Ad-hoc group consensus recommendations on the evaluation and quality control of molecular and serological diagnostics tests for SARS CoV-2 human infection*

Jorge E. Gomez Marin1,*, Jaime Castellanos2, Alfonso J. Rodriguez-Morales3, Jaime A. Cardona-Ospina4, Jorge Eduardo Forero Duarte5, Salim Mattar6, Germán Esparza7

Ad-hoc Group (Colombian Association of Infectious Diseases and Colombian Association of Virology)
of the Colombian consensus recommendations for the management of COVID 19

Abstract
A consensus working group formulates recommendations on the criteria to evaluate the diagnostic performance (size and criteria of selection of samples to determine sensitivity, analytical specificity, criteria for limit of detection, criteria for gold standard to evaluate serological assays) that should considered during the evaluation and validation/verification of diagnostic tests for SARS CoV-2 infection. A national external quality control program should be established to guarantee the suitability and performance of these diagnostic serological and molecular tests during this pandemic, that will have deep implications on clinical and public health decisions.

Key words: laboratory diagnosis; SARS CoV-2; COVID 19; laboratory quality control

Consenso de grupo Ad-hoc sobre recomendaciones para evaluación y controles de calidad para diagnóstico molecular y serológico de infección humana por SARS CoV-2

Resumen
Se formulan recomendaciones de un grupo de consenso de expertos sobre los criterios para evaluar el desempeño diagnóstico (tamaño y selección de muestras para sensibilidad y especificidad analítica, criterios para establecer límites de detección, criterios para establecer el estándar de oro para las serologías) que deberían ser tenidos en cuenta al evaluar y validar las pruebas diagnósticas para SARS CoV-2. Con el propósito de asegurar la calidad de las pruebas serológicas a utilizar en el país, se recomienda la participación en un programa de control de calidad externo, que garantice la idoneidad y desempeño en la realización de las pruebas diagnósticas serológicas y moleculares durante esta pandemia, ya que su uso tiene profundas implicaciones para las medidas de intervención clínicas individuales y de seguimiento y control en salud pública.

Palabras clave: Pruebas diagnósticas; SARS CoV-2, COVID 19, control calidad laboratorio

1 Research Group on Child Population (IPI), San Juan de Dios University Hospital, Armenia, Quindío, Colombia
Study Group on Molecular Parasitology (GEPMOL), Center for Biomedical Research, University of Quindío, Armenia, Quindío, Colombia. https://orcid.org/0000-0001-6472-3329

2 Director of Virology Laboratory, El Bosque University, Bogotá - Colombia. https://orcid.org/0000-0002-7052-7938

3 Public Health and Infection Research, Group Faculty of Health Sciences, Technological University of Pereira, Pereira, Risaralda, Colombia.
Biomedicine Research Group, Faculty of Medicine, Autonomous University Foundation of the Americas, Headquarters Pereira, Pereira, Risaralda, Colombia.
Public Health and Infection Research, Group Faculty of Health Sciences, Technological University of Pereira, Pereira, Risaralda, Colombia. https://orcid.org/0000-0001-9773-2192

4 Public Health and Infection Research, Group Faculty of Health Sciences, Technological University of Pereira, Pereira, Risaralda, Colombia.
Biomedicine Research Group, Faculty of Medicine, Autonomous University Foundation of the Americas, Headquarters Pereira, Pereira, Risaralda, Colombia.
Public Health and Infection Research, Group Faculty of Health Sciences, Technological University of Pereira, Pereira, Risaralda, Colombia.
Emerging Infections and Tropical Medicine Research Group, Institute for Research in Biomedical Sciences - Sci-Help, Pereira, Risaralda, Colombia. https://orcid.org/0000-0003-3996-2293

5 School of Microbiology, Antioquia University Medellin, Colombia. https://orcid.org/0000-0001-9289-2314

6 University of Cordoba, Tropical Biological Research Institute, Monteria, Colombia. https://orcid.org/0000-0003-0526-4630

7 Quality assurance program, PROASECA L SAS Colombia. https://orcid.org/0000-0002-8395-6803
* Corresponding author
Correo electrónico: jegomez@uniquindio.edu.co

* English version of the document originally published in Spanish by the Colombian Association of Infectious Diseases and the Colombian Association of Virology

http://dx.doi.org/10.22354/in.v24i3.867
Introduction

SARS-CoV-2 pandemic is a new viral agent for humans thus, it has been necessary to develop emergency diagnostic tests\(^1\). The first to be developed were those that detected the genetic material of the virus in respiratory secretions samples, which became the gold standard for diagnosis\(^2\). The recommended diagnostic protocols can be found on World Health Organization (WHO) website\(^3\). The United States Food and Drug Administration (FDA) has released a series of non-binding policies to authorize emergency marketing of diagnostic products that serve as a guide for clinical, industrial and FDA laboratories\(^4\). The American Society for Microbiology (ASM) has also presented recommendations for the evaluation of diagnostic tests\(^5\). Faced with this situation where there is an urgent need for diagnostic tests, it is crucial to establish clear guidelines that define the gold standards for the sensitivity and specificity values of the diagnostic products that are being marketed in Colombia, and in this way reduce the possibility of scarce resources being invested in tests with suboptimal performance\(^6\). For this reason, the Colombian Association of Infectious Diseases (ACIN) and the Colombian Association of Virology (ACV) created a working group to give an expert opinion on the evaluation and quality control of diagnostic, serological and molecular tests. These diagnostic tests include both commercial kits and the primer sequence recommendations offered by some molecular test reagent package providers. The in-house protocol of the Charité Institute of Virology in Berlin is also discussed; this protocol was standardized by the National Institute of Health (INS) in Colombia. Being knowledgeable on the appropriate information has profound implications as public health measures, critical patient management decisions, the evolution of treatment cases and isolation behaviors, are based on diagnostic tests that must comply with the highest quality standards. This document is an informal experts’ recommendations consensus on what the external evaluation, validation and control criteria for diagnostic tests for SARS CoV-2 infection should be and it is intended for test producers, the laboratories that perform them, and for the evaluators of government regulatory agencies at the national level. Section 2 specifies the indications of the different types of tests available and Section 3 presents the evaluation criteria and external quality control recommendations. Considering that the SARS CoV-2 pandemic is dynamic, and that scientific evidence is being generated permanently, this document will be subject to updates and adjustments as new information appears.

Defining the scope of the tests

The purpose of the diagnostic tests for SARS-CoV-2 may be (Figure 1):

- Viral RNA detection
- Detection of viral nucleoprotein (antigen)
- Detection of specific IgM, IgG or IgA type antibodies or tests that simultaneously detect IgG / IgM antibodies
- Detection of viral nucleoprotein (antigen)
- Detection of specific IgM, IgG or IgA type antibodies or tests that simultaneously detect IgG / IgM antibodies
- Viral RNA detection

Each one has its own principles and detect different analytes (nucleic acids or virus proteins or antibodies produced by the host as part of its immune response), therefore making sensitivity and specificity comparisons between them would not be accurate. Likewise, its usefulness and application depend on specific clinical diagnosis contexts, population screening and occupational health.

Additionally, sensitivity and specificity properties may vary from one test to another, therefore, when recommendations based on antibody kinetics at different stages of infection are required, it is necessary to consider that kinetics are test dependent, some may have higher sensitivity at certain times, thus any inference regarding the emergence of antibodies at different times of the infection, must take into account the detection method.

The Figure 2 shows the sensitivity evolution curves and the 95% confidence intervals (95% CI) calculated from various studies\(^7\)-\(^10\) that have supported indications on behavior and recommendations for testing at different stages of infection\(^11\)-\(^13\). However, these studies have overly broad estimates of 95% CI. Only one of them presented confidence intervals that allow estimating the uncertainty of the recommendations. That is why it is critical to establish sufficient sample sizes to be able to define recommendations on what tests should be used and at what time should they be done.

Applications for detection of the SARS CoV-2 viral genome by reverse transcriptase polymerase chain reaction (RT-PCR)

The SARS-CoV-2 RNA screening test requires an RT-qPCR test and the goal is to:

a. Detect contacts of confirmed symptomatic or asymptomatic cases
b. Follow up on positive cases, evaluate or confirm suspicious cases and define isolation behaviors, or in the confirmation of initial negative cases.
c. Perform a differential diagnosis with other respiratory infections or detect coinfections using panels for multiple respiratory pathogens

Normally, this test should have 100% specificity, but its sensitivity varies depending on critical factors such as the design of the target genes to be amplified, the duration of follow-up of negative cases, and factors in sample processing, therefore which its negativity does not rule out an infection at the beginning\(^14\)-\(^16\).

Tests that detect viral antigen

There are immunochromatographic tests that detect virus proteins, their applications would be similar to those that detect RNA. However, it still has not been established when they are positive and would be the operational characteristics of these tests. They will not be discussed in this consensus as there is scarce information on them.
Applications of serological tests for detection of IgG antibodies or combined IgG/IgM tests

There are different methods for IgG detection such as immunochromatography, indirect immunofluorescence (IFI) and ELISA; the latter seems to be more sensitive than the others. The possible applications of these type of tests are:

a. Population screening to find out the percentage of population exposed and make decisions about partial or definitive quarantine or containment measures.
b. Select the population that can return to work by identifying those with positive antibody tests. In this case, it is necessary to wait for the results of an evaluation based on the evidence of the protection quality of antibodies. The possible applications of these type of tests are:

c. Seroprevalence studies to determine epidemiological variables of interest in public health, especially in studies of disease burden and costs, such as the attack rate and the expansion factor.

Applications of serological tests for detection of IgM antibodies

a. Detection of patients with recent infection
b. Classification of infection status in conjunction with measurement of specific IgG in acute or convalescent cases.

Criteria recommendations for test assessments

For RT-qPCR SARS CoV 2 testing

The group considers that FDA criteria can be taken as base recommendations with the following details:

- **Limit of Detection (LoD):** To establish this, one must inoculate artificial (synthesized) RNA or RNA from a biological specimen, quantified, in a patient sample (bronchoalveolar lavage or sputum) or in viral transport medium. It is recommended to perform serial 1:2 dilutions with three replicates per dilution and then confirm the final amplification with 20 replicates. LoD is the lowest concentration at which the 19/20 replicate is positive.
- **Analytical sensitivity:** *In silico* analysis of primer and probe sequences should be performed.
- **Cross reactivity (analytical specificity):** The list of organisms should be verified according to Table 1, which should be analyzed *in-silico* and if possible *in-vitro*.
- **Microbial interference analysis:** If the in-silico analysis shows homology ≥ 80% with other microorganisms, interference studies with probes should be carried out for the microorganism with which there is homology. A minimum of three replicates of a 3x inoculum of the LoD concentration of SARS-CoV-2 and with a high level of interference from the whole microorganism or nucleic acid must be made.

Figure 1. Clinical and laboratory characteristics of COVID 19 infection. Image kindly provided by Infectoweb (Bogotá).

Authors: Luis Pablo Lesport, Javier Africano, Álvaro A. Faccini-Martínez y Carlos Eduardo Pérez.
Table 1. List of recommended organisms for cross-reactivity analysis

<table>
<thead>
<tr>
<th>Pathogens of the same virus family</th>
<th>Organisms with high probability of circulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human coronavirus 229E</td>
<td>Adenovirus (e.g. C1 Ad. 71)</td>
</tr>
<tr>
<td>Human coronavirus OC43</td>
<td>Human metapneumovirus (hMPV)</td>
</tr>
<tr>
<td>Human coronavirus HKU1</td>
<td>Parainfluenza 1-4</td>
</tr>
<tr>
<td>Human coronavirus NL63</td>
<td>Influenza A and B</td>
</tr>
<tr>
<td>SARS-coronavirus (does not circulate in Colombia)</td>
<td>Enterovirus (example: EV68)</td>
</tr>
<tr>
<td>MERS-coronavirus (does not circulate in Colombia)</td>
<td>Respiratory syncytial virus</td>
</tr>
<tr>
<td>Rhinovirus</td>
<td>Chlamydia pneumoniae</td>
</tr>
<tr>
<td></td>
<td>Hemophilus influenzae</td>
</tr>
<tr>
<td></td>
<td>Legionella pneumophila</td>
</tr>
<tr>
<td></td>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td></td>
<td>Streptococcus pneumoniae</td>
</tr>
<tr>
<td></td>
<td>Streptococcus pyogenes</td>
</tr>
<tr>
<td></td>
<td>Bordetella pertussis</td>
</tr>
<tr>
<td></td>
<td>Mycoplasma pneumoniae</td>
</tr>
<tr>
<td></td>
<td>Pneumocystis jirovecii (PIP)</td>
</tr>
<tr>
<td></td>
<td>Pool of human nasal lavage representing the respiratory tract microbiota</td>
</tr>
<tr>
<td></td>
<td>Candida albicans</td>
</tr>
<tr>
<td></td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td></td>
<td>Staphylococcus epidermidis</td>
</tr>
<tr>
<td></td>
<td>Staphylococcus salivarius</td>
</tr>
</tbody>
</table>

* For in vitro assays, concentrations of 10⁶ CFU/mL or higher for bacteria and 10⁵ PFU/mL or higher for viruses are recommended.

- **Interference analysis with endogenous substances:** Boom method and column based extraction methods recommended.
- **Use of clinical samples:** It is recommended to test with at least 30 randomly selected reactive and 30 non-reactive clinical samples. Samples can be manufactured using clean clinical matrices artificially contaminated with synthetic RNA. LoD results should be reported with serial dilutions in base 2.
- **Viral RNA extraction kits:** Obtaining quality viral RNA is a critical step in the qRT-PCR test. There are different extraction methods that can be manual or automated, with magnetic beads, different types of columns or with the addition of lysis buffer that releases the nucleic acid. Whatever the extraction method, it is recommended that they be evaluated and optimized for the extraction of SARS-CoV-2 in different types of samples.

**Recommendations for the evaluation of serological tests for IgG, IgM or IgA**

Antibody tests seek to detect the humoral immune response of patients whose titer increases as the infection progresses (9,17). They offer the possibility of detecting active disease of several days of evolution, but they do not exclude the possibility of continuing the transmission of the virus (17,22). The data on the course of the infection has shown that the antibodies begin to be produced on day 6 of the onset of symptoms and that a decrease in viral load occurs simultaneously (9,23,24). Although the gold standard test should be carried out with plaque reduction by neutralization (PRNT) tests, it requires the isolated SARS CoV-2 virus and a viral culture. Given the limitation to cultivate this virus, this group considers that an alternative gold standard would be to perform validation with the following groups of sera and estimating sufficient sample sizes to establish the applicability of the test under different conditions:

- **Serum selection criteria to establish test sensitivity:** All tests must have a sensitivity evaluation using serum samples, in the same proportion of symptomatic and asymptomatic cases and age and gender ranges, similar to what is found in the general population. The sera should correspond to people with confirmed RT-qPCR infection and the post-exposure days should be specified, according to the following groups: day 1 to 8, day 9 to 15 and day 16 to 30, after estimated contact with positive cases for SARS-CoV-2.

- **Serum selection criteria to establish specificity:** All tests must establish their specificity with sera from the same population prior to the report of initiation of cases in the country. The sera must correspond to the characteristics of the population in which the tests will be applied.

- **Sample size to establish sensitivity and specificity:** The tests should be performed with a number of true positive and true negative samples estimated from a unilateral test, for which it would be necessary to estimate the probable false positives in that test and the expected prevalence in the sample structured by the researcher. A value of positive likelihood ratios (LR) should be chosen considering the level of alpha and beta errors considered appropriate. The sensitivity and specificity reports must include the 95% CI values of each one of them.

**Quality control programs for diagnostic tests for SARS CoV-2 infection**

**Quality control program for the detection of SARS CoV-2 by RT-PCR test detection of SARS-CoV-2**

Internal and external quality controls must be carried out by the national reference laboratory in the institutions that are performing home RT-PCR tests. It should be done at least every 200 tests with a panel of samples artificially inoculated with control plasmid containing the product insert to be amplified.

Commercial matrices for external quality control could be used once they are validated and available.
Quality control program for serological tests
All serological tests must be carried out with internal positive and negative controls in each assembly. External quality controls must be carried out with positive and negative samples sent to each laboratory that performs serological diagnostic tests.

External control can be done monthly if there is a high volume, bi-monthly or quarterly if the volume of samples is low.

Conclusions
The diagnosis of SARS CoV-2 infection is the cornerstone for making appropriate decisions for the management and control of the pandemic. The viral RNA amplification tests (RT-qPCR) and antibodies detection against the virus have different scopes and their implementation and interpretation should be adjusted to the clinical or epidemiological context. Validation, verification and quality control of in-house tests and commercial diagnostic tests are mandatory to guarantee reliable and timely results. For molecular amplification tests, the FDA recommendations must be followed, and for serological tests, validations will be performed with sera that meet criteria for sensitivity, specificity, and an adequate sample size that allows the performance of the test to be properly calculated.

The WHO on April 8, 2020 made a position statement, by its group of experts, in light of the evidence available at the time, on rapid and serological tests, including ELISA, and established that for now these tests should only be used for research purposes. The statement reads “These tests should not be used for clinical decisions, until there is evidence available for specific indications”28. On April 13, the Colombian Ministry of Health issued guidelines on the use of diagnostic tests in Colombia following the line of the WHO, recommending that serological tests should only be for epidemiological studies and not for clinical decisions making29. Thus, the consensus recommendations seek to provide guidance to obtain the information required and high-quality clinical diagnostic tests.

References

Figure 2. 95% confidence intervals were calculated from the sensitivity values of four studies (7–10) that have supported most descriptions of antibody kinetics. Only the study by Zhao et al (a) calculated the confidence intervals, the other values (b, c, d) were calculated by the members of this consensus based on the data presented by the authors. It is observed that the confidence values are very wide and do not allow to define with certainty, for example, if the detection of IgM or IgG is more sensitive during the first 8 days of infection.