Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Clinical Epidemiology 138 (2021) 182–188

COMMENTARY

Evaluating tests for diagnosing COVID-19 in the absence of a reliable reference standard: pitfalls and potential solutions

Daniël A. Korevaar[a], Julie Toubiana[b], Martin Chalumeau[b,c], Matthew D.F. McInnes[d,e], Jérémie F. Cohen[b,c,*]

[a] *Department of Respiratory Medicine, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands*
[b] *Department of General Pediatrics and Pediatric Infectious Diseases, Necker-Enfants malades University Hospital, Assistance Publique-Hôpitaux de Paris (AP-HP), Université de Paris, Paris, France*
[c] *Centre of Research in Epidemiology and Statistics (CRESS), INSERM, EPOPé team, Université de Paris, Paris, France*
[d] *Departments of Radiology and Epidemiology, University of Ottawa, Ottawa, Canada*
[e] *Clinical Epidemiology Program, the Ottawa Hospital Research Institute, Ottawa, Canada*

Accepted 29 July 2021; Available online 3 August 2021

1. Introduction

Diagnostic tests play a crucial role in the management of the COVID-19 pandemic, helping to contain the spread of SARS-CoV-2 by detecting and isolating cases, enabling contact tracing, and guiding public health decisions about the initiation of lockdowns, thereby protecting people at increased risk of severe disease and our healthcare systems. Likewise, treatment decisions and enrollment in therapeutic trials require diagnostic confirmation. Because testing for SARS-CoV-2 is currently being done on a massive scale worldwide, false-positive and false-negative results may have considerable adverse downstream consequences [1]. Despite progress made in the last decades in our understanding of the complexity of medical test evaluation, assessing the value of diagnostic tests for COVID-19 still poses considerable challenges [1-6].

One major challenge is that multiple target conditions can be defined. For example, the target condition may be (past or present) SARS-CoV-2 infection, infectiousness, or COVID-19 (i.e., the acute disease caused by SARS-CoV-2). Additionally, COVID-19 has a broad clinical spectrum, which may vary from asymptomatic and mildly symptomatic cases of SARS-CoV-2 infection, to cases of severe pneumonia with or without acute respiratory distress syndrome (ARDS) and multiorgan failure [7]. Furthermore, SARS-CoV-2 may also trigger severe post-infectious processes such as myocarditis, multiorgan failure, and Kawasaki-like illness, notably in children (referred to as 'Paediatric inflammatory multisystem syndrome temporally associated with COVID-19′ (PIMS-TS)) [8].

It is important that researchers carefully define the target condition before initiating a test accuracy study of a test for COVID-19 [2,9]. However, errors in such studies may still occur if the final diagnosis entirely relies on detecting SARS-CoV-2. For example, patients with respiratory symptoms due to pulmonary embolism or community-acquired pneumonia due to *Streptococcus pneumoniae* may be misclassified as COVID-19 if they are concomitant carriers of SARS-CoV-2. Yet, carriage is still highly relevant to detect as part of contact tracing strategies because it may warrant isolation to contain the spread of SARS-CoV-2.

What makes test accuracy studies of COVID-19 tests particularly challenging is the absence of a reliable clinical reference standard. Because of this, the results of most test accuracy studies of COVID-19 tests are of limited value to clinicians and policymakers and may represent a considerable waste of research resources [10]. In this commentary article, we will discuss our views on the extent of this problem and provide potential solutions, based on our experience as clinicians and researchers, and with examples from published literature.

2. A reliable reference standard for diagnosing COVID-19 is currently not available

Tests for diagnosing COVID-19 include clinical signs and symptoms, laboratory tests such as molecular and antigen detection assays, sometimes done at the point of care, imaging tests such as chest CT, electronic noses, and multivariable clinical prediction models (Box 1). The most usual approach to inform clinicians and policymakers about test performance is by conducting test accuracy studies. In such

Conflict of interests: None of the authors have interest to disclose.

* Corresponding author: Phone: +33 1 42 34 55 80,Fax: +33 1 43 26 89 79.
E-mail address: jeremie.cohen@inserm.fr (J.F. Cohen).

https://doi.org/10.1016/j.jclinepi.2021.07.021

Box 1. Examples of index tests and reference standards for COVID-19

1/ Index tests
Signs and symptoms
- Fever
- Respiratory complaints (cough, dyspnea)
- Anosmia
- Recent contact with a patient with SARS-CoV-2 infection
Laboratory tests
- Rapid nucleic acid amplification tests
- Rapid antigen detection tests
- Antibody tests
- Inflammatory markers: C-reactive protein, procalcitonin
- Electronic nose
Imaging tests
- Chest X-ray
- Chest CT
- Chest ultrasound
Multivariable prediction models that combine several features to estimate the risk of COVID-19
- 'COVID-19 early warning score'
- 'Corona-Score'
2/ Reference standards
- Single RT-PCR
- Repeat RT-PCR
- Viral culture
- Combination of RT-PCR and other criteria (clinical, epidemiological, imaging, laboratory data)

Box 2. Main limitations of RT-PCR testing for SARS-CoV-2

1/ Pre-analytical and analytical considerations
- Diagnostic yield depends on the quality and type of the clinical sample
- Diagnostic yield depends on when the sample is taken in the course of the disease
- Variability in molecular targets across commercial kits
- No consensual cycle threshold (Ct) value to define test positivity
2/ Clinical interpretation
- RT-PCR may detect nonviable viral particles and low viral loads of unclear clinical significance
- RT-PCR may not distinguish between carriage and SARS-CoV-2 infection

studies, the results of a test under evaluation (the 'index test') are compared to those of another test that is supposed to distinguish between patients with and without the clinical target condition with a high level of certainty (the 'reference standard'). Typical outcomes in test accuracy studies are estimates of clinical sensitivity and specificity. Hundreds of test accuracy studies on COVID-19 tests have already been conducted, as well as systematic reviews to summarize them [11-16].

In these test accuracy studies, laboratory-based real-time reverse-transcription polymerase chain reaction (RT-PCR) is widely used as the reference standard for diagnosing COVID-19. However, this test has important shortcomings (Box 2), which threaten the validity of these studies. An imperfect reference standard will, by definition, lead to index test results that are wrongly classified as 'false positives' or 'false negatives', or wrongly classified as 'true positives' or 'true negatives', a problem referred to as 'reference standard error bias' [17]. This will result in either over- or under-estimation of the sensitivity or specificity of the index test under investigation. First, as for other respiratory viruses, the detection of SARS-CoV-2 by RT-PCR is strongly influenced by the quality, site, and timing of sampling [4]. For example, analytical sensitivity of RT-PCR is different for bronchoalveolar lavage fluid, nasopharyngeal swabs, and throat swabs [18]. In addition, in most cases, the probability that SARS-CoV-2 becomes detectable by RT-PCR in nasopharyngeal samples peaks around the onset of symptoms of COVID-19 and then gradually declines within 2 to 3 weeks [1,4,19]. Hence, most

Table 1. Risk of bias assessment for the reference standard in Cochrane reviews of diagnostic tests for COVID-19

Author [Reference]	Test(s)	No. of studies	No. of participants*	No. (%) of studies at high risk of bias	No. (%) of studies at unclear risk of bias	No. (%) of studies at low risk of bias
Struyf et al. [11]	Signs and symptoms	44	26,884	3 (7)	8 (18)	33 (75)
Stegeman et al. [12]	Routine laboratory tests	21	70,711	14 (67)	5 (24)	2 (10)
Dinnes et al. [13]	Rapid, point-of-care antigen and molecular-based tests	78	24,087	66 (85)	6 (8)	6 (8)
Deeks et al. [14]	Antibody tests	54	15,976	17 (31)	24 (24)	13 (24)
Islam et al. [15]	Thoracic imaging tests	51	19,775	20 (39)	20 (39)	11 (22)
TOTAL	-	248	157,433	120 (48)	63 (25)	65 (26)

* or samples
Table updated on May 20th, 2021

experts recommend repeating RT-PCR testing in patients with a sustained intermediate or high clinical suspicion of COVID-19 when the initial result is negative, especially in hospital settings [20]. Second, there is technical variability among the (more than) 150 RT-PCR kits for SARS-CoV-2 that have been approved and provided with FDA's emergency use authorization (EUA) label [5,21]. Notably, these kits rely on various molecular targets encoding structural (e.g., envelope (*E*), nucleocapsid (*N1, N2, N3*), and spike (*S*) genes) and nonstructural proteins (e.g., RNA-dependent RNA polymerase (*RdRp*) and open reading frame 1 segments 1a and 1b (*Orf1*) genes). RT-PCR analytical sensitivity is higher for certain combinations of nucleic acid targets than others. For example, in a recent study comparing 6 molecular kits approved for detecting SARS-CoV-2, the limit of detection after 40 thermal cycles ranged from a viral RNA concentration of 484 to 7744 copies/mL, a 16-fold difference [22]. Third, RT-PCR is not a binary test: most tests provide a quantitative result reflecting the number of amplification cycles after which a signal becomes detectable ('cycle threshold', Ct), but the Ct value that has the highest clinical relevance is still a matter of debate and may vary across commercial kits and between laboratories using the same kits [23-25]. Low viral loads may reflect noninfectious states (e.g., nonviable viral RNA fragments) and low-risk viral shedding of unclear clinical significance, which may lead to overdiagnosis. Yet, decreasing the positivity threshold for the Ct value may result in missing infectious cases with low viral loads. Fourth, SARS-CoV-2 is mutating over time, resulting in genetic variations across circulating viral strains. If such mutations occur in the genetic sequences targeted by a given RT-PCR kit, this may potentially have a negative effect on test accuracy [26].

There are alternatives to RT-PCR testing. Researchers evaluating COVID-19 tests have used other reference standards, ranging from viral culture to various combinations of epidemiological data, clinical information, and testing results (Box 1). Unfortunately, each of these alternatives has its shortcomings as well, and may further amplify the problem by making evidence syntheses more challenging to conduct and interpret. For example, in a recent Cochrane systematic review of the accuracy of imaging tests for COVID-19, which included 51 studies, 47 used only RT-PCR as the reference standard (single RT-PCR, 2 studies; repeat RT-PCR in all patients with an initial negative test, 11 studies; repeat RT-PCR in at least some patients with an initial negative test, 17 studies; number of RT-PCR tests not reported, 17 studies) and 4 studies used a combination of RT-PCR and other criteria (i.e., clinical symptoms, information about infected household contacts, imaging tests, and laboratory tests) [15].

The series of systematic reviews from the Cochrane COVID-19 Diagnostic Test Accuracy Group currently encompasses a total of 248 test accuracy studies and 157,433 participants (Table 1) [11-15]. Details about the reference standard were poorly reported, and risk of bias about the reference standard was deemed high in 48% and unclear in 25%; the reference standard was judged at low risk of bias in only 26% of the included studies.

3. Potential solutions in the absence of reliable reference standard for COVID-19

The challenge of evaluating the diagnostic accuracy of a test in the absence of a reliable reference standard is not new, and there is an array of possible solutions that can be applied to COVID-19, although each has its advantages and limitations (Table 2). Below, we discuss several possible solutions, but others exist as well [27-29].

3.1. Panel-based reference standard

A first option is to define disease status based on the opinion of a panel of experts that use evidence from a combination of signs, symptoms, tests, and sometimes follow-up in an unstructured way to classify each patient as having COVID-19 or not. An example of such an adjudication committee is reported in a study where five board-certified specialists in respiratory and internal medicine were retrospectively invited to make a final diagnosis for each patient suspected of COVID-19 [30]. In a similar study, all patients with a positive chest CT, but a negative RT-PCR,

Table 2. Potential solutions in the absence of a reliable reference standard for COVID-19.

Method	Explanation	Advantages	Limitations
Panel-based diagnosis	Final diagnosis is made by a panel of experts.	May reflect clinical practice.	• No clear definition of the target condition. • Poor reproducibility.
Rule-based diagnosis	Final diagnosis is made through the formal combination of various pieces of information.	Easy to apply and reproducible.	May not reflect clinical practice.
Model-based diagnosis	Accuracy estimates are computed by a model, assuming that no single test is able to define disease status.	Provides estimates of sensitivity and specificity for multiple tests, which allows comparisons.	• No clinical definition of the target condition. • Complex modeling.
Studies of agreement between test results	None of the two tests is accepted as a reference standard and agreement is measured in a cross-tabulation of test results.	Reference standard not needed.	Not possible to say if discrepancies are due to errors from one test or the other.
Studies of clinical effectiveness	The focus is not on sensitivity and specificity but on other outcomes such as infection, hospitalization, and mortality rates, or test uptake and diagnostic yield.	• Reference standard not needed. • Outcomes matter to users.	• Time and resource-consuming. • Large sample sizes required.

were discussed in a multidisciplinary meeting to arrive at a final diagnosis [31]. Panel-based diagnosis has the advantage of reflecting clinical settings, where it is common practice that difficult cases are examined during multidisciplinary meetings. The disadvantage is that panel-based studies do not provide clear criteria for ruling-in or ruling-out the target condition, which may hamper reproducibility. Also, the level of expertise of the panel members is crucial in such approaches and may vary across settings and over time, as the evidence base increases. The panel may even be worse than the reference standard it is seeking to replace if members of the panel have insufficient expertise in diagnosing the target condition.

3.2. Rule-based reference standard

A second possibility is to classify patients using a clear and structured classification rule in the form of a list of criteria or score that may include the results of several tests. This procedure is sometimes referred to as a 'composite' reference standard. An example is the case definition of the US Centers for Disease Control and Prevention, which combines clinical, laboratory, and epidemiological information into a classification rule with three levels of likelihood of COVID-19 (e.g., suspect, probable, or confirmed); Box 3. These case definitions of COVID-19 are easy to apply and reproducible. However, they were designed for epidemiological surveillance purposes but may not be well suited for clinical management and test accuracy studies of COVID-19 tests. Another challenge for test evaluations is that many case definitions for COVID-19 have multiple diagnostic categories according to the likelihood of disease, while calculating sensitivity and specificity usually requires a binary reference standard. Also, rule-based studies may be impacted by 'incorporation bias': if the test under evaluation is incorporated as part of the composite reference standard, then sensitivity and specificity may both be overestimated [17].

3.3. Model-based reference standard

A third option are latent class models [32]. In such models, none of the tests is considered a reference standard: the sensitivity and specificity of each test are estimated from the analysis of the cross-classified results of the tests for which results are available. For COVID-19, latent class models may include, for example, clinical signs and symptoms, RT-PCR, serology, and chest CT. The model provides estimates of COVID-19 prevalence and estimates of the sensitivity and specificity of each test. Hartnack and colleagues recently reported an example of bayesian latent class modeling applied to COVID-19 diagnostic data [33]. Latent class models were also implemented in a diagnostic meta-analysis of salivary tests for SARS-CoV-2, which allowed to account for the imperfectness of the reference standard and the potential non-independence between results obtained with salivary and nasopharyngeal nucleic acid amplification tests [34]. A benefit of latent class models is that they take advantage of all the available information and provide sensitivity and specificity estimates for all tests included in the model, allowing comparisons between tests. Limitations are that they are complex methods that may require expert statistical knowledge, that erroneous assumptions regarding dependence between tests in patients with and without the target condition may lead to biased estimates of test accuracy, and that they do not rely on a clinical definition of the target condition, but only a statistical one. Due to these limitations, results of studies applying latent class models may be more difficult to interpret by readers without a statistical background.

Box 3. COVID-19 case definition form the US Centers for Disease Control and prevention*

Clinical criteria
In the absence of a more likely diagnosis:
- At least two of the following symptoms: fever (measured or subjective), chills, rigors, myalgia, headache, sore throat, nausea or vomiting, diarrhea, fatigue, congestion or runny nose
OR
- Any one of the following symptoms: cough, shortness of breath, difficulty breathing, new olfactory disorder, new taste disorder
OR
- Severe respiratory illness with at least one of the following: clinical or radiographic evidence of pneumonia, acute respiratory distress syndrome.
Laboratory criteria
Confirmatory laboratory evidence:
- Detection of severe acute respiratory syndrome coronavirus 2 ribonucleic acid (SARS-CoV-2 RNA) in a clinical or autopsy specimen using a molecular amplification test
Presumptive laboratory evidence:
- Detection of SARS-CoV-2 by antigen test in a respiratory specimen
Supportive laboratory evidence:
- Detection of specific antibody in serum, plasma, or whole blood
- Detection of specific antigen by immunocytochemistry in an autopsy specimen
Epidemiologic linkage
One or more of the following exposures in the prior 14 days:
- Close contact with a confirmed or probable case of COVID-19 disease;
- Member of a risk cohort as defined by public health authorities during an outbreak.
Case classification
Suspect
- Meets supportive laboratory evidence with no prior history of being a confirmed or probable case.
Probable
- Meets clinical criteria AND epidemiologic linkage with no confirmatory laboratory testing performed for SARS-CoV-2.
- Meets presumptive laboratory evidence.
- Meets vital records criteria with no confirmatory laboratory evidence for SARS-CoV-2.
Confirmed
- Meets confirmatory laboratory evidence.
*2020 Interim case definition, approved August 5, 2020

3.4. Agreement between tests instead of test accuracy

A fourth option is to compare a new test to an existing one by evaluating the level of agreement between the two tests, rather than reporting accuracy estimates that are unreliable due to the absence of a satisfactory reference standard. This has been done in several studies assessing RT-PCR tests based on salivary samples compared to the same tests using nasopharyngeal samples. One study, for example, found an overall agreement of 98% between RT-PCR done on salivary and nasopharyngeal samples. The authors concluded that « saliva is an acceptable alternative source for detecting SARS-CoV-2 nucleic acids » [35]. An advantage of this approach is that classification outcomes do not rely on a (potentially imperfect) reference standard. A downside is that raw agreement does not tell if discrepancies are due to errors from one test or the other. Therefore, this option is often only useful if a new test is meant to replace an existing test, for example because it is cheaper or less invasive, and researchers want to illustrate that the tests produce similar results in the majority of patients. Alternatively, such studies could compare detection rates between two tests, which may be a useful statistic if both tests are considered to have a specificity close to 100% (i.e., almost no false-positive results).

3.5. Clinical effectiveness

Finally, we may consider moving from diagnostic accuracy to clinical effectiveness [36]. In this framework, we would not be interested in classification outcomes such as sensitivity and specificity but would focus on patient-centered or population-centered outcomes such as infection, hospitalization, quality of life, and mortality rates. Here the goal is to develop and implement testing strategies that would prove beneficial for health and society. For example, several authors have argued that rapid point-of-care tests for SARS-CoV-2 might be effective in reducing viral community transmission despite having higher ana-

lytical limits of detection than conventional RT-PCR tests based on nasopharyngeal swabs, because of better uptake and shorter turnaround time that allow for repeat testing and timely isolation [37]. In a screening setting, individuals or groups could even be randomized to receive conventional or rapid tests for SARS-CoV-2. Here, we could assess outcomes such as participation, usability, positivity rate, and diagnostic yield, as done, for example, in colorectal cancer screening trials [38-40]. A drawback is that clinical effectiveness studies (including randomized trials) of medical tests are generally much more time- and resource-consuming than cross-sectional test accuracy studies, although this may be less of a problem in a pandemic setting due to the large number of potential study subjects and available funds.

4. Conclusions

COVID-19 tests play a central position in managing the disease worldwide and are being done at an unprecedented scale. However, evaluating the diagnostic accuracy of these tests is challenging. Although there are several reasons for this [2], one major issue is the lack of a high-quality reference standard. In this commentary article, we have argued that, because of this, the evidence provided by many test accuracy studies is difficult to interpret and may not suffice to decide with confidence which test is optimal in which setting. To avoid this waste of resources, research is urgently needed to help clarify which reference standard(s) for COVID-19 we should use in future test accuracy studies. It has become clear that relying on a single RT-PCR test is problematic to detect SARS-CoV-2 infection in symptomatic individuals as this reference standard may miss too many cases. A minimal requirement to minimize bias could be to ask for at least two negative RT-PCR results to define COVID-19-negatives. Depending on the target condition (e.g., COVID-19, infectiousness, carriage, immunological responses), and guided by the intended use population (e.g., diagnosis in symptomatic patients, targeted screening in contact tracing programs, mass screening in the general population), different reference standards may be needed. In the absence of an appropriate clinical reference standard, researchers could consider alternative techniques such as panel-, rule- and model-based methods, or measures of agreement. We need test accuracy studies that provide a more informative description of essential study features and test methods by following, for example, the STARD reporting guideline [41,42]. We also need studies that, beyond accuracy, evaluate the effectiveness of COVID-19 tests through outcomes that directly matter to patients, policymakers, and society.

Contributions

JFC initiated the project and wrote the initial draft of the manuscript. All authors provided a substantial contribution to the manuscript and approved the final version.

Funding

No specific funding was obtained for this work.

References

[1] Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. BMJ 2020;369:m1808.
[2] Doust JA, Bell KJL, Leeflang MMG, Dinnes J, Lord SJ, Mallett S, et al. Guidance for the design and reporting of studies evaluating the clinical performance of tests for present or past SARS-CoV-2 infection. BMJ 2021;372:n568.
[3] Bossuyt PM. Testing COVID-19 tests faces methodological challenges. J clin epidemiol 2020;126:172–6.
[4] Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. JAMA 2020;323:2249–51.
[5] Vandenberg O, Martiny D, Rochas O, van Belkum A, Kozlakidis Z. Considerations for diagnostic COVID-19 tests. Nat Rev Microbiol 2021;19:171–83.
[6] Tang YW, Schmitz JE, Persing DH, Stratton CW. Laboratory diagnosis of COVID-19: current issues and challenges. J Clin Microbiol 2020;58(6):e00512–20.
[7] Datta SD, Talwar A, Lee JT. A proposed framework and timeline of the spectrum of disease due to SARS-CoV-2 infection: illness beyond acute infection and public health implications. JAMA 2020;324:2251–2.
[8] Toubiana J, Poirault C, Corsia A, Bajolle F, Fourgeaud J, Angoulvant F, et al. Kawasaki-like multisystem inflammatory syndrome in children during the covid-19 pandemic in Paris, France: prospective observational study. BMJ 2020;369:m2094.
[9] Korevaar DA, Gopalakrishna G, Cohen JF, Bossuyt PM. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses. Diagn Progn Res 2019;3:22.
[10] Glasziou PP, Sanders S, Hoffmann T. Waste in covid-19 research. BMJ 2020;369:m1847.
[11] Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeflang MM, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. Cochrane database syst rev 2021;2:CD013665.
[12] Stegeman I, Ochodo EA, Guleid F, Holtman GA, Yang B, Davenport C, et al. Routine laboratory testing to determine if a patient has COVID-19. Cochrane database syst rev 2020;11:CD013787.
[13] Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane database syst rev 2021;3:CD013705.
[14] Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane database syst rev 2020;6:CD013652.
[15] Islam N, Ebrahimzadeh S, Salameh JP, Kazi S, Fabiano N, Treanor L, et al. Thoracic imaging tests for the diagnosis of COVID-19. Cochrane database syst rev 2021;3:CD013639.
[16] Tang YW, Schmitz JE, Persing DH, Stratton CW. Electronic and animal noses for detecting SARS-CoV-2 infection. Cochrane database syst rev 2021;6:CD015013.
[17] Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann inte med 2004;140:189–202.
[18] Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA 2020;323(18):1843–4.
[19] He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat med 2020;26:672–5.

[20] Hanson KE, Caliendo AM, Arias CA, Englund JA, Lee MJ, Loeb M, et al. Infectious diseases society of america guidelines on the diagnosis of COVID-19. Clin Infect Dis 2020 ciaa760.

[21] FDA. In 2021 vitro diagnostics EUAs [Available from: www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas].

[22] Wang X, Yao H, Xu X, Zhang P, Zhang M, Shao J, et al. Limits of detection of 6 approved RT-PCR kits for the novel SARS-Coronavirus-2 (SARS-CoV-2). Clin chem 2020;66:977–9.

[23] Singanayagam A, Patel M, Charlett A, Lopez Bernal J, Saliba V, Ellis J, et al. Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. Euro Surveill 2020;25(32):2001483.

[24] Bullard J, Dust K, Funk D, Strong JE, Alexander D, Garnett L, et al. Predicting infectious severe acute respiratory syndrome coronavirus 2 from diagnostic samples. Clin Infect Dis 2020;71:2663–6.

[25] Atkinson B, Petersen E. SARS-CoV-2 shedding and infectivity. Lancet 2020;395:1339–40.

[26] Ascoli CA. Could mutations of SARS-CoV-2 suppress diagnostic detection? Nat Biotechnol 2021;39:274–5.

[27] Reitsma JB, Rutjes AW, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J clin epidemiol 2009;62:797–806.

[28] Umemneku Chikere CM, Wilson K, Graziadio S, Vale L, Allen AJ. Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - An update. PLoS One 2019;14:e0223832.

[29] Axell-House DB, Lavingia R, Rafferty M, Clark E, Amirian ES, Chiao EY. The estimation of diagnostic accuracy of tests for COVID-19: A scoping review. J Infect 2020;81:681–97.

[30] Skalidis I, Nguyen VK, Bothorel H, Poli L, Da Costa RR, Younossian AB, et al. Unenhanced computed tomography (CT) utility for triage at the emergency department during COVID-19 pandemic. Am J Emerg Med 2021;46:260–5.

[31] Korevaar DA, Kootte RS, Smits LP, van den Aardweg JG, Bonta PI, Schinkel J, et al. Added value of chest computed tomography in suspected COVID-19: an analysis of 239 patients. Eur Respir J 2020;56(2):2001377.

[32] van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KG, de Groot JA. Latent class models in diagnostic studies when there is no reference standard–a systematic review. Am J Epidemiol 2014;179:423–31.

[33] Hartnack S, Eusebi P, Kostoulas P. Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests. J Med Virol 2021;93:639–40.

[34] Butler-Laporte G, Lawandi A, Schiller I, Yao M, Dendukuri N, McDonald EG, et al. Comparison of saliva and nasopharyngeal swab nucleic acid amplification testing for detection of SARS–CoV-2: a systematic review and meta-analysis. JAMA Intern Med 2021;181:353–60.

[35] McCormick-Baw C, Morgan K, Gaffney D, Cazares Y, Jaworski K, Byrd A, et al. Saliva as an alternate specimen source for detection of SARS-CoV-2 in symptomatic patients using cepheid xpert xpress SARS-CoV-2. J Clin Microbiol 2020;58(8):e01109–20.

[36] Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin chem 2013;58:1636–43.

[37] Mina MJ, Parker R, Larremore DB. Rethinking Covid-19 test sensitivity - a strategy for containment. New Eng J med 2020;383:e120.

[38] Grobbee EJ, van der Vlugt M, van Vuuren AJ, Stroobants AK, Mundt MW, Spijker WJ, et al. A randomised comparison of two faecal immunochemical tests in population-based colorectal cancer screening. Gut 2017;66:1975–82.

[39] Pilonis ND, Bugajski M, Wieszczy P, Rupinski M, Pisera M, Pawlak E, et al. Participation in competing strategies for colorectal cancer screening: a randomized health services study (PICCOLINO Study). Gastroenterology 2021;160:1097–105.

[40] Passamonti B, Malaspina M, Fraser CG, Tintori B, Carlani A, D'Angelo V, et al. A comparative effectiveness trial of two faecal immunochemical tests for haemoglobin (FIT). Assessment of test performance and adherence in a single round of a population-based screening programme for colorectal cancer. Gut 2018;67:485–96.

[41] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.

[42] Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ open 2016;6:e012799.