Diagnostic accuracy estimates for COVID-19 RT-PCR and Lateral flow immunoassay tests with Bayesian latent class models

Polychronis Kostoulas (pkost@uth.gr)
School of Health Sciences, University of Thessaly, Karditsa, Greece https://orcid.org/0000-0001-7051-1541

Paolo Eusebi
Health Planning Service, Regional Health Authority of Umbria, Perugia, Italy https://orcid.org/0000-0002-0715-6396

Sonja Hartnack
Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland
https://orcid.org/0000-0002-5757-5708

Research Article

Keywords: SARS-CoV-2, COVID-19, RT-PCR, IgG, IgM, Sensitivity, Specificity, Bayesian analysis, Latent class models.

Posted Date: June 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-33243/v1

Version of Record: A version of this preprint was published on March 31st, 2021. See the published version at https://doi.org/10.1093/aje/kwab093.

Diagnostic accuracy estimates for COVID-19 RT-PCR and Lateral flow immunoassay tests with Bayesian latent class models

Polychronis Kostoulas[a*], Paolo Eusebi[b], Sonja Hartnack[c]

[a] *School of Health Sciences, University of Thessaly, Karditsa, Greece.*
[b] *Health Planning Service, Regional Health Authority of Umbria, Perugia, Italy.*
[c] *Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland.*
**Correspondence: pkost@uth.gr*

Abstract

The objective of this work was to estimate the diagnostic accuracy of RT-PCR and Lateral flow immunoassay tests (LFIA) for COVID-19, depending on the time post symptom onset. Based on the cross-classified results of RT-PCR and LFIA, we used Bayesian latent class models (BLCMs), which do not require a gold standard for the evaluation of diagnostics[1]. Data were extracted from studies that evaluated LFIA (IgG and/or IgM) assays using RT-PCR as the reference method. The cross-classified results of LFIA and RT-PCR were analysed separately for the first, second and third week post symptom onset. Se_{RT-PCR} was 0.695 (95% probability intervals: 0.563; 0.837) for the first week and remained similar for the second and the third week. $Se_{IgG/M}$ was 0.318 (0.229; 0.416) for the first week and increased steadily. It was 0.755 (0.673; 0.829) and 0.927 (0.881; 0.965) for the second and third week, respectively. Both tests had a high to absolute *Sp*, with point median estimates for Sp_{RT-PCR} being consistently higher. Sp_{RT-PCR} was 0.990 (0.980; 0.998) for the first week. The corresponding value for $Sp_{IgG/M}$ was 0.962 (0.905; 0.998). Further, *Sp* estimates for each test did not differ between weeks. BLCMs provide a valid and efficient alternative for evaluating the rapidly evolving diagnostics for COVID-19, under various clinical settings and for different risk profiles.

Keywords*: SARS-CoV-2, COVID-19, RT-PCR, IgG, IgM, Sensitivity, Specificity, Bayesian analysis, Latent class models.*

1. Introduction

Over the past few months, there has been a need for rapid development of diagnostic tests that will efficiently detect SARS-Cov-2 infection. Real-time reverse transcriptase polymerase chain reaction (RT-PCR) tests, which detect the RNA of SARS-Cov-2, are considered as the reference[2] for a COVID-19 diagnosis. In addition, the development of serological assays detecting SARS-COV-2-specific IgM and/or IgG started immediately and is on-going[3] with a large portion of them being Lateral flow immunoassays (LFIA). These immunoassays are evaluated using RT-PCR as a gold standard[4–6]. However, it is known that RT-PCR is less than 100% sensitive[7] while false positive results can also occur[8]. Thus, if a new diagnostic test is evaluated assuming

RT-PCR as a perfect reference standard – although it is not – the evaluation of the new test may be biased.

In the absence of a gold standard, Bayesian latent class models (BLCMs), which do not require a priori knowledge of the infection status, are a valid alternative to classical test evaluation. In a BLCM setting, none of the tests is considered as a reference method and the *Se* and *Sp* for each test is estimated from the analysis of the cross-classified results of two or more tests in one or more populations. Latent models for diagnostic accuracy studies were introduced with the two-test, two-population model[9], which is often referred to as the Hui and Walter paradigm. The first thorough discussion on the applicability of these methods in diagnostic accuracy studies was given by Walter and Irwig[10] and their implementation within a Bayesian framework has been evolving for over 20 years[11–13] Recently, guidelines for the application and sound reporting of BLCMs in diagnostic accuracy studies, the STARD-BLCM statement, have been proposed[1,14]. STARD-BLCM is an adaptation of the STARD statement[15] for the absence of a reference test and the use of Bayesian estimation procedures. Currently, an EU-funded initiative has brought together experts from 43 countries with the aim to further develop and expand the application of BLCMs in biomedicine (www.harmony-net.eu).

To the best of our knowledge, BLCMs (or LCMs) have not been used for the evaluation of COVID-19 diagnostics despite the obvious advantages arising from the fact that there is no need for a gold standard. BLCMs can be advantageous since diagnostic processes for COVID-19 have been developed at an unprecedented pace and understanding of viral dynamics across the course of SARS-COV-2 infection is incomplete. The objective of this work was to estimate the diagnostic accuracy of RT-PCR and LFIA tests depending on the week post symptom onset with the use of BLCMs. Hence, Se/Sp estimates for both RT-PCR and IgG/M were obtained and for each of the first three weeks after the onset of symptoms.

2. Materials and Methods

2.1 Literature search and selection of studies – datasets.

A flow chart for the selection process is in Figure S1. We conducted the literature search using PubMed (https://www.ncbi.nlm.nih.gov/pubmed/), medRxiv (https://medrxiv.org/) and bioRxiv (https://www.biorxiv.org/) without any language restrictions. The search strategy and results for each database are in Table S1.

The following search terms were used: (“SARS-CoV-2” OR "SARS-CoV-2” OR “Coronavirus disease 2019” OR “COVID-19”) AND (“IgM” OR “IgG” OR “antibodies” OR “antibody” OR “serological” OR “serologic” OR “serology” OR “serum” OR "lateral flow").

The searches were concluded by April 30, 2020, and two researchers independently screened articles. Disagreements in the initial evaluation were resolved by consensus.

Eligible articles were required to meet the following criteria: (i) inclusion of COVID-19 cases (non-cases) confirmed (ruled-out) by RT-PCR or by a combination of RT-PCR and clinical findings; (ii) results concerning IgM and/or IgG antibodies using lateral flow immunoassay; (iii) availability of clinical information, in particular with respect to days from onset of symptoms; (iv) more than 7 days from RT-PCR testing.

In order to construct the two by two contingency table and obtain estimates for Se and Sp, we obtained the numbers of: Ab and RT-PCR positive (Ab+/RT-PCR +); Ab positive and RT-PCR negative (Ab+/RT-PCR–); Ab negative and RT-PCR positives (Ab–/ RT-PCR +); Ab and RT-PCR negative (Ab–/ RT-PCR –).

Initially, 449 non-duplicated records were screened, and 28 full-text resources were scrutinized. Finally, four studies[16–19] were identified that fulfilled criteria (i) to (iv) and cross-classified results could be extracted.

2.2 Bayesian latent class model for Se/Sp estimation in the absence of a reference test.

BLCMs do not use a gold standard (i.e. a reference test with perfect diagnostic accuracy) to determine the disease/infection status. For dichotomized test results, estimation of the *Se* and *Sp* of the tests is based on the cross-classified results. With two tests in two populations the model is fully identifiable because there are six degrees of freedom (i.e. three from each population) and six parameters to be estimated: the Se and Sp of each test and the true prevalence of disease/infection in each population. Here, we extend this model in a two-test (i.e. RT-PCR and IgG/M) four-population model (i.e. each study is considered a different population).

Briefly, we assume that for each of the *i* populations – in our case the four different studies – the cross classified results of the two tests follow an independent multinomial sampling distribution:

$$y_i \sim Multinomial\left(n_i, \left(p_{i_{11}}, p_{i_{12}}, p_{i_{21}}, p_{i_{22}}\right)\right)$$

with the multinomial cell probabilities being expressed as:

$$p_{i_{11}} = p_i Se_{RT-PCR} Se_{IgG/M} + (1-p_i)(1-Sp_{RT-PCR})(1-Sp_{IgG/M})$$
$$p_{i_{12}} = p_i Se_{RT-PCR}(1-Se_{IgG/M}) + (1-p_i)(1-Sp_{RT-PCR})Sp_{IgG/M}$$
$$p_{i_{21}} = p_i(1-Se_{RT-PCR})Se_{IgG/M} + (1-p_i)Sp_{RT-PCR}\left(1-Sp_{IgG/M}\right)$$
$$p_{i_{22}} = p_i(1-Se_{RT-PCR})(1-Se_{IgG/M}) + (1-p_i)Sp_{RT-PCR}Sp_{IgG/M}$$

Within a fully Bayesian estimation framework, Beta distributions *Be (a, b),* are used as priors for the parameters of interest: Se_{RT-PCR}, Sp_{RT-PCR} $Se_{IgG/M}$, $Sp_{IgG/M}$ and the prevalence p_i in each population.

2.3 Assessing conditional dependence.

Our model assumed that RT-PCR and LFIA are conditionally independent, an assumption which is expected to be valid because the two tests are based on a different biological principle[11]. Nevertheless, to account for the unlikely, yet existent, possibility of conditional dependence between RT-PCR and LFIA we also considered a model that captures conditional dependences. That is:

$$p_{i_{11}} = p_i\left(Se_{RT-PCR}Se_{IgG/M} + cdp\right) + (1-p_i)\left((1-Sp_{RT-PCR})(1-Sp_{IgG/M}) + cdn\right)$$
$$p_{i_{12}} = p_i\left(Se_{RT-PCR}(1-Se_{IgG/M}) - cdp\right) + (1-p_i)\left((1-Sp_{RT-PCR})Sp_{IgG/M} - cdn\right)$$
$$p_{i_{21}} = p_i\left((1-Se_{RT-PCR})Se_{IgG/M} - cdp\right) + (1-p_i)\left(Sp_{RT-PCR}(1-Sp_{IgG/M}) - cdn\right)$$
$$p_{i_{22}} = p_i\left((1-Se_{RT-PCR})(1-Se_{IgG/M}) + cdp\right) + (1-p_i)\left(Sp_{RT-PCR}Sp_{IgG/M} + cdn\right)$$

where *cdp* and *cdn* is the conditional covariance between the Ses and the Sps, respectively. Uniform priors were specified for *cdp* and *cdn* with their limits being directly affected by the magnitude of the *Se* and *Sp* values[20]:

$$cdp \sim Uniform\left((Se_{RT-PCR} - 1)(1-Se_{IgG/M}), \left(min\left(Se_{RT-PCR}, Se_{IgG/M}\right) - Se_{RT-PCR}Se_{IgG/M}\right)\right)$$
$$cdn \sim Uniform\left((Sp_{RT-PCR} - 1)(1-Sp_{IgG/M}), \left(min\left(Sp_{RT-PCR}, Sp_{IgG/M}\right) - Sp_{RT-PCR}Sp_{IgG/M}\right)\right)$$

2.4 Priors and sensitivity analysis

We have a two-test, four-subpopulation model, which is fully identifiable because the number of parameters to be estimated are eight (i.e. the Se and Sp of each test and the prevalence of SARS-Cov-2 infection in each population) for the independence model and ten (i.e. the two additional *cdp* and *cdn* parameters) and the degrees of freedom available from the data are twelve. In all alternative prior combinations a non-informative, uniform beta prior distribution, *Be* (1, 1), over the range from 0 to 1, was adopted for the Se_{RT-PCR}, $Se_{IgG/M}$ and the prevalence of SARS-Cov-2 infection in each population p_i.

For our primary analysis (Prior set I) Sp_{RT-PCR} was expected to have a median of 0.99, and it was thought to be at least 0.98 with 95% certainty, which corresponds to a Be (426.36,4.64). For $Sp_{IgG/M}$ the median was expected to be 0.98 and it was thought to be higher than 0.95 with 95 certainty. That is a Be (108.19,2.53).

Alternative prior combinations were: (i) fixing Sp_{RT-PCR} equal to 1 and using the same prior for $Sp_{IgG/M}$ (Prior set II) and (ii) assuming for both Sp_{RT-PCR} and $Sp_{IgG/M}$ an a priori median of 0.95 and a lower value of 0.90 with 95% certainty. This is a Beta (76.63, 4.35). The latter prior can be assumed to be weakly informative as it specifies a range of values that is rather wide given the values that the specificities that RT-PCR and LFIA tests are expected to have.

2.5 *Convergence diagnostics and software.*

We used a combination of checks because convergence diagnostics of the Markov Chain Monte Carlo (MCMC) are not fool proof. Specifically, the Raftery and Lewis method[21], the Gelman–Rubin diagnostic[22], autocorrelation checks and visual

inspection of the trace plots and summary statistics were used as recommended[23]. Parameter estimates were based on analytical summaries of 60,000 iterations of three chains after a burn-in adaptation phase of 10,000 iterations. All checks suggested that convergence occurred and autocorrelations dropped-off fast. Models were run in the freeware program JAGS[24] through R[25] using the rjags package[24]. Priors were generated with the PriorGen package[26].

The code is available at https://github.com/paoloeusebi/BLCM-Covid19.

Table 1. Cross-classified results of the RT-PCR (PCR) and the Lateral flow immunoassay tests detecting either IgG or IgM antibodies (IgG/M) against COVID-19.

Study	Week	PCR (+) IgG/M (+)	PCR (+) IgG/M (-)	PCR(-) IgG/M (+)	PCR (-) IgG/M (-)
A[16]	Week 1	1	7	0	0
B[17]		3	13	2	7
C[18]		3	24	4	5
D[19]		12	15	14	38
A[16]	Week 2	8	16	15	3
B[17]		6	0	1	1
C[18]		26	2	5	1
D[19]		28	8	14	38
A[16]	Week 3	17	6	41	4
B[17]		68	0	5	9
C[18]		30	1	5	2
D[19]		17	4	14	38

3. Results

A total of 448 studies were initially identified as studies on the evaluation of COVID-19 diagnostics and 28 of them provided access to full data that can be extracted. From these, 4 gave details on the cross classified of RT-PCR and LFIA results for each post symptom onset (Table S1 and Figure S1).

Cross classified results of the of the RT-PCR and the LFIA for each week from the onset of COVID-19 symptoms are in Table 1. *Se* and *Sp* BLCM estimates for each week are in Table 2. Se_{RT-PCR} remained similar for the first three weeks, while $Se_{IgG/M}$ increased week by week with non-overlapping probability intervals (i.e. which would be the equivalent of a statistically significant difference in classical statistics). Both tests were of high to absolute *Sp* that did not differ with point estimates for Sp_{RT-PCR} being consistently higher. Further, *Sp* estimates were similar for all weeks.

The same results were observed under the model that adjusted for the potential conditional dependence between the tests. There was no strong evidence of conditional dependence since covariance parameters, *cdp* and *cdn*, had probability intervals that included zero.

Finally, alternative prior specifications – prior set II & III – gave similar results (Table S2).

Table 2. Medians and 95% probability intervals (PrIs) for the *Se* and *Sp* of the RT-PCR and the Lateral flow immunoassays detecting IgG and/or IgM antibodies (IgG/M) against COVID-19. Model (A) assumes conditional independence while model (B) adjusts for the potential dependencies between the Ses (cdp) and the Sps (cdn) of the tests.

Model	Parameter	Week 1	Week 2	Week 3
A	$Se_{RT\text{-}PCR}$	0.695 (0.563; 0.837)	0.694 (0.612; 0.777)	0.674 (0.607; 0.739)
	$Se_{IgG/M}$	0.318 (0.229; 0.416)	0.755 (0.673; 0.829)	0.927 (0.881; 0.965)
	$Sp_{RT\text{-}PCR}$	0.990 (0.980; 0.998)	0.990 (0.979; 0.997)	0.989 (0.978; 0.997)
	$Sp_{IgG/M}$	0.962 (0.905; 0.998)	0.978 (0.947; 0.998)	0.978 (0.944; 0.998)
B	$Se_{RT\text{-}PCR}$	0.773 (0.654; 0.874)	0.710 (0.626; 0.789)	0.671 (0.602; 0.736)
	$Se_{IgG/M}$	0.364 (0.268; 0.467)	0.779 (0.695; 0.852)	0.926 (0.872; 0.970)
	$Sp_{RT\text{-}PCR}$	0.990 (0.980; 0.998)	0.990 (0.979; 0.997)	0.989 (0.978; 0.997)
	$Sp_{IgG/M}$	0.973 (0.932; 0.998)	0.979 (0.947; 0.998)	0.978 (0.945; 0.999)
	cdp	-0.092 (-0.139; -0.039)	-0.038 (-0.068; 0.000)	-0.001 (-0.022; 0.03)
	cdn	0.004 (0.000; 0.012)	0.004 (0.000; 0.012)	0.003 (0.000; 0.011)

4. Discussion

We used BLCMs to estimate the diagnostic accuracy of RT-PCR and LFIA tests for SARS–CoV-2 infection depending on the time from the onset of symptoms. BLCMs do not require the presence of a reference test and thus allow for the simultaneous *Se* and *Sp* estimation of both tests. They provide a valid and efficient alternative to classical test evaluation[1,9]. Importantly, in either model (i.e. the conditional independence and conditional dependence model) the degrees of freedom provided by the data (i.e. 12) exceeded the number of parameters that had to be estimated (i.e. 8 and 10 for the conditional independence and dependence model, respectively), satisfying a necessary condition for identifiability. Further, sensitivity analysis revealed that under alternative prior specifications our results were similar (Table S2) without differences in the estimates between the two model structures and alternative prior sets. Finally, the assumption of conditional independence was valid because covariance estimates had, under any prior combination, probability intervals that included zero. Conditional independence is expected to hold when the tests, as in our case, are based on a different biological principle[11]. This is, to the best of our knowledge, the first study using BLCMs for the evaluation of COVID-19 diagnostics. This may be due to the absence of suitable data: despite the vast literature on the evaluation diagnostic tests for SARS–CoV-2 (Table S1) only four studies were identified with adequate information to extract cross-classified results for different time periods from the onset of symptoms.

Our $Se_{RT\text{-}PCR}$ estimates were moderate to high and in line with current evidence[27]. Further, we demonstrated that $Se_{RT\text{-}PCR}$ remains constant throughout the first three weeks after the onset of symptoms: probability intervals are largely overlapping suggesting no significant differences between the first, second or third week. Nevertheless, point estimates seem to indicate a decline which is higher for the third week. This is in accordance with evidence of decreasing viral load as the infection progresses and especially during the third week after the onset of symptoms[28,29].

The $Se_{IgG/M}$ estimates were low for the first week and show a steep increase to moderate in the second week that further continued resulting to high *Se* values for the third week (Table 2). Weekly $Se_{IgG/M}$ estimates had non-overlapping probability intervals which is equivalent to a statistically significant increase – in a frequentist setting – with time. At the early stages of SARS–CoV-2 infection IgG/M assays are likely to have false negative results and miss cases due to the fact that a detectable antibody response to SARS–CoV-2 infection can take more than ten days after the onset of symptoms[28]. The subsequent increase is in line with published findings[30]. Further, an increase in IgG and/or IgM during the first three weeks is also recorded[19,31–33]. The median seroconversion time is expected to occur 10 and 12 days post symptom onset for IgG and IgM, with a rapid increase after day 6 that can be followed by a decline in viral load[34]. The latter observation of increasing positive detection rate for IgG and/or IgM with a steady and potentially slight decrease for SARS-CoV-2 viral load has also been observed elsewhere[35,36]. $Se_{IgG/M}$ is higher than $Se_{RT\text{-}PCR}$ after the second week, which is also in line with recent evidence that the sensitivity of antibody assays overtook the RNA test on day 8 after the onset of symptoms[37]. Further, other authors also found a steep increase for antibodies, particularly in the second week, that is accompanied by a slight decrease in the probability of detection with nasopharyngeal swabs/ bronchoalveolar/sputum PCR over the first three weeks after symptom onset[38].

The $Sp_{RT\text{-}PCR}$ estimate was close to unity and steady across time, but false positive results can occur[8]. There is scarcity of *Sp* estimates for RT-PCR methods because they are considered as the reference standard for the evaluation of diagnostic tests for SARS–CoV-2 infection. False positive RT-PCR results are only assumed to occur as a result of sample contamination or the high threshold cycle (Ct) values[39]. Nevertheless, we do not believe that the estimated false positive rate could only be due to contamination issues. In studies comparing RT-PCR results to chest-CT a substantial number of samples was found chest-CT negative but RT-PCR positive[40,41]. Given that chest-CT has emerged as a valid test for early diagnosis of SARS–CoV-2 infection and its combination with RT-PCR is suggested[39], the perfect *Sp* of RT-PCR is at best in question. Undoubtedly, though, $Sp_{RT\text{-}PCR}$ is close to unity, but the possibility of false positive results should not be ruled out. The latter will be of great importance at the next steps in the fight of COVID-19 pandemic and the case of screening healthy or low prevalence populations. In such instances, false positive results can occur and should be accounted for to avoid unnecessary interventions.

Finally, $Sp_{IgG/M}$ was also close to perfect, but with median estimates consistently lower than those for $Sp_{RT\text{-}PCR}$ but not statistically different. False positive results can be due to cross-reactions, which have been observed in diagnostic evaluation studies that

were based on a reference standard from healthy individuals or individuals that have diseases unrelated to SARS–CoV-2 infection[42]. Cross reactivity between SARS–CoV-2 IgM assays and the rheumatoid factor IgM (RF-IgM) has also been observed[43].

A point of criticism for our analysis might have been that target variable bias can be a serious issue when BLCMs are applied in acute infection data because the time period during which the different targeted conditions (in our case presence of viral particles and IgG or IgM antibodies) are both detectable is narrower[1,11]. In such cases, the infection status that is detected by the BLCMs is limited to the individuals with simultaneous presence of RNA viral particles and IgG/M antibodies. Here, we expect such bias to be low because we narrowed our selection of cases in a period where both targets (i.e. viral particles for RT-PCR and IgG/M antibodies for LFIA) co-exist. This may not be true earlier in the course of SARS–CoV-2 infection when viral particles are present, but antibodies have not yet been produced or later when the infection may be cleared out, but antibody levels are high.

BLCMs provide a flexible and valid estimation framework to readily evaluate tests for COVID-19 and provide Se/Sp estimates without the need of a reference method. This facilitates the rapid evaluation of diagnostics depending on the clinical setting and the duration of SARS–CoV-2 infection, as in our case. In light of a continuously evolving pandemic and the influx of new epidemiological data, BLCMs can provide a framework for *Se/Sp* estimates that will be specific to different risk profiles and will allow for the interpretation of test outcomes according to the relevant epidemiological situation in each case.

5. References

1. Kostoulas, P. *et al.* Reporting guidelines for diagnostic accuracy studies that use Bayesian latent class models (STARD-BLCM). *Stat. Med.* **36**, (2017).
2. Cheng, M. P. *et al.* Diagnostic Testing for Severe Acute Respiratory Syndrome–Related Coronavirus-2: A Narrative Review. *Ann. Intern. Med.* (2020).
3. Petherick, A. Developing antibody tests for SARS-CoV-2. *Lancet* (2020) doi:10.1016/s0140-6736(20)30788-1.
4. Jia, X. *et al.* Clinical significance of IgM and IgG test for diagnosis of highly suspected COVID-19 infection. *medRxiv* (2020).
5. Paradiso, A. V. *et al.* Clinical meanings of rapid serological assay in patients tested for SARS-Co2 RT-PCR. *medRxiv* (2020).
6. Li, Z. *et al.* Development and Clinical Application of A Rapid IgM-IgG Combined Antibody Test for SARS-CoV-2 Infection Diagnosis. *J. Med. Virol.* (2020) doi:10.1002/jmv.25727.
7. Xiao, A. T., Tong, Y. X. & Zhang, S. False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: Rather than recurrence. *Journal of Medical Virology* (2020) doi:10.1002/jmv.25855.
8. Lan, L. *et al.* Positive RT-PCR Test Results in Patients Recovered from COVID-19. *JAMA - Journal of the American Medical Association* (2020) doi:10.1001/jama.2020.2783.
9. Hui, S. L. & Walter, S. D. Estimating the Error Rates of Diagnostic Tests. *Biometrics* (1980) doi:10.2307/2530508.
10. Walter, S. D. & Irwig, L. M. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *J. Clin. Epidemiol.* (1988) doi:10.1016/0895-4356(88)90110-2.
11. Branscum, A. J., Gardner, I. A. & Johnson, W. O. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev. Vet. Med.* (2005) doi:10.1016/j.prevetmed.2004.12.005.
12. Collins, J. & Huynh, M. Estimation of diagnostic test accuracy without full verification: A review of latent class methods. *Stat. Med.* (2014) doi:10.1002/sim.6218.
13. Kostoulas, P., Nielsen, S. S., Browne, W. J. & Leontides, L. Sample size estimation to substantiate freedom from disease for clustered binary data with a specific risk profile. *Epidemiol. Infect.* **141**, (2013).
14. Kostoulas, P. *et al.* STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Prev. Vet. Med.* **138**, (2017).
15. Bossuyt, P. M. *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann. Intern. Med.* (2003) doi:10.7326/0003-4819-138-1-200301070-00012-w1.
16. Garcia, F. P. *et al.* Rapid diagnosis of SARS-CoV-2 infection by detecting IgG and IgM antibodies with an immunochromatographic device: a prospective single-center study. *medRxiv* (2020) doi:10.1101/2020.04.11.20062158.
17. Liu, Y. *et al.* Diagnostic Indexes of a Rapid IgG/IgM Combined Antibody Test for SARS-CoV-2. *medRxiv* (2020).
18. Pan, Y. *et al.* Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. *J. Infect.* (2020) doi:10.1016/j.jinf.2020.03.051.
19. Whitman, J. D. *et al.* Test performance evaluation of SARS-CoV-2 serological

assays. *medRxiv* (2020) doi:10.1101/2020.04.25.20074856.
20. Dendukuri, N. & Joseph, L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* **57**, 158–167 (2001).
21. Raftery, A. E. & Lewis, S. M. Comment: One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Stat. Sci.* (1992) doi:10.1214/ss/1177011143.
22. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* (1992) doi:10.1214/ss/1177011136.
23. Best, N. G. & Cowles, M. K. CODA: Convergence diagnosis and output analysis software for Gibbs sampling output. *MRC Biostat. Unit, Cambridge Univ.* (1997).
24. Plummer, M. rjags: Bayesian graphical models using MCMC. *R package version 3-13* (2016) doi:http://cran.r-project.org/package=rjags.
25. Team, R. C. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (2016).
26. Kostoulas, P. PriorGen: Generates Prior Distributions for Proportions. *R Packag. version* **1**, (2018).
27. Al-Tawfiq, J. A. & Memish, Z. A. Diagnosis of SARS-CoV-2 Infection based on CT scan vs. RT-PCR: Reflecting on Experience from MERS-CoV. *J. Hosp. Infect.* (2020) doi:10.1016/j.jhin.2020.03.001.
28. To, K. K.-W. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* (2020) doi:10.1016/s1473-3099(20)30196-1.
29. Yang, Y. *et al.* Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. *medRxiv* (2020) doi:10.1101/2020.02.11.20021493.
30. Adams, E., Ainsworth, M., Anand, R. & Andersson, M. Antibody testing for COVID-19: A report from the National COVID Scientific Advisory Panel. *medRxiv* (2020) doi:10.1101/2020.04.15.20066407v2.
31. Long, Q. *et al.* Antibody responses to SARS-CoV-2 in COVID-19 patients: the perspective application of serological tests in clinical practice. *medRxiv* (2020) doi:10.1101/2020.03.18.20038018.
32. Zhang, J. *et al.* Serological detection of 2019-nCoV respond to the epidemic: A useful complement to nucleic acid testing. *medRxiv* (2020) doi:10.1101/2020.03.04.20030916.
33. Guo, L. *et al.* Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa310.
34. Lou, B. *et al.* Serology characteristics of SARS-CoV-2 infection since exposure and post symptom onset. *Eur. Respir. J.* (2020) doi:10.1183/13993003.00763-2020.
35. Jin, Y. *et al.* Diagnostic value and dynamic variance of serum antibody in coronavirus disease 2019. *Int. J. Infect. Dis.* (2020) doi:10.1016/j.ijid.2020.03.065.
36. Zhao, J. *et al.* Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa344.
37. Gao, Y. *et al.* Evaluation the auxiliary diagnosis value of antibodies assays for detection of novel coronavirus (SARS-Cov-2) causing an outbreak of pneumonia (COVID-19). *medRxiv* (2020) doi:10.1101/2020.03.26.20042044.

399 38. Sethuraman, N., Jeremiah, S. S. & Ryo, A. Interpreting Diagnostic Tests for
400 SARS-CoV-2. *JAMA - Journal of the American Medical Association* (2020)
401 doi:10.1001/jama.2020.8259.
402 39. Tahamtan, A. & Ardebili, A. Real-time RT-PCR in COVID-19 detection:
403 issues affecting the results. (2020).
404 40. Ai, T. *et al.* Correlation of Chest CT and RT-PCR Testing in Coronavirus
405 Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*
406 (2020) doi:10.1148/radiol.2020200642.
407 41. Bai, H. X. *et al.* Performance of radiologists in differentiating COVID-19 from
408 viral pneumonia on chest CT. *Radiology* (2020)
409 doi:10.1148/radiol.2020200823.
410 42. Lassaunière, R. *et al.* Evaluation of nine commercial SARS-CoV-2
411 immunoassays. *medRxiv* (2020).
412 43. Wang, Q. *et al.* A method to prevent SARS-CoV-2 IgM false positives in gold
413 immunochromatography and enzyme-linked immunosorbent assays. *J. Clin.*
414 *Microbiol.* (2020).
415

Declarations:

Competing interests: The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- COVID19supplement.pdf