of the point of care tests. Performing this task can be a major problem in congested collection centers, especially during pandemic peaks.

In this study we assessed the diagnostic accuracy of a SARS-CoV2 immunochemiluminescent assay (CLIA), implemented on an automatic immunochemistry analyzer, in routine operation, comparing it with a rt-PCR assay, in order to discuss its reliability and applicability in diagnostic practice.

Subject, materials and methods

The diagnostic accuracy study was conducted on in- and outpatients visited at the St. Bortolo general hospital in Vicenza, Italy, during the third peak of the SARS-CoV-2 pandemic. About 250,000 people refer to this hospital. In the period from January 29 to February 2, 2022, this study enrolled 379 patients (233 females, mean ± SD age 48 ± 18 years, range 10–97 years and 146 males, 50 ± 20 years, range 13–95 years). At the request of their doctor, the hospital wards or the Health Authority, these patients had turned to the swab center located inside the hospital for the detection of SARS RNA of the COV 2 virus with rt-PCR technique (molecular swab) for both screening (health employee surveillance, close contacts) and for follow up after the infection. No clinical selection was made, both symptomatic and asymptomatic subjects were enrolled. Patients hospitalized for COVID for more than 10–12 days or positive asymptomatic patients tested exclusively for epidemiological surveillance were equally enrolled in the same way as highly symptomatic patients, so as to have a wide range of viral loads.

Nasopharyngeal swabs were collected by nurses in Copan Universal Transport Medium (UTM) and examined within 4 h from collection. The rt-PCR analysis was performed using the Cobas® SARS-CoV-2 assay (Roche Diagnostics International AG, Rotkreuz, Switzerland) on the Cobas 6800 Roche platform, following the manufacturer’s instructions.

The Cobas® SARS-CoV-2 test is a dual-target test that includes both the specific detection of the SARS-CoV-2 ORF 1 a/b gene (target 1) and the pan-Sarbecovirus E gene (target 2). When both targets are recognized, the sample is classified as positive, while it is negative if one or both targets are not recognized after 40 amplification cycles (40 Cycle Threshold, CT).

The subsequent investigation to evaluate the antigen test was carried out on the leftover material, after the samples were anonymized. The samples were analyzed for the detection of SARS-CoV-2 antigen by a CLIA method, SARS-CoV-2 Antigen (CoV2Ag) on Atellica IM Analyzer (Siemens Healthineers Tarrytown, NY 10591 USA).

Both tests, rt-PCR and antigen search, were performed within 24 h.

Results

The detection of SARS COV 2 virus by the rt-PCR technique was positive in 134 samples, negative in 232 samples, indeterminate (1 only positive target out of 2) in 11, unsuitable in 2.

Unsuitable samples were excluded from the study, while 9 out of 11 indeterminate samples were considered negative because tested in the follow up of a previous infection and with a new negative test performed within the next day. The other 2 cases were excluded, being “close contacts” and given the impossibility
of performing another test within one day. The study therefore involved a total of 375 patients: 134 positives (35.7%, 76 w., 58 m.) and 241 negatives (64.2%, 151 w., 86 m.) according to the rt-PCR test.

The antigen test was positive in 99 samples (26.4%, 62 female, 37 male) and negative in 276 (73.6%, 169 female, 107 male).

The index value of Ag positive samples inversely correlates with the cycle threshold (CT) needed for the virus detection, both considering all the samples and limiting the analysis only to positive ones by PCR test (semi-logarithmic Pearson’s correlation coefficient, r=-0.94) (Figure 1).

The median CT in the samples identified as positive for the SARS-CoV-2 antigen test was 26, while the negative antigen samples showed a median of 40 CT by RT-PCR, meaning a statistically significant difference (Mann Whitney test p<0.001) (Figure 2).

Taking the rt-PCR as a reference, the overall sensitivity (Sn) of the antigen test is 68% and the Specificity (Sp) 97%. Sn and Sp are higher in women (74 and 97% respectively) than in men (60 and 92%), while age does not seem to affect Sn and Sp (Table 1).

Analysis of the required CT data to detect viral RNA by rt-PCR showed that the ability of the antigen test to recognize positive subjects increases with decreasing CT (table 2). In particular, Sn was 100% for CT until 29, 95% until 30 CT, then 85, 74, 72, 68%, for 31–35 CT respectively. The overall Sn of the test is higher than 80% considering the samples with CT≤31 as positive and reaches 100% for CT≤29.

Discussion

The antigenic tests introduced on the market are numerous and heterogeneous. The so-called first-generation tests are based on the LFIA principle, meaning that they do not require any type of instrument and the detection of the result is based on the visual inspection of a coloured reaction, providing a qualitative result. The second-generation assays are also based on the immunochromatography principle, but the signal generated by the Ag/Ab reaction is not colorimetric as in the previous case, but fluorescent and is not read by the human eye but by dedicated instrumentation. Third-generation tests produce a fluorescent signal too, but use a microfluidic technique; finally, fourth-generation tests are chemiluminescence immunoassays and are implemented on highly automated instruments, therefore intended for use in the analysis laboratory only [8].

The diagnostic accuracy of antigen tests is globally difficult to evaluate. In fact, it depends not only on the analytical characteristics of the test itself, but also on the population studied and on the state of the disease in those subjects, which is reflected in the viral load. The predictive value of the test also

Figure 1: Logarithmic correlation (log(\(y\))=−0.279 + 8.281; r=0.94; p<0.001) between SARS-CoV-2 antigen index and CT in 134 positive nasal swabs by viral SARS-CoV-2 rt-PCR detection.
depends on the prevalence of the disease and can therefore change depending on the different “moments” of the pandemic. Studying more than 500 patients who arrived consecutively in the emergency room both in red code (COVID symptomatology) and in green code (asymptomatic), the study “Increase” [9] found a low global Sn (48%) of the RATs compared to NAATs. The Sn however increases up to 74% considering only red-code patients and drops to 33% considering asymptomatic patients only.

An important meta-analysis performed on more than 130 studies reports very different sensitivities among different tests, depending on the type of test but, above all, on the population studied. The best performances are obtained in the evaluation of symptomatic individuals, with the test performed in the first days from the onset of the disease [10] showing an excellent correlation for third generation tests.

The meta-analysis by Khalid et al. [11] reports even better results but confirms the differences in sample type and disease status. The Sn of the antigen tests considered as a whole is around 70% (Sp 98%) but it is higher if the swabs are collected at the nasopharyngeal level as opposed to those collected from throat or salivary samples. Samples obtained from symptomatic patients show greater Sn (82%) than those obtained from asymptomatic patients (68%).

There is agreement in the literature that the third and fourth generation tests, regardless of how they are compared, show better performance than those of previous generations [4, 12, 13].

Although the tests based on the amplification of viral nucleic acids remain the gold standard for the diagnosis of SARS CoV-2 infection, the strategies for containing the epidemic cannot be based on such tests alone. The diagnostic delay has a negative impact.
on the possibility of isolating the positive subject and therefore stopping the spread of the virus. The number of people to analyze/test at some peaks of the pandemic exceeds the analytical capacity of the system; the overall costs of epidemic screening, diagnosis and monitoring campaigns are not sustainable in the long term.

On the other hand, cheaper tests may fail to recognize the patient’s positivity, especially in particular contexts, such as in asymptomatic subjects. Furthermore, in periods of low COVID prevalence, these tests could show too many false positives, causing an even greater damage overall.

Some recent meta-analyses [14, 15] actually report good Sp for antigen tests, so the problem of false positives appears to be limited. These studies also show that the Sn of antigen tests increases if the samples have a high viral load. In fact, Sn is inversely correlated to the decrease in gene amplification cycles which are necessary to identify the target (CT).

Our study confirms the reliability of the fourth-generation antigen test in recognizing negatives, with a high Sp (96.7%). Conversely, Sn appears less robust than reported in literature data (68%). Sn and SP are higher in women than in men, while age does not seem to affect Sn and Sp, contrary to other reports (Table 1) [16]. However, the low numerosity of our age classes could justify our data.

On the overall lower sensitivity of the test compared with other studies, it can be argued that our series is extremely heterogeneous and not randomized. Many samples were taken from patients hospitalized for COVID for more than 10–12 days. In this type of patient, the viral load is typically low. Other samples were collected by hospital employees for epidemiological surveillance, and most of them were asymptomatic. Basically, we introduced an involuntary selection bias of subjects with presumed low viral load, as was then effectively confirmed by the rt-PCR analysis. This explains the overall low Sn of the test, if evaluated by comparing the results obtained with a PCR test deemed positive for CT≤35. The test would appear not to reach the minimum requirements set by the WHO [17]. However, the selection of the cohorts on which to apply the antigen test is a fundamental element in defining the minimum requirements. In fact, if we considered in our study the rt-PCR positive samples for CT≤31, the Sn of the test would then rise to 84.9% and reach 100% if we considered the samples with CT≤29.

The ability of the antigen test to identify positive samples, according to the molecular test positivity, is strongly correlated with the viral load. If we considered only the positive samples in which 29 CT were sufficient to recognize the targets when performing rt-PCR, the Sn would reach 100%.

Clearly, the overall test Sn will be particularly affected by the percentage of positive patients, but also by the high, medium and low charge in the overall positivity range. For example, two recent studies evaluating rapid chemiluminescence antigenic tests [13–18] reported an overall higher Sn of the test (80 and 90% respectively, vs. 68% in our series), but the percentage of positive samples with CT<30 was also higher than that in our study.

In fact, the latest indications from the European Commission for health and food safety [19] recommend antigen tests having a good performance, i.e. Sn>80% and Sp>97% in the general population or Sn>90% on samples with CT≤25. For evaluation in the general population, the document also indicates the subdivision of the samples: 40% at very high charge (17–25 CT), 40% at high charge (25–30 CT) and 20% at low charge (30–36 CT).

Corman [20] and Kohmer [11] have shown that antigen tests can recognize as positive the samples which contain a viral load of at least 10E6 copies of

<table>
<thead>
<tr>
<th>Cycle threshold value (CT)</th>
<th>14–29</th>
<th>14–31</th>
<th>14–33</th>
<th>14–39</th>
<th>≥40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (85/85)</td>
<td>94.9%</td>
<td>94.0%</td>
<td>93.0%</td>
<td>92.7%</td>
<td></td>
</tr>
<tr>
<td>Specificity (234/241)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97.1%</td>
</tr>
<tr>
<td>VPP 100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPN 100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPP 92.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPP 85.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPP 64.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cycle threshold value (CT)</th>
<th>14–29</th>
<th>14–31</th>
<th>14–33</th>
<th>14–39</th>
<th>≥40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (85/85)</td>
<td>94.9%</td>
<td>94.0%</td>
<td>93.0%</td>
<td>92.7%</td>
<td></td>
</tr>
<tr>
<td>Specificity (234/241)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97.1%</td>
</tr>
<tr>
<td>VPP 100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPN 100.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPP 92.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPP 85.9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VPP 64.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RNA/ml. Infectious disease studies have indicated that this viral load is the threshold of transmission of the disease, which corresponds to 25 CT [21]. This data is particularly relevant, as antigen tests could play a very significant role in containing the spread of the virus, associating the positivity to the transmissibility of the infection.

An analysis of the history of the virus suggests that the viral load increases in the first days after the infection, only to decrease 5–8 days after the onset of symptoms [22]. Thus, in symptomatic patients, the viral load is high in the first 5–7 days. In this setting, the purpose of the diagnostic strategy is to quickly identify the positive patients in order to isolate them. If the molecular test is not available or the organization cannot guarantee a rapid response, then an antigen test that guarantees a Sn of at least 80 and 97% Sp may be acceptable (WHO criteria). In the peak phases of the epidemic, it can be assumed that the prevalence of infection in subjects undergoing swabs may reach 25–50%. Under these conditions, with the Sn and Sp hypothesized by the WHO, the Positive Predictive Value of these tests would be around 90–96% [23].

For this reason, WHO [13], CDC [24] and the European Center for Disease Prevention and Control (ECDPVC) [25], support and recommend antigen testing in these epidemic settings.

When comparing Sn and Sp with the pre-test probability, proportional post-test probabilities are obtained. In the case of the antigen test for COVID, the analysis is especially complicated because the viral loads of the test samples are likely to have different concentrations at different peak or down phases of the outbreak.

In case the prevalence of the disease exceeds 36% (peak of the pandemic), and hypothetically using an antigen test with the Sn and Sp characteristics suggested by the WHO, Peeling et al. [23] show that the NPV would drop to less than 90%, generating an unacceptable number of FNs. In this scenario, it would be necessary to repeat the negative results samples with an rt-PCR method.

To evaluate the impact of fourth-generation antigen test performances in the real world of our health district, we retrospectively retrieved data of COVID positivity in the period January/April 2022. During the last wave of the epidemic, the positivity to SARS-CoV-2 exceeded 35% of the samples measured, while in the downstream phases of the epidemic, the positivity among the samples measured usually remains between 5 and 10% (selections for symptomatic, case contact checks, rechecks of positives).

Since Se and Sp depend on the viral load of the sample, we calculated the accuracy for the different rt-PCR identification CTs, and estimated the PPV and NPV for pre-test probability of 5, 15, 25, and 35% (Table 2).

In situations with a positive pre-test probability of 35%, an assay such as the one we are using, with Se=67.9% and Sp=97.1% is associated with a PPV equal to 92.7%, and an NPV=84.9%. Therefore, even in our setting, during peak phases of the pandemic it is necessary to repeat a negative test, using an rt-PCR test.

Calculating the PPV and NPV at lower pre-test probabilities of disease, 25, 15 and 5%, we obtain respectively: 88.6%–90.1%; 80.5%–94.5%; 55.2%–98.3% and the framework thus changes completely.

With a low pre-test probability, the negative predictive value (NPV) looks rather good. Conversely, with high viral loads, it is necessary to repeat positive tests, given that nearly 1 in 2 could be false positives, even if symptomatic or higher-risk subjects are selected. However, selecting patients for higher viral load probability significantly increases the pre-test probability and contributes to shifting the terms of evaluation. For example, if a population has a general prevalence of 5% of virus spread, in the subgroup of symptomatic subjects (fever, respiratory symptoms) there will be a much higher prevalence of infections, meaning the predictive values must be estimated against this pre-test probability.

These data indicate that antigen testing can be a valuable tool in epidemic management, in peak phases of infections, even in the general screening phase.

In the same way, in low prevalence surveillance conditions, negative tests have a low risk of being false negatives, but only negative tests definitely exclude contagion in symptomatic or high-risk subjects. On the other hand, positive tests must always be reconfirmed.

If we consider patients beyond 8–9 days from the onset of symptoms, the Sn of the RDTs compared with the NAAT decreases significantly. This can be explained by the presence of a low viral load which is not recognized by antigen tests but also by the presence of sub genomic fractions of viruses that can persist in patients for a long time after infection (up to 22 days). However, the NAAT positivity does not indicate the ability of the virus to actively replicate [26]. The antigen test in this context can recognize a “subject no longer infectious” even if still “infected”.

If, from a strategic point of view, the goal is to recognize patients who could spread the virus, i.e.
those with a high viral load, antigen tests, at least those of the third and fourth generation, show a high Sn (>96%).

Moreover, despite their lower Sn and Sp, compared to the reference molecular tests, even simulated analyses acknowledge a role for antigen tests in symptomatic patients where the results of the molecular are not readily available in at least 85% of patients [7].

As for the screening of asymptomatic patients, WHO [17], CDC [24] and CDPVC [25] recommend the use of antigen tests only among asymptomatic high-risk patients (contacts of confirmed cases or healthcare professionals) where the probability of disease is greater. In fact, if the prevalence of the disease is too low, the number of false positives, despite the high Sp of the antigenic tests, can become too high. The fourth-generation antigen test that we evaluated has shown high Sp, exceeding the WHO requirements. It should thus partially reduce this risk, at least in theory.

From an organizational standpoint, the use of fourth generation antigen tests has the downside of not providing an immediate response after the swabs have been collected, unlike the use of previous generation tests that can be placed in the clinic. In fact, the fourth-generation tests are immunoassays based on the chemiluminescence principle and must be implemented on instruments present in the laboratories. However, these are highly automated tools capable of coping with large workloads with a very favourable personal/number of tests ratio, not only compared to POCT antigen tests but also to 3rd generation antigen tests. The turnaround times are also very short compared to molecular diagnostics, especially compared to large volumes of analysis. Furthermore, this approach allows to reduce staff presence at swab centers, who only have to execute the swab without doing the actual analysis.

Fourth generation antigenic tests are also more accurate than LFIA antigenic tests. This is due to three reasons: the analytical phase is more standardized than in POCT antigenic ones; the detection Sn is higher; the interpretation of the test is not subjective.

Furthermore, the fourth-generation tests are significantly cheaper than the third-generation ones.

In summary, the fourth-generation antigen test tested in our laboratory has a low global Sn but in line with the performances of other similar methods, given the heterogeneous population on which it was tested.

As suggested by the main Guidelines [27], the key indications for using this analytical method should be the screening of symptomatic patients in the first days from the onset of symptoms, when most antigenic tests are able to detect the high viral load. The screening should continue for the follow-up of positive patients, where, rather than detecting the presence of viral RNA, it is more useful to detect which patients are still able to spread the virus.

**Research funding:** None declared.

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

**Informed consent:** Not applicable.

**Ethical approval:** Not applicable.

**References**


