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.

Clinical Biochemistry 88 (2021) 18–22

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

Setting minimum clinical performance specifications for tests based on disease prevalence and minimum acceptable positive and negative predictive values: Practical considerations applied to COVID-19 testing

Tze Ping Loh [a,*], Sarah J Lord [b,c], Katy Bell [d], Mary Kathryn Bohn [e], Chun Yee Lim [f], Corey Markus [g], Hernan Fares Taie [h], Khosrow Adeli [e], Giuseppe Lippi [i], Sverre Sandberg [j,k,l], Andrea Horvath [m]

[a] Department of Laboratory Medicine, National University Hospital, Singapore
[b] School of Medicine, The University of Notre Dame Australia, Sydney, New South Wales, Australia
[c] National Health and Medical Research Council Clinical Trials Centre, Faculty of Medicine and Health, The University of Sydney, New South Wales, Australia
[d] Sydney School of Public Health, Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, New South Wales, Australia
[e] Clinical Biochemistry, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada
[f] Engineering Cluster, Singapore Institute of Technology, Singapore
[g] Metabolic Laboratory, Genetics and Molecular Pathology Directorate, SA Pathology, Women's and Children's Hospital Site, South Australia, Australia
[h] Fares Taie Instituto de Analisis, Mar del Plata, Argentina
[i] Section of Clinical Biochemistry, Department of Neuroscience, Biomedicine and Movement, University of Verona, Verona, Italy
[j] Norwegian Organization for Quality Improvement of Laboratory Examinations (Noklus), Haraldsplass Deaconess Hospital, Bergen, Norway
[k] Norwegian Porphyria Centre, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, Bergen, Norway
[l] Institute of Public Health and Primary Health Care, University of Bergen, Bergen Norway
[m] Department of Clinical Chemistry, New South Wales Health Pathology, Prince of Wales Hospital, Sydney, Australia

ARTICLE INFO

Keywords:
Performance specification
Sensitivity
Specificity
Negative predictive value
Positive predictive value
Prevalence
Evaluation
False positive
False negative

ABSTRACT

Objectives: Several guidelines for the evaluation of laboratory tests for severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) infection have recommended establishing an *a priori* definition of minimum clinical performance specifications before test selection and method evaluation.
Methods: Using positive (PPV) and negative predictive values (NPV), we constructed a spreadsheet tool for determining the minimum clinical specificity (conditional on NPV or PPV, sensitivity and prevalence) and minimum clinical sensitivity (conditional on NPV or PPV, specificity and prevalence) of tests.
Results: At a prevalence of 1%, there are no minimum sensitivity requirements to achieve a desired NPV of 60%-95% for a given clinical specificity above 20%. It is not possible to achieve 60–95% PPV even with 100% clinical sensitivity, except when the clinical specificity is near 100%. The opposite trend is seen in high prevalence settings (60%), where a relatively low minimum clinical sensitivity is required to achieve a desired PPV for a given clinical specificity, and a higher minimum clinical specificity is required to achieve a desired NPV for a given clinical sensitivity.
Discussion: The selection of laboratory tests and the testing strategy for SARS-CoV-2 involves delicate trade-offs between NPV and PPV based on prevalence and clinical sensitivity and clinical specificity. Practitioners and health authorities should carefully consider the clinical scenarios under which the test result will be used and select the most appropriate testing strategy that fulfils the a priori defined clinical performance specification.

* Corresponding author at: 5 Lower Kent Ridge Road, Singapore 119074, Singapore.
E-mail addresses: tploh@hotmail.com (T.P. Loh), Sally.Lord@nd.edu.au (S.J. Lord), katy.bell@sydney.edu.au (K. Bell), marykathryn.bohn@sickkids.ca (M.K. Bohn), Chunyee.Lim@singaporetech.edu.sg (C.Y. Lim), Corey.Markus@sa.gov.au (C. Markus), laboratorio@farestaie.com.ar (H. Fares Taie), khosrow.adeli@sickkids.ca (K. Adeli), giuseppe.lippi@univr.it (G. Lippi), sverre.sandberg@noklus.no (S. Sandberg), andrea.horvath@health.nsw.gov.au (A. Horvath).

https://doi.org/10.1016/j.clinbiochem.2020.11.003
Received 10 October 2020; Accepted 3 November 2020
Available online 20 November 2020

1. Introduction

Molecular and serological tests play an important role in the diagnosis, monitoring and epidemiological investigation of many diseases. These analyses are often interpreted as positive or negative results. A novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has caused an ongoing pandemic that causes severe coronavirus disease 2019 (COVID-19) [1,2]. Infection by SARS-CoV-2 may be identified by detecting viral RNA with nucleic acid amplification tests (NAATs) or by viral antigens, or in retrospect by documenting the presence of an immune response against the virus, as reflected by the presence of specific anti-SARS-CoV-2 antibodies [3,4].

Several guidelines for the evaluation of laboratory tests for SARS-CoV-2 infection have recommended establishing an *a priori* definition of minimum clinical performance specifications before test selection and method evaluation [5–11]. The clinical performance of a test refers to the "ability of a biomarker to conform to predefined clinical specifications in detecting patients with a particular clinical condition or in a physiological state" [12]. The minimum clinical performance specifications can be defined as the minimum clinical sensitivity and specificity the test must attain to support conclusions that it is fit for intended purpose, such as for informing clinical (e.g. diagnostic work up) or population health decisions (e.g. for epidemiological investigation in a particular population, for estimating prevalence of current infection, or for monitoring the development of antibodies within a population).

Acknowledging the fact that no tests are able to achieve perfect clinical performance, these specifications can be set by defining the tolerable rates of false negative and false positive findings above which the consequences of acting on the test results are considered likely to result in more harm than good [13]. Clinical performance of a test does not necessarily translate to clinical effectiveness, which is a measure of the actual harm-benefit trade-off derived from the appropriate medical decisions or actions triggered by the test results. Thus, good clinical performance of a test is a prerequisite of, but does not guarantee, improved health outcomes. When trials of a test directly measuring health, public health or other societal outcomes (i.e. clinical effectiveness) are not available, the clinical performance levels of a test can be used as a proxy to infer health outcomes.

The prevalence of disease in a population exerts a considerable influence on the probability of disease in those with either a positive or negative test result [14]. For example, when a test with 95% sensitivity and 95% specificity is applied to individuals in a population with 1% disease prevalence, the probability of disease after a positive result (i.e. positive predictive value, PPV) is 16%, and the probability of no disease after a negative result (i.e. negative predictive value, NPV) is 99.5%. In other words, of 100 who test positive, only 16 would actually have the disease, while 84 would not have the disease (false positive); and of 10,000 who tested negative, 9995 would not have the disease, while 5 would actually have the disease (false negative). When disease prevalence is 60%, the PPV of the same test increases to 97% (of those who test positive, 97% actually have the disease) and NPV falls to 93% (of those who test negative, 93% are disease-free, while 7% actually have the disease).

To be clinically useful, a test must be able to deliver an acceptable number of true versus false positive and negative findings, so that acting on a positive result can be considered to offer a favorable benefit-harm trade-off. The PPV and NPV capture these trade-offs, thus setting minimum acceptable PPV and NPV can be used as the starting point for defining minimum sensitivity and specificity requirements in a specific population. While PPV and NPV are useful for selecting testing strategy, the calculation of post-test probability of disease using likelihood ratios is helpful to make clinical decisions for individual patients.

The aim of this paper is to illustrate how this knowledge can be applied to set *a priori* clinical performance specifications tailored to the intended use of tests in a specific population. We investigated the minimum clinical performance specifications for a range of minimum acceptable trade-offs between the number of false positives (FP) to true positives (TP), and false negatives (FN) to true negatives (TN), expressed as PPV and NPV, respectively. Given the prevalence of SARS-CoV-2 infection and COVID-19 can be expected to vary across populations and over time, we provide a tool that can be used to consider the implications for minimum sensitivity and specificity requirements.

2. Material and Methods

The minimum clinical performance specifications in populations with different prevalence are defined below:

Clinical sensitivity (TP rate) = TP/ (TP + FN)

Clinical specificity (TN rate) = TN/ (TN + FP)

We then used the following equations for our calculations [14]:

Prevalence = TP/ population at risk,

$$\text{PPV} = \text{sensitivity} \times \text{prevalence} / [\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})] \quad (1)$$

$$\text{NPV} = \text{specificity} \times (1 - \text{prevalence}) / [(1 - \text{sensitivity}) \times \text{prevalence} + \text{specificity} \times (1\text{- prevalence})] \quad (2)$$

Solving (1) for specificity, we may derive the equation for calculating the minimum clinical specificity that is needed for a desired PPV/ NPV (for a particular intended test application) and a given sensitivity and prevalence:

$$\textbf{Minimum clinical specificity} = 1 - [(1/\text{PPV} - 1)(\text{sensitivity} \times \text{prevalence})]/(1 - \text{prevalence}) \quad (3)$$

Or

$$\textbf{Minimum clinical specificity} = \text{NPV} \times (1 - \text{sensitivity}) \times \text{prevalence} / [(1\text{-prevalence}) \times (1\text{-NPV})] \quad (4)$$

Solving (2) for sensitivity, we may derive the equation for calculating the minimum clinical sensitivity that is needed for a desired NPV/ PPV (for a particular intended test use) and a given specificity and prevalence:

$$\textbf{Minimum clinical sensitivity} = 1 - [(1/\text{NPV-1})(1\text{-prevalence})(\text{specificity})]/\text{prevalence} \quad (5)$$

Or

$$\textbf{Minimum clinical sensitivity} = \text{PPV} \times (1\text{ - specificity}) \times (1\text{- prevalence}) / [(1\text{- PPV}) \times (\text{prevalence})] \quad (6)$$

Using decision-analytic principles, one can weigh the relative harms of a false positive and a false negative test result, to set the minimum acceptable PPV or NPV as follows:

For new tests intended to improve outcomes through detection of infection, the critical trade-off for defining minimum clinical performance characteristics can be defined by asking "how many individuals are you prepared to treat/isolate/contact trace unnecessarily (FP), in order to identify additional cases that truly require treatment/isolation/contact tracing (TP)?". This trade-off can be expressed as the minimum acceptable PPV by calculating the proportion (TP/TP + FP).

For new tests intended to improve outcomes through ruling out of infection, the critical trade-off can be defined by asking "how many individuals with infection are you prepared to miss who might transmit the virus further (FN), in order to rule out additional cases who truly do not have infection (TN)?". This trade-off can be expressed as the minimum acceptable NPV by calculating the proportion (TN/TN + FN).

Using equations (3–6), we constructed a spreadsheet (Microsoft, Richmond, WA, USA; see Supplemental Material) for determining the minimum clinical specificity (conditional on NPV/ PPV, sensitivity and prevalence) and minimum clinical sensitivity (conditional on NPV/ PPV, specificity and prevalence). We also plotted these values at different prevalences (1%, 10%, 60%) to illustrate the variation in minimum

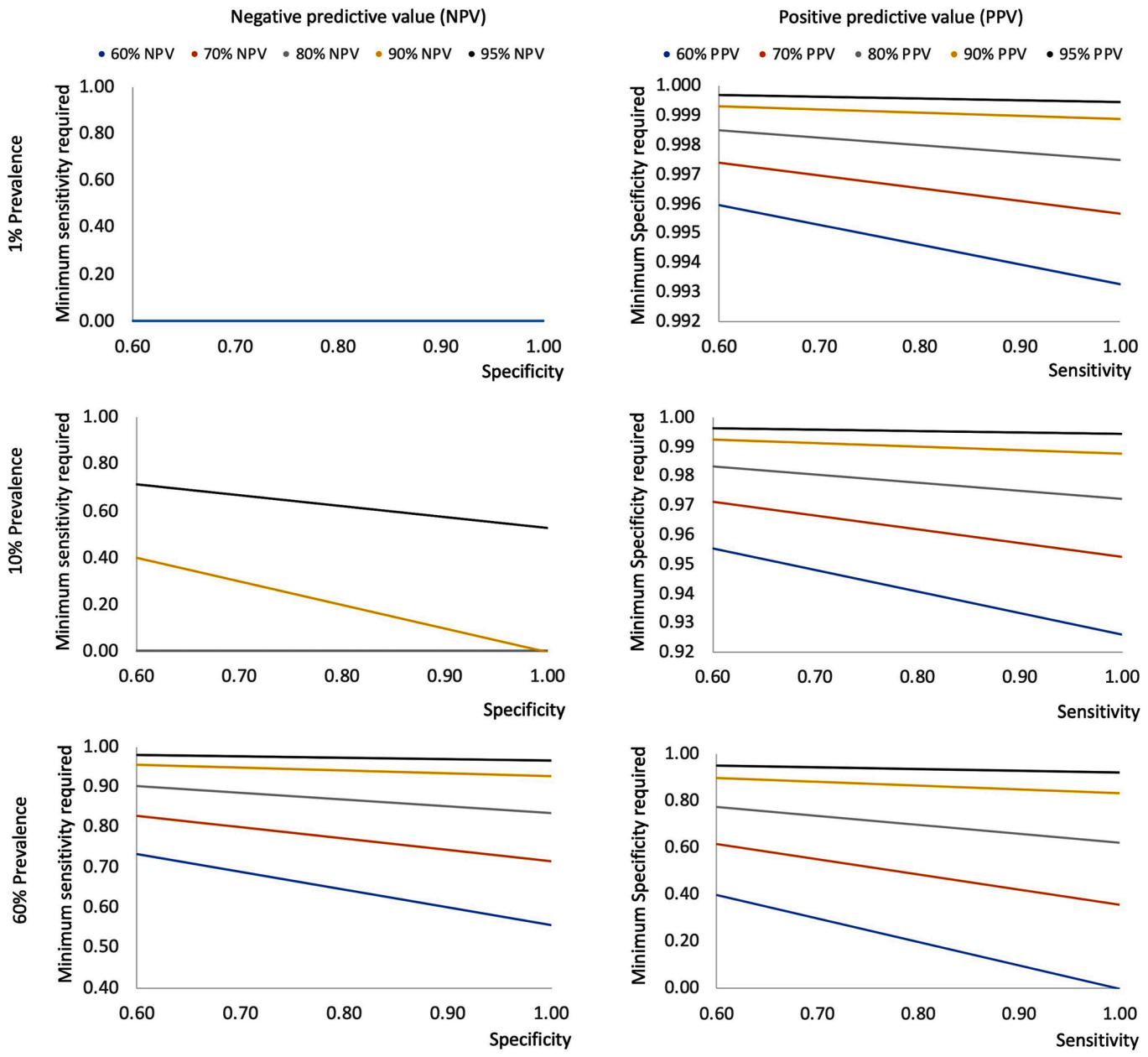

Fig. 1. The minimum clinical sensitivity (right panel) and minimum clinical specificity (left panel) for different prevalence, positive predictive value (PPV), negative predictive value (NPV). Note the scales of the axes are not equivalent in all panels.

sensitivity and specificity requirements across a range of desired NPV and PPV from 60 to 95%.

3. Results

At a prevalence of 1%, a test with any clinical sensitivity (i.e. a minimum clinical sensitivity of 0%) will be able to achieve a desired NPV in the range of 60%-95% for a given clinical specificity of 20% or higher (Fig. 1). On the other hand, near perfect minimum clinical specificity (>99%) is required to achieve 60–95% PPV even with a given clinical sensitivity of 100%. The opposite trend is seen in high prevalence settings (60%), where a relatively high minimum clinical sensitivity is required to achieve a desired NPV for a given clinical specificity, and a relatively lower minimum clinical specificity is required to achieve a desired PPV for a given clinical sensitivity.

As a practical example, consider the clinical use of a molecular test to rule out a COVID-19 infection among recent travel arrivals. A high rule out probability is desired to avoid inadvertent community spread of undetected COVID-19 cases, so *a priori* determined NPV is set as 99%. It is assumed that the prevalence of COVID-19 in the departure country is 5%, and the molecular test has 95% clinical specificity. Using the tool, the minimum clinical sensitivity required is 82% to achieve *a priori* NPV of 99%. In the same vein, assuming the prevalence of COVID-19 among close contacts is ~ 30%, a minimum clinical sensitivity of 98% is required to rule out a cross-infection with *a priori* NPV of 99%.

Consider another clinical scenario where a molecular test is used to rule in the diagnosis for initiation of an experimental COVID-19 treatment. A high rule in probability is desired to justify the risks of adverse effects, so *a priori* determined PPV is set as 90%. It is assumed that COVID-19 is endemic in the community (prevalence: 40%), and the clinical specificity is 95%. Using the tool, the minimum clinical sensitivity required is 68% to achieve *a priori* PPV of 95%.

4. Discussions

This report provides a practical way of setting clinical performance specifications dependent on the prevalence of the disease in question and using positive and negative predictive values. We provide worked examples and an interactive tool (Supplemental Material) to determine the minimum clinical performance required for a test in settings of varying SARS-CoV-2 prevalence. It is hoped that this tool will assist in deciding the most appropriate testing protocols across diverse clinical, healthcare, public health, and logistical conditions. To select the minimum clinical performance specification, one should first estimate the prevalence and define the desired PPV or NPV given the intended test use. The tool can then be used to estimate the minimum clinical sensitivity required for a range of set clinical specificities, and the minimum clinical specificity required for a range of set clinical sensitivities. Once the most appropriate testing strategy for a given target population has been chosen, clinicians may prefer to calculate the post-test probability of disease using likelihood ratios in order to make clinical decisions for individual patients [15].

The prevalence estimate depends on the target population where the test is intended to be used. For example, when the test is intended to be used as part of a seroprevalence study, then the prevalence should be estimated for the community under study. Conversely, when the test is applied to patients with symptoms compatible with COVID-19 or those with positive contact history, then the prevalence (and pre-test probability) will be much higher. The desirable PPV or NPV is dependent on end-user tolerance for false positive or false negative results respectively (e.g. 80% PPV reflects a tolerance of 1 false positive result for every 5 positive tests, while a PPV of 95% reflects a tolerance of 1 false positive result for every 20 positive tests).

The clinical performance specification is composed of both clinical sensitivity and clinical specificity. To determine the minimum requirement for one, the other needs to be considered accordingly. Most NAATs and serology tests for COVID-19 performed on the modern generation of analyzers claim clinical specificity of $\geq$ 95%, but with varying clinical sensitivity [16]. However, given high risk of bias in most diagnostic accuracy studies published to date, a range of clinical specificity beyond those claimed by the manufacturer should be allowed for when setting the minimum required clinical sensitivity [17,18].

A high NPV may be desired to avoid missing detection of SARS-CoV-2, which could otherwise put the population at risk of onward transmission of the infection. In relatively low prevalence scenarios ($\leq$5%) such as population-based screening of the general community, there is already a 95% probability that there is no disease. If this is considered to be adequate, then no test is necessary. We should also consider the difficulty in achieving a sufficiently high PPV in such low prevalence settings and the potential harm to individuals and society from false positive results. Nevertheless, the minimum clinical sensitivity required to achieve high NPV will increase in parallel with the prevalence, and this aspect must hence be carefully considered when planning large population-based screening programs.

For the clinical diagnosis of current SARS-CoV-2 infection, both a high PPV is necessary to avoid inappropriate clinical interventions (which are currently mostly experimental) and a high NPV to avoid missed diagnosis of infection. In higher prevalence scenarios such as patients presenting for healthcare with flu-like symptoms, both a high minimum clinical sensitivity and clinical specificity would be necessary. When testing is proposed to inform decisions where a rapid result is essential, such as participation or exclusion from employment or other activities, it is critical to define the consequence of a false positive or false negative result relative to action without the test, rather than relative to a 'gold' standard reference test that is not feasible within the required turn-around time. For example, the minimum acceptable PPV and NPV may be more modest if the alternative action is no testing with limitations on participation based on symptoms alone. This may open the possibility of selecting tests characterized by rapid turnaround time but having lower clinical sensitivity[19] or using combination testing. Importantly, the clinical performance of a test may change dynamically throughout the course of the infection [18–21]. Practitioners should therefore apply the test within the time window where the clinical sensitivity is highest to optimize the desired post-test probability. Pre-analytical factors such as sample collection techniques and suitable sample types may also influence the clinical performance of the test.

Table 1
Key elements of consideration when deciding on an *a priori* clinical performance specification.

Target population of test use
SARS-CoV-2 prevalence in the target population
Purpose of testing (i.e. rule in or rule out)
Minimum acceptable positive or negative predictive value
Minimum clinical sensitivity or clinical specificity

In summary, the selection of laboratory tests and the testing strategy for COVID-19 involves careful consideration of trade-offs between clinical sensitivity and clinical specificity, as well as NPV and PPV. Practitioners should consider the clinical scenarios under which the test result will be used and select the most appropriate testing strategy that fulfils the *a priori* defined clinical performance specifications (Table 1). In some cases, an individual laboratory test may not be able to fulfill the clinical requirement and a multi-modality testing strategy may need to be employed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinbiochem.2020.11.003.

References

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, et al., A Novel Coronavirus from Patients with Pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733.
[2] World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. https://www.who.int/dg/speeches/detail/who-directorgeneral-s-opening-remarks-at-the-media-briefing-on-covid-19—11-March-2020. (Accessed: 1 May 2020).
[3] M.K. Bohn, G. Lippi, A. Horvath, S. Sethi, D. Koch, M. Ferrari, et al., Molecular, serological, and biochemical diagnosis and monitoring of COVID-19: IFCC taskforce evaluation of the latest evidence, Clin. Chem. Lab. Med. 58 (2020) 1037–1052.
[4] H. Sung, K.H. Roh, K.H. Hong, M.W. Seong, N. Ryoo, H.S. Kim, et al., COVID-19 Molecular Testing in Korea: Practical Essentials and Answers From Experts Based on Experiences of Emergency Use Authorization Assays, Ann. Lab. Med. 40 (2020) 439–447.
[5] Bohn MK, Mancini N, Loh TP, Wang CB, Grimmler M, Gramegna M, et al. IFCC Interim Guidelines on Molecular Testing of SARS-CoV-2 Infection. Clin Chem Lab Med. 2020 Oct 7:/j/cclm.ahead-of-print/cclm-2020-1412/cclm-2020-1412.xml. [E-pub ahead of print].
[6] Bohn MK, Loh TP, Wang CB, Mueller R, Koch D, Sethi S, et al. IFCC Interim Guidelines on Serological Testing of Antibodies against SARS-CoV-2. Clin Chem Lab Med. 2020 Oct 7:/j/cclm.ahead-of-print/cclm-2020-1413/cclm-2020-1413.xml. [E-pub ahead of print].
[7] Public Health England. COVID-19: PHE assessment of laboratory tests and assays. Information, resources and guidance relating to the evaluation of COVID-19 tests. https://www.gov.uk/government/collections/covid-19-phe-assessment-of-laboratory-tests-and-assays. (Accessed 20 Aug 2020).
[8] U.S. Food & Drug Administration. EUA Authorized Serology Test Performance. https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance. (Accessed 20 Aug 2020).
[9] Egner W. Verification and validation methodology and sample sets for evaluation of assays for SARS-CoV-2 (COVID-19). https://www.rcpath.org/uploads/assets/

541a4523-6058-4424-81c119dd2ab0febb/Verification-validation-of-sample-sets-assays-SARS-CoV-2.pdf. Accessed 20 Aug 2020).
[10] Medicine & Healthcare Products Regulatory Agency. Target product profile: antibody tests to help determine if people have immunity to SARS-CoV-2: version 2. https://www.gov.uk/government/publications/how-tests-and-testing-kits-for-coronavirus-covid-19-work/target-product-profile-antibody-tests-to-help-determine-if-people-have-recent-infection-to-sars-cov-2-version-2. (Accessed 20 Aug 2020).
[11] Theel E, Filkins L, Palavecino E, Mitchell S, Campbell S, Pentella M, et al. Verification procedure for commercial serologic tests with Emergency Use Authorization for detection of antibodies to SARS-CoV-2. https://asm.org/ASM/media/Protocol-Images/ASM-Web-protocol-serology-062220-FINAL.pdf?ext=.pdf. (Accessed 20 Aug 2020).
[12] A.R. Horvath, S.J. Lord, A. StJohn, S. Sandberg, C.M. Cobbaert, S. Lorenz, et al., From biomarkers to medical tests: the changing landscape of test evaluation, Clin. Chim. Acta 427 (2014) 49–57.
[13] S.J. Lord, A. St John, P.M. Bossuyt, S. Sandberg, P.J. Monaghan, M. O'Kane, et al., Setting clinical performance specifications to develop and evaluate biomarkers for clinical use, Ann. Clin. Biochem. 56 (2019) 527–535.
[14] Lalkhen A.G., McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 2008;8:221-3.
[15] K.J.L. Bell, F. Stanaway, L. Irwig, A.R. Horvath, A. Teixeira-Pinto, C. Loy, Testing decisions in the Pandemic: How do we use imperfect tests for COVID-19 (SARS-CoV-2) to make clinical decisions? Med. J. Aus. (2020) (E-pub ahead of print).
[16] Public Health England. COVID-19: laboratory evaluations of serological assays. https://www.gov.uk/government/publications/covid-19-laboratory-evaluations-of-serological-assays. (Accessed 20 Aug 2020).
[17] J.J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, R. Spijker, S. Taylor-Phillips, et al. Antibody tests for identification of current and past infection with SARS-CoV-2 Cochrane Database, Syst. Rev. 2020, 6, CD013652.
[18] S. Mallett, A.J. Allen, S. Graziado, S.A. Taylor, N.S. Sakai, K. Green, et al., At what times during infection is SARS-CoV-2 detectable and no longer detectable using RT-PCR based tests? A systematic review of individual participant data, medRxiv 07 (13) (2020) 20152793.
[19] J. Lessler, R. Brookmeyer, T.M. Perl, An evaluation of classification rules based on date of symptom onset to identify health-care associated infections, Am. J. Epidemiol. 166 (2007) 1220–1229.
[20] I. Arevalo-Rodriguez, D. Buitrago-Garcia, D. Simancas-Racines, P. Zambrano-Achig, R.D. Campo, A. Ciapponi, et al., False negative results of initial RT-PCR assays for COVID-19: a systematic review, medRxiv 04 (16) (2020) 20066787.
[21] A. Weiss, M. Jellingsø, M.O.A. Sommer, Spatial and temporal dynamics of SARS-CoV-2 in COVID-19 patients: A systematic review and meta-analysis, EBioMedicine 58 (2020), 102916.