



Current performance of COVID-19 test methods and devices and proposed performance criteria

16 April 2020

Table of contents

Executive summary

1. Introduction

2. Scope

3. Methodology used

4. Overview of information on performance & proposed performance criteria

4.1. Detection of viral status

4.1.1. RT-PCR

A) Evaluation of evidence

B) Proposed performance criteria

4.1.2. Antigen tests

A) Evaluation of evidence

B) Proposed performance criteria

4.2. Detection of immunological status

4.2.1. Antibody tests

A) Evaluation of evidence

B) Proposed performance criteria

5. Conclusions and recommendations

Scientific terminology used

Annexes

Annex 1: Commercial devices

Annex 2: Scientific literature

Annex 3: Search on validation studies

DISCLAIMER

This working document has been produced as output of a dedicated project group consisting of representatives of the Commission services (DG SANTE, DG JRC, DG RTD), the European Centre for Disease Prevention and Control (ECDC) and several experts from *in vitro* diagnostics competent authorities and health technology assessment. DG JRC conducted the literature review and developed the proposed performance criteria. All other parties contributed input to the project and reviewed the document. DG SANTE coordinated the project group.

The information included in this working document is limited to what could be retrieved from online resources, following the strategy indicated in section 3 of the report, including information provided by members of the project group, up to 06 April 2020.

The correctness of information, such as listed performance data of the test methods and devices, has not been directly confirmed by checking raw experimental data or full technical documentation of the manufacturer (not accessible) or by own laboratory verification or any clinical validation studies. Therefore, the authors should not be deemed responsible for the validity of such data.

The evaluation has been conducted based on information available to the authors on 06 April 2020. Considering the rapidly evolving situation in relation to the development and commercialisation of test methods and diagnostic devices for COVID-19, the completeness of this report is limited to this date.

Because of the urgency of the request to provide this working document, it was not possible to consult with external experts and to perform an independent review of its content.

Executive summary

The coronavirus disease 2019, abbreviated as COVID-19, is a global pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Timely and accurate COVID-19 testing is an essential part of the management of the COVID-19 crisis.

In the EU, commercial *in vitro* diagnostic tests for COVID-19 are currently regulated by Directive 98/79/EC on *in vitro* diagnostic devices (the IVDD). As of 26 May 2022, the Directive will be replaced by Regulation (EU) 2017/746 on *in vitro* diagnostic devices (IVDR)¹. When assessing conformity with the legislation and prior to affixing the CE-mark, the manufacturer must evaluate the performance of the device and report the performance information in the instructions for use and technical documentation of the device. This is usually achieved by conducting performance studies. In addition to this, after being placed on the market the performance of devices may be validated, i.e. confirmed by additional testing that the manufacturer's specifications are indeed satisfied, e.g. in reference laboratories, academic institutions or national regulatory agencies. Such validation is not legally obligatory but highly recommended for public health decision making, especially in the context of the current COVID-19 crisis. Validation can be done not only for CE-marked devices but should also be performed for in-house laboratory protocols.

The aim of this working document was to collect and review publicly available information from manufacturers on commercially available devices for COVID-19² and to review performance assessment studies of test methods and devices for COVID-19 that have been performed by academic institutions, national regulatory agencies, international organisations, health technology assessment (HTA) bodies, reference laboratories, and similar organisations. Moreover, this report proposes performance criteria for different types of COVID-19 test methods and devices.

The tests used for COVID-19 can be classified into two groups. The first group contains tests that can **detect the presence of the virus itself (RNA and antigen tests)**. The main purposes of these tests are to support the diagnosis of patients with COVID-19-like symptoms, to screen for infections in crucial target groups like healthcare workers, and to test whether an individual recovered from COVID-19 is still infectious. The second group of tests **detect the immune response of the body against the SARS-CoV-2 virus, i.e. they report on past or ongoing infection with the virus (antibody tests)**. The immunity conferred by the antibodies is still under investigation. Once this is clarified, such antibody tests would be, together with the direct virus detection, an essential tool in the development of de-escalation strategies in which mobility and contact restrictions could be removed for people with proven immunity.

Literature review

RNA tests, based on a reverse transcriptase polymerase chain reaction (RT-PCR), are usually laboratory-based and need special equipment. Published in-house protocols (i.e. not based on commercial devices) generally perform well, in particular with low limits of detection and high analytical specificity. From the review of the literature, although several new methods are being developed, it is still recommended to use tests that explicitly declare the implementation of a WHO protocol, of which also validated versions are available. As for CE-marked devices, 78

¹ In the currently ongoing transition period, devices may also be placed on the market if they comply with the Regulation, according to its Article 110.

² It should be noted that full information on the manufacturer's performance evaluation of the device is contained in the technical documentation required by Annex III of the IVDD. Manufacturer technical documentation is not publicly available and was outside the scope of this study.

CE-marked RNA tests were identified. The performance parameters reported by the manufacturers were also overall good, in line with those in the published literature. Nevertheless it is difficult to link scientific publications to specific CE-marked devices as the latter do not disclose the RNA sequences detected by the test.

Antigen tests are available generally in a rapid test form that could be used at the point of care. These can potentially offer practical advantages compared to RNA tests for the purpose of reporting on the infectious status. However, the field of antigen tests for COVID-19 appears to be still relatively immature and information on their performance in the scientific literature is scarce. Only 13 antigen tests are CE-marked to date.

As regards information on **antibody tests** 101 CE-marked antibody devices were identified. Overall good levels of sensitivity and specificity are claimed (however not validated by third parties), with the exception of early infection when antibodies are only beginning to be produced. It is not clear how the performance of protocols reported in the literature translates to CE-marked tests. Tests that detect two antibody types at the same time (IgG and IgM) are superior to the ones testing for only one antibody.

In general, on the basis of the literature review it is difficult to recommend particular tests on the basis of independent studies, also because those usually do not mention specific devices.

Performance criteria

The proposed performance criteria for different types of COVID-19 devices are intended as additional guidance to the legally obligatory requirements defined in the IVDD (or IVDR). The proposals cover both analytical performance (relating to how well the marker of interest is detected) and clinical performance (relating to how well the device actually informs on patient status). They also include additional elements on descriptive information, quality control and safety measures. The **most critical performance parameters** for reliable decision-making are:

- **for identifying if a person is infected with SARS-CoV-2:** the **diagnostic sensitivity** of the RNA or antigen test, as false negative test results have to be avoided;
- **for identifying the persons who have developed an immune response against SARS-CoV-2:** the **diagnostic specificity** of the antibody test, as false positive test results have to be avoided.

The information collected in the working document clearly shows that currently available evidence on the reliability and comparability of most COVID-19 tests is limited and has to be expanded as soon as possible to ensure that these tests demonstrate suitability for their intended use. It should be kept in mind that in particular clinical studies are time and resource consuming and therefore an immense benefit would come from **pooling efforts** for test validation, including both sharing data and organising joint studies. A prerequisite for effective assessment of performance, both for the manufacturer in carrying out performance evaluation and for the laboratories validating that performance, is the **availability of necessary control samples and reference materials**. Some of these may be challenging to develop, such as positive virus samples that would be needed for antigen tests. In addition, it is necessary to discuss approaches to **enable more meaningful comparisons and standardisation**, including both methods to achieve this and availability of comparator data sets. Discussions among regulators and all stakeholders are necessary to tackle these issues in a pragmatic way to ensure that COVID-19 devices of the highest reasonably achievable performance level are available in the EU.

1. Introduction

The coronavirus disease 2019, abbreviated to COVID-19, is a global pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *In vitro* diagnostic tests play an essential role for a rapid and effective response to this crisis as they contribute to patient screening, diagnosis, monitoring/treatment, as well as epidemiologic recovery/surveillance (see Figure 1).

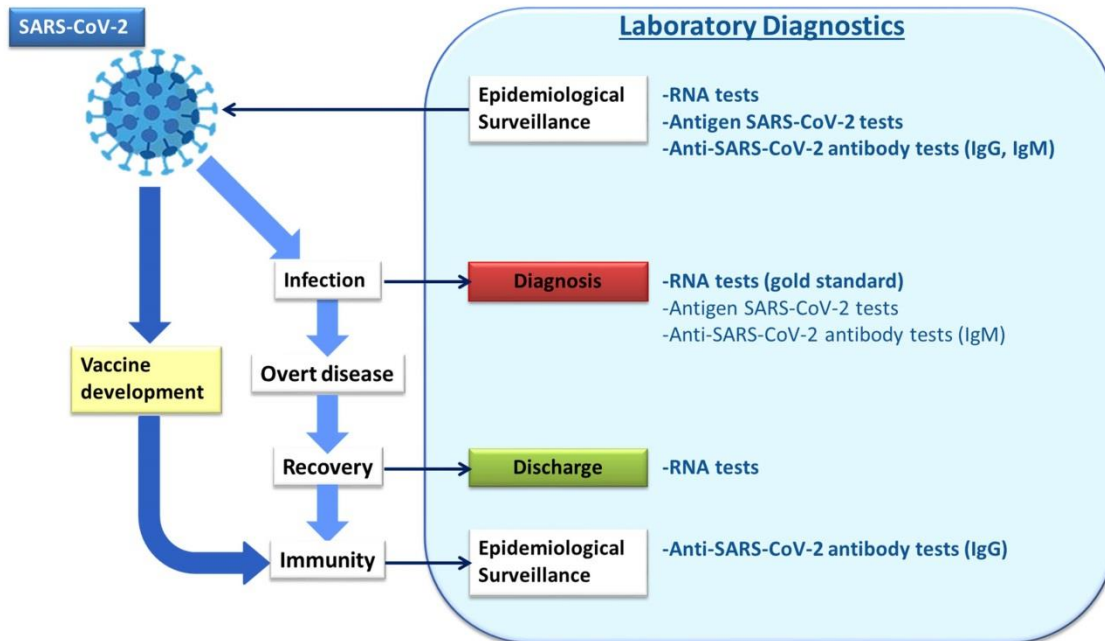


Figure 1: Testing in the context of the COVID-19 disease

At present real-time reverse transcription polymerase chain reaction (RT-PCR) is the gold standard for diagnosing suspected cases of COVID-19. These methods, targeting viral ribonucleic acid (RNA), are currently used as the preferred approach to identify the SARS-CoV-2 virus directly.

The RNA contained in this virus is generally detectable in respiratory specimens during the early and acute phases of infection. Whilst positive results are indicative of the presence of SARS-CoV-2 RNA, a clinical correlation with the patient history and other diagnostic information is necessary to determine the infection status of the patient. Positive results do not rule out an additional bacterial infection or a co-infection with other viruses.

Negative results do not preclude a SARS-CoV-2 infection and should not be used as the sole basis for treatment or other patient management decisions. For instance, negative RT-PCR results from throat swabs occur in the later infection phase when the virus has migrated into the lung. Negative results should be combined with clinical observations, patient history and epidemiological information.

Parts of the coronavirus, in particular several proteins on its surface, are able to act in the body of an infected person as antigens, i.e. they can cause an immunological reaction. In the context of a test, the detection of those ‘foreign’ proteins can also be used as a signal of a viral infection. The antibodies produced by the patient’s adaptive immune system to recognise and neutralise

the antigens can also be detected. These antibodies are secreted into the blood and mucosa. Antigens are likely detectable at a much earlier stage in the COVID-19 infection process than the antibodies produced as part of an immune response. Antigen tests give information on the patient's coronavirus infection status and determine whether a person should be isolated. Antibody tests provide information on a patient's immune status. After infection, the first antibodies to appear in the blood are of the IgM type. They initiate the first line of defence. IgA antibodies, which appear around the same time as the IgM, are mainly present in the mucosa and at lower concentrations in the blood.³ IgG antibodies appear later, and further control the infection. Typically IgM antibodies disappear in several weeks to months, but IgG could remain present in the blood for many years, or even the rest of the person's life, and play a role in protective immunity. It has been found that COVID-19 patients produce antibodies in a similar pattern. Figure 2 presents results from a study on anti-SARS-CoV-2 which shows IgM and IgG levels over time. Antibody tests would give negative results in the early stages of disease.

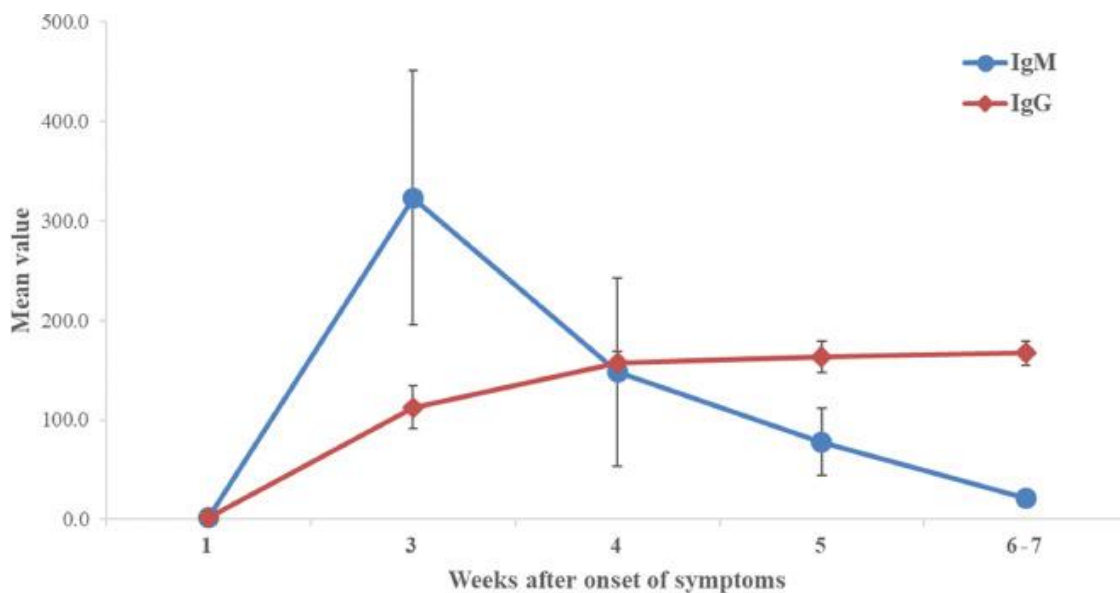


Figure 2: Estimated timeline of IgM and IgG antibody levels to SARS-CoV-2 from the onset of symptoms (adapted from ⁴)

Both antigen and antibody testing is based on the application of immunological reactions to capture and detect either SARS-CoV-2 components or the patient's antibodies, respectively. Such tests are collectively called immunoassays. The present value of immunoassays (both antibody and antigen tests) for COVID-19 diagnosis and monitoring is heavily debated. These tests are currently not recommended for the diagnosis of suspected COVID-19 cases by the European Centre for Disease Prevention and Control (ECDC)⁵, the World Health Organization

³ Li Guo, Lili Ren, Siyuan Yang et al. Profiling Early Humoral Response To Diagnose Novel Coronavirus Disease (COVID-19) *Clinical Infectious Diseases* (2020), doi: <https://academic.oup.com/cid/article/doi/10.1093/cid/ciaa310/5810754>

⁴ Ai Tang Xiao, Chun Gao, Sheng Zhang, Profile of Specific Antibodies to SARS-CoV-2: The First Report, *Journal of Infection* (2020), doi: <https://www.sciencedirect.com/science/article/pii/S0163445320301389?via%3Dihub>

⁵ ECDC Rapid risk assessment: Coronavirus disease 2019 (Covid-19) pandemic: increased transmission in the EU/EEA and the UK, eighth update, <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-eighth-update>

(WHO), the Centers of Disease Control and Prevention (CDC) in the United States⁶ and other public health organizations. The ECDC has indicated that clinical trials are needed for the clinical validation of COVID-19 immunoassays before they can be safely and reliably used for medical or public health decision making.^{7,8} In addition, the Food and Drug Agency (FDA) of the United States has stated that results from immunoassays should not be used as the sole basis to diagnose or exclude a SARS-CoV-2 infection.⁹

It is too early for sound scientific evidence on the long-term protective immunity against SARS-CoV-2. However, the ECDC has advised the EU that immunoassays detecting specific antibodies against SARS-CoV-2 will play an important role in the future for epidemiological surveillance, evaluation of immunity and the outcome of vaccination studies.⁵

Table 1 illustrates in a simplified manner the correlation between different test results and the phase of infection. It describes an envisaged ‘ideal decision’ situation that is so far not validated and achieved by currently available tests and data therefrom.

Table 1: Envisaged indications on a person’s COVID-19 status from testing different targets

COVID-19 phase	RNA test*	Antigen test	IgM test	IgG test
<i>before infection</i>	negative**	negative**	negative	negative
<i>first phase of infection</i>	positive	positive later than RNA test	negative	negative
<i>second phase of infection</i>	positive	positive	positive***	negative
<i>last phase of infection</i>	positive	positive	positive***	positive***
<i>after infection</i>	negative**	negative**	positive and later negative	positive***

*with optimal specimen sampled

**diagnostic sensitivity of the test very important (see 4.1.1 & 4.1.2)

***diagnostic specificity of the test very important (see 4.2.1)

The combination of test results from RT-PCR, antigen and antibody testing can provide a clearer picture on the status of a patient. When the RT-PCR and the antigen test results are positive (independently from the presence of symptoms), this indicates that the person is

⁶ Centers of Disease Control and Prevention (CDC): Information for laboratories website: <https://www.cdc.gov/coronavirus/2019-nCoV/lab/index.html>

⁷ European Centre for Disease Prevention and Control (ECDC): An overview of the rapid test situation for COVID-19 diagnosis in the EU/EEA <https://www.ecdc.europa.eu/en/publications-data/overview-rapid-test-situation-covid-19-diagnosis-eueea>

⁸ In the EU regulatory framework for in vitro diagnostic medical devices, these trials, when carried out by the manufacturer for the purposes of CE-marking, are called ‘performance studies’

⁹ FDA Policy for Diagnostic tests for Coronavirus Disease-2019 during the public health emergency: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/policy-diagnostic-tests-coronavirus-disease-2019-during-public-health-emergency>

infected and the virus is present and replicating in the body. The presence of IgM and IgG, as detected by immunological tests, shows that the patient's immune system is reacting to a recent infection (IgM) and that the patient has developed (temporary) immunity to the virus (IgG).

The screening of persons with or without symptoms can be differentiated based on the purpose and the testing techniques. Testing based on RT-PCR or an antigen test is important for the identification of infectious cases, as it confirms that the virus is present in the body. This information is essential for assessing the risk of spreading of the virus and to understand the chains of infection.

Antibody testing, coupled with RT-PCR, will be important for surveillance and provides a tool for developing an exit strategy with selective restrictions as a reaction to the pandemic⁵. Antibody tests can indicate who has had the virus and therefore may be immune, although as stated previously sound scientific evidence on the long-term protective immunity against SARS-CoV-2 is still lacking. Surveillance based on antibody testing may reduce the burden on direct virus testing. In the short term, it could also contribute to inform whether people with a demonstrated immunity could be exempt from confinement measures.

Under normal conditions, any positive result detected during screening tests should undergo confirmatory testing using more reliable methods. In case of COVID-19, two different scenarios based on the epidemiological situation are proposed by WHO¹⁰: in areas with no or low virus circulation¹¹, suspected cases should be confirmed by a positive RT-PCR result for at least two different targets of the virus genome or a positive RT-PCR result for the presence of a beta-coronavirus that is further identified by partial or whole sequencing of the virus genome. However, in the situation where the virus is widespread and where a second confirmatory test might not be feasible or delay the medical decision, screening by RT-PCR of a single discriminatory target is considered sufficient.⁵

In the light of infrastructure limitations and supply shortages, access to reliable rapid diagnostic tests could alleviate the pressure on laboratories and expand testing capacities⁵. Rapid tests, as defined in the common technical specifications for IVD (revised Commission Decision 2002/364/EC¹²) are qualitative or semi-quantitative tests, which involve non-automated procedures and have been designed to provide a fast result. Both antigen and antibody tests can be found on the market in the form of rapid tests. Moreover, such rapid tests can be designed either for use by health professionals or by lay users. In the latter case they would fall under the category of self-tests, which have to undergo additional verification by a notified body before being placed on the market. Self-testing performed by non-professionals could be a tool to reduce the burden on testing laboratories, but despite efforts by many companies to provide these tests, the EU notified bodies have not yet issued certificates for any such test, and Member State competent authorities are generally not in favour of their use for COVID-19 at this stage.

Point-of-care tests (POCT), also termed near-patient tests, carried out by professionals can provide quicker results than laboratory tests. Some tests of this kind have been CE-marked.

¹⁰ <https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117>

¹¹ According to ECDC daily reports, low circulation EU countries (less than 0.2% of reported cases/1000 inhabitants) are at present: Greece, Poland, Slovakia, Hungary, Bulgaria

¹² Commission Decision 2002/364/EC on common specifications for in vitro diagnostic devices, OJ L 131, 16.5.2002, p. 17

2. Scope

This report has been established under the mandate to collect and review performance assessment studies of test methods and devices for COVID-19 relevant to the EU that have been performed by academic institutions, national regulatory agencies, international organisations, health technology assessment (HTA) bodies, reference laboratories, and similar organisations, and to suggest performance criteria for different types of COVID-19 test methods and devices. The report does not provide a quality ranking of specific tests or devices with respect to their reliability or performance. Moreover, it does not rank commercial devices regarding their usefulness for the different tasks in COVID-19 diagnostics.

Readers should bear in mind that this report is based on the information available to the authors up to 06 April 2020.

Devices placed on the market in the EU must comply with the relevant requirements of Directive 98/79/EC on *in vitro* diagnostic medical devices (IVDD)¹³. As of 26 May 2022, the Directive will be replaced by Regulation (EU) 2017/746 (IVDR)¹⁴ and during the currently ongoing transition period it is also possible to place on the market devices which comply with the IVDR according to its Article 110. The performance criteria proposed in this document are intended as additional temporary emergency guidance, in view of the COVID-19 pandemic, that cannot replace the essential requirements given in the IVDD or IVDR. Ultimately, manufacturers are responsible for bringing the corresponding devices in full conformity with all requirements of the IVDD or IVDR.

¹³ OJ L 331, 7.12.1998, p.1

¹⁴ OJ L 117, 5.5.2017, p. 176

3. Methodology used

The evidence collected regarding the performance of test methods and devices for COVID-19 testing and diagnostics included:

- published peer-reviewed journal articles and non-peer-reviewed manuscript preprints;
- industry documentation on product technical specifications or instructions for use;
- non-peer reviewed published studies or assessment reports by national/international regulatory agencies, including WHO EUL prequalification of IVD;
- preliminary results from validation studies by clinical or reference laboratories reported to open access clinical research registries.

Search strategy for scientific literature

The literature search to identify documents describing the use (or evaluation) of methods for SARS-CoV-2 detection was performed on April 4th, 2020, in three resources:

- Scopus (<https://www.scopus.com>), peer-reviewed articles;
- bioRxiv (<https://www.biorxiv.org/>), preprints;
- Europe PMC (<https://europepmc.org/>), peer-reviewed articles and preprints.

Scopus

The search was performed using the following string:

TITLE-ABS-KEY (covid OR "sars-cov-2" OR "2019-nCoV")

The results (1028 articles) were downloaded as a tab-separated document.

Using a script, each entry was scanned (title, abstract, author keywords and index keywords) for any of the following strings: “detection”, “diagnos”, “polymerase”, “immuno”.

The results (218 articles) were extracted and reviewed one by one to identify relevant articles based on the title (and, when necessary, the abstract). The final selection contained 13 articles.

bioRxiv

The bioRxiv site provides a link listing COVID-19 SARS-CoV-2 preprints from both medRxiv and bioRxiv (<https://connect.biorxiv.org/relate/content/181>).

The JSON version of this list (1146 articles, 886 medRxiv, 260 bioRxiv on April 4th) was retrieved by a script and each entry was scanned (title and abstract) for any of the following strings: “detection”, “diagnos”, “polymerase”, “immuno”.

The results (312 articles) were extracted and reviewed one by one to identify relevant articles based on the title (and, when necessary, the abstract). The final selection contained 62 articles.

Europe PMC

The search was performed using the following string:

("2019-nCoV" OR "COVID-19" OR "SARS-CoV-2") AND (detect* OR diagnos*) AND (FIRST_PDATE:2020)

The results (1391) were downloaded as a tab-separated document. Each entry was reviewed one by one to identify relevant articles based on the title. The final selection contained 83 articles.

Final list

The final list was obtained by combining the results of the three searches, removing the duplicates (based on the DOI and/or title). The final list contained 101 unique articles (publications and preprints). Additional articles were suggested by the members of the Project Group. The total number of scientific articles assessed was 120 (see Annex 2).

Documentation on test methods and devices

The search was performed trying to cover the broader space and therefore different approaches were applied.

At first, depository websites were investigated which are dedicated to collect information on existing and emerging tests developed for the purpose or adapted to the detection of COVID-19:

- Tests commercially available or in development for the diagnosis of COVID-19 from the FIND¹⁵ webpage <https://www.finddx.org/covid-19/pipeline/>;
- 360Dx¹⁶ web page Coronavirus Test Tracker: Commercially Available COVID-19 Diagnostic Tests (<https://www.360dx.com/coronavirus-test-tracker-launched-covid-19-tests>).

Secondly, the EMM (Europe Media Monitor)-finder was used to identify news articles as of 1st January 2020 containing ‘COVID-19’ or ‘SARS-CoV-2’ in combination with ‘detection’ or ‘test’ or ‘diagnostics’ or ‘method’ or ‘measurement’. The EMM-finder is a specific adaptation of the JRC tool for text mining of media news and allows searching also back in time in the database.

Finally, similar searches were performed using Google to identify the largest possible number of Covid-19 diagnostic tests available in the market.

A parallel search was conducted using the PitchBook¹⁷ tool. The search, for deals from 15 November 2019, was made by looking for “covid” OR “covid-19” OR “covid 19” OR “SARS-CoV-2” OR “novel coronavirus” OR “novel corona virus” OR “2019-ncov” OR “2019ncov”. The above search was expanded with “sars” OR “mers” OR “ace2” OR “soluble ace 2” OR “sars-cov” OR “mers-cov”. Lastly, a search was performed in “diagnostic equipment” AND “virus” with deals from 15 November 2019.

The manufacturers obtained from the financial tool were manually screened to verify if they were producing COVID-19 devices. When this was the case and if they were not already in the list, they were added.

¹⁵ FIND (<https://www.finddx.org/>) is the Foundation for Innovative New Diagnostics, a global non-profit organization driving innovation in the development and delivery of diagnostics to combat major diseases affecting the world’s poorest populations.

¹⁶ 360Dx (<https://www.360dx.com/>) is published by GenomeWeb, an independent online news organization based in New York. 360Dx was launched in 2016, to cover emerging economic and technological trends in the clinical diagnostic market.

¹⁷ PitchBook (<https://pitchbook.com/>) is a financial data and software company that provides comprehensive data on the private and public markets (venture capital, private equity, mergers and acquisitions).

The final compilation of devices, together with the additional information that was gathered for each is found in Annex 1. The table contains the name of the manufacturer, the name of the device, the type of method (PCR, immunological, digital, control panel, CRISPR), the status of the device (“commercialised” or “in development”), an indication of the speed and the regulatory frame for which they are fit for. The indication of the speed is using the 7 categories defined by FIND, “rapid diagnostic test” that include immunoassays providing results in less than 20 min, the classical immunoassays without a label, the “manual NAT” are PCR methods based on nucleic acid amplification tests that are more time consuming, “Automated lab-based, near-POC NAT or POC NAT” are automatic PCR tests that allow near point-of-care or near-patient laboratory testing, finally those classified as “manual or automatic” because they exist in both versions. Finally, the regulatory frame classified the devices as “research use only”, “in development”, “proof of concept” and approved by the different country/region regulatory bodies (US FDA, China, Korea, India or CE-IVD (EU compliance label according to the IVD Directive).

From this list, each method/device obtained an internal unique number. A selection was made for further analysis based on criteria to focus on those devices accessible in the EU. For this reason, products labelled as for “research use only” or under development have been ignored as well as products fulfilling other regulatory frameworks than the EU IVD Directive. Those with the CE-IVD marking were further classified and split in three tables depending on the methodology (Annex 2). In total, 78 devices based on RT-PCR (or variants e.g. CRISPR and LAMP), 101 for the detection of antibodies and 13 for the detection of antigens were assessed.

4. Overview of information on performance & proposed performance criteria

4.1. Detection of viral status

4.1.1. RT-PCR

A) Evaluation of evidence

The most crucial information concerning RT-PCR based methods developed for the detection of SARS-CoV-2 are the sequences of the oligonucleotides (primers and probe) used for the amplification of the cDNA. That information allows to establish a RT-PCR based method, as it targets a precise sequence in the genome of the virus.

For many of the RT-PCR devices matching the selection criteria described in Section 3, information supplied by the manufacturer could be found in the form of technical specifications or instructions for use. This information contained some relevant information on the device performance, compiled in the annexed table and summarised below.

It is important to note that, except for a few cases, no information on the actual sequences of the primers and probes in the device could be found. This makes the complementation of performance information from what is published in the scientific literature for an individual device conditional on the explicit mentioning of the device in the ‘materials and methods’ part of a publication.

Literature review – general observations

A significant number of new real-time PCR protocols to detect the presence of SARS-CoV-2 RNA have been published recently in addition to those recommended by WHO¹⁸. Most of the studies tested and compared RT-PCR methods also on clinical samples from patients, some on different multiple samples coming from large numbers of patients.

Instances of new or improved methods have also been described. For example, ART69 (Annex 2) compared a new set of primers for the RdRp gene with the method recommended by WHO developed at the Charité Universitätsmedizin Berlin, Germany. The method shows *in vitro* a very low limit of detection (LOD) with viral RNA transcripts (11.2 RNA copies/reaction). In patient samples, the new assay detected SARS-CoV-2 RNA in 42 out of 273 (15.4%) additional specimens that were tested negative by the Charité assay. The new assay was significantly more sensitive than the Charité assay for the detection of SARS-CoV-2 RNA in nasopharyngeal swab, saliva, and plasma specimens.

A particularity of the assays described in the scientific literature is that attempts are being made to develop and implement methods based on novel techniques, such as digital chamber PCR, digital droplet PCR, LAMP and CRISPR.

A few papers applied the RT-PCR protocol from the CDC method based on the ORF1Ab and N gene to digital PCR (dPCR) (ART90, ART68 and ART42) or used a commercial kit for dPCR (ART104) to test patient samples (close to 100 in each study) previously tested by real-time PCR. In these studies, among all the confirmed positive samples, some that were tested false negative by RT-PCR were corrected by the more sensitive dPCR assay. The digital PCR

¹⁸ WHO, 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance>

shows an improved lower limit of detection, sensitivity and accuracy compared to real-time PCR for low viral load diagnosis. This methodology showed to reduce false negative detection in samples from the lower respiratory tract, especially in cases of low viral load samples. This technique requires more sophisticated instruments.

CRISPR technology is widely known for its use in gene editing techniques. But in recent years, CRISPR has been also used for the *in vitro* detection of nucleic acids, thereby emerging as a powerful technology for molecular diagnostics.^{19,20}

What made CRISPR revolutionary was its ability to recognise specific sequences of DNA. The same can be done for viral genetic material so that the reaction emits a fluorescence signal on cleavage or could be alternatively detected on a paper strip by lateral flow in a portable manner.²¹

However, it is important to note that all of the published studies on CRISPR are only at a proof of concept stage, where the technique is only tested on plasmid positive controls or spiked human samples. None of the published studies have analysed real patient samples. At present, only two kits are available using this technology but both are still “in development” phase; one of them is made available for “Research Use Only”.

LAMP is an isothermal variation of PCR (meaning it does not require dedicated and expensive thermal cycling equipment) and is therefore commonly used for point-of-care testing (POCT) due to its high sensitivity, rapid reaction and simple operation. The result is evaluated by a colour change that does not require specialised personnel.

All the published papers presented LAMP combined with a reverse transcription assay (RT-LAMP) which had been developed for the detection of multiple respiratory RNA viruses to avoid the need for expensive technologies. The LAMP technique is highly specific through the combined detection of three genes of the coronavirus. It is claimed by manufacturers that this would possibly increase the accuracy of detection to almost 99%. However, it is reported that the use of one of the three genes alone would reduce the diagnostic sensitivity or specificity of the test, causing false positive or false negative results.

Application aspects

RT-PCR methods require specific laboratory equipment and various sets of reagents. As a rule, these tests need to be performed by trained laboratory personnel who are proficient in performing the methods. In addition, several scientific publications underline the importance of combining CT scans and clinical observation with RT-PCR results for the follow-up of COVID-19 patients, in order to improve the discovery rate of the disease and to evaluate a patient discharge from the hospital. Furthermore, the testing of different kinds of swabs and samples from different body fluids is recommended in order to increase the potential detection of the virus by RT-PCR at different stages of the disease.

The time needed to perform the test, when reported, has been compiled in the annexed table, and varies from one to a few hours. Part of the reported differences comes from the different starting point, i.e. whether the method includes the steps to extract and purify the viral RNA from the original sample or only starts from a previously purified sample. The assay protocols

¹⁹ Gootenberg JS, Abudayyeh OO, Lee JW, et al. Nucleic acid detection with CRISPRCas13a/C2c2.2017. *Science*;356(6336):438–442

²⁰ Li SY, Cheng QX, Wang JM, et al. CRISPR-Cas12a-assisted nucleic acid detection. 2018. *Cell Discov.* 4:20

²¹ Myhrvold C, Freije CA, Gootenberg JS, et al. Field-deployable viral diagnostics using CRISPR-Cas13. 2018. *Science* 360(6387):444-448

generally rely on preparing tube-by-tube reactions; how many can then be run at once is varying according to the instrumental setup (e.g. 96 wells and more). Proofs of concept of testing workflows for lab-based surveillance have been proposed using pools of several RNA samples from different patients in order to increase the throughput (ART04, ART59). This approach needs further validation, but it would reduce costs while monitoring the epidemic in the population.

Despite the high analytical sensitivity and specificity of real-time PCR (see below), a study on 610 hospitalized patients from Wuhan in February 2020 (ART97) showed that RT-PCR test results of pharyngeal swab specimens were variable. This study concluded that this method should not be considered as the only indicator for the presence of the coronavirus.

Quality control

The information relevant to the controls included for each device, based on the components and protocol, are compiled (when found) in the annexed table.

In general devices include a positive control material, i.e. a purified template known to contain the amplified target(s) (e.g. a pseudovirus RNA). Many also include an internal control method, i.e. a method that amplifies a target and produces a signal whether or not the sample contains SARS-CoV-2 RNA (e.g. human RNase P gene, Rpp30 gene). Negative controls are usually described as reactions performed without adding a sample to the reaction mix (e.g. distilled water, IVT synthetic mix).

Analytical performance

Limit of Detection (LOD)

Some information made public by the manufacturers includes indication of the LOD of the methods, sometimes providing additional details on how the determination was performed. Identified values are compiled in the annexed table, reaching as low as a few copies of the viral RNA per reaction. This is in line with the generally very low LOD achievable by RT-PCR-based methods and is a main advantage of these approaches.

For most of the publications reviewed, only a partial in-house validation of the method was performed. Sensitivity measurements are not always done when testing a method or comparing different method performances. Moreover, when sensitivity is tested, the LOD assessment is rarely performed according to the minimum performance criteria described below.

(Analytical) Specificity

Depending on the devices, information was sometimes included in the public documentation regarding specificity, describing:

- bioinformatics analysis to confirm that the oligonucleotides would bind correctly to all known SARS-CoV-2 RNA isolates whose sequences have been compiled in repositories;
- bioinformatics analysis to confirm that the oligonucleotides would not bind to nucleic acid sequences of other microorganisms, including viruses close to SARS-CoV-2;
- laboratory experiments with samples from other microorganisms to demonstrate the lack of signal produced by the device in these instances.

The annexed table reports when such information was available, in particular for laboratory experiments that are of most practical relevance. The compiled information is in line with the generally very high analytical specificity achievable by RT-PCR-based methods and is a main advantage of these approaches.

In the literature, specificity of the methods is only sometimes tested with panels (often commercially available) of human viruses and pathogen RNA spiked into clinical samples, in order to exclude cross-reactivity. It should be noted that a study has highlighted some nucleotide mismatches (that may adversely affect the reactions efficiency) on the primers' annealing sites of methods listed by WHO (Annex 2, ART31).

PCR efficiency, robustness, precision

With very rare exceptions, no information was found on efficiency, robustness and precision as defined below.

Exceptions include Novacyt's genesig Real-Time PCR COVID-19 device, whose instructions for use include a description of the reproducibility and repeatability analysis performed in the laboratory, as well as the two devices from Genomica Sau, for which it is reported that a validation study was performed at the National Center for Microbiology (Instituto de Salud Carlos III, Spanish National Reference Center) with a panel of 80 samples. It is unclear for the latter whether the validation included considerations of robustness.

Similarly, the efficiency of RT-PCR reactions with synthetic target RNA transcripts is often not reported as measured in scientific studies and the effect of potential interfering substances has only been reported in one publication (Annex 2, ART37). The reproducibility of the method is sometimes assessed using different RNA extraction kits, which may lead to a different sensitivity of the method influencing the results with false negatives, especially in samples with low viral load.

Clinical performance

For a few of the devices assessed, information was found about the results of clinical performance studies, compiled in the annexed table.

When available, manufacturer claims on both the diagnostic sensitivity and specificity were very optimistic (in the range of 96-100%).

Linking the scientific literature with specific commercial devices

Finding additional information in the scientific literature on the performance of specific commercial devices is difficult due to the lack of information regarding the actual sequences of the oligonucleotides that are comprising the devices' main components, as explained above. Without this information, it is not possible to link relevant scientific studies, even if the sequences are almost always made explicit in the publications, unless the article explicitly mentions the device used in the study.

Exceptions to this include manufactured devices that specifically mention the implementation of one of the WHO recommended methods, as the primers and probes for those have been made public and have been extensively used in research and testing laboratories until now. These are:

Manufacturer	Test	WHO method explicitly referenced	
Idrop Inc.	1copy™ COVID-19 qPCR Kit	Charité protocol (2 nd version)	Reference
AB ANALITICA Srl	REALQUALITY RQ-2019-nCoV	Charité protocol (2 nd version)	Reference
Diatheva SRL	COVID-19 PCR DIATHEVA Detection kit	Charité protocol (2 nd version)	Reference
Sentinel CH	STAT-NAT® Covid-19 B	Charité protocol (2 nd version)	Reference
Sentinel CH	STAT-NAT® Covid-19 HK	Hong Kong Faculty of Medicine protocol	Reference
BioGX	SARS-CoV-2 Open System Reagents for BD MAX	US CDC protocol	Reference

The protocols suggested by WHO are currently the most recurrent ones in the literature, because some of their primer/probe sets are commercially available. However, many of them present background noise that render the interpretation of results difficult at low viral load and a proper validation of these RT-PCR methods is recommended. So far, among WHO suggested protocols only the ones from the Institut Pasteur, the Hong Kong Faculty of Medicine and the Charité were in-house validated and good analytical specificity and LODs were reported (Institute Pasteur, LOD: 100 cp/reaction; Charité for target E, LOD: 3.9 cp/reaction and for target RdRp, LOD: 3.6 cp/reaction). The performances of all WHO suggested protocols (except the one from Institut Pasteur) have been evaluated in several papers and, although all methods showed high analytical specificity, the Charité method targeting E and RdRp genes and the method of CDC (US) targeting N1 and N2 regions have the best sensitivity. However, not all methods were compared under the same conditions and not for all of them the analytical specificity and sensitivity were properly assessed.

Some scientific publications include studies evaluating or comparing explicitly named commercial CE marked devices found in the attached table. In one study, devices from DAAN Gene Co., Sansure Biotech, Chaozhou HybriBio Biochemistry Ltd were analysed and shown to perform slightly better than another device, Bioperfectus (Annex 2, ART37). It should, however, be noted that the resulting LOD values were higher than the ones provided by the manufacturers.

Commercial devices are sometimes used for comparison purposes in studies showing the performance of novel methods. For comparisons with new LAMP-based methods, ART47 uses a SARS-CoV-2 device from Shanghai BioGerm Medical Biotechnology Co. Ltd and a device from DAAN Gene Co., while ART49 uses the Liferiver Novel Coronavirus (2019-nCoV) Real Time Multiplex RT-PCRT device. In ART38, 57, 68 and 69 devices are mainly used in comparison with other developed RT-PCR and dPCR methods or to confirm the results coming from the newly developed methods. In these studies, no particular evaluation of the devices is included since it was not part of the main objectives of the work.

B) Proposed performance criteria

Both commercial diagnostic devices and test methods (e.g. protocols) based on the RT-PCR principle have been taken into account for this report. Test methods require a more detailed consideration, as each individual reagent (e.g. Master Mix, primers/probe, control samples) has to be purchased or prepared by the user, whereas such reagents will often be included as parts of the devices so that the corresponding information does not have to be necessarily disclosed.

The analytical and clinical performance criteria are described below. In addition to this, guidance on descriptive information, quality control and safety measures is given.

Descriptive information

Essential requirements are laid down in Annex I of the IVDD and IVDR. More specifically, for the testing of SARS-CoV-2 virus RNA, the following descriptive information is of particular importance:

- the nature of the test (qualitative, semi-quantitative);
- the measured target (the specific RNA fragment from SARS-CoV-2 amplified);
- the type of specimen that can be tested, e.g. oropharyngeal swab sample, its pretreatment (e.g. required dilution), stability and storage
- the purpose (e.g. the results are for the identification of SARS-CoV-2 RNA);
- the indication whether any particular training is required (e.g. the test is intended for use by trained laboratory personnel who are proficient in performing real-time RT-PCR assays).

The general principle of the method should be described (e.g. RNA isolated and purified from upper and lower respiratory specimens is reverse transcribed to cDNA and subsequently amplified in a real-time PCR instrument).

If the RNA extraction step is not part of the device, a list of the RNA extraction kits (name of the instrument/manufacture, name of the extraction kit, catalogue no.) for which the devices/tests have been proven to perform reliably should be provided.

A complete list of reagents either provided (in case of devices) or needed (in case of test methods) should be provided as well as the necessary consumables and equipment. Information concerning the control of efficacy of the sample preparation and the absence of inhibitors in the PCR reaction should be provided.

For reagents the expiration date and instructions concerning their storage and handling (e.g. reconstitution of lyophilised material and reagents) should be known.

Quality control

The required quality controls should be described. Extraction and amplification should be controlled by an internal control in each specimen. Positive and negative controls should be analysed in parallel with the specimens in each run to guarantee the validity of the test and the correct interpretation of the results.

Safety measures

It is of the utmost importance that proper biosafety guidelines are followed by clinical laboratories when handling samples from suspected COVID-19 patients. The laboratory

biosafety guidance related to the coronavirus disease issued by the World Health Organization should be taken into account where applicable.²²

Analytical performance

Limit of detection (LOD)

The LOD can be determined by limiting dilution studies using sufficiently characterised samples and should be provided.

Since no virus isolates with a quantified amount of the SARS-CoV-2 are currently available, assays designed for detection of the SARS-CoV-2 RNA could be tested with characterised stocks of *in vitro* transcribed RNA containing the target of interest of a calculated titer (RNA copies/ μ L) spiked into a diluent consisting of a suspension of human cells in viral transport medium (VTM) to mimic a clinical specimen. Such studies are generally performed in two steps. In a preliminary step, an approximate LOD for an assay is determined by testing triplicate samples of RNA purified using a defined extraction method. The approximate LOD is determined by extracting and testing 10-fold serial dilutions of characterized stocks of *in vitro* transcribed RNA. A confirmation of the predetermined LOD is then performed at a chosen dilution of the spiked RNA samples with a minimum of 20 extracted replicates. The LOD is determined as the lowest concentration where $\geq 95\%$ (19/20) of the replicates are positive. More commonly, this is done by at least five half-logarithmic dilutions around the pre-determined LOD, tested in replicates of at least 24 samples.

(Analytical) Specificity

The analytical specificity of the assay (also named inclusivity testing) is assessed *in silico* by aligning the oligonucleotide primer and probe sequences to all publically available nucleic acid sequences of SARS-CoV-2 in GenBank as of a particular date. The percentage of identity of the amplicon should be evaluated and a potential mismatch should be reported. The risk of a single mismatch resulting in a significant loss in reactivity and potentially a false negative result should be evaluated. Experience has shown that a proper design of the primers and probes with melting temperatures > 60 °C and PCR run conditions of the assay with annealing temperature at 55 °C could tolerate one to two mismatches.

A search for significant homologies with other SARS coronaviruses, the Bat SARS-like coronavirus genome and the human genome and human microflora should be performed to evaluate and predict potential false positive RT-PCR results.

The *in silico* assessment should be complemented by testing the assays against RNA extracted from real samples (isolates/clinical specimen of other human coronaviruses, MERS coronaviruses, SARS coronavirus, human influenza, etc.) to confirm a negative result.

Robustness

The effect of other potential interfering substances should be investigated. The equivalence of the LOD of the assays should be evaluated with different enzyme master mixes on serial dilutions of *in vitro* transcribed RNA.

²² <https://www.who.int/publications-detail/laboratory-testing-for-2019-novel-coronavirus-in-suspected-human-cases-20200117>

Retrospective positive and negative clinical respiratory specimens (when available), extracted by different extraction kits, should be tested to evaluate the equivalent performance of the extraction kits.

Precision

Both repeatability (i.e. testing the same sample under the same conditions) and reproducibility (i.e. testing the same sample under variable conditions such as different reagent kits, days, different analysts, or different instruments) should be assessed. In the case of qualitative test results the precision parameters can be expressed as the percentage of agreement.

Clinical performance

The analytical performance criteria are usually evaluated on a number of well-defined laboratory samples and extreme patient samples. The next step should be the clinical performance evaluation of the diagnostic test accuracy. The determination of diagnostic accuracy should be performed in (clinical) studies using head-to-head comparison between results from one or more RT-PCR tests under assessment and those of the reference RT-PCR test in the target population intended to be tested. The diagnostic accuracy is composed of:

- Diagnostic sensitivity: Proportion of those individuals with the target condition (infected individuals with reference SARS-CoV-2 RNA true positive specimen) who test positive with the RT-PCR test;
- Diagnostic specificity: Proportion of those individuals without the target condition (infection-free individuals with reference SARS-CoV-2 RNA true negative specimen) who test negative with the RT-PCR test.

The diagnostic sensitivity of a specified RT-PCR test is the most crucial parameter for the identification of persons who are infected by SARS-CoV-2, as false negative test results have to be as low as possible.

The number of patients (sample size) to be included in the trial should be determined by statistical power calculation for a desired precision level of the accuracy estimates. The study report should provide the estimates of diagnostic sensitivity and specificity with confidence intervals. The STARD 2015 (STAndards for Reporting Diagnostic accuracy studies) should be followed²³. It is recommended to report the 95 % confidence interval for the estimates of both the diagnostic sensitivity and the diagnostic specificity.

The clinical performance of a test method or device should be evaluated for all the intended conditions of use, e.g. each type of specimen mentioned should be evaluated.

4.1.2. Antigen tests

A) Evaluation of evidence

There are two main types of immunological tests, namely ELISA (enzyme-linked immunosorbent assay) and LFA (lateral flow assays). ELISA tests are more complex and require trained technicians operating under sterile laboratory conditions, but tend to be more

²³ Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799. doi:10.1136/bmjopen-2016012799

sensitive. LFA are point-of-care formats that have been developed with simplicity and portability in mind. They can be simple and easy-to-produce devices.

Literature and product review

At present rapid antigen tests are not fully developed and the literature is scarce. A few papers have described an antigen microarray, but this was rather fundamental research and not an available commercial device (Annex 2, ART05 and ART111).

We have identified very few (13) commercial antigen tests and, according to our analysis, they are currently at a too early stage of development, i.e. they do not meet the proposed performance criteria described in Section 4.2.1B.

B) Proposed performance criteria

The proposed performance criteria on the descriptive information and the analytical performance are identical to the ones for antibody tests and are therefore presented together in Section 4.2.1B. The clinical performance criteria described below are specific for the antigen tests:

Clinical performance

The proposed clinical performance criteria for SARS-CoV-2 antigen detection are similar to those for SARS-CoV-2 RNA detection by RT-PCR (Section 4.1.1B). The determination of diagnostic accuracy should be performed in clinical studies using head-to-head comparison between results from one or more antigen tests under assessment and those of the reference RT-PCR test in the target population intended to be tested. The clinical performance criterion for a SARS-CoV-2 antigen test under validation is the diagnostic accuracy, composed of:

- Diagnostic sensitivity: Proportion of those individuals with the target condition (infected individuals with reference SARS-CoV-2 RNA true positive specimen) who test positive with the antigen test;
- Diagnostic specificity: Proportion of those individuals without the target condition (infection-free individuals with reference SARS-CoV-2 RNA true negative specimen) who test negative with the antigen test.

For antigen tests the diagnostic sensitivity is the most crucial parameter for the correct identification of persons who are infected by SARS-CoV-2, as false negative test results must be as low as possible.

The number of patients (sample size) to be included in the trial should be determined by statistical power calculation for a desired precision level of the accuracy estimates. The study report should provide the estimates of diagnostic sensitivity and specificity with confidence intervals. The STARD 2015 (STANDards for Reporting Diagnostic accuracy studies) should be followed²⁶. It is recommended to report the 95 % confidence interval for the estimates of both the diagnostic sensitivity and the diagnostic specificity.

The clinical performance of a test method or device should be evaluated for all the intended conditions of use, e.g. each type of specimen mentioned should be evaluated.

4.2. Detection of immunological status

4.2.1. Antibody tests

A) Evaluation of evidence

The majority of the devices are targeting the immunoglobulins IgG and IgM in a combined manner (54 devices). These devices were designed to display the presence/absence of one of the two immunoglobulins or of both in one single test. The remaining devices are targeting only IgG or only IgM. We found only one case where the test was directed to IgA.

Literature review

Many of the currently available tests have been used on larger numbers of patients and their results tend to be similar (e.g. regarding sensitivity and specificity). Tests using the virus spike protein to capture the antibodies seem to be more sensitive than the ones using the nucleocapsid protein. Tests that detect both IgG and IgM at the same time are superior to the ones testing for only one antibody. One study (Annex 2, ART108) even claimed to measure total antibodies (TotAb) using a commercially available assay and reached an even better sensitivity. Overall, ELISA-type tests seem to provide better results than flow-strip tests. Although the latter are faster and potentially suitable for point-of-care, the former seem to be more sensitive and reliable.

Most of the early studies reported high sensitivity and specificity for the method, whereas the only EU study found (Annex 2, ART75) reported disappointing results when using the tests on the hospital floor. This may be related to the rather late antibody response (see Section 1) making the method less suitable for the triage of patients. In addition, earlier publications are often vague on the device they used and often do not mention how long after symptom onset the samples were taken.

Application aspects

In contrast to RT-PCR methods, the vast majority of antibody test methods do not require complex laboratory equipment and various sets of reagents. Only a few of them require large instruments for reading the results. Therefore, they are often suitable for applications outside a laboratory.

The reported time to perform a test has been compiled in the annexed table. In the majority of specifications for the analysed devices, there is a claim of being "quick" and the time of execution to obtain the final qualitative result (test positive/test negative) was reported to be between 8 and 20 minutes.

The main difference among the rapid tests is the format: either a dipstick enclosed in a cassette or a simple strip or deep strips, fingertips and cards that have to be either inserted in a portable reader or in a tube with reagents to reveal the results. These all appear to be practical formats, easy to use in terms of portability, size, and practicability.

The assay protocols generally include a few steps: sample taking, sample reaction, result visualization and interpretation. Usually a drop of sample is taken (e.g. whole blood and serum/plasma) and is brought to the reacting strip. After some minutes, the qualitative result is visualized and interpreted (positive/negative).

In the minority of evaluated devices, the time range was up to 120 minutes because the test (the classical ELISA) procedure was more demanding, meaning that it required additional laboratory resources (e.g. anticoagulants, non-portable reader, plate incubator, trained personnel).

For many of the searched devices the time of operation was not explicitly reported or found. However, it was claimed that they are "quick".

In all cases, the cost was only available upon request for the devices.

Quality Control

For the majority of devices an internal control was included. The control consisted of a built-in line to monitor procedural mistakes and reagent defects. In the case of the classical ELISA (up to 120 min), a positive and a negative control are included.

Analytical performance

Limit of Detection (LOD)

Only very limited information was available on the LOD of the analysed devices. It should be noted that almost all of them provided a qualitative (positive/negative for presence/absence) result and do not quantify the antibody amount further.

Specificity

Cross-reactivity was not reported for the majority of the analysed devices.

Efficiency, robustness, precision

No information on the efficiency, robustness and precision as defined in Section 4.2.1B was found.

Clinical performance

No information was found in the sources searched about the results of clinical performance studies. When available, claims about both the diagnostic sensitivity and specificity were variable in the overall range of 81-98%.

The annexed table lists the available information.

B) Proposed performance criteria

The analytical and clinical performance criteria are described below. In addition to this, guidance on descriptive information, quality control and safety measures is given.

Descriptive information

Essential requirements are laid down in Annex I of the IVDD. More specifically, for the testing of human antibodies against SARS-CoV-2 or SARS-CoV-2 specific antigen respectively, the following descriptive information is of particular importance:

- Test type: e.g. manual heterogeneous immunoassay, automated immunoassay, rapid test, point-of-care test (POCT);
- nature of the test result: quantitative, semi-quantitative or qualitative ;
- measured target: SARS-CoV-2 specific antigen or human antibodies against SARS-CoV-2. In case of antibodies the immunoglobulin class should be specified, i.e. IgG, IgM or IgA;
- individuals intended to be tested: e.g. patients with suspected SARS-CoV-2 infection, individuals who have been vaccinated, general population;
- type of specimen that can be tested: e.g. oropharyngeal swab sample, whole blood, serum or plasma (EDTA or Heparin), its pretreatment (e.g. required dilution), stability and storage ;
- required qualification level of the staff needed, indication whether any particular training is required;
- guidance on the interpretation of results: e.g. cut off, grey zone, result not being suitable as a sole basis of diagnosis, further testing needed to obtain a reliable result, testing on follow-up samples taken after a recommended time period;
- potential limitations: e.g. possible reasons for false negative or false positive results, known cross-reactions.

Moreover, a short description of the method principle should be provided including the following information:

- Solid support and vessel where the immunoassay takes place (e.g. microplate strips with pre-coated wells, test cassette);
- detection principle (e.g. enzyme-linked colorimetry, fluorescence, colloidal gold);
- if necessary, the reading interval time, i.e. first time point when a reliable result can be read until the time point beyond which the read result is no longer reliable.

A complete list of reagents, either provided (in case of devices) or needed (in case of test methods and devices), should be described as well as the necessary consumables and equipment. The composition of this list depends on the test type:

- For manual immunoassays:
 - Reagent kits including calibrators and positive and negative controls
 - Plate type
 - ELISA washers and readers
 - Common laboratory equipment like tips, pipettes, tubes...
- For automated tests:
 - Automated platform or instrument (including software)
 - Reagent kits
 - 96 well plates and/or dedicated sample cups
- For rapid tests:
 - Test cassettes
 - Reagents
 - Detection system (if applicable)

For reagents, the expiration date and instructions concerning their storage and handling (e.g. reconstitution of lyophilised material and reagents) should be stated.

Quality control

The design of the quality controls depends on the test format and the type of test results:

- For manual and automated immunoassays tests:

Positive and negative controls should be analysed in parallel with the specimens in each run. In case of quantitative or semi-quantitative immunoassays the measurement results obtained for the controls shall fall within predetermined limits. In case of devices, these controls should be included in the device or clear information about required controls should be provided.

- For rapid tests and POCT:

These tests should include a migration control line. A test result is only valid, if the control line is visible. In addition, positive and negative controls should be tested under specific circumstances (e.g. a new lot of tests, a new operator, a new test environment).

Safety measures

It is of the utmost importance that proper biosafety guidelines are followed by clinical laboratories when handling samples from suspected COVID-19 patients. The laboratory biosafety guidance related to the coronavirus disease issued by the World Health Organization should be taken into account where applicable.⁹

Analytical performance

Limit of detection (LOD)

The LOD can be determined by testing a dilution series of one or more samples with a known amount of the measured target. Ideally, certified reference materials should be used. As long as these are not available for the SARS-CoV-2 antigen and SARS-CoV-2 antibodies, respectively, in-house developed control materials should be used to allow a consistent LOD determination over various batches of the devices or reagents.

Tests that provide qualitative test results do not often use a numerical value for their assay cut-off. In the absence of a suitable reference material it might not be feasible to estimate the concentration of the measured target at the assay's cut-off. Nevertheless, the way in which the cut-off was selected to obtain a reliable differentiation between positive and negative specimens should be described.

(Analytical) Selectivity

Cross-reactivity refers to the potential of false positive results due to present (for antigen and antibody tests) or past (for antibody tests only) infections or vaccinations (for antibody tests only) that are not linked to the SARS-CoV-2 virus.

The effect of the following infections or vaccinations should be evaluated:

- Infections with the common human pathogenic coronaviruses like HCoV-HKU1, -NL63, -OC43, or -229E;
- infections with influenza viruses and other respiratory viruses;

- vaccination against influenza viruses;
- acute bacterial pneumonia.

Moreover, the effect of past infections with the closely related virus strains SARS-CoV (-1) and MERS-CoV could also be investigated.

The potential for wrong testing results (both false negative and false positive) arising from interferences from at least the substances/conditions listed below should be investigated:

- Samples with autoantibodies such as rheumatoid factor and anti-nuclear antibodies (ANA);
- samples with elevated IgG and IgA levels;
- samples from pregnant women, especially multipara (women who had more than one pregnancy);
- samples with high concentrations of haemoglobin (haemolytic), triglycerides (lipaemic) and bilirubin (icteric);
- samples with human antibodies against components of the expression system used to produce the antigens or antibodies present in the reagents of the immunoassay;
- samples of individuals treated with relevant medicines like
 - antiviral and antibacterial drugs,
 - common anti-inflammatory drugs (acetylsalicylic acid, paracetamol, ibuprofen)
 - common anti-hypertensive drugs,
 - common anti-diabetic drugs,
 - drugs currently used against COVID-19 in clinical studies (e.g. hydroxychloroquine)

Robustness

Robustness refers to the capacity of an immunoassay to remain unaffected by small variations in the test parameters. Therefore, the effects of the following parameters should be evaluated:

- incubation time;
- temperature.

Clear limits should be set for those parameters that have a significant impact on the outcome of the testing results.

Precision

Both repeatability (i.e. testing the same sample under the same conditions) and reproducibility (i.e. testing the same sample under variable conditions such as different reagent kits, days, different analysts, or different instruments) should be assessed.

For the presence/absence testing these precision parameters can be expressed as the percentage of agreement.

Clinical performance

The analytical performance criteria are usually evaluated on a number of well-defined laboratory samples and extreme patient samples. The next step in the validation process should be a clinical performance study which mimics as much as possible a real-life situation. In this study a number of both known positive and negative samples are tested and the following parameters should be determined:

- Diagnostic sensitivity, i.e. the percentage of the true positive samples that gave a positive result with the antibody test;
- Diagnostic specificity, i.e. the percentage of the true negative samples that gave a negative result with the antibody test.

The determination of the diagnostic sensitivity and specificity requires well-characterized samples, i.e. true negative and true positive samples. This is challenging in the specific case of antibodies against the SARS-CoV-2 virus, as there is no reference antibody test available at the time of writing this report. Currently, it is recommended to evaluate the diagnostic sensitivity in a group of individuals with a present infection (several days after onset of COVID-19 symptoms) or a past infection with the SARS-CoV-2 virus. Both the present and the past infection should have been proven by a positive result with a reference RT-PCR method during the infection. The diagnostic specificity of antibody tests would require the availability of specimens from individuals that had never been in contact with the SARS-CoV-2 virus. Considering the worldwide spread of the virus it is recommended to use specimens which were collected before November 2019.

Once a reference anti-SARS-CoV-2 antibody immunoassay would be established, the next step in the validation process should be the clinical performance evaluation of the diagnostic accuracy of the antibody test under assessment for the two distinct intended uses: (a) SARS-CoV-2 infection detection/diagnosis and (b) determination of specific antiviral immunity.

(a) SARS-CoV-2 antibody tests used for indirect detection of a SARS-CoV-2 infection:

The determination of diagnostic accuracy should be performed in clinical studies using head-to-head comparison between results from one or more antibody tests under assessment and those of the reference RT-PCR test in the intended to be tested target population.

The diagnostic accuracy is composed of:

- Diagnostic sensitivity: Proportion of those individuals with the target condition (infected individuals with reference SARS-CoV-2 RNA true positive specimen sampled at least several days after onset of the COVID-19 symptoms) who test positive with the antibody test;
- Diagnostic specificity: Proportion of those individuals without the target condition (infection-free individuals with reference SARS-CoV-2 RNA true negative specimen and without any history of SARS-CoV-2 infection) who test negative with the antibody test.

For SARS-CoV-2 antibody tests intended for indirect detection of a SARS-CoV-2 infection the diagnostic sensitivity is the most crucial parameter, as the false negative rate should be as low as possible.

(b) SARS-CoV-2 antibody tests used for determination of the immune status against SARS-CoV-2:

The determination of diagnostic accuracy of antibody tests of interest should be performed in clinical studies using head-to-head comparison between results from one or more antibody tests under assessment and those of the reference antibody test in the intended to be tested target population. The diagnostic accuracy is composed of:

- Diagnostic sensitivity: Proportion of those individuals with the target condition (immune individuals with reference SARS-CoV-2 antibody test true positive specimen) who test positive with the antibody test;
- Diagnostic specificity: Proportion of those individuals without the target condition (non-immune individuals with reference SARS-CoV-2 antibody test true negative specimen) who test negative with the antibody test.

For SARS-CoV-2 antibody tests intended for the determination of the immune status against SARS-CoV-2 the diagnostic specificity is the most crucial parameter, as the false positive rate should be as low as possible.

The number of patients (sample size) to be included in the study should be determined by statistical power calculation for a desired precision level of the accuracy estimates. The study report should provide the estimates of diagnostic sensitivity and specificity with confidence intervals. The STARD 2015 (STAndards for Reporting Diagnostic accuracy studies) should be followed²⁶. It is recommended to report the 95 % confidence interval for the estimates of both the diagnostic sensitivity and the diagnostic specificity.

5. Conclusions and recommendations

Literature review of RT-PCR tests

Performance parameter information summarised in this report was self-reported by the manufacturer or distributor of the device, with no access to details or raw data for the studies that created these quality parameters.

Unless and until the results of independent validation studies are made available for some of the other devices in the annexed table, we would recommend the use of the ones that explicitly declare the implementation of a WHO protocol, in view of the availability of information on the previous use of methods based on the same sets of primers and probes.

Literature review of antigen tests

From the information that was retrievable from the literature and other sources, current antigen tests are so far not accompanied by sufficient proof of evidence regarding their performance characteristics.

Literature review of antibody tests

Currently, a comparison of the available antibody tests (devices) is not possible, because there is an almost complete lack of proper validation and standardisation among antibody targeting methods. Reported performance parameters are difficult to compare. Even in the literature analysed it is, for example, not always mentioned how long after infection the samples were taken, and ‘true positive’ and ‘true negative’ samples are often chosen in a different manner. Information on performance parameters is usually self-reported by the manufacturer or distributor of the device, with no access to details or raw data for the studies that produced these quality parameters.

It was found in the analysis of close to 100 immunoassay devices that data on false negatives, false positives and cross-reactivity are practically never reported. Only a few of them mentioned ‘cross reactivity’ with other viruses that could be associated with false positive results.

General

- Currently a rapidly expanding number of SARS-CoV-2 RNA tests (mostly by RT-PCR) and tests for antibodies against the coronavirus (mostly immunoassays) are appearing in the literature and on the market. A small number of antigen tests are also available ;
- There is a clear mismatch between the currently existing or reported quality assurance information about the COVID-19 tests/devices and the performance criteria proposed above, which are based on the principles of good analytical (testing) practice and corresponding international standards such as ISO/IEC 17025 and ISO 15189 ;
- There is an urgent need to properly assess (i.e. validate) the performance of existing and emerging test methods targeting the viral RNA, SARS-CoV-2 as antigen or its antibodies (see also Annex 3);
- The most critical performance parameters for reliable diagnostic decisions are:

- for identifying if a person is infected with SARS-CoV-2: the diagnostic sensitivity of the RNA or antigen test, as false negative test results have to be avoided;
 - for identifying the persons who have developed an immune response against SARS-CoV-2: the diagnostic specificity of the antibody test, as false positive test results have to be avoided.
- For that purpose, as well as for the comparison of the performance of different tests, the required quality benchmarks, i.e. well-characterised reference (control) materials mimicking real patient samples, and reference test methods have to be inventoried, verified or established as soon as possible. In addition, more proficiency testing exercises should be organised allowing laboratories to demonstrate their COVID-19 testing competence.

Scientific terminology used

Limit of detection (LOD)

The LOD represents the lowest concentration of the measured target (number of virus RNA fragments, antigens from the SARS-CoV-2 virus or antibodies against the SARS-CoV-2 virus) at which approximately 95% of the replicate measurements on samples containing the target give a positive result.

Precision

The term precision refers to the closeness of agreement between independent test results obtained under stipulated conditions. Both repeatability (i.e. measuring the same sample under the same conditions) and reproducibility (i.e. measuring the same sample under variable conditions such as different reagent kits, days, different analysts, or different instruments) are important performance characteristics.

Selectivity

Selectivity refers to the extent to which the test can be used to determine particular analytes in mixtures or matrices without interferences from other components of similar behaviour.

(from: IUPAC in Pure Appl. Chem. 73, 1381-1386 (2001))

(analytical) Specificity

Capability of a measuring system, using a specified measurement procedure, to provide measurement results for one or more measurands which do not depend on each other nor on any other quantity in the system undergoing measurement.

(from EN ISO 18113-1:2011)

(analytical) Sensitivity

Quotient of the change in an indication of a measuring system and the corresponding change in a value of a quantity being measured.

(from ISO/IEC Guide 99:2007)

Robustness

Capacity of an analytical method to remain unaffected by small but deliberate variations in method parameters.

(from ISO 18158:2016)

Diagnostic sensitivity

Ability of an *in vitro* diagnostic examination procedure to identify the presence of a target marker associated with a particular disease or condition.

It is also defined as percent positivity in samples where the target marker is known to be present.

Diagnostic sensitivity is expressed as a percentage (number fraction multiplied by 100), calculated as $100 \times \frac{\text{number of true positive values (TP)}}{\text{sum of the number of true positive values (TP) plus the number of false negative values (FN)}}$, or $100 \times \frac{\text{TP}}{\text{TP} + \text{FN}}$. This calculation is based on a study design where only one sample is taken from each subject.

(from EN ISO 18113-1:2011)

Diagnostic specificity

Ability of an IVD examination procedure to recognise the absence of a target marker associated with a particular disease or condition.

It is also defined as percent negativity in samples where the target marker is known to be absent.

Diagnostic specificity is expressed as a percentage (number fraction multiplied by 100), calculated as $100 \times \frac{\text{number of true negative values (TN)}}{\text{sum of the number of true negative plus the number of false positive (FP) values}}$, or $100 \times \frac{\text{TN}}{\text{TN} + \text{FP}}$. This calculation is based on a study design where only one sample is taken from each subject.

(from EN ISO 18113-1:2011)

Please see also Commission Decision 2002/364/EC on common specifications for *in vitro* diagnostic devices, OJ L 131, 16.5.2002, p. 17.

Annexes

Annex 1: Commercial devices

Annex 2: Scientific literature

Annex 3: Search on validation studies