MDCG 2021-21 Rev.1

Guidance on performance evaluation of SARS-CoV-2 *in vitro* diagnostic medical devices

Revision 1 – February 2022

This document has been endorsed by the Medical Device Coordination Group (MDCG) established by Article 103 of Regulation (EU) 2017/745. The MDCG is composed of representatives of all Member States and it is chaired by a representative of the European Commission.

The document is not a European Commission document and it cannot be regarded as reflecting the official position of the European Commission. Any views expressed in this document are not legally binding and only the Court of Justice of the European Union can give binding interpretations of Union law.

MDCG 2021-21 Revision 1 changes	
Tables 1 and 2	Footnote on vaccinated individuals revised
Tables 4 and 5	Footnote on specimen types added
Tables 6 and 7	1st column revised, title 3rd column revised
All tables	Minor editorial clarifications

Introduction

This guidance document concerns performance evaluation of SARS-CoV-2 *in vitro* diagnostic medical devices (IVDs) in the context of conformity assessment under either Directive 98/79/EC or Regulation (EU) 2017/746. It covers devices for detection or quantification of SARS-CoV-2 nucleic acid, antigens and also detection or quantification of antibodies against SARS-CoV-2. These devices are collectively referred to as SARS-CoV-2 IVDs. The guidance is addressed to all interested parties, including notably the manufacturers, as well as notified bodies and competent authorities, authorised representatives, other market operators, professional and patient associations.

The content of this guidance document is envisaged to form the basis for common specifications to be adopted according to Article 9 of Regulation (EU) 2017/746 in the coming months. The content may be adapted to take account of changing circumstances and increasing scientific and technical knowledge, as the COVID-19 pandemic continues to evolve.

The terms “IVD”, “device”, “assay” and “test” are used interchangeably in this text.

General considerations

The general principles in this section should be taken into account for the performance evaluation of SARS-CoV-2 IVDs.

The following terms are being used in this guidance document:

- diagnostic sensitivity means the ability of a device to identify the presence of a target marker associated with SARS-CoV-2;
- true positive means a specimen known to be positive for the target marker and correctly classified by the device;
- false negative means a specimen known to be positive for the target marker and misclassified by the device;
- diagnostic specificity means the ability of a device to recognise the absence of a target marker associated with SARS CoV-2;
- false positive means a specimen known to be negative for the target marker and misclassified by the device;
- true negative means a specimen known to be negative for the target marker and correctly classified by the device;
- the limit of detection (LOD) means the smallest amount of the target marker that can be precisely detected, the LOD is part of analytical sensitivity of the device;
- analytical specificity means the ability of the method to determine solely the target marker;

- nucleic acid amplification techniques (NAT) – methods of detection and/or quantification of nucleic acids by either amplification of a target sequence, by amplification of a signal or by hybridisation;
- rapid tests means qualitative or semi-quantitative in vitro diagnostic medical devices, used singly or in a small series, which involve non-automated procedures and have been designed to give a fast result;
- robustness of an analytical procedure means the capacity of an analytical procedure to remain unaffected by small but deliberate variations in method parameters and provides an indication of its reliability during normal usage;
- cross-reactivity (or cross-reaction) means the ability of non-target analytes or markers to cause false-positive results in an assay because of similarity, e.g. the ability of non-specific antibodies binding to a test antigen of an antibody assay, or the ability of non-target nucleic acids to be reactive in a NAT assay;
- interference means the ability of unrelated substances to affect the results in an assay;
- whole system failure rate means the frequency of failures when the entire process is performed as prescribed by the manufacturer;
- first line assay means a device used to detect a marker or analyte, and which may be followed by a confirmatory assay. Devices intended solely to be used to monitor a previously determined marker or analyte are not considered first line assays;
- confirmatory assay means a device used for the confirmation of a reactive result from a first line assay;
- supplemental assay means a device that is used to provide further information for the interpretation of the test result of another assay;
- virus typing assay means a device used for typing with already known positive samples, not used for primary diagnosis of infection or for screening;
- 95% positive cut-off value for NAT assays means the analyte concentration where 95% of test runs give positive results following serial dilutions of an international reference material, where available, e.g. a World Health Organisation (WHO) International Standard or reference material calibrated against the WHO International Standard; this value describes the limit of detection (LOD) for NAT devices.

Overall considerations

Performance evaluations of SARS-CoV-2 IVDs should be carried out in direct comparison with a state-of-the-art device. The device used for comparison should be one bearing CE marking, if on the market at the time of the performance evaluation. For anti-SARS-CoV-2 tests, the new device should have an overall performance at least equivalent to that of the state of the art device of the same type, e.g. considering claims based on target antigens used and immunoglobulin classes detected.

Devices used for determination of status of samples used in performance evaluations of SARS-CoV-2 IVDs should be state-of-the-art devices bearing CE marking.

Performance evaluations of SARS-CoV-2 IVDs should be performed on a population equivalent to the European population.

If discrepant results are identified as part of a performance evaluation, these results should be resolved as far as possible, by one or more of the following: evaluation of the

discrepant sample in further devices; use of an alternative method or marker; a review of the clinical status and diagnosis of the patient; testing of follow-up samples.

As part of the required risk analysis the whole system failure rate leading to false-negative results should be determined in repeat assays on low-positive specimens.

Sensitivity and specificity

Positive specimens used in the performance evaluation should be selected to reflect different stages of the respective disease(s), different antibody patterns, different genotypes, different subtypes, mutants, *etc*.

For SARS-CoV-2 IVDs intended by the manufacturer to be used with serum or plasma, positive specimens should include 25 positive 'same day' fresh serum samples (≤ 1 day after sampling).

Seroconversion panels should start with a negative bleed(s) and should reflect narrow bleeding intervals as far as possible. Where this is not possible, manufacturers should provide a justification in the performance evaluation report.

Negative specimens used in a performance evaluation should be defined so as to reflect the target population for which the device is intended, such as blood donors, hospitalised patients, pregnant women, *etc*.

Specificity should be calculated using the frequency of repeatedly reactive (i.e. false positive) results in individuals negative for the target marker.

For SARS-CoV-2 IVDs intended by the manufacturer to be used with serum and plasma, the performance evaluation should demonstrate serum to plasma equivalency. This should be demonstrated for at least 25 positive donations for sensitivity and 25 negative donations for specificity.

Anti-SARS-CoV-2 IVDs intended by the manufacturer for testing body fluids other than serum or plasma, e.g. urine, saliva, etc., should meet the same requirements for sensitivity and specificity as serum or plasma devices. The performance evaluation should test samples from the same individuals in both the devices to be approved and in a respective serum or plasma device.

In the case of IVDs for SARS-CoV-2 detection from secretions of the respiratory tract, their performance on all claimed specimen types should be compared to NAT tests on nasopharyngeal swabs.

Interference and cross-reactivity

The manufacturer should select the potential interfering substances to be evaluated taking account of the composition of the reagents and configuration of the device. The manufacturer should include specimens such as, where applicable: those representing related infections; those from multipara, i.e. women who have had more than one pregnancy, or rheumatoid factor (RF) positive patients; those containing human antibodies to components of the expression system, for example anti-*E. coli*, or anti-yeast.

Anticoagulants

For SARS-CoV-2 IVDs intended for use with plasma, the performance evaluation should verify the performance of the device using all anticoagulants which the manufacturer indicates for use with the device. This should be demonstrated for at least 50 plasma specimens per anticoagulant (25 positive and 25 negative).

Batch testing

For SARS-CoV-2 antigen and antibody tests, the manufacturer's batch testing criteria should ensure that every batch consistently identifies the relevant antigens, epitopes, and antibodies and is suitable for the claimed specimen types.

Self-tests

SARS-CoV-2 IVDs for self-testing should meet the same requirements for sensitivity and specificity as respective devices for professional use. Relevant parts of the performance evaluation should be carried out (or repeated) by appropriate lay persons to validate the operation of the device and the instructions for use. The lay persons selected for the performance evaluation should be representative of the intended users groups.

Specific considerations

The following tables set out specific considerations for various types of SARS-CoV-2 IVDs.
Table 1 refers to the following first-line assays (including rapid tests) for antibodies against SARS-CoV-2 (anti-SARS-CoV-2): IgG-only, IgG combined with IgM and/or IgA, and total antibody.
Table 2 refers to assays for detection of anti-SARS-CoV-2 IgM and/or IgA (including rapid tests).
Table 3 refers to confirmatory or supplementary assays for anti-SARS-CoV-2.
Table 4 refers to antigen SARS-CoV-2 tests, including rapid antigen tests.
Table 5 refers to nucleic acid amplification techniques (NAT) assays for SARS-CoV-2 RNA.
Tables 6 and 7 refer to additional requirements for SARS-CoV-2 antigen and antibody self-tests respectively. They are intended for devices which have already undergone a performance evaluation for professional use.

Table 1: **First-line assays (including rapid tests) for anti-SARS-CoV-2: total antibody, IgG-only, IgG combined [1]with IgM and/or IgA**

Parameter	**Specimen**	**Anti-SARS-CoV-2 IgG, IgG combined, and total Ab**	**Acceptance criteria**
Diagnostic sensitivity	Positive specimens	≥400 including samples from early infection and post seroconversion[2] (within the first 21 days and after 21 days following the onset of symptoms); including samples from asymptomatic or subclinical and mildly symptomatic (outpatient treatment) individuals; including samples with low and high titers; including samples from vaccinated individuals if appropriate[3]; consideration of genetic variants	≥90% sensitivity[4] for samples taken >21 days after onset of symptoms[5]; overall sensitivity including the early infection phase should be comparable to other CE-marked[6] tests
	Seroconversion panels	As far as available	Seroconversion sensitivity comparable to other CE-marked tests
Analytical sensitivity	Reference preparations	WHO International Standard (IS) for anti- SARS-CoV-2 (NIBSC code 20/136); WHO International Reference Panel (RP) for anti-SARS-CoV-2 antibodies (NIBSC codes 20/140, 20/142, 20/144, 20/148, 20/150)	IS: for titre determinations / quantitative[7] result output; RP: all antibody assays
Specificity	Negative specimens[8]	≥400 samples from non-infected and non-vaccinated individuals[9]	>99% specificity[10]
		≥200 hospitalised patients (without SARS-CoV-2 infection)	Potential limitations for specificity should be determined
		≥100 in total potentially interfering (e.g. rheumatoid factor, pregnant women, etc.) and cross-reacting blood specimens: including antibodies against endemic human coronaviruses 229E, OC43, NL63, HKU1 and other pathogens of respiratory diseases such as influenza A, B, RSV etc.	

[1] Performance claim of the combined overall result; separate claims for IgM and/or IgA see table below.
[2] Details on the time interval between sampling and onset of symptoms (or time of infection, if available) should be provided.
[3] The manufacturer should provide a justification of the suitability and timing for sensitivity evaluation of the relevant antibodies in vaccinated individuals.
[4] Based on confirmed positive SARS-CoV-2-NAT result.
[5] Because sensitivity may vary or decrease over time, claims for sensitivity shall be specified in relation to the time between sampling after symptom onset or on the initial PCR diagnosis and the test.
[6] CE-marked under Regulation (EU) 2017/746 as class D. During the transition phase, reference is made to EU and ECDC SOTA guidance and current scientific literature.
[7] Quantitative assays if they are also first-line assays.
[8] Negative specimens should be from individuals with no history of SARS-CoV-2 infection (if available pre-pandemic).
[9] Individuals vaccinated with an antigen different from that used in the respective test may be included, if appropriate.
[10] False-positive results should be resolved by retesting in other SARS-CoV-2 serologic assays, if necessary with different test design and antigen coating than the initial test, and/or confirmatory testing.

Table 2: **Assays (including rapid tests) for anti-SARS-CoV-2: IgM and/or IgA detection**

Parameter	**Specimen**	**Anti-SARS-CoV-2 IgM and IgA**	**Acceptance criteria**
Diagnostic sensitivity	Positive specimens	≥200[1] samples with a significant proportion from the early phase of the infection (within 21 days after onset of symptoms) compared to samples past seroconversion (>21 days after onset of symptoms); including samples from asymptomatic, subclinical, mildly symptomatic (outpatient treatment) individuals; including freshly[2]vaccinated individuals if appropriate; consideration of genetic variants	≥80% sensitivity for samples taken during the first 21 days after symptom onset; overall sensitivity should be comparable to other CE-marked tests of the same type (i.e. IgM and/or IgA)
	Seroconversion panels	As far as available	Seroconversion sensitivity comparable to other CE-marked tests
Analytical sensitivity	Standards	N/A	N/A

Specificity	Negative specimens	≥200 samples from non-infected and non-vaccinated individuals	≥98% specificity[3]
		≥100 from hospitalised patients (without SARS-CoV-2 infection)	Potential limitations for specificity should be determined
		≥100 in total potentially interfering (e.g. rheumatoid factor, pregnant women, etc.) and cross-reacting blood specimens; antibodies against endemic human coronaviruses 229E, OC43, NL63, HKU1 and other pathogens of respiratory diseases such as influenza A, B, RSV etc.	

Where applicable, the footnotes of Table 1 should apply.
[1] In case of combination tests, 200 per marker IgM and IgA.
[2] The manufacturer should provide a justification of the suitability and timing for sensitivity evaluation of IgM and IgA in vaccinated individuals.
[3] Clarification of false-positive results may additionally include testing for presence of other anti-SARS-CoV-2 antibody types (IgA, IgG, total antibody).

Table 3: **Confirmatory or supplemental[1]assays for anti-SARS-CoV-2**

Parameter	**Specimen**	**Anti-SARS-CoV-2**	**Acceptance criteria**
Diagnostic sensitivity	Positive specimens	≥200 including samples pre and post seroconversion (within the first 21 days and after 21 days following the onset of symptoms)	Correct determination as “positive” (or “indeterminate”)
	Seroconversion panels/ low titre panels	as far as available	

Analytical sensitivity	Standards	N/A	N/A
Diagnostic specificity	Negative specimens[2]	≥200 from non-infected / non-vaccinated population	No false-positive results; correct determination as "negative" (or "indeterminate")
		≥200 from hospitalised patients (without SARS-CoV-2 infection) ≥50 potentially interfering and cross-reacting samples in total: antibodies against endemic human coronaviruses 229E, OC43, NL63, HKU1 and other pathogens of respiratory diseases such as influenza A, B, RSV etc.; including samples with indeterminate or false-positive results in other anti-SARS-CoV-2 assays	

[1] E.g. immunoblot providing antigens different from those used in the initial antibody test.
[2] Negative specimens should be from individuals with no history of SARS-CoV-2 infection (if available pre-pandemic).

Table 4: **Antigen assays (including rapid tests) for SARS-CoV-2**

Parameter	**Specimen**	**SARS-CoV-2 antigen**	**Acceptance criteria**
Diagnostic sensitivity	Positive specimens	≥100[1] NAT positive samples[2] from early infection within the first 7 days after symptom onset[3]; samples should represent naturally occurring viral loads[4]; consideration of genetic variants[5] consideration of variations in specimen collection and/or specimen handling[6]	Detection of >80% (rapid tests); detection of >85% (lab-based assays[7]); relative to SARS-CoV-2-NAT[8,9]

Analytical sensitivity	Standards	As soon as available	Establishment of a limit of detection[10]
Diagnostic specificity	Negative specimens	≥300 from non-infected individuals	Specificity >98% (rapid tests) Specificity >99% (lab-based[7])
		≥100 from hospitalised patients ≥50 potentially interfering and cross-reactive samples in total: including virus-positive samples of endemic human coronaviruses 229E, OC43, NL63, HKU1; influenza A, B, RSV, and other pathogens of respiratory diseases, eligible for differential diagnosis; including bacteria[11] present in the sampling area	Potential limitations for specificity should be determined

[1] If the device is intended to be used for more than one specimen type, 100 samples shall be required for each specimen type. If this is not possible in exceptional circumstances (e.g. if specimen collection is very invasive), the manufacturer shall provide a justification and evidence of matrix equivalence.
[2] Sampling should be matched for antigen and NAT testing, e.g., two simultaneous samples from each individual or optimally NAT- and antigen testing from the same sample (e.g. from the eluate of one swab); the buffer/transport medium should be compatible for both NAT and antigen testing; any volume change in the buffer/medium for sample uptake different from that of the proprietary assay, and/or between antigen and NAT test should be clearly communicated.
[3] Or time of infection, if known, taking into account the incubation time.
[4] I.e., without preselection; the viral loads and their distribution should be shown, e.g. characterized by Ct-values of RT-PCR; or transformed into viral load per ml or sample, if applicable.
[5] Depending on the design of the device and nature of the genetic variant. For the purpose of evaluation, at least 3 samples should be represented for each genetic variant.
[6] Specimen collection and extraction items such as swabs, extraction buffers, etc., should be part of the evaluation. If proprietary sampling/sample preparation is not included in the test kit, test performance should be investigated for an applicable range of sampling devices. If the sample is not tested immediately, e.g. after a certain transport time, stability of the antigen should be investigated.
[7] Other than rapid tests, i.e. formal laboratory-based assays e.g. enzyme immunoassay, automated tests, etc.
[8] The sensitivity of ≥80%, ≥85% respectively, should be for all specimen types claimed. All claimed specimen types should be compared with paired NAT results from nasopharyngeal specimens.
[9] The relationship between antigen test performance and NAT should be demonstrated; sensitivity may be shown relating to different viral load ranges and to the threshold of infectivity. The NAT and extraction method used should be described.

[10] Unless there is an available international standard, analytical sensitivity may be tested by dilution series of in-house virus preparations, comparatively with other antigen tests and NAT; if inactivated virus is used, the effect of inactivation and freeze/thawing on the antigen should be investigated.
[11] E.g. staphylococci and streptococci expressing protein A or G.

Table 5: **NAT assays for SARS-CoV-2 RNA**

Parameter	**Specimen**	**SARS-CoV-2 RNA qualitative**	**SARS-CoV-2 RNA quantitative**
Sensitivity			
Analytical Sensitivity: Limit of detection	WHO 1st International Standard SARS-CoV-2 RNA (NIBSC code 20/146; 7.70 Log10 IU/mL) Secondary standards calibrated against WHO IS	According to Ph. Eur. NAT validation guideline: several dilution series into borderline concentration; statistical analysis (e.g. Probit analysis) on the basis of at least 24 replicates; calculation of 95 % cut-off value	According to Ph. Eur. NAT validation guideline: several dilution series of calibrated reference preparations into borderline concentration; statistical analysis (e.g. Probit analysis) on the basis of at least 24 replicates; calculation of 95 % cut-off value as limit of detection (LOD)
Quantification limit; quantification features	WHO 1st International Standard SARS-CoV-2 RNA (NIBSC code 20/146; 7.70 Log10 IU/mL) Secondary standards calibrated against WHO IS		Dilutions (half-log10 or less) of calibrated reference preparations; determination of lower, upper quantification limit, limit of detection, precision, accuracy, “linear” measuring range, “dynamic range”. Synthetic target may be used as secondary standard to achieve higher concentration levels. Reproducibility at different concentration levels to be shown
Diagnostic Sensitivity: different SARS-CoV-2 RNA strains	Patient samples determined as SARS-CoV-2 RNA positive by comparator device from different regions and outbreak clusters; sequence variants	≥100[1]	

	Dilution series of SARS-CoV-2 positive cell cultures (isolates) may serve as potential substitutes		
Quantification efficiency	SARS-CoV-2 RNA positive patient samples from different regions and outbreak clusters; sequence variants with quantitative values obtained by comparator device Dilution series of SARS-CoV-2 RNA positive cell cultures may serve as potential substitutes		≥100
Inclusivity	In silico analysis[2]; at least two independent target gene regions in one test run (dual-target design)	Evidence of suitable assay design: primer/probe sequence alignments with published SARS-CoV-2 sequences	Evidence of suitable assay design: primer/probe sequence alignments with published SARS-CoV-2 sequences
Specificity			
Diagnostic specificity	SARS-CoV-2 RNA negative human specimens	≥500	≥100
In silico analysis[2]		Evidence of suitable assay design evidence (sequence alignments); regular check of primer/probe sequences against sequence data bank entries	Evidence of suitable assay design evidence (sequence alignments); regular check of primer/probe sequences against sequence data bank entries
Potential cross reaction	samples positive (various concentrations) for related human coronaviruses 229E, HKU1, OC43, NL63, MERS coronavirus; SARS CoV-1 if available; Influenza virus A, B; RSV; Legionella pneumophila; positive cell cultures may serve as potential substitutes	≥20 in total	≥20
Robustness			
Cross contamination		At least 5 runs using alternating high	At least 5 runs using alternating high

		positive (known to occur naturally) and negative samples	positive (known to occur naturally) and negative samples
Inhibition		Internal control preferably to go through the whole NAT procedure	Internal control preferably to go through the whole NAT procedure
Whole system failure rate leading to false-negative results: 99/100 assays positive		≥100 samples virus-spiked with 3 × the 95 % positive cut-off concentration (3 x LOD)	≥100 samples virus-spiked with 3 × the 95 % positive cut-off concentration (3 x LOD)

[1] If the device is intended to be used for more than one specimen type, 100 samples should be required for each specimen type. If this is not possible in exceptional circumstances (e.g. if specimen collection is very invasive), the manufacturer should provide a justification and evidence of matrix equivalence.
[2] The manufacturer should define frequency and document evidence of regular surveillance checks against updated data bank entries in a post-market performance follow-up plan and report.

Table 6: **Additional requirements for SARS-CoV-2 antigen Self-Tests[1]**

	Specimens[2]	Number of lay users	Criterion
Result interpretation	Interpretation of contrived tests[3] by lay users reflecting a range of results: • non-reactive • reactive • weak reactive[4] • invalid	≥100	Reading and interpretation of the contrived test results by 100 lay people; each lay person should be subjected to read the specified range of result reactivity levels; determination of concordance of lay reading of the same tests by professional readers
Diagnostic sensitivity	Lay users that are known antigen positive[5,6]	≥30	In comparison to the true infectious status, i.e. by RT-PCR; concordance of results with the professional test
Diagnostic specificity	Lay users that do not know their status[5]	≥60	Concordance of results with the professional test

[1] It is assumed that the underlying performance of the self-test has already been previously demonstrated with the evaluation/assessment of a professional test of the same design as the respective self-test under evaluation. In case for the self-use specimens in question there is no corresponding professional test variant, comparison should be made with the standard specimen type (e.g. nasopharyngeal swabs for antigen test, serum or plasma for antibody test) of the corresponding professional test.
[2] For each self-use specimen type claimed with the device (e.g. nasal, sputum, saliva, whole blood, etc.).
[3] Using whenever possible the original natural matrix of the respective specimen type.

[4] A higher proportion of the samples should be in the weak-positive range close to the cutoff or LoD of the test.
[5] Individuals unaware of the professional diagnostic result prior to self-testing, and performing the entire test procedure from specimen collection and specimen pre-treatment (swab, buffer extraction, etc.) to reading.
[6] Subjects up to about 7 days after symptom onset.

Table 7: **Additional requirements for SARS-CoV-2 antibody Self-Tests[1]**

	Specimens[2]	Number of lay users	Criterion
Result interpretation	Interpretation of contrived tests[3] by lay users reflecting a range of results: • non-reactive • reactive • weak reactive[4] • invalid	≥100	Reading and interpretation of the contrived test results by 100 lay people; each lay person should be subjected to read the specified range of result reactivity levels; determination of concordance of lay reading of the same tests by professional readers
Diagnostic sensitivity	Lay users that are known antibody positive[5]	≥100	With previous history of initial PCR confirmed infection for SARS-CoV-2; in comparison to a previous confirmed antibody result; concordance of results with the professional test
Diagnostic specificity	Lay users that do not know their status[5]	≥100	Concordance of results with the professional test

[1] It is assumed that the underlying performance of the self-test has already been previously demonstrated with the evaluation/assessment of a professional test of the same design as the respective self-test under evaluation. In case for the self-use specimens in question there is no corresponding professional test variant, comparison should be made with the standard specimen type (e.g. nasopharyngeal swabs for antigen test, serum or plasma for antibody test) of the corresponding professional test.
[2] For each self-use specimen type claimed with the device (e.g. nasal, sputum, saliva, whole blood, etc.).
[3] Using whenever possible the original natural matrix of the respective specimen type.
[4] A higher proportion of the samples should be in the weak-positive range close to the cutoff or LoD of the test.
[5] Individuals unaware of the professional diagnostic result prior to self-testing, and performing the entire test procedure from specimen collection and specimen pre-treatment (swab, buffer extraction, etc.) to reading.